Published in final edited form as:

Drug Alcohol Depend. 2015 March 01; 148: 27–33. doi:10.1016/j.drugalcdep.2014.11.026.

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

Bridget F. Grant^a, Rise B. Goldstein^a, Sharon M. Smith^a, Jeeseun Jung^a, Haitao Zhang^a, Sanchen P. Chou^a, Roger P. Pickering^a, Wenjun J. Ruan^a, Boji Huang^a, Tulshi D. Saha^a, Christina Aivadyan^b, Eliana Greenstein^b, and Deborah S. Hasin^{b,c,d}

^aLaboratory of Epidemiology and Biometry, Division of Clinical and Biological Research, National Institute of Alcohol Abuse and Alcoholism, 5635 Fishers Lane, Rockville, Maryland 20852, USA, phone: (301) 443-7370

^bNew York State Psychiatric Institute, 722 West 168th Street, New York, New York, 10032, USA

^cDepartment of Psychiatry, College of Physicians and Surgeons, Columbia University, 1051 Riverside Drive #123, New York, New York, 10032, USA

^dDepartment of Epidemiology, Mailman School of Public Health, 722 West 168th Street, New York, New York 10032, USA

Abstract

Background—The purpose of this study was to assess the test-retest reliability of substance disorder and psychiatric modules in the Alcohol Use Disorder and Associated Disabilities Interview Schedule, Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) Version (AUDADIS-5).

Methods—Kappa and intraclass correlation coefficients were calculated for DSM-5 substance use and psychiatric disorder diagnoses and dimensional criteria scales using a test-retest design among 1006 respondents drawn from the National Epidemiologic Survey on Alcohol and Related Conditions-III (NESARC-III).

Correspondence to: Bridget F. Grant.

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.

Disclaimer: The views and opinions expressed in this paper are those of the authors and should not be construed to represent the views of any of the sponsoring organizations, agencies or the U.S. government

Contributions

Drs. Grant, Hasin and Goldstein wrote the first/second draft of the manuscript. Drs. Smith, Jung, Zhang, Chou, Huang, Pickering, Ruan, Goldstein, Saha commented on subsequent revisions of the manuscript and/or collected or cleaned the data. Drs. Grant, Saha, Aivadyan and Greenstein, and Pickering conducted data analysis.

Conflict of Interest

No conflict of interest declared.

Results—Reliabilities of substance use disorder diagnoses and associated criteria scales were generally good to excellent, while reliabilities for mood, anxiety and trauma and stress-related disorders and associated scales were generally in the fair to good range.

Conclusions—The observed reliability of the DSM-5 diagnoses and dimensional scales for the substance use and psychiatric disorders found in this study indicates that the AUDADIS-5 can be a useful tool in various research settings, particularly in studies of the general population, the target population for which it was designed.

Keywords

alcohol use disorder; drug use disorder; reliability; mood disorders; anxiety disorders; psychometrics; National Epidemiologic Survey on Alcohol and Related Conditions-III; Alcohol Use Disorder and Associated Disabilities Interview Schedule-5

1. INTRODUCTION

The Alcohol Use Disorder and Associated Disabilities Interview Schedule – Diagnostic and Statistical Manual of Mental Disorders (AUDADIS) is a fully structured interview designed to assess alcohol, drug and mental disorders according DSM criteria in both clinical and general populations (Grant et al., 2001). The AUDADIS provides detailed measurement of alcohol and drug use disorders, other psychiatric disorders, and many risk factors. The version of the AUDADIS that measured DSM-IV criteria, the AUDADIS-IV, was initially used in the U.S. National Institute on Alcohol Abuse and Alcoholism's (NIAAA) National Longitudinal Alcohol Epidemiologic Survey in 1991–1992 (NLAES: Grant et al., 2004). The AUDADIS-IV was subsequently used in the NIAAA 2001–2002 Wave 1 National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) and its 3-year follow-up, conducted in 2004–2005 (Grant et al., 2003a, 2004a). The AUDADIS-IV measures in these national surveys formed the basis for hundreds of publications. The AUDADIS-IV has also been used in clinical (Hasin et al., 2013; 2014), high-risk prospective (Hasin et al., 1996; 2007) and genetic studies (Meyers et al, 2013).

