



Published in final edited form as:

*Addict Behav.* 2008 January ; 33(1): 41–53.

## The Validity of Cocaine Dependence Subtypes

Henry R. Kranzler, M.D.<sup>1,\*</sup>, Marsha Wilcox, Ed.D, Sc.D.<sup>2</sup>, Roger D. Weiss, M.D.<sup>3</sup>, Kathleen Brady, M.D., Ph.D.<sup>4</sup>, Victor Hesselbrock, Ph.D.<sup>1</sup>, Bruce Rounsaville, M.D.<sup>5</sup>, Lindsay Farrer, Ph.D.<sup>2</sup>, and Joel Gelernter, M.D.<sup>5</sup>

1 Department of Psychiatry, University of Connecticut School of Medicine, Farmington, CT

2 Departments of Medicine (Genetics Program), Epidemiology, and Biostatistics, Boston University Schools of Medicine and Public Health

3 Department of Psychiatry, Harvard Medical School, Boston, MA, and McLean Hospital, Belmont, MA

4 Department of Psychiatry, Medical University of South Carolina, Charleston, SC

5 Department of Psychiatry, Yale University School of Medicine, New Haven, CT and VA CT Healthcare Center, West Haven, CT

### Abstract

Cocaine dependence (CD) is a multifactorial disorder, variable in its manifestations, and heritable. We examined the concurrent validity of homogeneous subgroups of CD as phenotypes for genetic analysis. We applied data reduction methods and an empirical cluster-analytic approach to measures of cocaine use, cocaine-related effects, and cocaine treatment history in 1393 subjects, from 660 small nuclear families. Four of the six clusters that were derived yielded heritability estimates in excess of 0.3. Linkage analysis showed genomewide significant results for two of the clusters. Here we examine the concurrent validity of the six clusters using a variety of demographic and substance-related measures. In addition to being differentiated by a variety of cocaine-related measures, the clusters differed significantly on measures that were independent of those used to generate the clusters, i.e., demographic features and prevalence rates of co-morbid substance use and psychiatric disorders. These findings support the validity of the methods used to derive homogeneous subgroups of CD subjects and the resulting CD subtypes. Independent replication of these findings would provide further validation of this approach.

### Keywords

Cocaine Dependence; Subtyping; Cluster Analysis; Heritability; Phenotype

## 1. INTRODUCTION

Cocaine use is widespread in the U.S., producing a variety of adverse medical and neuropsychiatric effects (Wolff and O'Donnell, 2004; Karch, 2005; Nnadi et al., 2005; Substance Abuse and Mental Health Services Administration, 2004). Cocaine dependence (CD), as a broad diagnostic entity, is a complex, heterogeneous, multifactorial disorder that includes cognitive, behavioral, and physiologic features. One way to reduce the heterogeneity

\*Correspondence to: Department of Psychiatry, University of Connecticut School of Medicine, Farmington, CT 06030-2103, Phone 860-679-4151, Fax 860-679-1316, Email: kranzler@psychiatry.uchc.edu.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

of CD is the use of a typologic (or subtyping) approach (Babor and Dolinsky, 1988; Epstein, 2001). The validity of such an approach can be evaluated in terms of the utility of the subtypes for understanding etiology, presentation, natural history, or response to treatment of individuals with CD.

Vulnerability to the development of CD varies among individuals. Studies in animals and humans have examined the relative contributions of environmental and genetic factors in the etiology of substance dependence (Uhl et al., 1995). Elucidating the genetic basis of CD would represent major progress in understanding the etiology of the disorder and could contribute substantially to the effort to develop efficacious medications to treat the disorder. This effort has, to date, been largely unsuccessful (Kosten et al., 2005). The failure to identify efficacious medications to treat cocaine dependence may reflect an inadequate understanding of the heterogeneity of the disorder or an incomplete understanding of the pathophysiology of the disorder, with inadequate specification of potential medication-responsive dimensions.

Twin studies have shown that cocaine and other stimulant dependence is genetically influenced (Tsuang et al., 1996; Kendler and Prescott, 1998; Kendler et al., 2003). Although these studies considered CD as a single diagnostic entity, there appear to be multiple subtypes of CD (Weiss and Mirin, 1986). If the broader set of CD subjects could be decomposed into valid subgroups, members of which were more similar phenotypically to each other than to members of other groups, these more homogenous subgroups could provide a basis for more powerful genetic analysis.

Univariate approaches for subtyping CD have focused on co-occurring psychopathology (Rounsaville et al., 1991), family history of substance abuse (Roehrich and Gold, 1988) and personality dimensions (Craig and Olson, 1992; Ball et al., 1998). Multivariate approaches to subtyping CD and other drug dependence (Ball et al., 1995; Feingold et al., 1996; Cohen, 1999; Basu et al., 2004) have used an empirical (k-means) clustering technique that was first applied to differentiate alcoholics into Type A (i.e., low-risk/severity) and Type B (i.e., high-risk/severity) subtypes (Babor et al., 1992). The present study describes a multivariate cluster analytic approach that has been refined from those used previously to yield cluster assignments for use in a genome-wide linkage analysis.

In our genome-wide linkage study of CD (Gelernter et al., 2005), we examined CD as a unified entity and used cluster analysis to identify subgroups that may vary in terms of heritability and of the genes underlying their vulnerability. Based on a six-cluster solution, we found interesting linkage results with two of these clusters. In this manuscript, we provide evidence for the concurrent validity of these clusters by comparing the subtypes on a variety of cocaine-related features, as well as the prevalence of co-morbid substance use and psychiatric disorders.

