Published in final edited form as:
Addiction. 2011 June ; 106(6): 1146–1154. doi:10.1111/j.1360-0443.2011.03390.x.

Empirically Derived Subtypes of Opioid Use and Related Behaviors

Grace Chan[1], **Joel Gelernter**[2], **David Oslin**[3], **Lindsay Farrer**[4], and **Henry R. Kranzler**[3]

[1]Department of Psychiatry, University of Connecticut School of Medicine, Farmington, CT, USA

[2]Departments of Psychiatry, Genetics, and Neurobiology, Yale University School of Medicine, New Haven, CT and VA CT Healthcare Center, West Haven, CT, USA

[3]Department of Psychiatry, University of Pennsylvania School of Medicine and the Philadelphia VAMC, Philadelphia, PA, USA

[4]Departments of Medicine (Genetics Program), Neurology, Genetics & Genomics, Epidemiology, and Biostatistics, Boston University Schools of Medicine and Public Health, Boston, MA, USA

Abstract

Aims—To identify and validate homogeneous subtypes of opioid use and related behaviors.

Design—Family-based and case-control genetic studies of opioid and/or cocaine dependence.

Settings—Clinical and general community samples from Connecticut, Massachusetts, Pennsylvania, and South Carolina.

Participants—4,061 individuals (2,003 individuals from 835 families and 2,058 unrelated individuals) recruited to participate in genetic studies.

Measurements—The computer-assisted Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA) was used to assess participants' demographics, medical history, substance use behaviors, and other psychiatric disorders.

Findings—Five homogeneous subtypes were identified, which differed on opioid-related measures, demographics, and prevalence rates of co-morbid substance use and psychiatric disorders. Heritability estimates for the two most severely affected subtypes exceeded 0.60.

Conclusions—An empirical approach based on opioid use and related behaviors can yield homogeneous subtypes that could be of value in gene finding for opioid dependence.

Keywords

Opioid dependence; Subtypes; Phenotype; Multiple correspondence analysis; *K*-means clustering; Hierarchical clustering; Heritability

Correspondence to: Henry R. Kranzler, Department of Psychiatry, University of Pennsylvania School of Medicine, Treatment Research Center, 3900 Chestnut St., Philadelphia, PA 19104 kranzler_h@mail.trc.upenn.edu .

Conflict of interest declaration: Drs. Chan, Gelernter, and Oslin have no disclosures. Dr. Farrer received a research grant from Eisai Pharmaceuticals and consultant fees from Novartis Pharmaceuticals. Dr. Kranzler has been a paid consultant for Alkermes and Gilead. He has received research support from Merck. He also reports associations with Eli Lilly, Janssen, Schering Plough, Lundbeck, Alkermes, GlaxoSmithKline, Abbott, and Johnson & Johnson, as these companies provide support to the ACNP Alcohol Clinical Trials Initiative (ACTIVE) and Dr. Kranzler receives support from ACTIVE.

INTRODUCTION

According to the 2006 National Survey on Drug Use and Health [1], 15.1% of the US population aged 12 years or older reported having used an opioid drug at some time in their lives, 0.8% met criteria for a lifetime opioid use disorder, and 0.4% were treated for an opioid-related problem in the past year. Although opioid dependence (OD) is highly heritable [2], few genes contributing to the risk for OD have been identified. Gene finding in OD is complicated by the fact that it is a complex disorder whose etiology, clinical presentation, course, and treatment outcome vary widely. Because of this complexity, homogeneous subtypes of OD may be more useful phenotypes than the DSM-IV diagnosis of OD and may also help to improve treatments for the disorder [3].

Efforts to define substance dependence (SD) subtypes have employed univariate and multivariate approaches [4, 5]. To identify cocaine dependence (CD) subtypes, Kranzler et al. [6] employed a 3-step clustering procedure first described by Lebart, Morineau and Warwick [7]. This multivariate approach to the use of categorical variables involves data reduction and both hierarchical and non-hierarchical techniques. Two genome-wide linkage studies used this approach to identify regions harboring genes for CD [8] and OD [9] subtypes. We sought to replicate a subtype analysis that employed this procedure to refine phenotypes in a linkage study of OD [9].

The purpose of this study was to identify homogeneous phenotypes using a wide range of opioid use behaviors (OUBs) in both affected individuals (i.e., those meeting criteria for DSM-IV OD) and unaffected ones. We included OUBs that were associated with specific DSM-IV OD diagnostic criteria and other important clinical features, including route of opioid administration, age of onset of the heaviest opioid use period, and opioid abuse treatment history. However, we reserved non-OUB characteristics, which might also differentiate OD subtypes, such as demographics and co-morbid substance and psychiatric disorders, to validate the subtypes externally. This empirical subtyping approach rests on theories that emphasize the multi-faceted nature of substance use and related behaviors [4,5]. That is, similar to studies of alcohol use behaviors in affected individuals, which have helped to refine a complex phenotype [10], OUBs may be multi-dimensional. Thus, the identification of homogeneous groups based on OUBs may help to guide research on the etiology, prevention, and treatment of OD. Further, with rapid developments in genotyping methods and the capacity to sequence large numbers of whole genomes, advances in phenotyping are needed to enhance genotype-phenotype correlation [11].

