Test-retest reliability of DSM-5 substance disorder measures as assessed with the PRISM-5, a clinician-administered diagnostic interview

Deborah Hasin[a,b,c,*], **Dvora Shmulewitz**[a,b], **Malka Stohl**[a], **Eliana Greenstein**[a], **Stephanie Roncone**[a], **Efrat Aharonovich**[a,b], **Melanie Wall**[a,b]

[a]New York State Psychiatric Institute, 1051 Riverside Dr, New York, NY 10032, USA

[b]Department of Psychiatry, Columbia University Medical Center, 1051 Riverside Dr, New York, NY, 10032, USA

[c]Department of Epidemiology, Mailman School of Public Health, Columbia University, 722 W 168th St, New York, NY, 10032, USA

Abstract

Aim: In DSM-5, the definitions of substance use disorders (SUD) were changed considerably, yet little is known about the reliability of DSM-5 SUD and its new features.

Methods: The test-retest reliability of DSM-5 SUD and DSM-IV substance dependence (SD) was evaluated in 565 adult substance users, each interviewed twice by different clinician interviewers using the semi-structured Psychiatric Research Interview for Substance and Mental Disorders, DSM-5 version (PRISM-5). DSM-5 SUD and DSM-IV SD criteria were assessed for past year and lifetime, yielding diagnoses and severity levels for alcohol, tobacco, cannabis, cocaine, heroin, opioids, sedatives, hallucinogen, and stimulant use disorders. Cohen's and intraclass correlation coefficients (ICC) assessed reliability for categorical and graded outcomes, respectively. Factors potentially influencing reliability were explored, including inpatient vs. community participant, days between interviews gender, age, race/ethnicity, and SUD severity.

Results: DSM-5 SUD diagnoses had substantial to excellent reliability for most substances (κ = 0.63–0.94), and moderate for others (hallucinogens, stimulants, sedatives; κ = 0.50 – 0.59). For graded outcomes (DSM-5 SUD mild, moderate, severe; criteria count 0–11), reliability was substantial to excellent (ICC = 0.74 – 0.99). Comparisons of DSM-5 SUD and DSM-IV SD reliability showed few significant differences. Reliability of the DSM-5 craving criterion was

excellent for heroin ($\kappa = 0.84 - 0.95$) and moderate to substantial for other substances ($\kappa = 0.49 -$ 0.76). The only factor influencing reliability of SUD was severity, with milder disorders significantly more likely to be discordant between the interviews.

Conclusion: Reproducibility is crucial to good measurement. In a large sample using rigorous methodology, diagnoses and dimensional measures from clinician-administered interviews for DSM-5 SUD were generally highly reliable.

Keywords

Reliability; DSM-5; Substance use disorders; Dependence; Test-restest

1. Introduction

Substance use and substance use disorders (SUD) are leading causes of preventable morbidity and mortality worldwide (Alcohol Collaborators, 2018; Vigo et al., 2019; Rehm and Shield, 2019; Kariisa et al., 2019; Scholl et al., 2018). Research to identify preventable risk factors and effective treatments for SUD requires reliable, valid SUD diagnoses and dimensional measures. Definitions in the Diagnostic and Statistical Manual of Mental Disorders (DSM) are widely used for clinical, research, and training purposes in the U.S. DSM has been revised several times, with the fifth revision, DSM-5, published in 2013 (American Psychiatric Association, 2013). Some disorders changed little in DSM-5 compared to earlier versions, but SUD changed considerably (Hasin et al., 2013). In DSM-III-R (American Psychiatric Association, 1987) and DSM-IV (American Psychiatric Association, 1994), two main substance disorders were defined, abuse and dependence. Categorical and dimensional measures of DSM-IV dependence were reliable in general population and clinical samples (Hasin et al., 2006a). For example, in the semi-structured, clinician-administered Psychiatric Research Interview for Substance and Mental Disorder (PRISM), moderate-to-excellent test-retest reliability and validity was demonstrated for dependence in patients treated for psychiatric and/or substance use disorders (Hasin et al., 2006b; Hasin et al., 1996; Torrens et al., 2004; Hasin et al., 1998) and HIV (Morgello et al., 2006). However, across studies, abuse was less reliable, although its criteria were unidimensional with the dependence criteria (Hasin et al., 2013). Therefore, in a major change, DSM-5 combined most criteria of the two disorders into a single disorder, SUD (Fig. 1). Additional DSM-5 changes included new criteria: craving, a withdrawal criterion for cannabis use disorders, and three "abuse" criteria for tobacco use disorder. Little is known about the reliability of DSM-5 SUD or these new additions.

While most DSM-5 substance disorder changes were based on empirical analyses, the data that were used were 10–20 years old (Hasin et al., 2013). Since 2013, a study of lifetime DSM-5 SUD in clinical and community participants found moderate to excellent test-retest inter--rater reliability (Denis et al., 2015). However, the study did not address current (past-year) diagnoses, and further, this was a secondary analysis of data collected over 15 years ago. In recent years, many aspects of substance use patterns (Hasin et al., 2019a; Hasin et al., 2019b; Hasin, 2018; Martins et al., 2017; Kerridge et al., 2019; Saha et al., 2016), prescriptions (Frenk et al., 2015; Centers for Disease Control and Prevention, 2018), potency (Spencer et al., 2019; Hedegaard et al., 2018; Spindle et al., 2019; Chandra et al., 2019;

ElSohly et al., 2016; Mehmedic et al., 2010), and laws (Hasin, 2018; Cerda et al., 2019) have changed. Therefore, reliability information is needed from more recent samples. In a 2012–2013 study of test-retest inter-rater reliability, 1006 adult household residents were assessed by lay interviewers using a fully-structured interview designed to produce DSM-5 substance disorder categorical and dimensional measures (Grant et al., 2015a), the Alcohol Use Disorder and Associated Disabilities Schedule-5 (AUDADIS-5). Past-year DSM-5 substance diagnoses had moderate to excellent reliability, with higher reliabilities for dimensional measures. However, this study did not examine lifetime diagnoses, and some substance disorders were too rare to examine (e.g. heroin). Also, the AUDADIS-5 study did not provide information about semi-structured clinician-administered procedures, which are commonly used in clinical and genetic studies. Therefore, in a study using the PRISM-5 (DSM-5 version), we addressed these questions: (1) What is the test-retest inter-rater reliability of DSM-5 current and lifetime SUD binary diagnoses, severity indicators, and new criteria, i.e., craving, cannabis withdrawal, new tobacco criteria? (2) How is DSM-5 reliability affected by days between the interviews (range 1–21) and participant characteristics, i.e., inpatient vs community; sex, age, race/ethnicity; and SUD severity. Furthermore, because of the extensive scientific literature based on the DSM-IV dependence diagnoses, we also examine how the reliability of DSM-5 SUD compares to DSM-IV dependence.

2. Methods

2.1. Sample and procedures

Participants were adults age 18 years with substance use (cannabis, cocaine, opioid use or binge drinking [5 drinks or 4 drinks for men and women respectively) in the prior 30 days or the 30 days prior to inpatient admission, who endorsed at least one DSM-5 substance use disorder criteria in pre-study screening. In a suburban inpatient addiction program, potentially eligible participants were informed about the study by hospital staff or posted flyers; those interested met with an on-site research coordinator who described the study, screened for eligibility and obtained informed consent. In an urban medical center, potentially eligible community participants were recruited through two types of advertisements. Newspapers ads invited individuals to call a research coordinator for a brief explanation of the study and eligibility prescreening. Facebook ads provided a brief explanation of the study and a link to click to an online prescreening survey. Potentially eligible participants via both type of advertisements then came to the research offices to meet with a research coordinator for further screening and consent. Exclusion criteria were minimal: non-English-speaking; impairment in hearing, vision, or cognition that precluded participation; plans to leave the area within six months (required for follow-ups, not covered in this report); or actively psychotic, suicidal, or homicidal, based on inpatient medical records and items adapted from the Addiction Severity Index and AUDADIS-5.

