Published in final edited form as:

Drug Alcohol Depend. 2015 August 1; 153: 229–235. doi:10.1016/j.drugalcdep.2015.05.019.

Inter-Observer Reliability of DSM-5 Substance Use Disorders*

Cécile M. Denis¹, Joel Gelernter², Amy B. Hart¹, and Henry R. Kranzler^{1,3}

¹Center for Studies of Addiction, Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104

²Departments of Psychiatry (Division of Human Genetics), Neurobiology, and Genetics, Yale University School of Medicine, New Haven, CT and VA Connecticut Healthcare System, West Haven, CT 06516

³VISN4 MIRECC, Philadelphia VAMC, Philadelphia, PA 19104

Abstract

Aims—Although studies have examined the impact of changes made in DSM-5 on the estimated prevalence of substance use disorder (SUD) diagnoses, there is limited evidence of the reliability of DSM-5 SUDs. We evaluated the inter-observer reliability of four DSM-5 SUDs in a sample in which we had previously evaluated the reliability of DSM-IV diagnoses, allowing us to compare the two systems.

Methods—Two different interviewers each assessed 173 subjects over a 2-week period using the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA). Using the percent agreement and kappa (κ) coefficient, we examined the reliability of DSM-5 lifetime alcohol, opioid, cocaine, and cannabis use disorders, which we compared to that of SSADDA-derived DSM-IV SUD diagnoses. We also assessed the effect of additional lifetime SUD and lifetime mood or anxiety disorder diagnoses on the reliability of the DSM-5 SUD diagnoses.

Results—Reliability was good to excellent for the four disorders, with κ values ranging from 0.65 to 0.94. Agreement was consistently lower for SUDs of mild severity than for moderate or

*Supplementary material can be found by accessing the online version of this paper at <http://dx.doi.org> and by entering doi:...

Correspondence to: Henry R. Kranzler, Treatment Research Center, Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, 3900 Chestnut Street, Philadelphia, PA 19104; Telephone: 215-386-6662; Facsimile: 215-386-6770; kranzler@mail.med.upenn.edu.

Conflict of interest

Cécile Denis, Joel Gelernter, and Amy Hart have no disclosures to make. Henry Kranzler has been a consultant or advisory board member for Alkermes, Lilly, Lundbeck, Otsuka, and Pfizer and is a member of the American Society of Clinical Psychopharmacology's Alcohol Clinical Trials Initiative, which is supported by AbbVie, Ethypharm, Lilly, Lundbeck, and Pfizer.

Contributors

Cécile Denis, Joel Gelernter, Amy Hart, and Henry Kranzler had full access to all of the data in the study and take responsibility for the integrity of the data and accuracy of the data analysis. *Study concept and design*: Henry Kranzler, Joel Gelernter. *Analysis and interpretation of data*: Cécile Denis, Henry Kranzler, Amy Hart. *Statistical analysis*: Cécile Denis. *Obtained funding*: Henry Kranzler, Joel Gelernter. *Study supervision*: Henry Kranzler. All authors contributed to and approved the final manuscript.

Role of funding source

The sponsors had no role in the study design, collection, analysis or interpretation of the data, or in the writing of the manuscript.

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.

severe disorders. DSM-5 SUD diagnoses showed greater reliability than DSM-IV diagnoses of abuse or dependence or dependence only. Co-occurring SUD and lifetime mood or anxiety disorders exerted a modest effect on the reliability of the DSM-5 SUD diagnoses.

Conclusions—For alcohol, opioid, cocaine and cannabis use disorders, DSM-5 criteria and diagnoses are at least as reliable as those of DSM-IV.

Keywords

Substance use disorder; DSM-5; reliability; mood disorder; anxiety disorder

1. INTRODUCTION

The fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), published in May, 2013, included several changes to the Substance Use Disorders (SUD) diagnostic category, which were intended to address limitations in the DSM-IV SUD diagnoses. Specifically, DSM-5 omitted the criterion related to “Illegal Acts,” added a “Craving” criterion, and combined abuse and dependence into a single SUD diagnosis with three severity levels: mild (2–3 criteria), moderate (4–5 criteria), and severe (6 or more criteria) (Hasin et al., 2013).

Several studies supported the decision to omit the “Illegal Acts” criterion on the basis of low endorsement rates (Gilder et al., 2014; Hasin et al., 2012; Kerridge et al., 2011; Saha et al., 2012). Although the Craving criterion did not provide additional information in the context of the other changes (Keyes et al., 2011), it was added to ensure that the DSM-5 SUD diagnosis was comparable to the International Classification of Diseases (ICD) diagnosis, which includes craving as a diagnostic criterion (Hasin et al., 2013; Keyes et al., 2011; Mewton et al., 2013). Studies that used Item Response Theory (IRT) analyses consistently demonstrated a single dimension underlying the 11 DSM-5 SUD criteria, the severity of which varied on a continuum, justifying the use of a single diagnosis of SUD (Borges et al., 2010; Gilder et al., 2014; Hasin et al., 2012; Kerridge et al., 2011; Preuss et al., 2014; Saha et al., 2012). Studies that examined the impact of these changes on the estimated prevalence of SUDs showed that the performance of the DSM-5 criteria was similar to that of the DSM-IV criteria, though DSM-5 captured DSM-IV “diagnostic orphans” (i.e., individuals meeting only 1 or 2 DSM-IV dependence criteria and no DSM-IV abuse criteria; Kelly et al., 2014; Mewton et al., 2013; Peer et al., 2013; Proctor et al., 2012).