METHOD

Subject Recruitment

The study sample was an aggregate of individuals recruited to participate in family-based and case-control genetic studies of OD or CD. Recruitment was conducted at five sites: the University of Connecticut Health Center (UConn; Farmington, CT), Yale University School of Medicine (Yale; New Haven, CT), the University of Pennsylvania School of Medicine (Penn; Philadelphia, PA), McLean Hospital (McLean; Belmont, MA), and the Medical University of South Carolina (MUSC; Charleston, SC). The protocol and informed consent forms were approved by the institutional review board at each site, and a Certificate of Confidentiality was obtained from the National Institute on Drug Abuse.

Families were recruited to include sibling pairs concordant for a lifetime DSM-IV [12] diagnosis of OD and/or CD. Unrelated individuals with OD and/or CD were recruited as cases, and control subjects were screened to exclude those with a lifetime DSM-IV substance use disorder. Subjects with a primary diagnosis of a major psychotic illness were

excluded. Once a case or family was recruited, we invited all additional siblings and parents to participate, regardless of affection status. In some families other members (e.g., aunts, uncles, grandparents) participated and in others only one member of a sibpair completed the assessments.

This analysis included 4,061 participants (UConn: 1,675; Yale: 1,620; Penn: 285; McLean: 259; and MUSC: 222) with complete OD assessments. Included were 2,003 subjects from 835 families and 2,058 unrelated individuals. Of the families, 96 (11.5%) had ≥ two members with OD only, 455 (54.5%) had ≥ two members with CD only, 226 (27.1%) had ≥ two members with both OD and CD, and 64 (7.7%) had only 1 member affected by the same disorder(s). Families from case-control studies included fewer than two participating members with OD or CD. The sample included 838 probands, 1,074 siblings, 89 parents, and 28 other family members from the sibling-pair studies and 1,836 cases and 196 controls from the case-control studies.

Subjects' mean age was 40.1 years (range 17–79, SD = 9.3) and 46.3% were women. Most subjects (57.8%) were never married, 28.5% were widowed, separated, or divorced, and 13.8% were married at the time of the interview. Based on self-report, the sample was 48.1% African-American (AA), 33.3% European-American (EA), 10.0% Hispanic, and 8.6% Native American or other. Few subjects (6.1%) had completed only grade school; 37.5% had some high school; 28.8% had completed high school; and 27.5% went beyond high school.

Assessment Procedures

Interviews were conducted with the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA), a computer-assisted interview that yields lifetime DSM-IV diagnoses of Axis I disorders and antisocial personality disorder (ASPD) [13, 14]. The test-retest and inter-rater reliabilities for SSADDA-derived diagnoses were fair to excellent both for lifetime SD disorders ($\kappa = 0.94$ and $\kappa = 0.91$, respectively, for OD and $\kappa = 0.48 – 0.92$ for other substance dependence disorders (except sedative dependence due to extremely low prevalence)) and lifetime psychiatric diagnoses ($\kappa = 0.43 – 0.76$) [14]. Table 1 displays the lifetime prevalence of DSM-IV disorders for the sample as a whole and separately by sex. More than 80% and 45% of individuals met lifetime criteria for CD and OD, respectively, as would be expected given the ascertainment on these disorders. Generalized Estimating Equations (GEE) Wald Type 3 χ^2-tests with Bonferroni correction for multiple comparisons ($p < 0.05/16 = 0.0031$) showed that men were significantly more likely to receive a diagnosis of dependence on cocaine, alcohol, opioids, cannabis, and other substances (i.e., PCP, hallucinogens, inhalants, solvents, or combinations such as "speedballs"), ASPD, and compulsive gambling. Women were more likely to receive a diagnosis of MDE, PTSD, OCD, agoraphobia, and panic disorder.

Measures

The opioid drugs section of the SSADDA contains 23 questions on behaviors related to a wide range of legal and illicit opioids. Each question contains several key and follow-up items, yielding 210 variables, the majority of which are categorical. We used 69 key variables (see Supplementary Tables 1 and 2) to generate homogeneous subgroups. Because the computerized version of the SSADDA skips out of the opioid section when the respondent reports never having used an opioid more than 10 times (lifetime), subsequent items were coded as "Obligate No" or "Obligate Missing." These response options form their own categories. There were 313 possible categories for the 69 key categorical variables.

Data Analysis

We conducted three analyses. The first analysis used a three-step LMW (Lebart, Morineau, and Warwich [7]) clustering procedure to identify homogeneous phenotypic groups. Step 1 of this analysis was data reduction, which used multiple correspondence analysis [MCA; 15-19], the counterpart of principal components analysis for categorical data, to treat each of the 313 categories as a binary variable and reduce them to a lower-dimensional space. The number of dimensions retained was guided by the Benzécri adjusted cumulative percentage (i.e., the percentage of variance explained) [20]. The output of MCA was the (continuous) coordinates of the retained dimensions for each of the 4,061 subjects. Step 2 of the first analysis was preliminary clustering, in which multiple *k*-means clustering [21-24] partitioned the 4,061 subjects into intermediate clusters based on the retained continuous variables from MCA. Each *k*-means application started with a different, randomly chosen seed with k=50 being much larger than the expected number of homogeneous subtypes. Subjects were iteratively classified into k partitions with respect to the nearest partition means. These *k*-means solutions were cross tabulated to identify groups of subjects who shared OUBs, and these stable groups and all remaining individuals formed the intermediate cluster solution. Step 3 of the first analysis was the final clustering step, in which hierarchical clustering [25-28] was used to build a hierarchy of intermediate clusters based on Ward's aggregation criterion, yielding a dendrogram and statistics [e.g., cubic clustering criterion (CCC), R^2, pseudo F and pseudo t^2], which helped to determine the final number of clusters [28-31]. Although we chose the number of clusters that were associated with peak CCC and relatively high R^2, pseudo F and low pseudo t^2, ultimately, we took the interpretability of cluster assignment into consideration in determining the final homogeneous groupings. Finally group characteristics were used to assign labels to the final set of clusters. GEE Wald Type 3 χ^2-tests were used to test whether the resultant groups differed significantly on 33 variables (demographic characteristics, OUBs, and the prevalence of substance use and psychiatric disorders), with Bonferroni correction for multiple comparisons ($p < 0.05/33 = 0.0015$).