Initial interviews were conducted with 588 participants (150 in-patients; 438 community participants). Of these, 565 (96 %) participated in retest interviews (149 inpatients, 416 community participants) and constituted the analytic sample. Participants received $50 for the initial interview and $35 for the re-test. All interviews were conducted between

05/11/2016 and 06/30/2019. Procedures were approved by Institutional Review Boards of New York State Psychiatric Institute and South Oaks Hospital. Based on participant or site preference, interviews were face-to-face (564 for test, 434 for retest) or by telephone (1 for test, 131 for retest). The mean time for baseline interviews was 74.66 min (SD = 35.91) and the mean time for retest interviews was 65.72 min (SD = 35.57). No participant indicated any discomfort in being interviewed over the telephone. Retest interviewers were performed by a different interviewer who always blind to the initial interview. Prior to starting, retest interviewers told participants that the purpose was to help understand the quality of the initial interview, that they did not know participants' responses in the first interview and that participants should respond with "whatever answer seems right to you today. Don't try to make your answers the same as last time, or different - just give the answer that seems right to you now." The mean test–retest interval was 5.27 days (s.d. = 4.26), median = 5, range, 1–21 days, interquartile range = 5.

2.2. Diagnostic interview

The PRISM-5 interview is a semi-structured, computer-assisted interview designed for clinician interviewers. PRISM-5 covers the symptoms and criteria of substance and psychiatric disorders in detail, and based on this information, produces DSM diagnoses via computer algorithms. "Semi"-structured indicates that PRISM-5 items are initially asked as written. However, in contrast to fully structured interviews, PRISM-5 interviewers may then ask their own unstructured follow-up probes to obtain more information and clarify responses. PRISM-5 assesses DSM-IV and DSM-5 criteria for substance use disorders, and DSM-5 criteria for psychiatric disorders (to be reported elsewhere). PRISM-5 differs from other diagnostic interviews by assessing substance disorders first, and by providing more detailed symptom data. Substances include alcohol, cannabis, cocaine, hallucinogens, heroin, opioid painkillers, sedatives/tranquilizers, stimulants, and tobacco (cigarettes).

2.3. Interviewers, training and supervision

Interviewers had at least a master's degree in psychology or social work, and an average of 4.5 years of clinical experience with substance and/or psychiatric patients (range, 1–10 years). PRISM-5 training includes a manual, 2-day workshop, practice interviewing, role-playing, and certification. To become certified, trainees record mock interviews reviewed by PRISM trainers using structured interviewer scoresheets. Trainees only became interviewers after 2 recordings were rated satisfactory or better. Supervision of the study interviewers was conducted by trained, highly experienced supervisors with clinical masters degrees (psychology or social work). Their mean years of clinical experience utilizing the PRISM interview in research settings was 7.6 years (range: 3–10), and they had a mean of 6 years of supervisory experience (range: 2–8). After certification, PRISM-5 supervisors met weekly with interviewers (22 over the course of the study). PRISM supervisors also reviewed recordings of 10 % of the interviews, scoring interviewer performance on the same interviewer scoresheet used to rate trainee interviewers for certification. Review of the recordings indicated occasional typical issues in such interviewing that are routinely addressed in supervision (e.g., reading questions exactly as written) but that overall, the PRISM-5 interviews were largely conducted in a standardized way according to the PRISM-5 training procedures.

2.4. Measures

Screening.—The PRISM-5 substance disorder module begins with screening questions that ask about use of each substance at least 6 times within any 12-month period. For substances available for medical use, the screening questions ask about non-medical use (without a prescription or other than prescribed, for example, to get high). Among those screened positive, the substance disorder criteria are assessed.

Diagnoses and severity measures (Fig. 1).—Diagnoses for current (past-year) DSM-5 SUD are made among participants endorsing 2 criteria of the 11 criteria in the past 12 months. Past diagnoses require 2 criteria of the 11 criteria within a 12-month period prior to the past year. Lifetime diagnoses require a current or past diagnosis. Two severity indicators are also generated: dimensional criteria count (range, 0–11), and DSM-5 severity (no disorder, mild, moderate and severe disorder; 0–1, 2–3, 4–5, and 6 criteria, respectively). DSM-IV substance dependence was diagnosed when 3 of the DSM-IV criteria were endorsed within a 12-month period, with current and lifetime designations made similarly to DSM-5.

SUD criteria new to DSM-5 (Craving, Cannabis Withdrawal, Tobacco).—*Craving.* For each substance and timeframe, the craving criterion is assessed using two questions used previously in national surveys (Grant, 1992; Muthen et al., 1993; Keyes et al., 2011; Grant et al., 2016; Grant et al., 2015b) and earlier PRISM (Hasin et al., 2012) versions: feeling a very strong desire or urge to use the substance; and wanting to use it so badly one couldn't think of anything else. A positive response to either question was considered positive for the craving criterion. *Cannabis withdrawal.* DSM-IV did not include withdrawal as a criterion for cannabis, but it was added in DSM-5. PRISM-5 assessment of cannabis withdrawal is done similarly to assessment of withdrawal from other substances. Participants are first asked about lifetime occurrence of each specific withdrawal symptom when the substance effects were wearing off. For cannabis, these include being irritable, angry or aggressive; decreased weight or appetite; physical symptoms; nervousness or anxiety; restlessness; sleep disturbance; and low mood or depression. If at least three of these occurred, participants were asked if they occurred at the same time and were associated with impaired functioning. If the lifetime withdrawal criterion was met, participants were asked if the syndrome occurred in the last twelve months. Consistent with DSM-5, the withdrawal criterion could also be met by using the substance to relieve or avoid withdrawal symptoms. *New tobacco criteria.* DSM-IV included nicotine dependence, but not abuse. In DSM-5, tobacco use disorder was defined by DSM-IV nicotine dependence criteria plus the "abuse" criteria (Fig. 1) of hazardous use, use despite social/interpersonal problems, and neglecting social roles in order to smoke. These additional criteria were included in the PRISM-5 with questions similar to those for other substances.

2.5. Statistical analysis

Analyses were carried out for each substance, for past year and lifetime timeframes, using SAS 9.4.

2.5.1.   Test-retest inter-rater reliability—For binary variables (e.g., diagnoses, individual criteria), Cohen's kappa κ (Cohen, 1960) indicated chance-corrected agreement between test and retest interviews. Diagnoses with prevalence too low to yield stable results (<2% in test or retest interviews, e.g., club drugs) were not analyzed. For severity measures (DSM-5 severity, dimensional criteria counts), the intraclass correlation coefficient (ICC) was used. We used the standard one-way random effects ANOVA model to derive ICCs (Fleiss et al., 2003; Shrout and Fleiss, 1979), using the SAS *intracc* macro (SAS Institute Inc., 2020), and calculated 95 % confidence intervals (CI) using 100 bootstrapped samples. Cohen's κ and ICC values range from +1.00 to −1.00, with zero indicating chance-level agreement. Their interpretation is as follows (Davies and Fleiss, 1982): 0–0.20, poor; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and 0.81–1.00, excellent reliability (Landis and Koch, 1977).