The studies that led to the development of DSM-5 focused primarily on psychometric properties of the DSM-5 SUD diagnoses, rather than assessing their reliability. The DSM-IV substance dependence diagnoses have been shown to be reliable, with substance abuse diagnoses less so (Hasin et al., 1997, 2006; Pierucci-Lagha et al., 2007, 2005). Because DSM-5 criteria include 10 of the 11 DSM-IV abuse and dependence criteria, DSM-5 SUD criteria are also likely to be reliable (Hasin et al., 2013). Moreover, the empirical basis for the changes embodied in DSM-5 would predict that the reliability of SUD diagnoses would be at least as good as for DSM-IV diagnoses (Kraemer et al., 2012). However, a formal evaluation of the reliability of the DSM-5 SUD diagnoses is needed, as evidence of the reliability of DSM-5 would increase the utility of DSM-5 for research and policy purposes

(Hasin et al., 2013). To date, the only published study of the reliability of any DSM-5 SUD was from the DSM-5 field trials in the United States and Canada, where the reliability of the DSM-5 diagnosis of alcohol use disorder was considered to be good (intraclass kappa (κ) = 0.40, 95% CI = 0.27–0.54) (Regier et al., 2013).

The primary objectives of the current study were 1) to examine the inter-observer reliability of DSM-5 alcohol, opioid, cocaine, and cannabis use disorder diagnoses and criteria measured using the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA) and 2) to compare the reliability of these diagnoses with that of SSADDA-derived DSM-IV diagnoses in the same sample of 173 individuals recruited from a variety of settings. Consistent with the results of other studies, we found that SUDs commonly co-occurred with other SUDs (Regier et al., 2013; Stinson et al., 2005) and psychiatric disorders (Compton et al., 2007; Conway et al., 2006). Because diagnostic criteria for one disorder could mimic those of another disorder, thereby confounding the diagnoses, we also evaluated the effect of comorbid disorders on the reliability of DSM-5 SUD diagnoses.

2. MATERIAL AND METHODS

2.1 Sample

The study was a secondary analysis of data from a study conducted in 2004–2005 in a sample of 173 subjects in which the inter-observer reliability of DSM-IV SUD diagnoses (Pierucci-Lagha et al., 2007, 2005) was evaluated. The subjects were recruited from addiction treatment facilities at the University of Connecticut Health Center (UConn) and the Yale University School of Medicine (Yale) (n=103), inpatient and outpatient psychiatric services at UConn (n=29), and community respondents recruited through advertisements in local media (n=41). The advertisement invited individuals to participate in a study testing a psychiatric interview. Enrolled participants were at least 18 years old, without clinical evidence of severe psychiatric disorder that would limit their capacity to provide accurate information (e.g., schizophrenia, gross cognitive impairment), able to read and write in English, and willing to provide informed consent to participate. All procedures were approved by each institution's Institutional Review Boards.

2.2 Assessment instrument

Subjects were interviewed with the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA), a computer-assisted interview that elicits data on demographics, SUDs and other psychiatric disorders. The SSADDA was adapted from the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA; Bucholz et al., 1994, 1995). Following an intensive training program and using quality assurance procedures, non-clinician interviewers administered the SSADDA. Although the SSADDA was originally designed to elicit DSM-IV substance abuse and dependence criteria, the modest changes in criteria made it possible to make DSM-5 diagnoses as well. For each substance, the craving criterion was assessed by the following question: "Have you ever had such a strong desire for (*substance*) that it was hard to think of anything else?" The remaining 10 DSM-5 criteria consist of all seven DSM-IV dependence criteria plus three of the four DSM-IV abuse criteria (the Illegal Acts criterion having been excluded from DSM-5).

Although the SSADDA measured craving and DSM-IV abuse criteria, it did not assess whether these criteria clustered within the same 12-month period as the other criteria (i.e., the criteria from the DSM-IV dependence diagnosis). However, a previous study that we conducted in 7,543 individuals with an alcohol, cocaine, opioid or cannabis use disorder showed a very modest effect of the clustering requirement, indicating that the lack of information on clustering for DSM-5 is unlikely to have an impact on the criterion endorsement and prevalence of DSM-5 SUD diagnoses (Peer et al., 2013).

2.3 Study design

Twelve non-clinicians (nine with a bachelor's degree and three with a master's degree as well) conducted the interviews. All interviewers underwent a three-day training program delivered by a highly experienced SSADDA interviewer. The trainer then carefully reviewed the last three of 10 practice interviews to ensure proficiency before the interviewers were approved to conduct study interviews. This is based on the standard SSADDA interviewer training protocol.

One hundred seventy-three subjects (103 from addiction treatment facilities, 49 from psychiatric facilities, and 21 from the community) were interviewed twice by different interviewers either from the same center (n=136, 78.6%) or from the other center (n=37, 21.4%). The two interviews were separated by approximately two weeks (mean=13.5 days, SD=5.9). To reduce potential inflation of the reliability estimates, interviewers were prohibited from discussing the interviews with one another. Interviewers that conducted the assessments reviewed (edited) their own interviews. The interviews were then reviewed in detail (cross-edited) by experienced interviewers who were not otherwise involved in the study.