The reliability of the AUDADIS-IV was extensively tested, including U.S. clinical (Canino et al., 1999; Hasin et al., 1997) and general population (Grant et al., 1995, 2003b; Ruan et al., 2008) samples, and in other countries in the National Institutes of Health/World Health Organization Reliability and Validity Study (Chatterji et al., 1997; Vrasti et al., 1998). In the U.S. general population, test-retest reliabilities for DSM-IV nicotine dependence ($\kappa = 0.60\text{--}0.63$) alcohol and drug use disorders ($\kappa = 0.66\text{--}0.91$) and personality disorders ($0.67\text{--}0.71$) were good to excellent (Grant et al., 1995, 2003a; 2008). Similarly high reliability coefficients were found in treatment settings for substance abuse (Hasin et al., 1997) and primary care (Canino et al., 1999). Reliability studies conducted in Romania, India, and Australia also yielded kappas between 0.57 and 0.96 for DSM-IV alcohol and drug use disorders (Chatterji et al., 1997; Vrasti et al., 1998). Reliability coefficients derived from U.S. general population samples for DSM-IV mood ($\kappa = 0.58\text{--}0.65$) and anxiety ($\kappa = 0.40\text{--}0.77$) disorders were somewhat lower, but generally fair to good. Reliability of dimensional measures of substance use disorders (counts of positive diagnostic criteria)

generally exceeded the reliability of the corresponding binary diagnoses (Grant et al., 2003; Hasin et al., 1997).

The AUDADIS-5 (Grant et al., 2011) was designed to assess the diagnostic definitions embodied in the DSM-Fifth Edition (DSM-5; American Psychiatric Association, 2013). The AUDADIS-5 was used in NIAAA's most recent large-scale national survey of alcohol, drug and psychiatric disorders, the 2012–2013 National Epidemiologic Survey on Alcohol and Related Conditions-III (NESARC-III) (Grant et al., 2014).

Reliability (consistency) is a fundamental psychometric property of any measure and thus, important to assess. Given the many changes in diagnostic definitions from DSM-IV to DSM-5, the reliability of AUDADIS-5 diagnoses and corresponding dimensional scales could not be inferred from DSM-IV measures. This study was conducted to provide information on the reliability of AUDADIS-5 measures of substance use disorders and selected mood, anxiety and trauma, stress-related and personality disorders using rigorous methods in a large, general population sample.

2. METHOD

2.1 Sample

Participants in the test-retest study were a subsample of NESARC-III participants. The NESARC-III target population was the non-institutionalized civilian population 18 years or older living in the 50 U.S. states and the District of Columbia, including residents of selected group quarters such as group homes and dormitories for workers (Grant et al., 2014). Multistage probability sampling was used to randomly select respondents. Primary sampling units (PSUs) were individual counties or groups of contiguous counties, secondary sampling units (SSUs) consisted of groups of Census-defined blocks, and in the third stage households within the sampled SSUs were selected. The last sampling stage involved the random selection of eligible adults within sampled households. Minority persons (i.e., Hispanic, Black, Asian) were given higher probabilities of selection than nonminority household members. Further, in households with four or more eligible minority persons, two respondents were selected. The total sample size of NESARC-III was 36,309. The screener response rate was 72.0% and the person response rate was 84.0%. The total NESARC-III response rate was 60.1%, comparable to the majority of national surveys currently conducted in the U.S. (Centers for Disease Control and Prevention, 2013; Substance Abuse and Mental Health Administration, 2013).

Each respondent completing the AUDADIS-5 was asked to participate in a face-to-face retest interview. Of the 36,309 NESARC-III respondents, 25,769 consented to the re-interview. Re-interview respondents were sampled each week between June, 2012 and July, 2013 using an algorithm designed to increase the selection of respondents with psychopathology. Screening questions to route respondents into modules covering alcohol use disorders, drug use disorders, nicotine use disorders and each mood, anxiety and trauma and stress-related disorder were coded as “positive” or “negative”. Respondents with the greatest number of positive screening questions, and therefore completed modules, were given higher probabilities of selection into the reliability study. Using this algorithm,

reliability participants were then randomly selected each week in order to manage the caseload of retest interviews. This procedure was followed until the target reliability sample size of 1006 was achieved. The target sample size was determined by results of prior NESARC survey test-retest designs as optimal for obtaining sufficient power (80%) for effect sizes (prevalences) greater than about 3%. Of the 1048, respondents selected for the reliability re-interview, 1006 completed the reinterview and 92 refused, for a response rate 92%.