## 2. METHODS

### 2.1. Subjects

We recruited 1393 subjects, from 660 small nuclear families. Of these families, 482 had at least 2 siblings with CD, 207 had at least 2 siblings with opioid dependence, and 156 had at least 2 siblings with both CD and opioid dependence. The average age of subjects was 39.2 years (range 17-79) and 51.8% were women. The majority of subjects (57.1%) were never married, 27% were divorced, separated, or widowed, and 15.9% were married. The ethnic/racial distribution of the sample was 49.6% African-American (AA), 33.0% European-American (EA), 12.6% Hispanic, and 4.8% Native American, Pacific Islander or members of other minority groups (according to subject self-report). With respect to level of education, 7.3% had only completed grade school; 39.5% had some high school, but no diploma; 30.7% had completed high school; and 22.6% had education beyond high school.

The most common lifetime DSM-IV (American Psychiatric Association, 1994) substance use and psychiatric disorders are shown in Table 1, both in the aggregate and separately by sex. Nearly 90% of individuals were cocaine dependent, as might be expected given that ascertainment was based primarily on that diagnosis. Nicotine dependence was the next most common diagnosis, with approximately two-thirds of individuals receiving that diagnosis, followed by alcohol dependence and opioid dependence, each with a prevalence of about 45%, and cannabis dependence, occurring in just over one-quarter of the sample. Major depressive episode (MDE) was the most common psychiatric disorder, with a prevalence of nearly 15%, followed by antisocial personality disorder (ASPD), which was diagnosed in about 12% of individuals. The frequency of posttraumatic stress disorder (PTSD) and compulsive gambling each approached 10%.

There was no sex difference in the prevalence of CD or nicotine dependence, or the less commonly diagnosed sedative dependence and stimulant (other than cocaine) dependence. However, men were significantly more likely to receive diagnoses of alcohol dependence, opioid dependence, and other substance (i.e., PCP, hallucinogens, inhalants, solvents, or combinations such as “speedballs”) dependence, as well as ASPD and compulsive gambling. In contrast, women were more likely to receive a diagnosis of MDE, PTSD, panic disorder, or agoraphobia.

## 2.2. Procedure

Small nuclear families were recruited for participation in genomewide linkage studies of CD and opioid dependence. All subjects gave written, informed consent to participate, using procedures approved by the institutional review board at each participating site. Inclusion required the participation of a sibling pair, both members of which were affected with a lifetime diagnosis of either CD, opioid dependence, or both. Additional family members, including parents and other siblings (irrespective of their having a lifetime substance dependence diagnosis), were also invited to participate. Recruitment was conducted at substance abuse treatment programs and through clinical referrals and advertisements in local media in Connecticut (through sites at Yale University and the University of Connecticut); Boston, MA (McLean Hospital) and Charleston, SC (Medical University of South Carolina). Data were submitted electronically to a database at Boston University, where analysis was conducted.

## 2.3. Assessments

Phenotypic information was obtained through administration of the SSADDA (Pierucci-Lagha et al., 2005), which includes a separate section for the diagnosis of cocaine dependence. Questions from the Addiction Severity Index (McLellan et al., 1992) were added to the SSADDA, to allow the estimation of lifetime measures of substance (including cocaine) use.

The SSADDA was formatted for computer-assisted administration, incorporating automated skip patterns and logical data-entry checks. The reliability of the SSADDA has been shown to be good-to-excellent for all of the major substance dependence diagnoses (Pierucci-Lagha et al., 2005). Specifically, the test-retest and inter-rater reliabilities for CD dependence diagnosed using criteria from the Diagnostic and Statistical Manual of Mental Disorders, 4<sup>th</sup> edition (DSM-IV; American Psychiatric Association, 1994) were 0.92 and 0.83, respectively. Items ( $n = 68$ ) from the cocaine use disorder section of the SSADDA were used to generate the clusters. These included age of onset and frequency and intensity of cocaine use, route of cocaine administration, occurrence of psychosocial and medical consequences of cocaine use, quit attempts, and cocaine abuse treatment sought and received.

## 2.4. Data Analysis

All items from the cocaine use disorders section of the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA; Pierucci-Lagha et al., 2005) were used initially in the development of subtypes. Multiple correspondence analysis (MCA) (Greenacre, 1984; Lebart et al., 1984), a non-parametric data reduction method, was used to identify the underlying dimensions in the data from the CD section of the SSADDA (as briefly described previously in Gelernter et al., 2005). Binary symptoms included cocaine use characteristics, cocaine-related effects, and cocaine treatment history. Each study participant with phenotypic data was assigned a score on each of the retained dimensions (data not shown). An iterative k-means partitioning, using nearest centroid sorting (Hirano et al. 2003), was then conducted using several different starting points and a larger than expected number of clusters ( $k=50$ ). Starting with a number of clusters that is far higher than a number we would accept as a final solution allows the identification of stable small groups. The next stage involved cross-classification of the results of the k-means clustering and retention of the groups that consistently clustered together. These groups and the remaining observations were then used in an agglomerative hierarchical clustering, with Ward's method (Ward 1963) serving to identify the final cluster structure. Ward's method avoids the idiosyncrasies that can occur with different starting seeds for k-means clustering by identifying intact groups for the hierarchical clustering that are joined prior to the agglomerative process (Hirano et al. 2003). It thereby retains the strengths of both types of clustering while mitigating the weaknesses. The selection of the final subgroup structure was based on a comparison of the within-to-between-group variation on items used to form the groups and group profiles on other variables. We considered solutions with larger numbers of clusters, though with more than six clusters, single individuals or small groups of outliers emerged, making it impossible to interpret the solution.