In the second analysis, all participants were assigned scores that were the natural logarithm of the probabilities of membership in each of the identified clusters. The probabilities were estimated by stepwise logistic regressions (with $\alpha = 5\%$) of each cluster on all 69 cluster input variables.

In the third analysis, we estimated the heritability (h^2) of each of the homogeneous subgroups [32], using assigned scores from the 751 multiple-member EA, AA, or Hispanic families (i.e., a total of 1,801 individuals) with at least two participating members. Sex, age and race were used as covariates only in this analysis; the preceding analysis did not include these covariates because although they were important features differentiating the subtypes, they were not OUBs *per se*.

We used SAS 9.2 [33] to conduct all analyses, except the heritability estimation, which was performed using Sequential Oligogenic Linkage Analysis Routines [SOLAR; 34]. Questions concerning the methods employed should be addressed to the corresponding author (HRK).

RESULTS

Overview of Subtype Analysis Results

In step1 of the first analysis, the MCA reduced the 313 binary dimensions associated with the 69 opioid-related categorical SSADDA items to 10 continuous dimensions, which explained more than 99% of the variance. In step 2 of the first analysis, we conducted 10 *k*-means analyses based on the 10 MCA continuous output variables. Each *k*-means analysis partitioned the 4,061 subjects into 50 mutually exclusive groups. Of the 50^{10} cells, only 973

were non-empty. There were 287 stable groups that contained multiple subjects. The largest stable group included all 1,460 individuals who had never used an opioid. A second stable group included 220 individuals who first used opioids at age 17-29 but had used them fewer than 11 times lifetime. A third stable group included 32 individuals who first used opioids before age 19, had used them daily for a total of over 3,400 times, including by injection. Step 3 of the first analysis yielded a hierarchy of clusters from 1 to 973. Supplementary Figure 1 shows the resultant dendrogram with R^2. There was no local peak CCC, pseudo F suggested few clusters, while pseudo t^2 suggested that the number of clusters was 3-8. Using the dendrogram, the statistics, and the interpretability of the subtypes as criteria, we chose a five-subtype solution. The subtypes, which are shown in Table 2, differed significantly on age, sex, race, education, and marital status. The mean age was greatest in Group 3 and lowest in Group 5. Moreover, subjects in Groups 2–5 were comprised, in the majority, of unmarried, EA men with less education than those in Group 1.

Opioid Use Behaviors and Other Substance Use and Psychiatric Disorders by Group

The five groups differed significantly on lifetime opioid use characteristics, opioid-related effects, treatment history (Table 3), and lifetime prevalence of most substance use disorders and panic disorder (Table 4).

Group 1 (Low-level or Non-opioid user) comprised 2,382 individuals (58.7% of the sample) with an estimated heritability of 0.31 ($p < 10^{-5}$). Only 38.7% of Group 1 had ever used an opioid, with a mean age of first use of 24.8 (SD = 8.7) years, which was significantly later than for Groups 2, 4, and 5 (GEE Wald $\chi^2_{(4)} = 553.49$, $p < 0.001$). Less than 20% of this group had a lifetime OD diagnosis. More than 80% of the group did not complete the opioid drugs section because of inadequate opioid usage and the group was excluded from the χ^2 tests in Table 3. Although more than 75% of individuals in Group 1 were diagnosed with CD, this group had the lowest prevalence of all other SD and psychiatric disorders (except ASPD).

Group 2 (Moderate Opioid Users) comprised 405 individuals (10.0% of the sample) and had the second latest age of onset of both any and heaviest opioid use, and the lowest percentage of daily or almost daily opioid use, intravenous opioid use, any opioid-related effects and opioid abuse treatment (Table 3). Individuals in Group 2 used opioids only moderately. During their heaviest periods of opioid use, individuals in Group 2 spent the least on opioids daily (GEE Wald $\chi^2_{(3)} = 180.70$, $p < 0.001$; Group 2: mean (SD) = $46.6 (78.6), Group 3 = $97.1 (107.1), Group 4 = $141.8 (120.5), and Group 5 = $157.5 (163.6)), and they used opioids on the fewest days per month (GEE wald $\chi^2_{(3)} = 414.51$, $p < 0.001$; Group 2: mean (SD) = 17.6 (11.2), Group 3 = 29.1 (3.8), Group 4 = 28.6 (4.8), and Group 5 = 29.4 (3.1)).