To reduce respondent re-test burden, two shortened versions of the AUDADIS-5 were used. Both versions included all alcohol and drug use modules. One version additionally included tobacco use disorder, affective disorders (major depression, mania, and dysthymia) and conduct and antisocial personality disorders (PDs). The other version included the anxiety and stress-related disorders (i.e., panic disorder, agoraphobia, social anxiety, specific phobia, and generalized anxiety and posttraumatic stress disorders) and borderline and schizotypal PDs. The algorithm for selection into the test-retest reliability study randomly assigned respondents to one of these two shortened versions. Of the respondents, 502 were reinterviewed with the version that included the tobacco and affective disorders, while the remaining 504 were re-interviewed with the version that included the anxiety, trauma and stress-related disorders.

All respondents gave informed consent to participate in the full national survey and in the test-retest sub-study. All procedures in the NESARC-III protocol, including both the full national survey and the test-retest reliability study, were approved by the National Institutes of Health and Westat Institutional Review Boards.

2.2 Test retest design

The test-retest design of the present study was similar to all prior test-retest studies of the AUDADIS-IV. No interviewer interviewed the same respondent twice. Interviewers administering the retest interviews were always blind to the results of the initial interview. Interviewer assignments during the initial test and subsequent retest were randomized among the staff. There were 970 interviewers who administered the test interview and 274 interviewers who administered the retest interview. The test - retest interval ranged from 1 to 10 weeks, with an average of 2.86 weeks. All interviewers had on average 5 years of field experience working on health-related and other surveys, and each interviewer completed a 1-day self-study and participated in a 4-day in-class training session. Ten percent of each interviewer's cases were also re-interviewed to validate respondent answers to sociodemographic and other key survey variables and to ensure the survey interview procedures were being followed.

2.3 Diagnostic assessment

In the NESARC-III, the AUDADIS-5 assessed diagnoses according to the DSM-5 criteria. The DSM-5 diagnoses assessed in the test-retest study included alcohol use disorder and other drug-specific use disorders for tobacco, cannabis, cocaine, other stimulants, sedatives/ tranquilizers, hallucinogens, inhalants/solvents, opioids, club drugs and heroin. DSM-5 mood disorders included major depressive and manic episode, and dysthymia (persistent

depressive disorder). DSM-5 anxiety disorders assessed included panic disorder, agoraphobia, social phobia, specific phobia and generalized anxiety disorder. The AUDADIS-5 also included DSM-5 diagnoses of posttraumatic stress disorder, conduct disorder, and antisocial, borderline and schizotypal PDs. While many sections of the DSM-5 were not finalized until 2012, the criteria for the disorders assessed in the NESARC-III were essentially in final form in 2011, and hence were incorporated into the AUDADIS-5, allowing for the test of these finalized diagnoses.

2.4 Statistical analyses

The present study used the standard statistical methodology used in prior test-retest studies of the AUDADIS-IV to derive reliability coefficients. For dichotomous data elements, e.g., diagnoses, kappa was used as the reliability coefficient, defined as a measure of pairwise agreement corrected for chance (Fleiss, 1981; Shouten, 1980). For continuous measures, intraclass correlation coefficients (ICC) were used. ICCs were used primarily to examine the reliability of dimensional diagnostic criteria scales, and also to examine reliability of ages at first onset and most recent episode for each psychiatric disorder. McNemar's test, a test for paired comparisons of binary variables, was used to determine if the prevalences of diagnoses differed between test and retest interviews.