Binary logistic regression was used to estimate the probability of cluster membership for each study participant in each of the clusters. Variables selected for clustering were used in the estimation of the probability of cluster membership. Each study participant was assigned a score on each of the retained dimensions using a procedure similar to the assignment of factor scores. The natural logarithm of the probability of membership in each group was the dependent measure in the quantitative trait analyses. Chi-square analysis was used to compare clusters on categorical measures and analysis of variance (ANOVA) was used to compare clusters on continuous measures.

SPAD software (DECISIA: Paris, France; <http://eng.spad.eu>) was used for both the MCA and the clustering algorithms [see Ambrogi et al. (2005) for an alternate, compatible approach]. SAS software (SAS Institute, 2001) was used for subsequent analyses, including cluster profiling. Heritability of the log of probability of group membership was computed using Sequential Oligogenic Linkage Analysis Routines (SOLAR; Almasy and Blangero, 1998), with sex as a covariate. We used linkage analysis methods to identify regions of chromosomes that harbor genes influencing risk for CD; those are beyond the scope of the present report and are described in detail in Gelernter et al. (2005).

## 3. RESULTS

### 3.1. Overview of Cluster Analysis Results

As shown in Table 2, cluster analysis yielded a 6-cluster solution, with the following distribution of individuals (% of the total) across the 6 clusters: Cluster 1 ["Heavy, Cocaine Use Predominant,"  $N = 336$  (24.1%)], Cluster 2 ["Heavy, Mixed Drug Injector,"  $N = 303$  (21.8%)], Cluster 3 ["Heavy Cocaine Use, Later Onset,"  $N = 350$  (25.1%)], Cluster 4 ["Moderate Cocaine and Opioid Use,"  $N = 258$  (18.5%)], Cluster 5 ["Low Cocaine and Opioid Abuse,"  $N = 104$  (7.5%)], and Cluster 6 ["Opioid Abuse,"  $N = 42$  (3%)]. Comparison of the

clusters on demographic features showed significant differences among clusters on age, sex, race, and marital status. Clusters 1 and 3 contain predominantly AA women who were never married. Cluster 2 has the oldest subjects, contains the fewest women, and, together with Clusters 5 and 6, the fewest AAs. Clusters 2 and 6 also contain the largest percentage of individuals who were divorced, separated, or widowed. Although Clusters 4, 5, and 6 are all roughly evenly divided by sex, with the majority of individuals in each having never married, these clusters differ substantially from one another on race, with Cluster 5 containing the largest proportion of Hispanics.

### 3.2. Features of Cocaine Use Behavior by Cluster

The clusters were clearly differentiated on cocaine use characteristics, cocaine-related effects, and on cocaine treatment history (Table 3).

**3.2.1. Heavy Cocaine Use Clusters**—Consistent with the prevalence of CD in the clusters, Clusters 1, 2, and 3 each included a high percentage of individuals who reported having used cocaine daily or almost daily (96.2%, 96.4% and 92.0%, respectively). However, members of Clusters 1 and 2 reported earlier ages of initial cocaine use and of heaviest use than members of Cluster 3. Members of Cluster 1 were more likely than those of Clusters 2 or 3 to endorse having gotten higher and stayed higher longer than others when they first started to use cocaine. With respect to other measures of cocaine-related effects, including adverse ones, Clusters 1 and 2 were comparable to one another and generally higher than all other clusters, including Cluster 3. Subjects in Clusters 1 and 2 were more likely than those in Cluster 3 to have received formal treatment for cocaine abuse, but comparable proportions of the three clusters reported ever having attended a self-help group meeting due to cocaine abuse.

Cluster 2 members were predominantly intravenous cocaine users who progressed to their periods of heaviest cocaine use sooner after initiating cocaine use than did the members of Clusters 1 or 3 (6.8 vs. 8.3 years for both Clusters 1 and 3). The high rate of intravenous drug use is consistent with the high rate of opioid dependence in this cluster.

As shown in Table 3, members of Cluster 3 were more likely than those in Clusters 4, 5, or 6 to endorse a variety of cocaine use characteristics and cocaine-related effects. However, Cluster 3 individuals reported a later age of initial cocaine use than did Cluster 4 members, and their heaviest use occurred later than it did for members of Clusters 4, 5, or 6.

**3.2.2. Low-to-Moderate Cocaine Use Clusters**—Although more than 90% of Cluster 4 members endorsed having used cocaine daily or almost daily, the prevalence of cocaine dependence was lower in this cluster than for the heavy cocaine use clusters. However, Cluster 4 had a higher prevalence of opioid dependence and a higher rate of intravenous cocaine use than did Clusters 1 or 3. More than one-third of Cluster 5 members met criteria for CD, though fewer than 3% endorsed daily or almost daily use of cocaine. However, more than 40% of this cluster met criteria for opioid dependence and nearly 20% endorsed intravenous cocaine use. Both Clusters 4 and 5 contained substantial proportions of individuals who endorsed cocaine-related effects and cocaine treatment histories.

**3.2.3. Occasional or No Cocaine Use Cluster**—Members of Cluster 6 reported the latest onset of both cocaine use and heavy cocaine use among the clusters. Cluster 6 also consistently had the lowest proportion of members who endorsed cocaine-related effects and cocaine treatment. However, this cluster had the second-highest prevalence of opioid dependence.

### 3.3. Substance Dependence and Psychiatric Disorders by Cluster

As can be seen in Table 4, more than 99% of individuals in Clusters 1-3 met lifetime DSM-IV diagnostic criteria for CD, a significantly higher percentage than that in Clusters 4, 5, or 6 (78%, 38%, and 7%, respectively, all significantly different from one another). Based on these findings and other characteristics described below, we identified these groups as Heavy Cocaine Use (Clusters 1-3), Low-to-Moderate Cocaine Use (Clusters 4 and 5), and Occasional or No Cocaine Use (Cluster 6) clusters.