Nearly 90% of individuals in Group 2 were diagnosed with CD, but less than 45% had OD (a significantly lower prevalence than for Groups 3–5). Group 2 endorsed the fewest DSM-IV OD criteria (GEE Wald $\chi^2_{(3)} = 1212.46$, $p < 0.001$; Group 2: mean (SD) = 2.0 (2.5), Group 3 = 6.2 (1.2), Group 4 = 6.5 (1.2), and Group 5 = 6.6 (1.0)), and were least likely to have nicotine and stimulant dependence, MDE, PTSD, social phobia, agoraphobia, and panic disorder than Groups 3–5 (Table 4). The heritability of Group 2 membership was 0.40 ($p < 10^{-10}$).

Group 3 (Heavy, Late-onset Opioid User) included 290 subjects (7.1% of the sample) and all reported having used opioids daily or almost daily and had lifetime OD, with more than 80% of individuals endorsing the presence of all seven DSM-IV OD criteria. The majority of Group 3 had used opioids intravenously. Group 3 reported a later onset of any and heaviest opioid use, but a significantly lower rate of arrests or trouble with the police due to opioid use than Groups 4 and 5, with the rate of other opioid-related effects similar among

Groups 3–5. Group 3 had a lower prevalence of cannabis, sedative, and other SD disorders, ASPD, and PTSD than Groups 4 and 5 (Table 4). The heritability for Group 3 was 0.36 ($p < 10^{-8}$).

Group 4 (Heavy, Early-onset, Highly Co-morbid Opioid Users) included 296 individuals (7.3% of the total); nearly all had used opioids daily or almost daily and more than 65% had used opioids intravenously. Group 4 had a later age of onset of opioid use and a lower prevalence of intravenous opioid use, but a higher prevalence of all other SD and psychiatric disorders than individuals in Group 5. Nearly all individuals in Group 4 had lifetime OD, and a rate of endorsement for each DSM-IV OD criterion > 85%. The heritability of this group was 0.66 ($p < 10^{-27}$).

Group 5 (Heavy, Early-onset Opioid Users) included 688 individuals (16.9% of the sample). Although similar in many respects to Groups 3 and 4, Group 5 reported a significantly earlier age of onset of any and heaviest opioid use than all other groups and was significantly more likely to inject opioids intravenously, have a strong desire for opioids, and be arrested due to opioid use. The prevalence rates for substance use and psychiatric disorders in Group 5 were intermediate between those of Groups 3 and 4. Each of the DSM-IV OD criteria was endorsed by > 90% of individuals in Group 5, and 76.3% endorsed all seven criteria. Heritability for this group was 0.61 ($p < 10^{-20}$).

DISCUSSION

Using the 3-step LMW clustering procedure, all 4,061 participants in genetic studies of OD and CD were assigned to five mutually exclusive groups with distinct OUBs. The largest group (Group 1) comprised non-opioid users and low-level opioid users. The heavy opioid users were clustered into three groups with unique characteristics: early-onset (Group 5), late-onset (Group 3), and early-onset with co-morbid disorders (Group 4). Almost all of the subjects in these three groups met criteria for DSM-IV OD, which suggests that OD is a heterogeneous diagnosis. However, the two most severe early-onset groups (Groups 4 and 5) showed the highest estimated heritability (> 0.60). The estimated heritability for the late-onset group (Group 3) was somewhat lower (0.36). Overall, despite comparable severity of OD in these groups, genetic risk varies for subgroups with the disorder.

Consistent with our earlier analysis based on a much smaller family-based sample that employed different software for part of the analysis, there were five distinct subtypes identified among subjects participating in genetic studies of OD and CD. Also consistent with that analysis, the proportion of opioid non-users/low-level users, moderate users and heavy users were approximately 60%, 10% and 30%, respectively [9]. However, the present subtype findings combined in Group 1 individuals from two clusters in the earlier analysis: non-opioid users (Cluster A, $h^2 = 0.61$) and low-level opioid users (Cluster B, $h^2 = 0.40$). Moreover, the current analysis differentiated three groups of heavy opioid users compared to the two identified previously [9]. Specifically, the heavy-opioid use Cluster E ($h^2 = 0.40$) from the earlier analysis (which yielded a lod score of 3.06 on chromosome 17 for EA and AA subjects combined) was divided into early- and late-onset groups in this study. The larger sample available in the present study may have resulted in a more refined subtype analysis, as evidenced by the higher h^2 that resulted, with the estimates for Groups 4 and 5 exceeding 0.60.

This study has a number of strengths. First, the 3-step LMW clustering procedure produced meaningful and homogeneous subgroups and was ideal for clustering a large number of individuals (4,061) on a large number of categorical variables (69 variables with 313 associated categories). An initial data-reduction step reduced the number of dimensions

from 313 to 10, followed by a preliminary clustering step to reduce the number of individuals from 4,061 to 973 intermediate groups. The final clustering step produced a sequence of nested cluster solutions. Unlike latent class analysis (LCA), this procedure does not require users to specify the number of clusters in advance and the sequence of nested cluster solutions provides insight into how intermediate groups are merged. This procedure has been used to identify homogeneous groups of risk profiles for cardiovascular disease [35] and of farms using socio-economic profiles [36]. Although LCA has certain advantages over the 3-step LMW clustering procedure, we chose to employ the same technique as in our previous study [9]. Second, the instrument used to collect phenotypic information, the SSADDA, has been shown to yield reliable diagnostic criteria and diagnoses for a variety of SD and psychiatric disorders [13, 14]. Third, our analysis is based on a large sample with diverse opioid use and opioid-related problem severity. Fourth, the groups differed on a variety of demographic features and co-morbid substance use and psychiatric disorders, which were not used in the 3-step LMW clustering procedure [37]. Fifth, all groups showed significant heritability, with the most severe, earlier-onset groups having the highest heritability estimates, which is convergent with the group assignment.