2.5.2.   Differences between DSM-IV dependence and DSM-5 substance use disorder—Given the non-independence of DSM-IV dependence and DSM-5 substance use disorder within individuals, differences in their κ s and ICCs were tested by generating 100 bootstrapped samples (Efron and Tibshirani, 1986; DiCiccio and Efron, 1996; Efron, 1979), calculating the difference in κ or ICC between DSM-IV dependence and DSM-5 substance use disorder within each sample, and then the mean difference (and 95 % bootstrapped CI) in κ or ICC across the bootstrapped samples. Reliability of DSM-IV and DSM-5 measures was considered significantly different if the 95 % CI of the mean difference did not include 0.0.

2.5.3.   Factors potentially affecting reliability—We assessed whether type of participant (inpatient vs. community), days between test and retest interviews, telephone vs. in-person interviews, participant demographic characteristics (gender, age, race/ethnicity), and severity of the SUD in the initial interview were factors associated with great likelihood of discordant DSM-5 substance disorder diagnoses in the test and retest interviews. A dichotomous variable was created indicating concordant or discordant SUD diagnoses across the test and retest interviews. This variable was used as a binary outcome in logistic regression models with the factor of interest as the main predictor. For analyses of whether diagnostic severity was associated with discordance between the test and retest interviews, only participants with at least one criterion for that substance in the initial interview were included. Gender, age, and race/ethnicity were included as control variables in all logistic regression models when assessing these factors, except when the demographic characteristic itself was being assessed.

3.   Results

Sample.—Overall characteristics (Supplementary): 69.4 % were men, 37.8 % <40 years old, 72.4 % non-white; 80.0 % single; and 25.7 % employed; 54.5 % had only high school education; 84.3 % had public health insurance, and 8.5 % were in group housing or homeless. In-patients were younger; more likely to be white; more likely to be employed, yet also more likely to be homeless. The prevalence of lifetime substance use varied, with high rates of alcohol, cannabis, cocaine and tobacco (96.5 %, 85.5 %, 72.9 % and 79.5 %,

respectively) and lower rates of hallucinogens, heroin, non-medical use of prescription opioids, sedatives and stimulants (33.8 %, 35.0 %, 37.2 %, 28.7 %, and 15.8 %, respectively).

3.1. Reliability of DSM-5 substance use disorders (Table 1)

For current (past-year) disorders, reliability was substantial to excellent for alcohol, tobacco, cannabis, cocaine, heroin, and prescription opioids use disorders ($\kappa = 0.67 – 0.94$), and moderate ($\kappa = 0.50 – 0.59$) for hallucinogen, stimulant, and sedative use disorders. Reliability of lifetime disorders was substantial to excellent for alcohol, tobacco, cannabis, cocaine, heroin, prescription opioid and sedative use disorders ($\kappa = 0.63 – 0.92$), and moderate for hallucinogen and stimulant use disorders ($\kappa = 0.53$). Reliability of the individual DSM-5 criteria for each substance is shown in Supplementary Tables 2a (current) and 2b (lifetime); with few exceptions, these were within the same range as the reliabilities of the full diagnoses.

3.2. Reliability of severity measures (Table 1)

Whether assessed by DSM-5 severity (mild, moderate, severe) or by the dimensional measure (range, 0–11), for most substance disorders, reliability of current severity measures was excellent ($ICC = 0.81 – 0.97$), while reliability of current hallucinogen and sedative disorder severity was substantial to excellent ($ICC = 0.74\ 0.83$). For lifetime dimensional measures, reliability was excellent ($ICC\ 0.81 – 0.99$) for most substances, and substantial for DSM-5 hallucinogen and stimulant severity ($ICC = 0.76, 0.75$). For reference, the prevalence of each of the substance disorders by severity level in the test and retest interviews is shown in Supplementary Table 10.

3.3. Reliability of DSM-5 craving (Table 2)

Reliability of the current (past-year) craving criterion was excellent for heroin ($\kappa = 0.89$), substantial for alcohol, tobacco, cannabis, cocaine, and opioids ($\kappa = 0.66 – 0.76$), and moderate for current sedatives and hallucinogens ($\kappa = 0.55 – 0.57$). Reliability of the lifetime craving criterion was substantial to excellent for heroin, alcohol, tobacco, cannabis, cocaine, opioids and sedatives ($\kappa = 0.67 – 0.95$), and moderate for hallucinogens and stimulants ($\kappa = 0.50 – 0.54$). Reliabilities for the two separate craving questions were similar to the reliability of the overall craving criterion.

3.4. Reliability of DSM-5 cannabis withdrawal and tobacco "abuse" criteria (Table 3)

The reliability of lifetime cannabis withdrawal was moderate ($\kappa = 0.55$), as was the cannabis withdrawal syndrome and use of cannabis to avoid or relieve withdrawal symptoms ($\kappa = 0.46 – 0.55$). Kappas for individual lifetime cannabis withdrawal symptoms were also moderate ($\kappa = 0.48 – 0.58$). Reliability for past-year cannabis withdrawal indicators was also in the moderate range ($\kappa = 0.41 – 0.53$). For the three new tobacco criteria, reliability was fair for social/interpersonal problems due to use ($\kappa = 0.34 – 0.37$), moderate for neglected major roles in order to smoke ($\kappa = 0.42 – 0.52$), and substantial for hazardous use ($\kappa = 0.67 – 0.73$).

3.5. Reliability of DSM-IV substance dependence and DSM-5 substance use disorder compared (Table 4)

Table 4 shows the pairs of mean κs and ICCs for DSM-IV substance dependence and DSM-5 substance use disorder, binary and dimensional measures, generated across the 100 bootstrapped samples, the mean differences between the pairs, and the 95 % CIs around these differences (Supplementary Table 3 shows the reliability coefficients for DSM-IV substance dependence generated the standard way, which were identical or nearly-identical to those in Table 4). κs for current and lifetime DSM-IV dependence ranged from moderate to excellent (κ = 0.45 – 0.92). Of the nine potential pairs of κs for current binary DSM-IV and DSM-5 diagnoses in Table 4, eight had prevalences that could be tested (hallucinogens were too rare). Six of the eight did not differ significantly. The two pairs that differed were tobacco (DSM-IV, κ = 0.67; DSM-5, κ = 0.79) and cannabis (DSM-IV, κ = 0.62; DSM-5, κ = 0.74), with DSM-5 κs higher. For lifetime diagnoses, of the nine pairs of κs, only tobacco differed significantly (DSM-IV, κ = 0.64; DSM-5, κ = 0.74), with κ higher for DSM-5.

For current dimensional measures, three pairs of ICCs differed significantly (alcohol, tobacco, cannabis). While the ICCs were slightly higher for DSM-5 measures, all were in the excellent range (ICC = 0.87 – 0.91). For lifetime dimensional measures, five pairs of ICCs differed significantly (alcohol, tobacco, cannabis, cocaine, hallucinogens), four within the excellent range (ICC = 0.86 0.93) and only hallucinogen use disorder measures crossing a boundary (substantial to excellent; DSM-IV ICC = 0.75; DSM-5 ICC = 0.81). Of note, while none of the significant differences between pairs of DSM-IV and DSM-5 reliability coefficients were large, all reflected slightly higher reliability for DSM-5.