Significant differences exist among the clusters on the lifetime prevalence of all other substance dependence diagnoses (Table 4). The Heavy Cocaine Use clusters showed a higher prevalence of nearly all substance dependence diagnoses than the other three clusters (which did not differ consistently from one another). Of the Heavy Cocaine Use clusters, Cluster 2 showed the highest prevalence of all categories of substance dependence, particularly opioid dependence; Clusters 4, 5, and 6 also showed a high prevalence of opioid dependence, despite having a lower prevalence of the other substance dependence diagnoses. The prevalence of cannabis dependence was highest in Clusters 1 and 2, with Clusters 3 and 4 being intermediate, and Clusters 5 and 6 having the lowest prevalence of this disorder.

Subjects in Clusters 1 and 2 also had the highest prevalence of major depressive episode (MDE), antisocial personality disorder (ASPD), and posttraumatic stress disorder (PTSD), with no consistent pattern of differences among the other clusters on these disorders. Differences in the prevalence of compulsive gambling, panic disorder, and agoraphobia showed a similar pattern, but were less pronounced, possibly due to the lower overall prevalence of these disorders.

Although not a co-morbid disorder, cocaine-induced paranoia (CIP) is a psychopathologic feature that occurs commonly among individuals with CD (Satel et al., 1991a; Brady et al., 1991; Rosse et al., 1995; Bartlett et al., 1997; Cubells et al., 2005). There is evidence that CIP is genetically influenced (Gelernter et al., 1994; Cubells et al., 2000). The prevalence of CIP varied significantly across clusters in a pattern similar to the co-morbid substance use and psychiatric disorders (Cluster 1: 87.5%, Cluster 2: 78.2%, Cluster 3: 52.9%, Cluster 4: 38.4%, Cluster 5: 35.6%, Cluster 6: 4.8%) [ $\chi^2_{(5)} = 295.8, p < 0.001$ ].

### 3.4. Heritability of the Clusters

All of the clusters showed significant heritability, with the estimate for each of the first four clusters exceeding 30% (Table 5). Although the size of the clusters was not fully explained by their size, the two smallest clusters (Clusters 5 and 6) showed the lowest heritability estimates.

### 3.5. Linkage Analysis of the Clusters

As reported previously, we (Gelernter et al., 2005) conducted linkage analysis on a sub-sample of 986 individuals from the present study sample using CD diagnosis, CIP, and cluster membership as phenotypes. Interestingly, of the phenotypes examined (which included DSM-IV CD, CIP, and cluster membership), the strongest linkage results were observed for cluster membership. These findings included a lod score of 4.66 for membership in Cluster 1 on chromosome 12 (in EAs only) and a lod score of 3.35 for membership in Cluster 4 on chromosome 18 (in AAs only). In addition, there was suggestive evidence of linkage for Cluster 1 on chromosome 3 (in EAs only), for Clusters 1 or 3 on chromosome 6 (in the total sample), and for Cluster 3 on chromosome 2 (in AAs only).

## 4. DISCUSSION

DSM-IV CD may not represent the optimal CD-related phenotype for genetic mapping; other phenotypes might identify more genetically homogeneous sets of subjects. Based on demographic characteristics and cocaine use histories, we used a cluster analytic approach to identify phenomenologically distinct CD clusters. Subjects in Clusters 1, 2, and 3 were characterized by a history of heavy cocaine use, a high degree of CD severity (measured by the vast majority of individuals in these clusters having endorsed enough cocaine dependence symptoms to receive a diagnosis of CD). A preponderance of individuals in these clusters also reported having participated in both self-help recovery programs and formal treatment for cocaine abuse. In addition, although these measures were not included as variables in the cluster analysis, members of the heavy cocaine use clusters (particularly those in Clusters 1 and 2) were substantially more likely to meet diagnostic criteria for a variety of other substance dependence and co-morbid psychiatric disorders.

Cluster analysis, which is used to classify a set of observations into two or more mutually exclusive groups based on a set of interval measures, results in a solution in which members of the groups share properties in common (Stockburger, 1998). Cluster analysis allows many choices about the nature of the algorithm for combining groups and, although it will always produce a grouping, the structure of the resultant solution will vary with the method employed, the cases included in the analysis, and the quality of the data used in the analysis. To address these concerns, we applied a multi-staged, iterative approach to identify cluster membership, which effectively repeated the clustering to achieve stable groups. The advantages of the modified strategy are that the clustering is repeated with several starting points and two different methods are used to obtain stable clusters, which should increase the generalizability of the cluster results. In addition, careful quality assurance procedures were employed in data collection using the SSADDA (Pierucci-Lagha et al., 2005), which is likely to have reduced variance attributable to poor reliability of the assessment. These factors may help to explain why the six-cluster solution obtained here differs from the two-cluster solution obtained previously in samples of cocaine- and other drug-dependent individuals analyzed using a simple k-means clustering procedure (Ball et al., 1995; Feingold et al., 1996; Ball et al., 1998; Basu et al., 2004). In addition to the different analytic approach, the current study sample differs from the clinical samples recruited in prior studies, in that the family-based recruitment approach led to the inclusion of a substantial number of individuals without drug dependence diagnoses.

The validity of the cluster solution was supported by its capacity to discriminate between variables that were not used to create the grouping (Stockburger, 1998). That is, the clusters differed significantly on demographic features, but also particularly on prevalence rates of co-morbid substance use and psychiatric disorders. The differences among the clusters were further confirmed by estimates of the heritability of the trait embodied in cluster membership, which was greatest for the heavy cocaine use clusters. Because the sample was ascertained using an affected sib pair strategy, the heritability estimates may be inflated; consequently, they were treated as directional, and were used to select traits to examine with a genomewide linkage analysis (Gelernter et al., 2005). Linkage analysis yielding genomewide significant findings in a sub-sample of individuals from the sample described herein using cluster membership as phenotypes provides further validation of the subtyping procedure.