An important limitation of this study is the absence of additional sources of information concerning the identified subtypes (e.g., based on follow-up studies or linkage analysis), which are needed to validate these findings. Second, the high prevalence of CD in the study sample limits the generalizability of the findings. Third, the SSADDA does not provide detailed information on the kinds of opioids that subjects used. There may be other subtypes of OD that were not captured in this sample (e.g., individuals whose use consists primarily of prescription opioids). Finally, Groups 4 and 5, which showed the highest heritabilities, included only 984 individuals, or 24% of the total sample, which could limit the use of this approach in genetic studies. Independent replication of these findings and refinement of the method are needed to address these limitations.

In conclusion, because it is heterogeneous, the DSM-IV diagnosis of OD phenotype may not be the best phenotype to identify genes contributing to the risk of OD or to elucidate the mechanism of such risk. This study highlights the presence of at least three subtypes of heavy opioid users: late-onset, early-onset, and early-onset with co-morbid disorders. Further study of these and other phenotypes should provide insight into factors that determine the risk for development of OD, which could enhance the early identification and treatment of the disorder.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank the interviewers and other research personnel that conducted the study, and the study participants at all of the recruitment/assessment sites. Marsha Wilcox, Ph.D. contributed to the initial effort to develop OD subtypes. Kathleen Brady, M.D., Ph.D. of the Medical University of South Carolina and Roger Weiss, M.D. of McLean Hospital and Harvard Medical School oversaw study recruitment at their respective sites. Laura Almasy, Ph.D. provided assistance in the use of SOLAR. John Farrell provided database management support. Raymond Anton, M.D. provided helpful comments on the manuscript.

This work was supported by NIH grants DA12849, DA12690, DA22288, DA15105, DA005186, AA03510, AA11330, AA13736, and GM08607 and the VA CT and Philadelphia VA Mental Illness Research, Education, and Clinical Centers (MIRECCs).

References

1. SAMHSA. Results from the 2006 National Survey on Drug Use and Health: National Findings. Department of Health and Human Services; Rockville, MD: 2007.
2. Saxon AJ, Oreskovich MR, Brkanac Z. Genetic determinants of addiction to opioids and cocaine. Harv Rev Psychiatry. 2005; 13(4):218–32. [PubMed: 16126608]
3. Ball SA, Nich C, Rounsaville BJ, Eagan D, Carroll KM. Millon Clinical Multiaxial Inventory-III subtypes of opioid dependence: validity and matching to behavioral therapies. J Consult Clin Psychol. 2004; 72(4):698–711. [PubMed: 15301655]
4. Basu D, Ball SA, Feinn R, Gelernter J, Kranzler HR. Typologies of drug dependence: comparative validity of a multivariate and four univariate models. Drug Alcohol Depend. 2004; 73(3):289–300. [PubMed: 15036551]
5. Moss HB, Chen CM, Yi HY. Subtypes of alcohol dependence in a nationally representative sample. Drug Alcohol Depend. 2007; 91(2-3):149–58. [PubMed: 17597309]
6. Kranzler HR, Wilcox M, Weiss RD, Brady K, Hesselbrock V, Rounsaville B, et al. The validity of cocaine dependence subtypes. Addict Behav. 2008; 33(1):41–53. [PubMed: 17582692]
7. Lebart, L.; Morineau, A.; Warwick, KM. Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Martices. John Wiley and Sons Inc.; New York: 1984.
8. Gelernter J, Panhuysen C, Weiss R, Brady K, Hesselbrock V, Rounsaville B, et al. Genomewide linkage scan for cocaine dependence and related traits: significant linkages for a cocaine-related trait and cocaine-induced paranoia. Am J Med Genet B Neuropsychiatr Genet. 2005; 136B(1):45–52. [PubMed: 15909294]
9. Gelernter J, Panhuysen C, Wilcox M, Hesselbrock V, Rounsaville B, Poling J, et al. Genomewide linkage scan for opioid dependence and related traits. Am J Hum Genet. 2006; 78(5):759–69. [PubMed: 16642432]
10. Babor TF, Hofmann M, DelBoca FK, Hesselbrock V, Meyer RE, Dolinsky ZS, et al. Types of alcoholics, I. Evidence for an empirically derived typology based on indicators of vulnerability and severity. Arch Gen Psychiatry. 1992; 49(8):599–608. [PubMed: 1637250]
11. Sabb FW, Burggren AC, Higier RG, Fox J, He J, Parker DS, et al. Challenges in phenotype definition in the whole-genome era: multivariate models of memory and intelligence. Neuroscience. 2009; 164:88–107. [PubMed: 19450667]
12. APA. Diagnostic and Statistical Manual of Mental Disorders. 4th ed. American Psychiatric Press Inc.; Washington, DC: 1994.
13. Pierucci-Lagha A, Gelernter J, Chan G, Arias A, Cubells JF, Farrer L, et al. Reliability of DSM-IV diagnostic criteria using the semi-structured assessment for drug dependence and alcoholism (SSADDA). Drug Alcohol Depend. 2007; 91(1):85–90. [PubMed: 17590536]
14. Pierucci-Lagha A, Gelernter J, Feinn R, Cubells JF, Pearson D, Pollastri A, et al. Diagnostic reliability of the Semi-structured Assessment for Drug Dependence and Alcoholism (SSADDA). Drug Alcohol Depend. 2005; 80(3):303–12. [PubMed: 15896927]
15. Abdi, H.; Valentin, D. Multiple correspondence analysis. In: Salkind, N., editor. Encyclopedia of Measurement and Statistics. SAGE; Thousand Oaks, CA.: 2007.
16. Carroll JD, Green PE, Schaffer CM. Interpoint distance comparisons in correspondence analysis. Journal of Market Research. 1986; 23(3):271–280.
17. Greenacre M, Hastie T. The geometric interpretation of correspondence analysis. Journal of the American Statistical Association. 1987; 82(398):437–447.
18. LeRoux, B.; Rouanet, H. Multiple Correspondence Analysis. Thousand Oaks: 2010. SAGE Quantitative Applications in the Social Sciences Series volume 163
19. Tenenhaus M, Young F. An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. Psychometrika. 1985; 50(1):91–119.
20. Benzecri JP. Sur le Calcul des taux d'inertie dans l'analyse d'un questionaire. Addendum et erratum á. Cahiers de l'Analyse des Données. 1979; 4:377–8. (The calculation of the rate of inertia in the analysis of a questionnaire).