2.4 Statistical analyses

We evaluated the inter-observer reliability of DSM-5 SUD diagnoses for alcohol, opioid, cocaine and cannabis using the kappa (κ) coefficient, a measure of chance-corrected agreement (Cicchetti and Sparrow, 1981; Cohen, 1960) and its 95% confidence interval (95% CI) to assess the precision of the estimates (Bland and Altman, 1986; Kraemer et al., 2012). We excluded sedative, stimulant, and hallucinogen use disorders from the analyses due to their low prevalence in the sample. Because the SSADDA does not include the "Craving" criterion for tobacco use disorder, that diagnosis was also excluded from the analyses.

Although the κ statistic has been criticized because it is affected by the prevalence of the disorder, κ remains the most common statistic to measure reliability in psychiatric diagnoses (Regier et al., 2013). Although Yule's Y is more independent of prevalence (Spitznagel and Helzer, 1985), it can be unstable when prevalence is very low or very high (Ridenour and Heath, 1999; Spitznagel and Helzer, 1985). For the present study, we calculated both κ and Yule's Y for the disorders with the lowest prevalence in our sample (opioid use disorder and cannabis use disorder) and found similar values with both approaches. Further, Bowker's test of symmetry showed that none of the κ values were asymmetric. Thus, κ appears to be a

valid measure of agreement in this sample (Sim and Wright, 2005; Spitznagel and Helzer, 1985).

To meet criteria for a DSM-5 SUD, subjects had to endorse at least two of 11 diagnostic criteria. We also evaluated the reliability according to the severity of the DSM-5 SUD diagnosis, i.e., mild (2–3 criteria met), moderate (4–5 criteria met), and severe (6 or more criteria met). We treated the severity of SUD as an ordinal variable: i.e., no SUD (0–1 criteria), mild SUD (2–3 criteria), moderate SUD (4–5 criteria), or severe SUD (6 or more criteria). We treated the change in severity between the first and the second interview to be the same irrespective of the magnitude of change (e.g., a change from mild to moderate was equivalent to a change from mild to severe). For each substance, we calculated the percentage of discrepancies between the first and the second interview across levels of severity (no, mild, moderate, severe) and used Cohen's weighted kappa to evaluate reliability (Cohen, 1968).

We considered reliability to be poor when the κ coefficient was 0.40, fair when it was 0.41–0.59, good when it was 0.60–0.74, and excellent above 0.75 (Fleiss et al., 2003). When discrepancies occurred, we evaluated the direction of change in the diagnosis between the first and second interviews with a McNemar test.

We also examined the reliability of the individual SUD items, using the entire sample irrespective of prior substance use history (in contrast to our analysis of DSM-IV criteria, which we limited to individuals with lifetime use of a substance; Pierucci-Lagha et al., 2007). Then, we compared the inter-observer reliability of DSM-IV SUD diagnoses (i.e., either dependence or abuse) and DSM-IV dependence alone (as previously reported; Pierucci-Lagha et al., 2007, 2005) with those of the DSM-5 SUD diagnoses (Donner et al., 2000). Finally, we assessed the effect of having one or more additional SUDs (other than tobacco use disorder) and a lifetime mood or anxiety disorder on DSM-5 SUD diagnostic reliability. Lifetime mood and anxiety disorders were assessed with the SSADDA, using DSM-IV criteria. All analyses were performed using JMP® 10.0 (SAS® Institute Inc., Cary, North Carolina).

3. RESULTS

3.1 Sample characteristics

The sample comprised 87 men (50.3%) and 86 women (49.7%), with an average age of 37.5 years (SD=10.4). Subjects were predominantly European-American (38.7%), African-American (43.9%) or Hispanic (11.2%), and never married (62.4%). Most subjects (68.2%) had at least a high school education (mean years of education=12.9, SD=2.4), and 19.2% reported having a current full-time job. Most subjects (75.3%) had at least one SUD, which most commonly involved alcohol, followed by cocaine, cannabis and opioids (Table 1). The majority of subjects with a SUD (80.8%) also had at least one other SUD (other than tobacco use disorder). About one-quarter (n=47, 27.3%) of the sample had at least one lifetime mood disorder (major depressive disorder, bipolar disorder) (n=28, 16.5%) and one-sixth had a lifetime anxiety disorder (panic disorder, agoraphobia, obsessive-compulsive disorder, post-traumatic stress disorder) (n=29, 16.8%).

As expected, given the different settings from which participants were recruited (i.e., addiction treatment facilities, psychiatric services, the community), the three sub-samples differed significantly on several variables (Supplemental Table S1¹). Patients in addiction treatment facilities were more likely to be male, African-American and less educated. Participants from the community were younger, more likely never to have married, more educated, and more likely to have a full-time job. As would be expected, participants with at least one SUD (alcohol, opioid, cocaine, or cannabis) were more likely to be enrolled in addiction treatment facilities, while participants enrolled in psychiatric services were more likely to have a psychiatric disorder (mood disorder, anxiety disorder).

3.2 Reliability of DSM-5 Substance Use Disorder diagnoses

Table 1 displays the inter-observer reliability estimates for the lifetime DSM-5 diagnoses of alcohol, opioid, cocaine and cannabis use disorders in 173 subjects. Across substances, the rate of endorsement for each criterion differed between the first and second interviews (Supplemental Table S2²), generally with more subjects endorsing criteria at the first interview. However, in no case was the modest change in prevalence statistically significant (McNemar test p-values all >0.06).