The timeframes assessed for substance use, anxiety and mood diagnoses and PTSD were past year and prior to the past year. ICC statistics for substance use disorder dimensional scales were also computed for these two time frames since criteria information was separately collected for each time frame on the AUDADIS-5. In contrast, the AUDADIS-5 collects information on the occurrence of anxiety and mood disorders and PTSD criteria on a lifetime basis prior to determining whether an episode occurred in the past year or prior to that time. Therefore, ICC estimates of dimensional scales associated with these disorders are based on lifetime occurrence of criteria symptoms. Reliability of the dimensional measure consisting of all PTSD criteria (criteria B–E) is reported, along with dimensional scales associated with PTSD component criteria of intrusion (criterion B), avoidance (criterion C), change in cognition/mood (criterion D), and change in arousal/activity (criterion E) since symptoms within each of these criterion components are extensive.

Kappa statistics for conduct disorder and antisocial, borderline and schizotypal PD diagnoses were assessed on a lifetime basis consistent with their definition in the DSM-5 (unchanged from DSM-IV). However, the timeframe associated with ICC statistics for dimensional criteria scales for these disorders differed: prior to the age of 15 years for conduct disorder; since age 15 for antisocial PD; and lifetime for borderline and schizotypal PDs. Kappa statistics for conduct disorder and the three PDs are reported for diagnoses with and without the requirement for social/occupation dysfunction or distress based on the clinical observation that individuals with PDs may lack insight into their conditions (Zimmerman, 1994). Onset and recency of conduct disorder and antisocial PD are not reported here because this was not directly assessed on the AUDADIS-5.

The reliability design assumed that interviewers were randomly drawn from a larger population of interviewers. We therefore used a one-way random effects ANOVA model to derive intraclass correlation coefficients (Shrout and Fleiss, 1979). Kappa and ICC values

share the same interpretation (Davies and Fleiss, 1982). Kappa and ICC values range from 1.00 (perfect agreement) to – 1.00 (total disagreement) with values of zero indicating agreement equivalent to chance. Excellent agreement is indicated by kappa or ICC values of 0.75 and above; fair to good agreement, from 0.40 to 0.74; and poor agreement, below 0.39 (Fleiss, 1981; Landis and Koch, 1977).

Despite the large sample size of this test-retest study relative to prior studies of other psychiatric assessment instruments (Easton et al., 1997; Helzer et al. 1981; Semler et al., 1987; Williams et al., 1992; Wittchen et al., 1994, 1995, 1996), the prevalence of some diagnoses was too low (<3.0%) to yield stable estimates according to our predetermined power calculations. Therefore, kappa statistics for rare diagnoses with prevalences at test and/or retest of less than 0.03 (or 3%) were not reported here, including those for tranquilizer/sedative, inhalant/solvent, hallucinogen, club drug and heroin use disorders. Kappa statistics for cocaine and other stimulant diagnoses were only reported for the prior to past year time frame due to low past year prevalences. However, since ICCs assessed substance use disorder dimensional criteria scales independently of diagnoses, ICCs were reported for all alcohol and drug classes for both time frames with the exception of inhalants and solvents for the past-year time frame as the result of low prevalence. With respect to mood and anxiety diagnoses and PTSD, no reliability statistics are reported for agoraphobia and for manic episode in the past year timeframe due to low prevalence.

3. RESULTS

Table 1 shows the distribution of sociodemographic characteristics of the reliability sample. The sample contained more males than females with most respondents classified as white, married or cohabitating, aged 25 to 64 years, with some college education and employed. There were no differences between the reliability sample ($n = 1006$) and the remaining NESARC-III sample ($n = 35,309$) with respect to sex ($\chi^2 = 0.03$, $p < 0.87$), education ($\chi^2 = 0.89$, $p < 0.64$) or employment status ($\chi^2 = 0.83$, $p < 0.36$). However, there were difference with respect to race-ethnicity ($\chi = 83.1$, $p < 0.001$) and age ($\chi^2 = 66.21$, $p < 0.001$), with the reliability sample containing a somewhat larger number of whites and fewer respondents aged 65 years and older. Given the high response rate of individuals selected for the reliability re-interview (92%), there were no differences between participants ($n=1006$) and refusals ($n=92$) with respect to these sociodemographic measures or extent of psychopathology as assessed with the screening algorithm.