21. Alsabti, K.; Ranka, S.; Singh, V. An efficient k-means clustering algorithm. Proceedings of the First Workshop on High Performance Data Mining; Orlando, FL: 1998.
22. Anderberg, MR. Cluster Analysis for Applications. Academic Press Inc.; New York: 1973.
23. Kanungo, T.; Mount, DM.; Netanyahu, NS.; Piatko, CD.; Silverman, R.; Wu, AY. An efficient k-means clustering algorithm: analysis and implementation; IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002. p. 881-892.Archived URL: http://www.webcitation.org/query?url=http%3A%2F%2Fwww.cs.umd.edu%2F%7Emount%2FProjects%2FKMeans%2Fpami02.pdf&date=2010-08-31
24. MacQueen, J. Some methods for classification and analysis of multivariate observations; Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability. 1967. p. 281-97.Archived URL: http://www.webcitation.org/query?url=http%3A%2F%2Fweb.inf.unibz.it%2Fdis%2Fteaching%2FDWDM%2Fproject2009%2FK-means.pdf&date=2010-09-02
25. Day WH, Edelsbrunner H. Efficient algorithms for agglomerative hierarchical clustering methods. Journal of Classification. 1984; 1(1):7–24.
26. Johnson SC. Hierarchical clustering schemes. Psychometrika. 1967; 32(3):241–254. [PubMed: 5234703]
27. Milligan GW. Ultrametric hierarchical clustering algorithms. Psychometrika. 1979; 44:343–46.
28. Ward, JH. Hierarchical grouping to optimize an objective function; Journal of the American Statistical Association. 1963. p. 236-44.Archived URL: http://www.webcitation.org/query?url=http%3A%2F%2Fiv.slis.indiana.edu%2Fsw%2Fdata%2Fward.pdf&date=2010-08-31
29. Calinski, T.; Harabasz, J. A dendrite method for cluster analysis; Communications in Statistics. 1974. p. 1-27.Archived URL: http://www.webcitation.org/query?url=http%3A%2F%2Fpdfserve.informaworld.com%2F72015_731222808_780133830.pdf&date=2010-09-02
30. Duda, RO.; Hart, PE. Pattern Classification and Scene Analysis. Wiley and Sons Inc.; New York: 1973.
31. Sarle, WS. SAS Technical Report A-108. SAS Institute Inc.; Cary, NC.: 1983. Cubic Clustering Criterion. Archived URL: http://www.webcitation.org/query?url=http%3A%2F%2Fsupport.sas.com%2Fdocumentation%2Fonlinedoc%2Fv82%2Ftechreport_a108.pdf&date=2010-09-02
32. Lynch, M.; Walsh, B. Genetics and Analysis of Quantitative Traits. Sinauer Associates Inc.; Sunderland, MA: 1998.
33. SAS. SAS release 9.2. SAS Institute; Cary, NC.: 2009.
34. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet. 1998; 62(5):1198–211. [PubMed: 9545414]
35. Wilcox M, Li Q, Sun Y, Stang P, Berlin J, Wang D. Genome-wide association study for empirically derived metabolic phenotypes in the Framingham heart study offspring cohort. BMC Proceedings. 2009; 3(Suppl 7):S53. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2795953/pdf/1753-6561-3-S7-S53.pdf. [PubMed: 20018046]
36. Menexes G, Angelopoulos S. Proposals for the financing and development of Greek farms based on a clustering method for categorical data. EuroMed Journal of Business. 2008; 3(3):263–285.
37. Windle M, Scheidt DM. Alcoholic subtypes: are two sufficient? Addiction. 2004; 99:1508–1519. [PubMed: 15585042]

Table 1

Lifetime prevalence of substance use and psychiatric disorders by sex [*N* (%)[a]]