The values of κ for the DSM-5 SUD diagnoses ranged from good for cannabis ($\kappa=0.65$) to excellent for alcohol ($\kappa=0.81$), cocaine ($\kappa=0.92$) and opioids ($\kappa=0.94$) (Table 1). Reliability across the different levels of severity of DSM-5 SUD (no, mild, moderate, or severe SUD) was lower than for the overall diagnoses, being fair for cannabis ($\kappa=0.49$), good for alcohol ($\kappa=0.61$) and cocaine ($\kappa=0.71$), and excellent for opioids ($\kappa=0.85$) (Table 1). The discrepancies between the first and second interviews were 3.0% for opioid, 17.9% for cocaine, 23.1% for alcohol and 27.4% for cannabis. Most of the discrepancies were between mild and moderate severity, explaining 77.5% of the discrepancies for alcohol, 60.0% for opioids, 56.5% for cannabis and 48.4% for cocaine.

3.3 Reliability of the individual DSM-5 Substance Use Disorder criteria

As was true for the diagnoses, values of κ were good to excellent for opioid and cocaine use disorder criteria taken individually (Table 2). For alcohol use disorder criteria, although most of the κ values were good, three criteria were only fair (“Great deal of time spent to obtain, use or recover,” “Continuing use despite persistent or recurrent physical or psychological problems,” and “Tolerance”). For cannabis use disorder criteria, all but one of the values of κ were fair, one was excellent (“Craving or strong desire or urge to use”), and one was poor (“Continuing use despite persistent, or recurrent physical or psychological problems”).

Individual criteria showed more variability in κ values than did the diagnoses across the substances (Table 2).

¹Supplementary material can be found by accessing the online version of this paper at <http://dx.doi.org> and by entering doi:...

²Supplementary material can be found by accessing the online version of this paper at <http://dx.doi.org> and by entering doi:...

3.4 Comparison of the reliability of DSM-5 Substance Use Disorders with DSM-IV diagnoses of Abuse or Dependence or Dependence only

Generally, the percent agreement between the first and second interviews was higher for the DSM-5 SUD diagnoses than for DSM-IV abuse or dependence diagnoses, or for DSM-IV dependence alone (Table 3). One exception to this was for cannabis, which showed similar agreement for DSM-5 cannabis use disorder and DSM-IV cannabis dependence only (83.3% versus 84.4%, respectively) (Table 3). Values of κ were also generally higher for DSM-5 SUD diagnoses than for DSM-IV abuse or dependence diagnoses and for DSM-IV dependence only (Table 3). However, none of the differences reached statistical significance.

3.5 The effect of additional substance use and lifetime mood and anxiety disorders on the reliability of DSM-5 Substance Use Disorder diagnoses

The presence of an additional SUD modestly decreased the percent agreement and values of κ between the two interviews (Table 4). Similarly, a mood or anxiety disorder modestly decreased the percent agreement and values of κ (Table 4). One exception to this was for opioid use disorder, where the presence of a mood or anxiety disorder increased the reliability (Table 4).

The presence of an additional SUD changed the reliability level for alcohol and cannabis. When an additional SUD was present, the κ coefficient for alcohol use disorder fell from excellent to fair, and the κ coefficient for cannabis use disorder fell from good to fair (Table 4).

4. DISCUSSION

The inter-observer reliability of the DSM-5 diagnoses of alcohol, opioid, cocaine and cannabis use disorders was good-to-excellent. When evaluated in relation to the severity of SUD (mild, moderate, or severe), the reliabilities, although good to excellent for all substances except cannabis, were modestly lower than for any SUD. Most of the discrepancies between the two interviews occurred between the categories of mild and moderate SUD severity: the greater the prevalence of mild SUD diagnoses, the poorer the concordance. This is consistent with previous findings, in which the only consistent predictor of lower reliability for DSM-IV alcohol and cocaine dependence diagnoses was low severity of dependence (Hasin et al., 1997).

Our finding that, overall, the reliability for cannabis use disorder criteria was lower than for the other substances is also consistent with previous studies showing that the reliability of DSM-IV dependence criteria ranged from good to excellent for all substances except cannabis (Bucholz et al., 1995; Feinn et al., 2009; Hasin et al., 2006; Pierucci-Lagha et al., 2007). This could be an artifact of the SSADDA, as unlike the alcohol, cocaine, and opioid sections, each of which is devoted to one substance, the cannabis section of the SSADDA is contained in a general “other drugs” section that elicits information on use of substances other than alcohol, cocaine, and opioids. Although in the “other drugs” section of the SSADDA all criteria are elicited independently for cannabis, stimulant, and sedative use, the lack of a section focusing specifically on cannabis could have reduced the reliability of the

cannabis use disorder diagnosis. A recent study in the general population using the Alcohol Use Disorder and Associated Disabilities Interview Schedule-5 (AUDADIS-5), which was designed to assess SUDs according DSM-5 criteria, showed that cannabis use disorder was among the least reliable diagnoses ($\kappa=0.41$) (Grant et al., 2015). Also, in our sample, most of the participants who met cannabis use disorder also had another SUD (i.e., 35% of the 39% that met criteria for a DSM-5 cannabis use disorder). Thus, in the context of an alcohol, cocaine or opioid use disorder, subjects may have misattributed their cannabis-related symptoms to the effects of these other substances.