Table 2 shows the kappa statistics for all DSM-5 substance use disorder for which prevalence permitted. Kappa statistics indicated good to excellent reliability for past year diagnoses of tobacco use disorder (kappa = 0.87), and fair to good reliability for alcohol and other drug-specific use disorders (kappa = 0.41–0.62) for each time frame. With few exceptions, the ICCs for DSM-5 substance use disorder dimensional scales were in the good to excellent range (substantially higher than the kappa estimates), mainly from 0.60 to 0.85 (Table 3).

Kappa statistics for DSM-5 mood and anxiety disorders and PTSD (Table 4) were generally in the fair range (0.40–0.51), with only two exceptions (past specific phobia and generalized

anxiety). The corresponding ICCs for the dimensional scales associated with these disorders were generally in the good to excellent range ($ICC = 0.50\text{--}0.79$; Table 5). As shown in Table 6, reliability of ages at first onset and most recent episode for each substance use disorder, mood and anxiety disorder and PTSD assessed in this study were good to excellent ($ICC = 0.61\text{--}0.85$).

Kappa statistics for conduct disorder and the three PDs without the requirement for social/occupational dysfunction or distress were in the fair range ($0.46\text{--}0.54$; Table 4). The corresponding kappas for these disorders were somewhat lower when social/occupational dysfunction or distress was required for a diagnosis: 0.24 for conduct disorder; 0.22 for antisocial PD; 0.46 for borderline PD; and 0.33 for schizotypal PD. ICCs for conduct disorder and all three PDs fell in the fair to good range ($0.60\text{--}0.70$), with reliability of age at first onset and age at most recent episode in the good range for borderline PD ($ICC = 0.61$) and schizotypal PD ($ICC = 0.65$).

Results of the McNemar tests indicated that among the substance use disorders with sufficient prevalence to analyze, only the prevalence of alcohol use and nicotine use disorders was significantly greater ($p > 0.01$) at test relative to retest interview. With respect to all other disorders assessed in this study, the prevalences of major depressive episode and antisocial, borderline and schizotypal PDs were also significantly greater at test compared with retest interview.

4. DISCUSSION

In a large, rigorous test-retest design in the general population, the AUDADIS-5 DSM-5 demonstrated fair to excellent reliability for the binary DSM-5 substance use disorders examined in this study, with reliability for binary AUDADIS-5 DSM-5 major mood, anxiety, trauma and stress-related and personality diagnoses generally fair. These reliability statistics are comparable to those in psychometric studies of the AUDADIS-IV diagnoses in general population (Chatterji et al., 1997; Grant et al., 1995, 2003b, 2008) and clinical (Canino et al., 1999; Hasin et al., 1997) samples, and for DSM-IV diagnoses assessed by other fully-structured interviews (e.g., Wittchen et al., 1995, 1996). Interestingly, the reliabilities of conduct disorder and antisocial, borderline and schizotypal PD diagnoses were greater when the social/occupational dysfunction or distress criterion was not required. This result is consistent with clinical literature recommending that diagnostic assessment of PDs take a longitudinal perspective and utilize multiple sources of data: interviews with target individuals and collateral informants, and extant records (Clark, 2007; Pedersen et al., 2013; Zimmerman, 1994). While unavoidable in a large national epidemiologic survey, PD diagnoses requiring impairment or distress that relied solely on respondent self-report may have suffered from distortions or instabilities in respondents' self-appraisals, including lack of insight into the effects of symptomatic behaviors on role functioning (Clark, 2007; Pedersen et al., 2013; Zimmerman, 1994). In the case of conduct disorder, respondents also needed to recall and report impairment or distress from symptoms that may have occurred many years before interview (Rueter et al., 2000).

It may be unrealistic to expect that the reliability of psychiatric diagnoses, based largely on self-reports or observations can be much greater than the reliability of diagnoses in other areas of medicine based on directly observed evidence. Some researchers (Regier, 2012) have proposed that the standard of evaluation of the test-retest reliability of psychiatric diagnoses be consistent to that observed for other medical conditions. Although examples in the medical literature of test-retest reliability studies are as rare as those reported for psychiatric disorders, the reliability of substance use disorder and other psychiatric disorders reported here were generally consistent to those reported for other medical conditions. For example, Kappa values between 0.36 and 0.60 have been associated with diagnoses of anemia based on conjunctival inspection (Wallace et al., 2000) and diagnoses of skin and soft-tissue infection (Martin et al., 2010) and the reliability of pelvic examinations was associated with Kappa values ranging from 0.07 to 0.26 (Close et al., 2001).