These findings lend support to the use of a data reduction and cluster-analytic strategy to provide more homogeneous subgroups of CD subjects for linkage analysis. Because the choice and implementation of the clustering algorithm, as well as data quality and the stability of the underlying groups can influence the likelihood of replicating the results of cluster analyses, we applied a multi-staged approach to identify cluster membership. The advantages of this

strategy are that the clustering is repeated with several starting points and two different methods are used to obtain stable clusters, increasing the generalizability of the cluster results.

This novel application of clinical subtyping to a molecular genetic investigation of drug dependence is consistent with prior efforts to explain the phenotypic variability among alcoholics (Cloninger et al., 1981; Babor et al., 1992) and drug abusers (Ball et al., 1995; Feingold et al., 1996; Ball et al., 1998; Basu et al., 2004). The approach satisfied a stringent criterion for validity of the cluster structure by demonstrating its utility as a phenotype for linkage analysis. Further analysis of the subtypes provided evidence of their concurrent validity. By reducing genetic heterogeneity, this subtyping approach represents a powerful and useful refinement for the diagnosis of CD. Replication of the findings reported here in a second sample of individuals with CD would add further to evidence of the validity and utility of this approach.

#### Acknowledgements

This work was supported by NIDA grants R01 DA12849, R01 DA12690, K24 DA15105, and K02 DA00326, and NIAAA grant K24 AA13736. We appreciate the efforts of the following SSADDA interviewers: Michelle Slivinsky, Michelle McKain, Deborah Pearson, Kevin Young (at Univ. CT); Alisha Pollastri, Yari Nunez, Matthew Madura, and Melyssa Pokrywa (at Yale); Victoria deMenil and Catherine Cogley (at McLean); and Heather Remy (at MUSC). Jennifer Blesso and John Farrell provided excellent database support. Alisa Manning, Deborah Cebrik, and Carolien Panhuysen assisted in statistical analyses and data management.

#### References

- Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *American Journal of Human Genetics* 1998;62:1198–1211. [PubMed: 9545414]
- Ambrogio F, Biganzoli E, Boracchi P. Multiple correspondence analysis in S-PLUS. *Computer Methods and Programs in Biomedicine* 2005;79:161–167. [PubMed: 15975690]
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4. Washington DC: American Psychiatric Press; 1994.
- Babor, TF.; Dolinsky, ZS. Alcoholic typologies: historical evaluation and empirical evaluation of some common classifications schemes. In: Rose, RM.; Barrett, J., editors. *Alcoholism: origins and outcome*. Raven Press; New York: 1988. p. 245-266.
- Babor TF, Hofmann M, DelBoca FK, Hesselbrock V, Meyer RE, Dolinsky ZS, Rounsaville B. Types of alcoholics: I. Evidence for an empirically-derived typology based on indicators of vulnerability and severity. *Archives of General Psychiatry* 1992;8:599–608. [PubMed: 1637250]
- Ball SA, Carroll KM, Babor TF, Rounsaville BJ. Subtypes of cocaine abusers: support for a Type A-Type B distinction. *Journal of Consulting and Clinical Psychology* 1995;63:115–124. [PubMed: 7896976]
- Ball SA, Kranzler HR, Tennen H, Poling JC, Rounsaville BJ, Rounsaville BJ. Personality disorders and dimension differences between type A and type B substance abusers. *Journal of Personality Disorders* 1998;12:1–12. [PubMed: 9573515]
- Bartlett E, Hallin A, Chapman B, Angrist B. Selective sensitization to the psychosis-inducing effects of cocaine: a possible marker for addiction relapse vulnerability? *Neuropsychopharmacology* 1997;16:77–82. [PubMed: 8981391]
- Basu D, Ball SA, Feinn R, Gelernter J, Kranzler HR. Typologies of drug dependence: comparative validity of a multivariate and four univariate models. *Drug and Alcohol Dependence* 2004;73:289–300. [PubMed: 15036551]
- Brady KT, Lydiard RB, Malcolm R, Ballenger JC. Cocaine-induced psychosis. *The Journal of Clinical Psychiatry* 1991;52:509–511. [PubMed: 1752853]
- Cloninger CR, Bohman M, Sigvardsson S. Inheritance of alcohol abuse: Cross-fostering analysis of adopted men. *Archives of General Psychiatry* 1981;38:861–868. [PubMed: 7259422]
- Cohen E. An exploratory attempt to distinguish subgroups among crack-abusing African-American women. *Journal of Addictive Diseases* 1999;18:41–54. [PubMed: 10507581]