3.6. Factors potentially affecting reliability

Of tests of factors potentially affecting reliability, neither participant type (inpatient; community), time (days) between test and retest interviews nor interview mode (in-person vs. telephone) were significant predictors of discordant interviews except for one out of 36 tested ORs (telephone vs. in-person for lifetime heroin; Supplementary Tables 4–6). Of 123 comparisons of participants in different gender, age and race/ethnicity groups, only 11 (8.9 %) were significant (Supplementary Tables 7, 8 and 9). These results suggested that these factors were not major predictors of diagnostic discrepancies. However, dimensional severity in the initial interview (range, 0–11 criteria) was a significant predictor of discordance (Supplementary Table 10). For most current and lifetime disorders, lower initial severity was associated with discordant diagnoses in the retest interviews (p = .03 – <.01); exceptions were hallucinogen and stimulant use disorders.

4. Discussion

In a large, rigorous study conducted in the context of 21st-century substance use patterns, attitudes, and policies, the test-retest reliability of DSM-5 substance use disorders and DSM-IV substance dependence was evaluated in 565 participants using the semi-structured PRISM-5 interview administered by trained clinician interviewers. Our findings indicated substantial to excellent reliability of current and lifetime diagnoses of DSM-5 alcohol, tobacco, cannabis, cocaine, heroin, and opioid use disorders. Dimensional measures for

DSM-5 disorders involving these substances were all excellent, and higher than for binary disorders. To produce useful scientific results, a measure's output must be reproducible across independent administrations. These strong reliability findings support the use of PRISM-5 measures of DSM-5 SUD measures in research, and they contribute to support for the DSM-5 substance disorder measures more generally. In this study, we also evaluated three new aspects of the DSM-5 substance disorder criteria: craving, cannabis withdrawal, and new tobacco criteria. We discuss each of these in turn.

Although psychometric support was equivocal (Hasin et al., 2013), craving was added to the DSM-5 criteria due to arguments for its conceptual and clinical relevance, e.g., association with relapse and recovery; craving-related neural activation (Sinha, 2013). Laboratory procedures and multi-item scales (Sinha and O'Malley, 1999; Sayette et al., 2000; Anton, 2000) to measure craving have long existed. However, such procedures are not feasible in large-scale studies requiring assessment of multiple criteria for multiple disorders. While the PRISM-5 craving items were previously used in clinical (Hasin et al., 2012) and national surveys (Grant, 1992; Muthen et al., 1993; Keyes et al., 2011; Grant et al., 2016; Grant et al., 2015b), this is the first information on their test-retest reliability. This was substantial to excellent, and one of the most reliable of the DSM-5 SUD criteria across substances (Supplementary Table 2). While this is reassuring about craving as an individual criterion, the results do not indicate whether the addition of craving improves the validity of the entire DSM-5 SUD diagnosis, which was the relevant evaluation standard advocated by the chair of the DSM-5 Scientific Review Committee (Kendler, 2013). Further studies should examine the addition of craving in this larger context.

The cannabis withdrawal criterion was also new in DSM-5. In DSMIII, DSM-III-R and DSM-IV (American Psychiatric Association, 1987; American Psychiatric Association, 1994), while dependence diagnoses included withdrawal for most substances, evidence for cannabis withdrawal was considered too inconclusive (Budney et al., 2004). Since then, extensive research has demonstrated a cannabis withdrawal syndrome (Hasin, 2018). We found the reliability of cannabis withdrawal to be moderate ($\kappa = 0.53 – 0.55$), similar to the reliability of withdrawal from most other substances (Supplementary Tables 2a, 2b). Assessing withdrawal briefly in diagnostic assessment is challenging, involving lists of substance-specific withdrawal symptoms, determining if these co-occurred, vagueness about the timeframe for onset, and differentiation between withdrawal, depression and anxiety symptoms. While moderate reliability for withdrawal from cannabis and most other substances is reassuring, developing improved brief measures of withdrawal syndromes would be a contribution.

The three new individual tobacco disorder criteria were modeled on the DSM-IV abuse criteria for other substances, added to align the DSM-5 TUD criteria with those for other substances, and because the 11 DSM-5 tobacco criteria were unidimensional and had higher information than the DSM-IV dependence criteria (Hasin et al., 2013; Shmulewitz et al., 2013). Reliabilities of the three new tobacco criteria ranged widely ($\kappa = 0.34 – 0.73$), within the range of reliabilities for all individual criteria across substances ($\kappa = 0.29 0.95$, Supplementary Tables 2a, 2b).

Comparing DSM-5 SUD and DSM-IV substance dependence reliabilities showed few significant differences, and of those that did differ statistically, κ s were all in the substantial range, suggesting little meaningful difference. The greater information and statistical power of dimensional measures produced more significant differences between DSM-IV and DSM-5 severity indicators than the binary comparisons. However, in almost all cases, ICCs were excellent, with differences only at the second decimal place. Results thus allay concerns that reliability of DSM-5 SUD is inferior to the reliability of DSM-IV dependence, and set the stage to compare the validity of DSM-IV and DSM-5 SUD measures.

We explored whether several factors affected reliability, including time (days) between test and retest, type of participant (inpatient vs. community), interview mode (telephone vs. in-person), gender, age, race/ethnicity, and. Of many exploratory tests, few were significant, suggesting that these factors had little impact on reliability and supporting the generalizability of the results. However, low SUD severity in the initial interview did predict discordant interviews at retest, consistent with earlier studies (Hasin et al., 2006b; Hasin et al., 1996; Denis et al., 2015; Grant et al., 2015a; Hasin et al., 1997a; Grant et al., 1995). In mild disorders, a difference of only one or two criteria can determine the presence or absence of a diagnosis, which is not the case for severe disorders.

Higher reliability for the dimensional measures than the binary diagnoses likely occurred for several reasons. First, unlike diagnoses, continuous measures do not impose arbitrary thresholds (Hasin et al., 2013). Second, since the dimensional measures cover a greater range of values, they are more informative than categorical measures. The fact that PRISM-5 generates these as part of its standard output is one of its advantages. We note that better reliability for the dimensional measures is consistent with AUDADIS-5 studies utilizing lay interviewers and fully-structured assessments (Grant et al., 2015a), supporting generalizability of these results across methods, and the growing movement towards dimensional assessment of psychopathology. While diagnoses are necessary for clinical and reimbursement purposes, dimensional measures reflecting the full severity range of the underlying phenomena add invaluable information for research.

Comparing the prior DSM-5 SUD reliability study using the lay-administered AUDADIS-5 with the present study using the clinician-administered PRISM-5, reliability was similar for current DSM-5 alcohol and tobacco use disorders, and appreciably higher for PRISM-5 diagnoses of cannabis and opioid use disorders. For current dimensional measures, the range for AUDADIS-5 was somewhat lower (ICC = 0.50 – 0.85) than for PRISM-5 (ICC = 0.76 – 0.96). To our knowledge, the AUDADIS-5 is not currently being disseminated for use in new studies. However, PRISM-5, its training and computer algorithms (including versions customized for various purposes) are available and in use in a number of studies.

Limitations of the current study are noted. Only inpatients and a convenience sample of community participants were included. While both types of samples are commonly included in research, future studies should include additional samples. Specific withdrawal symptoms were not assessed for the prior 12 months, but should be in future studies. Stimulants are not commonly used in the Northeast, but are more common in Western U.S. states, and reliability studies should be conducted there. Finally, this report focused on substance use

disorders, but the PRISM-5 also assesses a full range of DSM-5 affective, anxiety and personality disorders. Findings on the reliability of these PRISM-5 diagnoses will be presented separately.