We found that DSM-5 SUD diagnoses were more reliable than DSM-IV dependence diagnoses. This was due, in part, to the fact that there is a greater probability of endorsing 2 or more of the 11 criteria required for a DSM-5 SUD diagnosis than 3 or more of the 7 criteria required for a DSM-IV dependence diagnosis. DSM-5 diagnoses were also more reliable than the DSM-IV diagnosis of abuse or dependence. This is also consistent with the greater probability of endorsing 2 or more of 11 DSM-5 criteria than 3 or more of the 7 DSM-IV dependence criteria or 1 of the 4 DSM-IV abuse criteria (when the dependence diagnosis was not met). Because the κ statistic corrects for chance agreement and the large majority of criteria across the four substances (alcohol, opioids, cocaine, and cannabis) showed good-to-excellent reliabilities, it is unlikely that random effects can explain the differences in reliability between the two diagnostic systems. The lack of significant differences between rates of endorsement of DSM-5 criteria between the first and second interviews argues against an order effect as a contributor to the observed reliability findings.

Previous studies assessing the reliability of DSM-IV SUD diagnoses have generally shown that abuse is a less reliable diagnosis than dependence (Hasin et al., 1997; Pierucci-Lagha et al., 2005). As DSM-5 combines the criteria for DSM-IV abuse and dependence into a single disorder in which abuse and dependence are no longer defined as hierarchical, the potential for discordance between interviews was reduced. Further, we found that the reliability of the individual DSM-IV abuse criteria was comparable to that of the DSM-IV dependence criteria, all ranging from good to excellent, consistent with prior research (Hasin et al., 1997, 2006). As in previous studies of the reliability of DSM-IV SUD criteria (Bucholz et al., 1995; Pierucci-Lagha et al., 2007), we also found that the values of κ for the individual DSM-5 criteria appeared not to influence the reliability of DSM-5 diagnoses.

Finally, we found that the presence of additional substance use, mood, or anxiety disorders only modestly affected the reliability of the DSM-5 SUD diagnosis. Because our sample consisted mainly of subjects with multiple SUDs, subjects may have misattributed symptoms to one substance when they were, in fact, due to a different substance. Although theoretically this could reduce the reliability of some SUD criteria and diagnoses, it appeared that its effect here was negligible. The observed modest effect of mood and anxiety disorders on the reliability of SUD diagnoses was inconsistent with the findings of the DSM-5 Field Trials, which showed that the presence of co-occurring major depressive disorder and generalized anxiety disorder among participants with an alcohol use disorder reduced the reliability of the alcohol diagnosis (Regier et al., 2013).

Our study has limitations that require consideration. First, this was a secondary analysis of data collected previously for a study that assessed the reliability of DSM-IV criteria. Despite some major conceptual changes from DSM-IV to DSM-5, there were few changes to the SUD criteria. This enabled us to use the data to evaluate the reliability of DSM-5 criteria and to compare that with the reliability of DSM-IV diagnoses. Although the differences in reliability between DSM-IV and DSM-5 might not change much if the study were repeated, due to random variation the estimated reliability of SUDs defined by both DSM-IV and DSM-5 could be reduced. This is a theoretical issue that our data do not allow us to address. Second, the sample was recruited from different settings (addiction treatment facilities, psychiatric services, and the community). The resulting heterogeneity in the sample supports the external validity of the results. However, most of the participants recruited from settings other than addiction treatment facilities did not meet SUD criteria and most of the participants with an additional psychiatric disorder were recruited from psychiatric services. The sample was too small to evaluate statistically the impact of the recruitment settings on estimated reliability. Also, because we excluded individuals with severe psychiatric disorders, there was a low prevalence of such individuals in the sample, with the comorbidity largely restricted to major depressive disorder and anxiety disorders. This limited our capacity to assess the effect of more diverse or severe psychiatric disorders on the reliability of the DSM-5 SUDs. We also used only one diagnostic interview, the SSADDA, which limits our ability to generalize the findings regarding the reliability of DSM-5 SUDs diagnoses using other instruments. Finally, we limited the analysis to four DSM-5 SUDs.

In summary, we demonstrated that the reliability of DSM-5 SUD diagnoses is at least as good as for DSM-IV SUDs. Further research is needed to evaluate the reliability of DSM-5 SUD diagnoses among individuals with a variety of additional psychiatric disorders and sedative, stimulant, inhalant, hallucinogen, and tobacco use disorders and in different populations and with different diagnostic instruments to replicate and extend these findings.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank the COGA investigators for generously allowing them to modify the SSAGA interview for our use. We also thank the participants and interviewers at the participating sites and Joseph Cubells, M.D., Ph.D. and Lindsay Farrer, Ph.D. for their contributions to the initial data collection and management, respectively. This study was supported by NIH grants DA12422, DA12849, DA12690, DA028874, and AA13736.

References

- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986; 1:307–310. [PubMed: 2868172]
- Borges G, Ye Y, Bond J, Cherpitel CJ, Cremonte M, Moskalewicz J, Swiatkiewicz G, Rubio-Stipec M. The dimensionality of alcohol use disorders and alcohol consumption in a cross-national perspective. *Addiction*. 2010; 105:240–254. [PubMed: 20078482]
- Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI Jr, Reich T, Schmidt I, Schuckit MA. A new, semi-structured psychiatric interview for use in genetic linkage