- Craig RJ, Olson R. MMPI subtypes for cocaine abusers. *The American Journal of Drug and Alcohol Abuse* 1992;18:197–205. [PubMed: 1562015]
- Cubells JF, Feinn R, Pearson D, Burda J, Tang Y, Farrer LA, Gelernter J, Kranzler HR. Rating the severity and character of transient cocaine-induced delusions and hallucinations with a new instrument, the Scale for Assessment of Positive Symptoms for Cocaine-Induced Psychosis (SAPS-CIP). *Drug and Alcohol Dependence* 2005;80:23–33. [PubMed: 15894433]
- Cubells JF, Kranzler HR, McCance-Katz E, Anderson GM, Malison RT, Price LH, Gelernter J. A haplotype at the DBH locus, associated with low plasma dopamine - hydroxylase activity, also associates with cocaine-induced paranoia. *Molecular Psychiatry* 2000;5:56–63. [PubMed: 10673769]
- Epstein, EE. Classification of alcohol-related problems and dependence. In: Heather, N.; Peters, T.J.; Stockwell, TR., editors. *Handbook of alcohol dependence and related problems*. John Wiley & Sons; London: 2001. p. 47-70.
- Feingold A, Ball SA, Kranzler HR, Rounsaville BJ. Generalizability of the Type A/Type B distinction across different psychoactive substances. *The American Journal of Drug and Alcohol Abuse* 1996;22:449–462. [PubMed: 8841691]
- Gelernter J, Kranzler HR, Satel SL, Rao PA. Genetic association between dopamine transporter protein alleles and cocaine-induced paranoia. *Neuropsychopharmacology* 1994;11:195–200. [PubMed: 7865100]
- Gelernter J, Panhuysen C, Weiss R, Brady K, Hesselbrock V, Rounsaville B, Poling J, Wilcox M, Farrer L, Kranzler HR. Genomewide linkage scan for cocaine dependence and related traits: significant linkages for a cocaine-related trait and cocaine-induced paranoia. *The American Journal of Medical Genetics Part B, Neuropsychiatric Genetics* 2005;136:45–52.
- Greenacre, M. *Theory and applications of correspondence analysis*. New York: Academic Press; 1984.
- Hirano S, Sun X, Tsumoto S. Comparison of clustering methods for clinical databases. *Information Sciences* 2004;159:155–165.
- Karch SB. Cocaine cardiovascular toxicity. *Southern Medical Journal* 2005;98:794–799. [PubMed: 16144174]
- Kendler KS, Prescott CA. Cocaine use, abuse and dependence in a population-based sample of female twins. *The British Journal of Psychiatry* 1998;173:345–350. [PubMed: 9926041]
- Kendler KS, Jacobson KC, Prescott CA, Neale MC. Specificity of genetic and environmental risk factors for use and abuse/dependence of cannabis, cocaine, hallucinogens, sedatives, stimulants, and opiates in male twins. *The American Journal of Psychiatry* 2003;160:687–695. [PubMed: 12668357]
- Kosten, TR.; Ciraulo, DA. Cocaine and psychostimulants. In: Kranzler, HR.; Ciraulo, DA., editors. *Clinical manual of addiction psychopharmacology*. American Psychiatric Publishing, Inc; Washington, DC: 2005. p. 183-209.
- Lebart, L.; Morineau, A.; Warwick, KM. *Multivariate descriptive statistical analysis: Correspondence analysis and related techniques for large matrices*. New York: John Wiley; 1984.
- McLellan AT, Kushner H, Metzger D, Peters R, Smith I, Grissom G, Pettinati H, Argeriou M. The Fifth Edition of the Addiction Severity Index. *Journal of Substance Abuse Treatment* 1992;9:199–213. [PubMed: 1334156]
- Nnadi CU, Mimiko OA, McCurtis HL, Cadet JL. Neuropsychiatric effects of cocaine use disorders. *Journal of the National Medical Association* 2005;97:1504–1515. [PubMed: 16334497]
- Pierucci-Lagha A, Gelernter J, Feinn R, Cubells J, Pearson D, Farrer L, Kranzler HR. Diagnostic reliability of the Semi-structured Assessment of Drug Dependence and Alcoholism (SSADDA). *Drug and Alcohol Dependence* 2005;80:303–312. [PubMed: 15896927]
- Roehrich H, Gold MS. Familial addiction in cocaine abusers. *The Journal of Clinical Psychiatry* 1988;49:326. [PubMed: 3410837]
- Rosse RR, Alim TN, Johri SK, Hess AL, Deutsch SI. Anxiety and pupil reactivity in cocaine dependent subjects endorsing cocaine-induced paranoia: preliminary report. *Addiction* 1995;90:981–984. [PubMed: 7663321]
- Rounsaville BJ, Anton SF, Carroll K, Budde D, Prusoff BA, Gawin F. Psychiatric diagnosis of treatment-seeking cocaine abusers. *Archives of General Psychiatry* 1991;48:43–51. [PubMed: 1984761]
- SAS Institute (2001) SAS release 8.02, Cary, NC

- Satel SL, Southwick S, Gawin F. Clinical features of cocaine-induced paranoia. *The American Journal of Psychiatry* 1991a;148:495–499. [PubMed: 2006696]
- Stockburger, DW. Cluster Analysis. *Multivariate statistics: Concepts, models, and applications*. WWW Version 1.0. 1998. <http://www.psychstat.smsu.edu/multibook/mlt04.htm>. revised March
- Substance Abuse and Mental Health Services Administration, Office of Applied Studies. Drug Abuse Warning Network. Interim National Estimates of Drug-Related Emergency Department Visits. 2003 [on December 4 2005]. Dawn Series D-26, DHHS Publication No. (SMA) 04-3972 Rockville, MD, 2004. Accessed at <http://dawninfo.samhsa.gov/>
- Tsuang MT, Lyons MJ, Eisen SA, Goldberg J, True W, Lin N, Meyer JM, Toomey R, Faraone SV, Eaves L. Genetic influences on DSM-III-R drug abuse and dependence: A study of 3,372 twin pairs. *American Journal of Medical Genetics Part B, Neuropsychiatric Genetics* 1996;67:473–477.
- Uhl, GR.; Elmer, GL.; LaBuda, MC.; Pickens, RW. Genetic influences in drug abuse. In: Bloom, FE.; Kupfer, DJ., editors. *Psychopharmacology: The fourth generation of progress*. New York: Raven Press; 1995. p. 1793-1806.
- Ward JH. Hierarchical groupings to optimize an objective function. *Journal of the American Statistical Association* 1963;58:234–244.
- Weiss RD, Mirin SM. Subtypes of cocaine abusers. *The Psychiatric Clinics of North America* 1986;9:491–501. [PubMed: 3774602]
- Wolff AJ, O'Donnell AE. Pulmonary effects of illicit drug use. *Clinics in Chest Medicine* 2004;25:203–216. [PubMed: 15062611]