The study also had numerous strengths, including a large sample, varied participants in terms of age, sex, race/ethnicity and education, and rigorous blinded re-test procedures. The prevalence of many SUDs was sufficiently high to determine current as well as lifetime reliability. We conducted detailed analyses of new features of DSM-5 substance use disorders, including craving, cannabis withdrawal, and DSM-5 tobacco disorder criteria. Interviewers conducting retest interviews were always blind to the diagnoses produced in the initial interviews. Standardized training and certification procedures developed for earlier PRISM versions required only updating for PRISM-5. Further, this report provides reliability information not only on diagnoses and dimensional measures, but of all the individual diagnostic criteria for current and lifetime DSM-5 substance use disorders for multiple substances, considerable information about the criteria that serve as the building blocks of the diagnostic measures. Future studies should do the same.

This report examines reliability, a critical and fundamental psychometric property of a measure. This is a necessary step before investigating validity, which requires many additional variables and is considerably more complex. In an earlier study, PRISM-5 SUD measures showed moderate to excellent concordance (Hasin et al., 2015) with corresponding AUDADIS-5 measures. Determining correspondence between two methods of operationalizing a construct is one way to indicate validity (Hasin et al., 1997b). However, additional antecedent, concurrent and predictive validity information on the DSM-5 SUD measures is needed. Such studies exceed the scope of the present report, but are now underway using data from the participants reported here, and will be presented separately.

In conclusion, using rigorous methodology, DSM-5 substance use disorders and their dimensional measures were generally shown to have substantial to excellent reliability when assessed using the PRISM-5 interview administered by clinician interviewers. The study findings thus support the use of the DSM-5 definitions of substance use disorders in research. Further information about the PRISM-5 and its training will be available after September 2020: please see https://www.columbiapsychiatry.org/profile/deborah-hasin-phd

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

References

Alcohol Collaborators GBD, 2018 Alcohol use and burden for 195 countries and territories, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. Lancet 392 (10152), 1015–1035. [PubMed: 30146330]

American Psychiatric Association, 2013 Diagnostic and Statistical Manual of Mental Disorders, fifth edition American Psychiatric Association, Arlington, VA.

American Psychiatric Association, 1987 Diagnostic and Statistical Manual of Mental Disorders, third edition American Psychiatric Association, Washington, D.C. revised (DSM-III-R).

American Psychiatric Association, 1994 Diagnostic and Statistical Manual of Mental Disorders, fourth edition American Psychiatric Association, Washington, D.C.

Anton RF, 2000 Obsessive-compulsive aspects of craving: development of the Obsessive Compulsive Drinking Scale. Addiction. 95 (Suppl 2), S211–217. [PubMed: 11002915]

Budney AJ, et al., 2004 Review of the validity and significance of cannabis withdrawal syndrome. Am. J. Psychiatry 161 (11), 1967–1977. [PubMed: 15514394]

Centers for Disease Control and Prevention, 2018 U.S. Opioid Prescribing Rate Maps https://www.cdc.gov/drugoverdose/maps/rxrate-maps.html.

Cerda M, et al., 2019 Association Between Recreational Marijuana Legalization in the United States and Changes in Marijuana Use and Cannabis Use Disorder From 2008 to 2016. JAMA Psychiatry.

Chandra S, et al., 2019 New trends in cannabis potency in USA and Europe during the last decade. Eur. Arch. Psychiatry Clin. Neurosci 2008–2017.

Cohen J, 1960 A coefficient of agreement for nominal scales. Educ. Psychol. Meas 20 (1), 37–46.

Davies M, Fleiss JL, 1982 Measuring agreement for multinomial data. Biometrics. 38, 1047–1051.

Denis CM, et al., 2015 Inter-observer reliability of DSM-5 substance use disorders. Drug Alcohol Depend. 153, 229–235. [PubMed: 26048641]

DiCiccio T, Efron B, 1996 Bootstrap confidence intervals. Stat. Sci 11 (3), 189–228.

Efron B, 1979 Bootstrap methods: another look at jackknife. Ann. Stat 7, 1–26.

Efron B, Tibshirani R, 1986 Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat. Sci 1 (1), 54–75.

ElSohly MA, et al., 2016 Changes in Cannabis potency over the last 2 decades (1995–2014): analysis of current data in the United States. Biol. Psychiatry 79 (7), 613–619. [PubMed: 26903403]

Fleiss J, et al., 2003 Statistical Methods for Rates and Proportions, third edition John Wiley & Sons, Hoboken, New Jersey.

Frenk SM, et al., 2015 Prescription opioid analgesic use among adults: united States, 1999–2012. NCHS Data Brief (189), 1–8.

Grant BF, 1992 Prevalence of the proposed DSM-IV alcohol use disorders: United States, 1988. Br. J. Addict 87 (2), 309–316. [PubMed: 1555008]

Grant BF, et al., 2016 Epidemiology of DSM-5 drug use disorder: results from the national epidemiologic survey on alcohol and related Conditions-III. JAMA Psychiatry 73 (1), 39–47. [PubMed: 26580136]

Grant BF, et al., 1995 The Alcohol Use Disorder and Associated Disabilities Interview schedule (AUDADIS): reliability of alcohol and drug modules in a general population sample. Drug Alcohol Depend. 39 (1), 37–44. [PubMed: 7587973]

Grant BF, et al., 2015a 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. 148, 27–33. [PubMed: 25595052]

Grant BF, et al., 2015b Epidemiology of DSM-5 alcohol use disorder: results from the national epidemiologic survey on alcohol and related conditions III. JAMA Psychiatry 72 (8), 757–766. [PubMed: 26039070]

Hasin DS, 2018 US epidemiology of cannabis use and associated problems. Neuropsychopharmacology. 43 (1), 195–212. [PubMed: 28853439]

Hasin DS, et al., 2013 DSM-5 criteria for substance use disorders: recommendations and rationale. Am. J. Psychiatry 170 (8), 834–851. [PubMed: 23903334]

Hasin DS, et al., 1996 Psychiatric Research Interview for Substance and Mental Disorders (PRISM): reliability for substance abusers. Am. J. Psychiatry 153 (9), 1195–1201. [PubMed: 8780425]

Hasin D, et al., 1998 Psychiatric research interview for substance and mental disorders: phenomenologically based diagnosis in patients who abuse alcohol or drugs. Psychopharmacol. Bull 34 (1), 3–8. [PubMed: 9564191]

Hasin DS, et al., 2012 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. 122 (1–2), 28–37. [PubMed: 21963333]

Hasin DS, et al., 2015 The Alcohol Use Disorder and Associated Disabilities Interview Schedule-5 (AUDADIS-5): procedural validity of substance use disorders modules through clinical re-appraisal in a general population sample. Drug Alcohol Depend. 148, 40–46. [PubMed: 25604321]

Hasin D, et al., 2006a Substance use disorders: Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV) and International Classification of Diseases, tenth edition (ICD-10). Addiction. 101 (Suppl 1), 59–75. [PubMed: 16930162]

Hasin DS, et al., 2019a Alcohol use and binge drinking among U.S. men, pregnant and non-pregnant women ages 18–44: 2002–2017. Drug Alcohol Depend. 205, 107590. [PubMed: 31600616]

Hasin D, et al., 1997a The alcohol use disorder and associated disabilities interview schedule (AUDADIS): reliability of alcohol and drug modules in a clinical sample. Drug Alcohol Depend. 44 (2–3), 133–141. [PubMed: 9088785]

Hasin D, et al., 2006b Diagnosis of comorbid psychiatric disorders in substance users assessed with the Psychiatric Research Interview for Substance and Mental Disorders for DSM-IV. Am. J. Psychiatry 163 (4), 689–696. [PubMed: 16585445]

Hasin DS, et al., 2019b Time trends in US cannabis use and cannabis use disorders overall and by sociodemographic subgroups: a narrative review and new findings. Am. J. Drug Alcohol Abuse 45 (6), 623–643. [PubMed: 30870044]

Hasin D, et al., 1997b Nosological comparisons of alcohol and drug diagnoses: a multisite, multi-instrument international study. Drug Alcohol Depend. 47 (3), 217–226. [PubMed: 9306047]

Hedegaard H, et al., 2018 Drug overdose deaths in the United States, 1999–2017. NCHS Data Brief (329), 1–8.