Diagnosis[b]	Total (*N* = 4,061)	Male (*N* = 2,182)	Female (*N* = 1,879)	GEE Wald χ^2 (1 df)	*p*-value
Substance use disorders					
Cocaine dependence	3,293 (81.2)	1,808 (82.9)	1,485 (79.1)	11.72	< 0.001
Nicotine dependence	2,682 (66.1)	1,446 (66.3)	1,236 (65.8)	0.13	0.716
Alcohol dependence	1,933 (47.7)	1,138 (52.2)	795 (42.4)	37.54	< 0.001
Opioid dependence	1,837 (45.3)	1,080 (49.5)	757 (40.3)	32.61	< 0.001
Cannabis dependence	1,167 (28.8)	766 (35.2)	401 (21.4)	93.17	< 0.001
Sedative dependence	229 (5.7)	181 (8.3)	120 (6.4)	4.92	0.027
Stimulant dependence	301 (7.4)	137 (6.3)	92 (4.9)	3.68	0.055
Other substance dependence [c]	700 (17.3)	455 (21.0)	245 (13.1)	41.68	< 0.001
Psychiatric disorders					
ASPD [c]	541 (13.7)	392 (18.4)	149 (8.2)	83.49	< 0.001
MDE [c]	667 (17.0)	253 (11.8)	424 (23.2)	91.96	< 0.001
PTSD [c]	629 (15.7)	267 (12.4)	362 (19.6)	37.15	< 0.001
OCD [c]	96 (2.4)	29 (1.3)	67 (3.6)	20.41	< 0.001
Social Phobia	159 (4.0)	71 (3.3)	88 (4.8)	5.76	0.016
Agoraphobia	247 (6.2)	78 (3.6)	169 (9.2)	51.94	< 0.001
Panic disorder	280 (7.0)	106 (4.9)	174 (9.4)	29.05	< 0.001
Compulsive gambling	381 (9.5)	296 (13.8)	85 (4.6)	86.39	< 0.001

[a] Percentages are based on the number of subjects with known status for each diagnosis. The rate of missing data varied slightly across diagnoses.

[b] Disorders that differed significantly by sex at 0.05/16 = 0.0031 are in bold.

[c] Other substance dependence includes dependence on phencyclidine, hallucinogens, inhalants, solvents, or a combination of opioids and cocaine (i.e., “speedballs”); ASPD: Antisocial Personality Disorder; MDE: Major Depressive Episode; PTSD: Posttraumatic Stress Disorder; OCD: Obsessive-Compulsive Disorder.

Table 2

Demographic characteristics by group [*N* (%)]

Characteristic[b]	**Group 1 2,382 (58.7)**	**Group 2 405 (10.0)**	**Group 3 290 (7.1)**	**Group 4 296 (7.3)**	**Group 5 688 (16.9)**	**Test statistic**[c]
Age [mean (SD)]	40.8 (9.4)	39.4 (9.4)	42.4 (5.7)	39.0 (8.2)	37.8 (10.1)	$\chi^2_{(4)} = 84.21$
Sex [*N* (%)]						$\chi^2_{(4)} = 60.22$
Female	1,218 (51.1)	144 (35.6)	128 (44.1)	134 (45.3)	255 (37.1)	
Male	1,164 (48.9)	261 (64.4)	162 (55.9)	162 (54.7)	433 (62.9)	
Race [*N* (%)]						$\chi^2_{(12)} = 490.18$
AA	1,469 (61.7)	174 (43.0)	79 (27.2)	89 (30.1)	141 (20.5)	
EA	563 (23.7)	156 (38.5)	145 (50.0)	109 (36.8)	380 (55.2)	
Hispanic	161 (6.8)	36 (8.9)	42 (14.5)	59 (19.9)	108 (15.7)	
Other	188 (7.9)	39 (9.7)	24 (8.3)	39 (13.2)	59 (8.6)	
Education [*N* (%)]						$\chi^2_{(12)} = 80.18$
No HS[d]	117 (4.9)	23 (5.7)	9 (3.1)	30 (10.2)	70 (10.2)	
Some HS[d]	833 (35.0)	158 (39.0)	117 (40.5)	137 (46.4)	277 (40.3)	
HS graduate	703 (29.5)	105 (25.9)	88 (30.5)	80 (27.1)	194 (28.2)	
Beyond HS[d]	729 (30.6)	119 (29.4)	75 (26.0)	48 (16.3)	147 (21.4)	
Marital status [*N* (%)]						$\chi^2_{(8)} = 40.59$
Never married	1,365 (57.3)	248 (61.2)	138 (47.6)	170 (57.4)	425 (61.8)	
Married	372 (15.6)	48 (11.9)	35 (12.1)	30 (10.1)	74 (10.8)	
Div/Sep/Wid	645 (27.1)	109 (26.9)	117 (40.3)	96 (32.4)	189 (27.5)	

[a] GEE Wald χ^2 tests for age and sex; independent χ^2 tests for race, education and marital status because the GEE algorithm is not available for generalized logistic regression on multi-level categorical outcome variables.

[b] All demographic variables differed significantly by group at $p < 0.0015$ (i.e., 0.05/33).