Table 1

Lifetime Prevalence of Substance Use and Psychiatric Disorders by Sex [N (%)]<sup>1</sup>

Diagnosis	Total (N = 1393)	Male (N = 687)	Female (N = 706)	$\chi^2$ (1 df)	P-value
<b>Substance Use Disorders</b>					
Cocaine Dependence	1229 (88.2)	606 (87.8)	623 (88.2)	0.21	0.65
Nicotine Dependence	948 (68.1)	458 (66.7)	490 (69.4)	1.77	0.18
Alcohol Dependence	628 (45.1)	339 (49.3)	289 (41.0)	9.05	0.003
Opioid Dependence	626 (44.9)	347 (50.5)	279 (39.5)	15.82	<0.001
Cannabis Dependence	383 (28.3)	240 (34.9)	143 (20.3)	39.03	<0.001
Sedative Dependence	99 (7.1)	57 (8.3)	42 (5.9)	2.76	0.096
Stimulant Dependence	97 (7.0)	56 (8.2)	41 (5.8)	2.80	0.094
Other Substance Dependence	237 (17.0)	139 (20.2)	98 (13.9)	9.49	0.002
<b>Psychiatric Disorders</b>					
MDE	198 (14.6)	69 (10.0)	129 (18.8)	18.98	<0.001
ASPD	173 (12.4)	126 (18.3)	47 (6.8)	45.93	<0.001
PTSD	133 (9.6)	44 (6.4)	89 (13.1)	16.66	<0.001
Compulsive Gambling	127 (9.1)	93 (13.5)	34 (4.9)	32.5	<0.001
Panic Disorder	93 (6.7)	33 (4.8)	60 (8.8)	8.29	0.004
Agoraphobia	59 (4.2)	14 (2.0)	45 (6.7)	17.27	<0.001

<sup>1</sup> In some cases, percentage is increased due to missing data

Other Substance Dependence includes dependence on phenylcyclidine, hallucinogens, inhalants, solvents, or a combination of opioids and cocaine (i.e., “speedball’s”); MDE: Major Depressive Episode; ASPD: Antisocial Personality Disorder; PTSD: Posttraumatic Stress Disorder

Table 2

Demographic Characteristics

	Cluster 1(N = 336)	Cluster 2(N = 303)	Cluster 3(N = 350)	Cluster 4(N = 258)	Cluster 5(N = 104)	Cluster 6(N = 42)	Test Statistic
<b>Mean age in yr (SD)</b>	37.1 (7.7)	40.6 (7.4)	38.7 (6.6)	38.9 (8.7)	37.0 (9.6)	37.4 (10.0)	$F_{(5,1387)} = 7.45^{***}$
<b>Sex [N (%)]</b>							$\chi^2_{(5)} = 16.8^{**}$
Female	192 (57.1)	126 (41.6)	186 (53.1)	127 (49.2)	54 (51.9)	21 (50.0)	
Male	144 (42.9)	177 (58.4)	164 (46.9)	131 (50.8)	50 (48.1)	21 (50.0)	
<b>Race [N (%)]</b>							$\chi^2_{(15)} = 117.5^{***}$
AA	185 (55.1)	96 (31.7)	233 (66.6)	134 (51.9)	40 (38.5)	11 (26.2)	
EA	99 (29.5)	149 (49.2)	78 (22.3)	69 (26.7)	36 (34.6)	21 (50.0)	
Hispanic	40 (11.9)	46 (15.2)	26 (7.4)	36 (14.0)	22 (21.2)	6 (14.3)	
Other	12 (3.6)	12 (4.0)	13 (3.7)	19 (7.4)	6 (5.8)	4 (9.5)	
<b>Education [N (%)]</b>							$\chi^2_{(15)} = 24.8$
no HS	26 (7.7)	27 (8.9)	19 (5.4)	21 (8.1)	5 (4.8)	3 (7.1)	
some HS	143 (42.6)	121 (39.9)	136 (38.9)	109 (42.3)	27 (26.0)	18 (42.9)	
HS graduate	93 (27.7)	86 (28.4)	104 (29.7)	86 (33.3)	43 (41.4)	14 (33.3)	
> HS	74 (22.0)	69 (22.8)	91 (26.0)	42 (16.3)	29 (27.9)	7 (16.7)	
<b>Marital Status [N (%)]</b>							$\chi^2_{(10)} = 36.2^{***}$
Never married	214 (63.7)	148 (48.8)	204 (58.3)	162 (62.8)	66 (63.5)	23 (54.8)	
Married	59 (17.6)	44 (14.5)	52 (14.9)	27 (10.5)	15 (14.4)	3 (7.1)	
Div/Sep/Wid	63 (18.8)	111 (36.6)	94 (26.9)	69 (26.7)	23 (22.1)	16 (38.1)	

AA: African American, EA: European American, HS: high school, Div/Sep/Wid: Divorced, Separated, or Widowed

\*\* p < 0.01,

\*\*\* p < 0.001

Table 3  
Cocaine Use Characteristics, Cocaine-Related Effects, and Cocaine Treatment History by Cluster