- studies: a report on the reliability of the SSAGA. *J Stud Alcohol*. 1994; 55:149–158. [PubMed: 8189735]
- Bucholz KK, Hesselbrock VM, Shayka JJ, Nurnberger JI Jr, Schuckit MA, Schmidt I, Reich T. Reliability of individual diagnostic criterion items for psychoactive substance dependence and the impact on diagnosis. *J Stud Alcohol*. 1995; 56:500–505. [PubMed: 7475029]
- Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *Am J Ment Defic*. 1981; 86:127–137. [PubMed: 7315877]
- Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Measure*. 1960; 20:37–46.
- Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull*. 1968; 70:213–220. [PubMed: 19673146]
- Compton WM, Thomas YF, Stinson FS, Grant BF. Prevalence, correlates, disability, and comorbidity of DSM-IV drug abuse and dependence in the United States: results from the national epidemiologic survey on alcohol and related conditions. *Arch Gen Psychiatry*. 2007; 64:566–576. [PubMed: 17485608]
- Conway KP, Compton W, Stinson FS, Grant BF. Lifetime comorbidity of DSM-IV mood and anxiety disorders and specific drug use disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *J Clin Psychiatry*. 2006; 67:247–257. [PubMed: 16566620]
- Donner A, Shoukri MM, Klar N, Bartfay E. Testing the equality of two dependent kappa statistics. *Stat Med*. 2000; 19:373–387. [PubMed: 10649303]
- Feinn R, Gelernter J, Cubells JF, Farrer L, Kranzler HR. Sources of unreliability in the diagnosis of substance dependence. *J Stud Alcohol Drugs*. 2009; 70:475–481. [PubMed: 19371499]
- Fleiss, JL.; Levin, B.; Paik, MC. *Statistical Methods For Rates And Proportions*. 3. John Wiley & Sons; Hoboken, NJ: 2003.
- Gilder DA, Gizer IR, Lau P, Ehlers CL. Item response theory analyses of DSM-IV and DSM-5 stimulant use disorder criteria in an American Indian community sample. *Drug Alcohol Depend*. 2014; 135:29–36. [PubMed: 24200103]
- Grant BF, Goldstein RB, Smith SM, Jung J, Zhang H, Chou SP, Pickering RP, Ruan WJ, Huang B, Saha TD, Aivadyan C, Greenstein E, Hasin DS. The Alcohol Use Disorder and Associated Disabilities Interview Schedule-5 (AUDADIS-5): reliability of substance use and psychiatric disorder modules in a general population sample. *Drug Alcohol Depend*. 2015; 148:27–33. [PubMed: 25595052]
- Hasin D, Carpenter KM, McCloud S, Smith M, Grant BF. The Alcohol Use Disorder And Associated Disabilities Interview Schedule (AUDADIS): reliability of alcohol and drug modules in a clinical sample. *Drug Alcohol Depend*. 1997; 44:133–141. [PubMed: 9088785]
- Hasin D, Hatzenbuehler ML, Keyes K, Ogburn E. Substance use disorders: Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV) and International Classification of Diseases, tenth edition (ICD-10). *Addiction*. 2006; 101(Suppl 1):59–75. [PubMed: 16930162]
- Hasin DS, Fenton MC, Beseler C, Park JY, Wall MM. Analyses related to the development of DSM-5 criteria for substance use related disorders: 2. Proposed DSM-5 criteria for alcohol, cannabis, cocaine and heroin disorders in 663 substance abuse patients. *Drug Alcohol Depend*. 2012; 122:28–37. [PubMed: 21963333]
- Hasin DS, O'Brien CP, Auriacombe M, Borges G, Bucholz K, Budney A, Compton WM, Crowley T, Ling W, Petry NM, Schuckit M, Grant BF. DSM-5 criteria for substance use disorders: recommendations and rationale. *Am J Psychiatry*. 2013; 170:834–851. [PubMed: 23903334]
- Kelly SM, Gryczynski J, Mitchell SG, Kirk A, O'Grady KE, Schwartz RP. Concordance between DSM-5 and DSM-IV nicotine, alcohol, and cannabis use disorder diagnoses among pediatric patients. *Drug Alcohol Depend*. 2014; 140:213–216. [PubMed: 24793367]
- Kerridge BT, Saha TD, Smith S, Chou PS, Pickering RP, Huang B, Ruan JW, Pulay AJ. Dimensionality of hallucinogen and inhalant/solvent abuse and dependence criteria: implications for the Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition. *Addict Behav*. 2011; 36:912–918. [PubMed: 21621334]
- Keyes KM, Krueger RF, Grant BF, Hasin DS. Alcohol craving and the dimensionality of alcohol disorders. *Psychol Med*. 2011; 41:629–640. [PubMed: 20459881]