Kariisa M, et al., 2019 Drug overdose deaths involving cocaine and psychostimulants with abuse potential - United States, 2003–2017. MMWR Morb. Mortal. Wkly. Rep 68 (17), 388–395. [PubMed: 31048676]

Kendler KS, 2013 A history of the DSM-5 scientific review committee. Psychol. Med. (Paris) 43 (9), 1793–1800.

Kerridge BT, et al., 2019 Changes in the prevalence and correlates of cocaine use and cocaine use disorder in the United States, 2001–2002 and 2012–2013. Addict. Behav 90, 250–257. [PubMed: 30471553]

Keyes KM, et al., 2011 Alcohol craving and the dimensionality of alcohol disorders. Psychol. Med. (Paris) 41 (3), 629–640.

Landis JR, Koch GG, 1977 The measurement of observer agreement for categorical data. Biometrics. 33 (1), 159–174. [PubMed: 843571]

Martins SS, et al., 2017 Changes in US lifetime heroin use and heroin use disorder: prevalence from the 2001–2002 to 2012–2013 national epidemiologic survey on alcohol and related conditions. JAMA Psychiatry 74 (5), 445–455. [PubMed: 28355458]

Mehmedic Z, et al., 2010 Potency trends of Delta9-THC and other cannabinoids in confiscated cannabis preparations from 1993 to 2008. J. Forensic Sci 55 (5), 1209–1217. [PubMed: 20487147]

Morgello S, et al., 2006 Interrater reliability of the Psychiatric Research Interview for Substance and Mental disorders in an HIV-infected cohort: experience of the national NeuroAIDS tissue consortium. Int. J. Methods Psychiatr. Res 15 (3), 131–138. [PubMed: 17019897]

Muthen BO, et al., 1993 Factor analysis of ICD-10 symptom items in the 1988 national health interview survey on alcohol dependence. Addiction. 88 (8), 1071–1077. [PubMed: 8401161]

Rehm J, Shield KD, 2019 Global burden of disease and the impact of mental and addictive disorders. Curr. Psychiatry Rep 21 (2), 10. [PubMed: 30729322]

Saha TD, et al., 2016 Nonmedical prescription opioid use and DSM-5 nonmedical prescription opioid use disorder in the United States. J. Clin. Psychiatry 77 (6), 772–780. [PubMed: 27337416]

SAS Institute Inc. Sample 25031: Compute Six Intraclass Correlation Measures. http://support.sas.com/kb/25/031.html.

Sayette MA, et al., 2000 The measurement of drug craving. Addiction. 95 (Suppl 2), S189–210. [PubMed: 11002914]

Scholl L, et al., 2018 Drug and opioid-involved overdose deaths - United States, 2013–2017. MMWR Morb. Mortal. Wkly. Rep 67 (5152), 1419–1427. [PubMed: 30605448]

Shmulewitz D, et al., 2013 Validity of proposed DSM-5 diagnostic criteria for nicotine use disorder: results from 734 Israeli lifetime smokers. Psychol. Med. (Paris) 43 (10), 2179–2190.

Shrout PE, Fleiss JL, 1979 Intraclass correlations: uses in assessing rater reliability. Psychol. Bull 86 (2), 420–428. [PubMed: 18839484]

Sinha R, 2013 The clinical neurobiology of drug craving. Curr. Opin. Neurobiol 23 (4), 649–654. [PubMed: 23764204]

Sinha R, O'Malley SS, 1999 Craving for alcohol: findings from the clinic and the laboratory. Alcohol Alcohol. 34 (2), 223–230. [PubMed: 10344782]

Spencer MR, et al., 2019 Drug overdose deaths involving fentanyl, 2011–2016. Vital Stat. Rep 68 (3), 1–19.

Spindle TR, et al., 2019 Changing landscape of cannabis: novel products, formulations, and methods of administration. Curr. Opin. Psychol 30, 98–102. [PubMed: 31071592]

Torrens M, et al., 2004 Diagnosing comorbid psychiatric disorders in substance abusers: validity of the Spanish versions of the Psychiatric Research Interview for Substance and Mental Disorders and the Structured Clinical Interview for DSM-IV. Am. J. Psychiatry 161 (7), 1231–1237. [PubMed: 15229056]

Vigo D, et al., 2019 Burden of Mental, Neurological, Substance Use Disorders and Self-Harm in North America: A Comparative Epidemiology of Canada, Mexico, and the United States. Can. J. Psychiatry, 706743719890169.

	DSM-IV Substance Dependence	DSM-IV Substance Abuse[a]	DSM-5 Substance Use Disorder
Withdrawal syndrome *OR* use to avoid withdrawal[b]	✓		✓
Tolerance	✓		✓
Used in larger amounts or longer than expected	✓		✓
Persistent desire/repeated attempts to quit or cut down	✓		✓
A great deal of time spent using the substance	✓		✓
Continued use despite physical *OR* psychological problems	✓		✓
Important activities given up to use	✓		✓
Hazardous use		✓	✓
Social/interpersonal problems due to use		✓	✓
Neglected major roles to use		✓	✓
Legal problems due to use		✓	
Craving			✓
Binary diagnosis	≥3 criteria within a 12-month period	≥1 criteria, NO dependence diagnosis	≥2 criteria within a 12-month period

[a] Not in DSM-IV for tobacco

[b] DSM-IV cannabis, inhalant, and hallucinogen dependence diagnoses did not include a withdrawal criterion, so this was not included for these diagnoses. In DSM-5, cannabis withdrawal was included for cannabis use disorder.

Fig. 1.
DSM Substance Use Disorder Criteria.

Table 1

Test-retest reliability of PRISM-5 DSM-5 substance use disorders, N = 565.

DSM-5 Substance Use Disorder			Severity indicator	
% Test	% Retest	Kappa (95 % CI)	DSM-5 severity[b] ICC (95 % CI)	Dimensional severity[c] ICC (95 % CI)
Current (past year)				
Alcohol	65.3	62.5	0.69 (0.62, 0.75)	0.87 (0.84, 0.91)	0.90 (0.87, 0.92)
Tobacco	62.3	60.5	0.79 (0.74, 0.84)	0.90 (0.87, 0.92)	0.91 (0.87, 0.93)
Cannabis	43.9	42.3	0.74 (0.68, 0.79)	0.88 (0.85, 0.90)	0.90 (0.87, 0.92)
Cocaine	45.3	43.4	0.82 (0.78, 0.87)	0.91 (0.88, 0.93)	0.91 (0.88, 0.93)
Heroin	24.8	24.2	0.94 (0.90, 0.97)	0.97 (0.95, 0.98)	0.96 (0.95, 0.98)
Opioids	15.9	12.2[a]	0.67 (0.58, 0.76)	0.86 (0.81, 0.90)	0.87 (0.82, 0.92)
Sedatives	13.3	11.3	0.58 (0.48, 0.69)	0.79 (0.68, 0.85)	0.83 (0.73, 0.88)
Hallucinogens	3.5	2.7	0.50 (0.29, 0.70)	0.74 (0.37, 0.88)	0.76 (0.34, 0.90)
Stimulants	2.8	3.0	0.59 (0.39, 0.80)	0.81 (0.64, 0.92)	0.89 (0.77, 0.97)
Lifetime				
Alcohol	78.2	74.7[a]	0.63 (0.55, 0.70)	0.87 (0.83, 0.90)	0.93 (0.91, 0.94)
Tobacco	68.8	66.0[a]	0.74 (0.68, 0.80)	0.88 (0.84, 0.91)	0.91 (0.88, 0.93)
Cannabis	55.2	50.1[a]	0.65 (0.59, 0.71)	0.85 (0.81, 0.89)	0.89 (0.87, 0.91)
Cocaine	60.9	58.1	0.72 (0.66, 0.78)	0.87 (0.84, 0.90)	0.93 (0.91, 0.94)
Heroin	31.5	31.2	0.92 (0.88, 0.95)	0.97 (0.96, 0.98)	0.99 (0.98, 0.99)
Opioids	25.0	21.4[a]	0.65 (0.58, 0.73)	0.84 (0.79, 0.88)	0.90 (0.87, 0.93)
Sedatives	18.1	16.8	0.63 (0.54, 0.71)	0.81 (0.74, 0.87)	0.88 (0.82, 0.91)
Hallucinogens	11.9	8.8[a]	0.53 (0.42, 0.65)	0.76 (0.66, 0.83)	0.82 (0.73, 0.88)
Stimulants	6.2	7.1	0.53 (0.39, 0.67)	0.75 (0.62, 0.87)	0.84 (0.75, 0.91)