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

Table 3

Lifetime opioid use characteristics, opioid-related effects, and opioid treatment history for Groups 1–5 [*N* (%)]

Behaviors[b]	**Group 1**[c] **2,382 (58.7)**	**Group 2 405 (10.0)**	**Group 3 290 (7.1)**	**Group 4 296 (7.3)**	**Group 5 688 (16.9)**	$\chi^2_{(3)}$ **Test statistic**[a]
Opioid use characteristics						
Used an opioid	922 (38.7)	405 (100)	290 (100)	296 (100)	688 (100)	
Mean age of first opioid use in yr (SD)	24.8 (8.7)	22.0 (7.5)	26.6 (7.1)	20.2 (5.8)	18.8 (4.1)	346.16
Mean age of onset of heaviest opioid use in yr (SD)	29.4 (10.3)	26.8 (8.4)	34.0 (4.5)	26.0 (7.2)	25.1 (7.1)	625.60
Used opioids daily or almost daily	400 (16.8)	243 (60.0)	290 (100.0)	291 (98.3)	688 (100.0)	538.95
Injected opioids intravenously	193 (8.1)	131 (32.4)	190 (65.5)	192 (65.1)	609 (88.5)	302.00
Opioid-related effects						
Stayed high from opioids for a whole day or more	337 (14.2)	248 (61.2)	255 (87.9)	262 (88.5)	639 (92.9)	162.12
Strong desire for opioids made it hard to think of anything else	294 (12.3)	118 (29.1)	257 (88.6)	264 (89.2)	661 (96.1)	488.43
Opioid use interfered with work, school, or home life	299 (12.6)	163 (40.3)	267 (92.1)	296 (100.0)	678 (98.6)	741.31
Family members, friends, doctor, clergy, boss, or people at work or school objected to opioid use	267 (11.2)	134 (33.1)	255 (87.9)	279 (94.3)	660 (95.9)	465.77
Been arrested or had trouble with the police because of opioid use	183 (7.7)	77 (19.0)	165 (56.9)	210 (71.0)	555 (80.7)	332.81
Gave up or greatly reduced important activities due to opioid use	267 (11.2)	125 (30.9)	256 (88.3)	284 (96.0)	661 (96.1)	481.16
Opioid treatment history						
Ever treated for an opioid-related problem	314 (13.2)	120 (29.7)	277 (95.5)	261 (88.2)	678 (98.6)	451.99
Ever attended self-help group for opioid use	193 (8.1)	90 (24.1)	230 (79.6)	217 (73.8)	563 (81.8)	327.67

[a] GEE Wald χ^2 tests for all, except independent χ^2 test for used opioid daily or almost daily, and opioid use interfered with life due to numerical problem as a result of near complete separation.

[b] All behaviors differed significantly by group at $p < 0.0015$ (i.e., 0.05/33).

[c] Over 80% of Group 1 reported using opioids fewer than 11 times and thus skipped out of the rest of the opioid drugs section; this group was excluded from group comparison χ^2 tests.

Table 4

Lifetime prevalence of substance use and psychiatric disorders by group [*N* (%)]

Disorders[b]	**Group 1 2,382 (58.7)**	**Group 2 405 (10.0)**	**Group 3 290 (7.1)**	**Group 4 296 (7.3)**	**Group 5 688 (16.9)**	$\chi^2_{(4)}$ **Test statistic**[a]
Substance use disorders						
Cocaine dependence	1,850 (77.8)	359 (88.6)	238 (82.1)	270 (91.2)	576 (83.7)	52.82
Nicotine dependence	1,375 (57.8)	288 (71.1)	218 (75.2)	256 (86.5)	545 (79.2)	183.79
Alcohol dependence	1,027 (43.2)	237 (58.5)	148 (51.0)	183 (62.0)	338 (49.1)	63.82
Opioid dependence	394 (16.6)	171 (42.2)	290 (100.0)	295 (99.7)	687 (99.9)	2325.94
Cannabis dependence	580 (24.4)	155 (38.6)	83 (28.6)	135 (46.2)	214 (31.2)	80.57
Sedative dependence	84 (3.5)	33 (8.2)	22 (7.6)	34 (11.6)	56 (8.2)	229.80
Stimulant dependence	61 (2.6)	22 (5.5)	29 (10.0)	70 (24.0)	119 (17.4)	47.69
Other substance dependence [c]	161 (6.8)	74 (18.5)	49 (17.0)	139 (47.6)	277 (40.6)	479.21
Psychiatric disorders						
ASPD [c]	258 (11.0)	70 (18.0)	29 (10.2)	70 (24.9)	114 (17.1)	56.25
MDE [c]	352 (15.0)	65 (16.5)	57 (20.1)	63 (22.3)	140 (20.8)	19.38
PTSD [c]	331 (14.0)	57 (14.4)	43 (15.0)	77 (27.11)	121 (17.9)	34.00
OCD [c]	41 (1.7)	11 (2.8)	6 (2.1)	17 (6.0)	21 (3.1)	20.00
Social Phobia	68 (2.9)	16 (4.1)	14 (4.9)	20 (7.2)	41 (6.4)	21.95
Agoraphobia	114 (4.9)	26 (6.6)	23 (8.0)	34 (12.0)	50 (7.4)	26.19
Panic disorder	101 (4.3)	28 (7.1)	27 (9.5)	46 (16.2)	78 (11.5)	80.30
Compulsive gambling	199 (8.5)	43 (10.9)	30 (10.5)	40 (14.0)	69 (10.2)	10.63

[a] GEE Wald χ^2 tests for all, except independent χ^2 test for OD due to a numerical problem resulting from near-complete separation.

[b] Disorders that differed significantly by group at $p < 0.0015$ (i.e., 0.05/33) are in bold.

[c] Other substance dependence includes dependence on phencyclidine, hallucinogens, inhalants, solvents, or a combination of opioids and cocaine (i.e., “speedballs”).

ASPD: Antisocial Personality Disorder; MDE: Major Depressive Episode; PTSD: Posttraumatic Stress Disorder; OCD: Obsessive-Compulsive Disorder.