- Kraemer HC, Kupfer DJ, Clarke DE, Narrow WE, Regier DA. DSM-5: how reliable is reliable enough? *Am J Psychiatry*. 2012; 169:13–15. [PubMed: 22223009]
- Mewton L, Slade T, Teesson M. An evaluation of the proposed DSM-5 cannabis use disorder criteria using Australian national survey data. *J Stud Alcohol Drugs*. 2013; 74:614–621. [PubMed: 23739026]
- Peer K, Rennert L, Lynch KG, Farrer L, Gelernter J, Kranzler HR. Prevalence of DSM-IV and DSM-5 alcohol, cocaine, opioid, and cannabis use disorders in a largely substance dependent sample. *Drug Alcohol Depend*. 2013; 127:215–219. [PubMed: 22884164]
- Pierucci-Lagha A, Gelernter J, Chan G, Arias A, Cubells JF, Farrer L, Kranzler HR. Reliability of DSM-IV diagnostic criteria using the semi-structured assessment for drug dependence and alcoholism (SSADDA). *Drug Alcohol Depend*. 2007; 91:85–90. [PubMed: 17590536]
- Pierucci-Lagha A, Gelernter J, Feinn R, Cubells JF, Pearson D, Pollastri A, Farrer L, Kranzler HR. Diagnostic reliability of the Semi-structured Assessment for Drug Dependence and Alcoholism (SSADDA). *Drug Alcohol Depend*. 2005; 80:303–312. [PubMed: 15896927]
- Preuss UW, Watzke S, Wurst FM. Who/ISBRA Study on State, Trait Markers of Alcohol Use Dependence Investigators. Dimensionality and stages of severity of DSM-5 criteria in an international sample of alcohol-consuming individuals. *Psychol Med*. 2014; 44:3303–3314. [PubMed: 25065388]
- Proctor SL, Kopak AM, Hoffmann NG. Compatibility of current DSM-IV and proposed DSM-5 diagnostic criteria for cocaine use disorders. *Addict Behav*. 2012; 37:722–728. [PubMed: 22386237]
- Regier DA, Narrow WE, Clarke DE, Kraemer HC, Kuramoto SJ, Kuhl EA, Kupfer DJ. DSM-5 field trials in the United States and Canada, Part II: test-retest reliability of selected categorical diagnoses. *Am J Psychiatry*. 2013; 170:59–70. [PubMed: 23111466]
- Ridenour TA, Heath AC. A note on issues in meta-analysis for behavioral genetic studies using categorical phenotypes. *Behav Genet*. 1999; 29:155–162. [PubMed: 10547921]
- Saha TD, Compton WM, Chou SP, Smith S, Ruan WJ, Huang B, Pickering RP, Grant BF. Analyses related to the development of DSM-5 criteria for substance use related disorders: 1. toward amphetamine, cocaine and prescription drug use disorder continua using Item Response Theory. *Drug Alcohol Depend*. 2012; 122:38–46. [PubMed: 21963414]
- Sim J, Wright CC. The Kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther*. 2005; 85:257–268. [PubMed: 15733050]
- Spitznagel EL, Helzer JE. A proposed solution to the base rate problem in the kappa statistic. *Arch Gen Psychiatry*. 1985; 42:725–728. [PubMed: 4015315]
- Stinson FS, Grant BF, Dawson DA, Ruan WJ, Huang B, Saha T. Comorbidity between DSM-IV alcohol and specific drug use disorders in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Drug Alcohol Depend*. 2005; 80:105–116. [PubMed: 16157233]

Highlights

- We evaluated the reliability of DSM-5 substance use disorder (SUD) criteria
- DSM-5 alcohol, opioid, cocaine and cannabis use disorder diagnoses are reliable
- DSM-5 SUD criteria and diagnoses are at least as reliable as those of DSM-IV
- Reliability is lower for mild disorders than for those that are moderate or severe
- Additional SUD and mood/anxiety disorders have a modest effect on reliability

Table 1

Prevalence and inter-observer reliability for four Substance Use Disorder diagnoses (DSM-5 criteria)

	Prevalence n (%)		Reliability
	1 st interview	2 nd interview	κ (95% CI)
Alcohol Use Disorder			
Disorder present (2 criteria)	121 (69.9)	117 (67.6)	0.81 (0.72, 0.91)
No disorder (0–1 criterion)	52 (30.1)	56 (32.4)	0.61 (0.52, 0.69)
Mild disorder D (2–3 criteria)	35 (20.2)	37 (21.4)	
Moderate disorder (4–5 criteria)	22 (12.7)	24 (13.9)	
Severe disorder (6 criteria)	64 (37.0)	56 (32.4)	
Opioid Use Disorder			
Disorder present (2 criteria)	29 (17.5)	28 (16.9)	0.94 (0.87, 1.00)
No disorder (0–1 criterion)	137 (82.5)	138 (83.1)	0.85 (0.76, 0.95)
Mild disorder (2–3 criteria)	3 (1.8)	2 (1.2)	
Moderate disorder (4–5 criteria)	0 (–)	1 (0.6)	
Severe disorder (6 criteria)	26 (15.7)	25 (15.6)	
Cocaine Use Disorder			
Disorder present (2 criteria)	101 (58.4)	96 (55.5)	0.92 (0.86, 0.98)
No disorder (0–1 criterion)	72 (41.6)	77 (44.5)	0.71 (0.62, 0.79)
Mild disorder (2–3 criteria)	8 (4.6)	8 (4.6)	
Moderate disorder (4–5 criteria)	13 (7.5)	15 (8.7)	
Severe disorder (6 criteria)	80 (46.2)	73 (42.2)	
Cannabis Use Disorder			
Disorder present (2 criteria)	67 (39.9)	61 (36.3)	0.65 (0.53, 0.77)
No disorder (0–1 criterion)	101 (60.1)	107 (63.7)	0.49 (0.38, 0.59)
Mild disorder (2–3 criteria)	25 (14.9)	13 (7.7)	
Moderate disorder (4–5 criteria)	15 (8.9)	20 (11.9)	
Severe disorder (6 criteria)	27 (16.1)	28 (16.7)	

Note: κ = Kappa coefficient; 95% CI= 95% Confidence Interval

Table 2

Inter-observer reliability for the individual DSM-5 Substance Use Disorder criteria