CI = Confidence interval; ICC = Intraclass Correlation Coefficient. Interpretation of reliability coefficients: Fair: 0.21–0.40; moderate: 0.41–0.60; substantial: 0.61–0.80; excellent: 0.81–1.00.

[a] Prevalence differed significantly between the test and retest interviews (McNemar's test p < .05).

[b] DSM-5 severity: no disorder, 0–1 criteria; mild, 2–3 criteria; moderate, 4–5 criteria; severe, 6+ criteria.

[c] Dimensional severity is the total number of positive DSM-5 substance use disorder criteria (range, 0–11, except hallucinogens, whose criteria do not include withdrawal and whose range was therefore 0–10).

Table 2

Test-retest reliability of DSM-5 craving, by substance N = 565.

Substance	DSM-5 Craving criterion			Wanted to use so badly, couldn't think of anything else			Felt a very strong desire or urge to use		
	% Test	% Retest	Kappa (95 % CI)	% Test	% Retest	Kappa (95 % CI)	% Test	% Retest	Kappa (95 % CI)
Current									
Alcohol	52.4	48.7[a]	0.66 (0.60, 0.72)	27.4	27.3	0.64 (0.57, 0.71)	50.8	47.1[a]	0.65 (0.59, 0.71)
Tobacco	55.0	51.7[a]	0.73 (0.67, 0.78)	33.8	30.8	0.60 (0.53, 0.67)	54.2	51.2	0.71 (0.65, 0.77)
Cannabis	39.5	34.7[a]	0.72 (0.67, 0.78)	22.1	21.6	0.64 (0.57, 0.72)	38.2	34.2[a]	0.71 (0.64, 0.77)
Cocaine	38.6	36.5	0.76 (0.70, 0.81)	29.2	26.0[a]	0.71 (0.64, 0.77)	37.7	35.9	0.76 (0.70, 0.81)
Heroin	23.7	23.0	0.89 (0.85, 0.94)	20.5	18.9	0.84 (0.78, 0.90)	23.7	23.0	0.89 (0.85, 0.94)
Opioids	11.5	9.9	0.69 (0.60, 0.79)	10.1	8.0[a]	0.68 (0.57, 0.79)	11.5	9.4[a]	0.66 (0.56, 0.76)
Sedatives	9.7	8.3	0.57 (0.45, 0.69)	7.8	6.0	0.59 (0.45, 0.72)	9.6	8.1	0.56 (0.44, 0.68)
Hallucinogens	3.4	2.3	0.55 (0.34, 0.76)	1.4	1.1	[b]	3.4	2.1	0.57 (0.36, 0.78)
Stimulants	1.9	2.1	[b]	1.2	1.6	[b]	1.9	2.1	[b]
Lifetime									
Alcohol	64.6	60.4[a]	0.71 (0.65, 0.77)	37.2	35.8	0.67 (0.61, 0.74)	64.1	58.9[a]	0.71 (0.65, 0.77)
Tobacco	62.1	59.6	0.67 (0.61, 0.74)	39.3	37.5	0.60 (0.54, 0.67)	60.7	58.6	0.66 (0.60, 0.72)
Cannabis	51.5	43.0[a]	0.68 (0.62, 0.74)	31.0	27.1[a]	0.64 (0.57, 0.71)	50.4	42.1[a]	0.69 (0.63, 0.74)
Cocaine	56.5	54.7	0.73 (0.67, 0.78)	44.4	42.7	0.70 (0.64, 0.76)	54.3	53.5	0.73 (0.67, 0.78)
Heroin	31.0	30.6	0.95 (0.92, 0.98)	28.1	27.4	0.89 (0.85, 0.94)	30.6	30.6	0.94 (0.91, 0.97)
Opioids	23.2	19.1[a]	0.73 (0.66, 0.80)	19.6	15.6[a]	0.70 (0.62, 0.78)	22.7	18.2[a]	0.71 (0.64, 0.78)
Sedatives	15.2	14.3	0.68 (0.60, 0.77)	11.3	10.4	0.61 (0.50, 0.71)	14.7	13.8	0.65 (0.55, 0.74)
Hallucinogens	10.1	7.4[a]	0.50 (0.38, 0.63)	5.5	4.1	0.53 (0.37, 0.70)	9.4	7.1[a]	0.52 (0.39, 0.65)
Stimulants	5.8	4.8	0.54 (0.39, 0.70)	3.5	3.5	0.64 (0.46, 0.81)	5.8	4.4	0.49 (0.33, 0.65)

CI = Confidence interval.

Interpretation of reliability coefficients: Fair: 0.21–0.40; moderate: 0.41–0.60; substantial: 0.61–0.80; excellent: 0.81–1.00.

[a] Prevalence differed significantly between the test and retest interviews (McNemar's test p < .05).

[b]Kappa was not calculated if prevalence was <2% in test or retest.

Table 3

Reliability of new DSM-5 criteria: cannabis withdrawal and tobacco "abuse" criteria.

	Lifetime			Current[a]		
	% Test	% Retest	Kappa (95 % CI)	% Test	% Retest	Kappa (95 % CI)
Cannabis withdrawal						
Withdrawal criterion	36.6	29.4[c]	0.55 (0.48, 0.62)	24.2	20.5[c]	0.53 (0.44, 0.61)
Withdrawal syndrome[b]	16.6	14.7	0.46 (0.36, 0.56)	8.8	9.4	0.41 (0.29, 0.54)
3+ withdrawal symptoms	35.2	26.7[c]	0.57 (0.50, 0.64)	- -	- -	- -
Used cannabis or another drug to relieve or avoid cannabis withdrawal symptoms	31.7	25.0[c]	0.55 (0.47, 0.62)	21.6	17.5[c]	0.51 (0.42, 0.60)
Individual lifetime cannabis withdrawal symptoms						
Irritable, angry, or aggressive	25.8	22.3[c]	0.58 (0.51, 0.66)	- -	- -	- -
Lose weight, or have decreased appetite	24.2	20.4[c]	0.48 (0.39, 0.57)	- -	- -	- -
Feel nervous or anxious	32.0	24.4[c]	0.50 (0.42, 0.58)	- -	- -	- -
Feel restless, or like you couldn't sit still	29.0	21.6[c]	0.51 (0.43, 0.59)	- -	- -	- -
Sleep disturbance	40.4	29.9[c]	0.49 (0.42, 0.56)	- -	- -	- -
Feel down or depressed	30.8	23.4[c]	0.50 (0.42, 0.58)	- -	- -	- -
Physical symptoms[d]	25.1	20.2[c]	0.56 (0.48, 0.64)	- -	- -	- -
Tobacco "abuse" criteria						
Hazardous use	38.4	37.2	0.73 (0.67, 0.78)	28.7	26.7	0.67 (0.60, 0.74)
Social/Interpersonal problems	11.0	9.4	0.37 (0.25, 0.49)	8.3	7.1	0.34 (0.20, 0.48)
Neglect roles	21.9	17.9[c]	0.52 (0.43, 0.61)	13.6	11.5	0.42(0.31, 0.53)

CI = Confidence interval.