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Test Statistic
<b>Cocaine Use Characteristics</b>							
Age of first cocaine use <sup>1</sup>	20.4 (5.8)	19.2 (5.5)	22.5 (5.7)	21.3 (6.1)	23.5 (7.7)	23.8 (8.2)	F=15.6 <sup>**</sup>
Age of heaviest cocaine use <sup>1</sup>	27.1 (7.3)	27.7 (8.2)	29.3 (7.1)	27.7 (8.3)	27.4 (8.4)	28.0 (11.2)	F=3.2 <sup>*</sup>
Used cocaine daily or almost daily	96.1%	96.4%	92.0%	93.4%	2.9%	2.4%	$\chi^2=1881.8$ <sup>**</sup>
Injected cocaine intravenously	0	100%	5.4%	36.4%	19.2%	4.8%	$\chi^2=2055.5$ <sup>**</sup>
<b>Cocaine-Related Effects</b>							
Got higher and stayed higher longer than others	52.4%	30.5%	40.0%	30.6%	28.9%	4.8%	$\chi^2=1154.4$ <sup>**</sup>
Desire for cocaine made it hard to think of anything else	91.4%	83.2%	69.4%	36.1%	15.4%	2.4%	$\chi^2=1477.5$ <sup>**</sup>
Cocaine interfered with work, school, or home life	88.1%	79.8%	49.4%	8.9%	3.9%	4.8%	$\chi^2=752.1$ <sup>**</sup>
Family, friend, or doctor objected to cocaine use	91.7%	86.1%	79.7%	39.5%	26.9%	2.4%	$\chi^2=1456.0$ <sup>**</sup>
Cocaine-related legal problems	61.3%	63.0%	46.3%	25.2%	9.6%	2.4%	$\chi^2=1275.9$ <sup>**</sup>
Gave up social, recreational, or work activities due to cocaine use	94.6%	93.1%	82.6%	35.7%	25.0%	2.4%	$\chi^2=1588.2$ <sup>**</sup>
<b>Cocaine Treatment History</b>							
Ever treated for cocaine abuse	82.4%	76.2%	68.3%	38.4%	18.5%	2.4%	$\chi^2=1346.9$ <sup>**</sup>
Ever attended a self-help group meeting due to cocaine abuse	70.8%	69.0%	65.9%	29.8%	15.5%	0	$\chi^2=1312.4$ <sup>**</sup>

<sup>1</sup> Values are mean (SD)

\* p&lt;.01;

\*\* p&lt;.001

**Table 4**  
Lifetime Prevalence of Substance Use and Psychiatric Disorders by Cluster [N (%)]<sup>1</sup>

Diagnosis	Cluster 1 (N = 336)	Cluster 2 (N = 303)	Cluster 3 (N = 350)	Cluster 4 (N = 258)	Cluster 5 (N = 104)	Cluster 6 (N = 42)	$\chi^2$ (5 df)
<b>Substance Dependence</b>							
Cocaine Dependence	336 (100)	300 (99.0)	348 (99.4)	202 (78.3)	40 (38.5)	3 (7.1)	659.35***
Nicotine Dependence	253 (75.3)	234 (77.2)	241 (68.9)	144 (55.8)	52 (50.0)	24 (57.1)	55.61***
Alcohol Dependence	204 (60.7)	179 (59.1)	140 (40.0)	66 (25.6)	28 (26.9)	11 (26.2)	120.31***
Opioid Dependence	101 (30.1)	247 (81.5)	76 (21.7)	130 (50.4)	44 (42.3)	28 (66.7)	281.60***
Cannabis Dependence	142 (43.8)	91 (31.3)	79 (23.0)	50 (19.8)	16 (16.0)	5 (12.2)	66.56***
Sedative Dependence	30 (8.9)	50 (16.5)	7 (2.0)	9 (3.5)	2 (1.9)	1 (2.4)	66.80***
Stimulant Dependence	27 (8.0)	47 (15.5)	13 (3.7)	6 (2.3)	2 (1.9)	2 (4.8)	53.43***
Other SD	41 (12.2)	119 (39.3)	28 (8.0)	40 (15.5)	8 (7.7)	1 (2.4)	145.18*
<b>Psychiatric Disorders</b>							
MDE	65 (20.1)	64 (22.2)	35 (10.1)	17 (6.7)	12 (12.1)	5 (11.9)	40.14***
ASPD	58 (17.9)	44 (15.3)	32 (9.2)	31 (12.3)	2 (2.0)	6 (14.3)	23.88***
PTSD	49 (14.9)	33 (11.5)	27 (7.8)	16 (6.3)	5 (5.0)	3 (7.3)	18.94**
Compulsive Gambling	37 (11.3)	35 (12.3)	31 (9.0)	19 (7.5)	4 (4.0)	1 (2.4)	11.43*
Panic Disorder	26 (7.9)	31 (10.8)	17 (4.9)	11 (4.3)	6 (5.9)	2 (4.8)	12.66*
Agoraphobia	24 (7.5)	13 (4.5)	13 (3.8)	7 (2.7)	2 (2.0)	0 (0.0)	12.58*

<sup>1</sup>In some cases, percentage is increased due to missing data;

\* p<0.05

\*\* p<0.01

\*\*\* p<0.001

Other SD (Substance Dependence): dependence on phencyclidine, hallucinogens, inhalants, solvents, or opioids and cocaine (i.e., "speedballs"); MDE: Major Depressive Episode; ASPD: Antisocial Personality Disorder; PTSD: Posttraumatic Stress Disorder

**Table 5**

## Heritability Estimates for the Clusters

Cluster Number and Description	Heritability ( $h^2$ )	P-Value
1. Heavy, Cocaine Predominant	0.387	0.0000060
2. Heavy, Mixed Drug Injectors	0.438	0.0000006
3. Heavy, Later Onset	0.501	0.0000000
4. Moderate Cocaine and Opioid Abuse	0.320	0.0002900
5. Low Cocaine and Opioid Abuse	0.238	0.0025000
6. Opioid Abuse	0.145	0.0398000