DSM-5 Criteria	Alcohol κ (95% CI)	Opioids κ (95% CI)	Cocaine κ (95% CI)	Cannabis κ (95% CI)
Taken in larger amount or over a longer period	0.65 (0.54, 0.77)	0.67 (0.50, 0.85)	0.75 (0.65, 0.84)	0.57 (0.42, 0.71)
Persistent desire or unsuccessful efforts to cut down	0.63 (0.52, 0.75)	0.79 (0.66, 0.92)	0.83 (0.74, 0.91)	0.42 (0.21, 0.64)
Great deal of time spent to obtain, use or recover	0.54 (0.39, 0.68)	0.82 (0.70, 0.94)	0.77 (0.67, 0.86)	0.40 (0.24, 0.56)
Craving or strong desire or urge to use	0.69 (0.55, 0.82)	0.93 (0.84, 1.00)	0.69 (0.59, 0.80)	0.75 (0.60, 0.89)
Recurrent use resulting in a failure to fulfill major role obligations	0.72 (0.60, 0.83)	0.91 (0.81, 1.00)	0.60 (0.48, 0.73)	0.42 (0.23, 0.60)
Continued use despite persistent or recurrent social or interpersonal problems	0.60 (0.48, 0.72)	0.91 (0.83, 1.00)	0.79 (0.70, 0.88)	0.58 (0.43, 0.74)
Giving up or reducing important social, occupational, or recreational activities	0.70 (0.58, 0.81)	0.79 (0.65, 0.93)	0.74 (0.63, 0.84)	0.52 (0.35, 0.68)
Recurrent use in situations in which it is physically hazardous	0.74 (0.63, 0.84)	0.74 (0.58, 0.90)	0.75 (0.64, 0.86)	0.53 (0.38, 0.68)
Continued use despite persistent or recurrent physical or psychological problems	0.59 (0.45, 0.72)	0.85 (0.74, 0.97)	0.77 (0.67, 0.86)	0.31 (0.11, 0.51)
Tolerance	0.54 (0.42, 0.67)	0.74 (0.59, 0.88)	0.62 (0.50, 0.75)	0.57 (0.42, 0.72)
Withdrawal	0.66 (0.53, 0.79)	0.94 (0.86, 1.00)	0.75 (0.65, 0.85)	0.47 (0.31, 0.63)

Note: κ = Kappa coefficient; 95% CI= 95% Confidence Interval

Table 3

Comparison of inter-observer reliability between DSM-IV abuse or dependence, dependence alone and DSM-5 Substance Use Disorder diagnoses

	% Agreement			κ (95% CI)		
	DSM-IV (abuse/dependence)	DSM-IV dependence	DSM-5 SUD	DSM-IV (abuse/dependence)	DSM-IV dependence	DSM-5 SUD
Alcohol	90.2%	83.8%	91.9%	0.80 (0.71 – 0.89)	0.66 (0.54 – 0.77)	0.81 (0.72 – 0.91)
Opioid	97.1%	97.7%	98.2%	0.89 (0.80 – 0.98)	0.91 (0.82 – 1.00)	0.94 (0.87 – 1.00)
Cocaine	95.4%	91.3%	96.0%	0.91 (0.84 – 0.97)	0.83 (0.74 – 0.91)	0.92 (0.86 – 0.98)
Cannabis	80.3%	84.4%	83.3%	0.58 (0.45 – 0.70)	0.55 (0.40 – 0.70)	0.65 (0.53 – 0.77)

Note: κ = Kappa; 95% CI= 95% Confidence Interval; NS = Not significant

Table 4

Inter-observer reliability for the presence of a co-occurring lifetime Substance Use Disorder or Mood/Anxiety Disorder

	n	% Agreement	κ (95% CI)
Alcohol Use Disorder			
Whole sample	173	91.9%	0.81 (0.72, 0.91)
No other SUD	70	91.9%	0.81 (0.65, 0.97)
Other SUD	103	91.3%	0.53 (0.26, 0.79)
No Psychiatric Disorder	106	92.5%	0.84 (0.73, 0.94)
Mood/Anxiety Disorder	67	91.0%	0.75 (0.57, 0.94)
Opioid Use Disorder			
Whole sample	166	98.2%	0.94 (0.87, 1.00)
No other SUD	66	100%	1.00
Other SUD	100	97.0%	0.92 (0.84, 1.00)
No Psychiatric Disorder	102	97.1%	0.88 (0.75, 1.00)
Mood/Anxiety Disorder	64	100%	1.00
Cocaine Use Disorder			
Whole sample	173	96.0%	0.92 (0.86, 0.98)
No other SUD	70	100%	1.00
Other SUD	103	93.2%	0.63 (0.38, 0.88)
No Psychiatric Disorder	106	96.2%	0.92 (0.85, 1.00)
Mood/Anxiety Disorder	67	95.5%	0.89 (0.78, 1.00)
Cannabis Use Disorder			
Whole sample	168	83.3%	0.65 (0.53, 0.77)
No other SUD	66	98.4%	0.66 (0.04, 1.00)
Other SUD	102	73.5%	0.45 (0.28, 0.63)
No Psychiatric Disorder	104	84.6%	0.65 (0.49, 0.80)
Mood/Anxiety Disorder	64	81.3%	0.62 (0.43, 0.82)

Note: κ = Kappa value; 95% CI= 95% Confidence Interval; SUD= Substance Use Disorder (DSM-5 criteria) except Tobacco Use Disorder; Mood/Anxiety Disorder: at least one mood or anxiety disorder