Interpretation of reliability coefficients: Fair: 0.21–0.40; moderate: 0.41–0.60; substantial: 0.61–0.80; excellent: 0.81–1.00.

[a] Individual withdrawal symptoms were not assessed for the past-year time period, including those for cannabis withdrawal.

[b] 3+ withdrawal symptoms, clustering together in time, associated with self-acknowledged impairment.

[c] Prevalence differed significantly between the test and retest interviews (McNemar's test p < .05).

[d] stomach pain, shaking hands, sweating, chills, headache, fever.

Table 4

Differences between test-retest reliability of DSM-5 substance use disorder and DSM-IV substance dependence, N = 565.

	DSM-5 binary diagnosis Mean Kappa (95% CI)[c]	DSM-IV binary diagnosis	Mean difference, DSM-5 and DSM-IV kappa (95% CI)[d]	DSM-5 dimensional severity[a] Mean ICC (95% CI)[e]	DSM-IV dimensional severity[b]	Mean difference, DSM-5 and DSM-IV ICC (95% CI)[f]
Current						
Alcohol	0.68 (0.62, 0.74)	0.70 (0.64, 0.76)	−0.01 (−0.09, 0.04)	0.89 (0.87, 0.92)	0.88 (0.84, 0.90)	**0.02 (0.01, 0.03)**
Tobacco	0.79 (0.74, 0.83)	0.67 (0.61, 0.73)	**0.12 (0.04, 0.19)**	0.91 (0.87, 0.93)	0.89 (0.85, 0.91)	**0.02 (0.01, 0.03)**
Cannabis	0.74 (0.68, 0.81)	0.62 (0.56, 0.70)	**0.12 (0.04, 0.22)**	0.90 (0.87, 0.92)	0.87 (0.84, 0.89)	**0.03 (0.02, 0.04)**
Cocaine	0.82 (0.77, 0.87)	0.79 (0.74, 0.83)	0.04 (−0.01, 0.08)	0.91 (0.88, 0.93)	0.91 (0.88, 0.93)	0.00 (−0.00, 0.01)
Heroin	0.94 (0.90, 0.97)	0.92 (0.87, 0.95)	0.02 (−0.01, 0.05)	0.97 (0.95, 0.98)	0.97 (0.95, 0.98)	− 0.00 (−0.01, 0.01)
Opioids	0.67 (0.58, 0.75)	0.69 (0.60, 0.77)	−0.02 (−0.09, 0.04)	0.87 (0.82, 0.92)	0.86 (0.82, 0.91)	0.00 (−0.01, 0.02)
Sedatives	0.58 (0.47, 0.66)	0.60 (0.48, 0.72)	− 0.02 (−0.10, 0.08)	0.82 (0.73, 0.88)	0.81 (0.72, 0.88)	0.01 (−0.01, 0.02)
Hallucinogens	g_	g_	—	0.73 (0.34, 0.90)	0.67 (0.35, 0.84)	0.06 (−0.00, 0.13)
Stimulants	0.60 (0.38, 0.80)	0.57 (0.32, 0.73)	0.04 (−0.09, 0.20)	0.87 (0.77, 0.97)	0.85 (0.74, 0.95)	0.02 (−0.02, 0.06)
Lifetime						
Alcohol	0.62 (0.53, 0.69)	0.66 (0.62, 0.73)	− 0.04 (−0.11, 0.02)	0.93 (0.91, 0.94)	0.91 (0.89, 0.93)	**0.01 (0.00, 0.02)**
Tobacco	0.74 (0.69, 0.80)	0.64 (0.57, 0.70)	**0.10 (0.03, 0.16)**	0.91 (0.88, 0.93)	0.89 (0.86, 0.91)	**0.02 (0.01, 0.03)**
Cannabis	0.65 (0.58, 0.71)	0.59 (0.52, 0.65)	0.06 (−0.02, 0.14)	0.89 (0.87, 0.91)	0.86 (0.83, 0.89)	**0.03 (0.02, 0.04)**
Cocaine	0.72 (0.67, 0.78)	0.74 (0.69, 0.78)	−0.02 (−0.06, 0.04)	0.93 (0.91, 0.94)	0.92 (0.90, 0.94)	**0.01 (0.00, 0.02)**
Heroin	0.92 (0.88, 0.95)	0.92 (0.89, 0.96)	−0.01 (−0.04, 0.02)	0.99 (0.98, 0.99)	0.98 (0.97, 0.99)	0.00 (−0.00, 0.00)
Opioids	0.65 (0.57, 0.72)	0.67 (0.59, 0.75)	−0.03 (−0.06, 0.02)	0.90 (0.87, 0.93)	0.90 (0.87, 0.92)	0.01 (−0.00, 0.02)
Sedatives	0.62 (0.54, 0.69)	0.66 (0.55, 0.75)	− 0.05 (−0.10, 0.02)	0.88 (0.82, 0.91)	0.86 (0.80, 0.90)	0.01 (−0.00, 0.02)
Hallucinogens	0.53 (0.40, 0.64)	0.45 (0.30, 0.58)	0.08 (−0.05, 0.20)	0.81 (0.73, 0.88)	0.75 (0.66, 0.84)	**0.06 (0.03, 0.09)**
Stimulants	0.54 (0.41, 0.67)	0.59 (0.42, 0.70)	− 0.05 (−0.14, 0.06)	0.84 (0.75, 0.91)	0.83 (0.73, 0.90)	0.01 (−0.02, 0.03)

CI = Confidence interval; ICC = Intraclass Correlation Coefficient.

[a] Total number of DSM-5 substance use disorder criteria endorsed (of possible 11, except hallucinogens, which do not include withdrawal).

[b] Total number of DSM-IV substance dependence criteria endorsed (of possible 7, except cannabis and hallucinogens, which did not include withdrawal).

[c] Mean kappa from 100 bootstrapped samples, with empirical CI

[d] Mean difference calculated as the mean of ([DSM-5 κ] - [DSM-IV κ]) in each of the 100 bootstrapped samples, with empirical CI.

CI that do not include 0 indicate a significant greater κ for DSM-5 substance use disorder and are bolded.

[e] Mean ICC from 100 bootstrapped samples, with empirical CI

[f] Mean difference is calculated as the mean of ([DSM-5 severity ICC] − [DSM-IV severity ICC]) in each of the 100 bootstrapped samples, with empirical CI.

CI that do not include 0 indicate a significantly greater ICC for DSM-5 severity, and are bolded.

[g] Current hallucinogens not included because prevalence of DSM-IV hallucinogen dependence was <2% in the test and retest interviews. Interpretation of reliability coefficients: Fair: 0.21–0.40; moderate: 0.41–0.60; substantial: 0.61–0.80; excellent: 0.81–1.00.