Reliability was considerably higher when dimensional criteria scales for substance use disorders were examined, likely because continuous measures of substance use disorders do not impose thresholds that are psychometrically arbitrary (Hasin et al., 2013), and thus the continuous measures are more statistically informative than categorical measures. Moreover, mild or borderline cases of binary disorder variables impact categorical measures to a greater degree than their continuous counterparts, and the diagnostic threshold/borderline for substance use disorders is lower (hence, milder) for DSM-5 than it was for DSM-IV dependence.

The higher reliability of the dimensional measures of substance use disorders and other psychiatric disorders assessed in this study provides important support for the growing movement towards dimensional assessment of psychopathology (Helzer et al., 2006; Narrow and Kuhl, 2011; Regier, 2007). Dimensional measures can enable researchers and clinicians to indicate severity of psychiatric and substance use disorders and change over time in symptomatology, which is poorly represented by binary indicators of presence or absence of inherently dimensional conditions. The findings on dimensional measures are also highly relevant to phenotype development for the NESARC-III survey, in which DNA was collected from 24,381 participants that is currently being genotyped and prepared for genetic analyses. Although categorical diagnoses are useful for communication among clinicians and researchers and for policy decisions, corresponding dimensional measures should add invaluable information on symptomatology, reflecting the full severity range of the underlying phenomena for use in a number of research areas.

For most diagnoses, an order effect was observed, with the prevalence in the retest interview lower than in the initial test interview. However, only the prevalences of alcohol and nicotine use disorder, major depressive episode and antisocial, borderline and schizotypal PDs were significantly greater at test than retest interviews. An order effect has been observed in many other psychiatric test-retest reliability studies (Bromet et al., 1986; Helzer et al., 1981), including those conducted on the AUDADIS-IV (Grant 1995; Ruan 2008). That significant order effects were only found for alcohol and nicotine use disorders, major depressive episode and the three PDs and not for other substance use and other psychiatric disorders examined in this study may reflect the unique nature of these AUDADIS-5 modules. Unlike all other substance use disorders that are assessed together in a single module, alcohol and

nicotine use disorders are assessed in separate modules more extensive than the one used to collect information on other drug use disorders. Similarly, the major depressive episode module is more lengthy and complicated than modules used to assess other mood and anxiety disorders and PTSD. In view of this, the order effect for these disorders may have resulted, in part, from boredom or fatigue on the part of respondents who responded to these more extensive AUDADIS-5 modules within a short administration-readministration time interval (< 10 weeks). Further, the PD diagnoses were assessed on a lifetime basis consistent with their DSM-5 definitions. The lifetime assessment of these disorders coupled with the early onset of PDs may have contributed to the observed order effect. Future research should focus on this important, yet not clearly understood, phenomenon.

Although reliability of DSM-5 measures of substance use disorders and other psychiatric disorders is a fundamental and critical psychometric property, validity of the diagnoses is also crucial. Establishing the validity of substance abuse and psychiatric measures is considerably more complex than determining reliability, involving the correspondence of diagnostic and dimensional measures with a variety of external validators. Hence, investigation of AUDADIS-5 validity goes beyond the scope of the present report. Validity analyses of the AUDADIS-5 measures are underway, and findings will be presented shortly.

Limitations of the study are noted. First, the present sample had relatively few cases of some disorders, and hence, prevalence was too low to report on the reliability of some disorders. Such low prevalence is unavoidable in general population samples. To understand the reliability of the omitted AUDADIS-5 diagnoses, future studies should be conducted in samples with higher prevalences, e.g., clinical populations, as was done for the AUDADIS-IV (Canino et al., 1999; Hasin et al., 1997). Second, while presenting important information on overall diagnoses and dimensional measures, the present study did not examine individual items or criteria. Such examination, while beyond the scope of the present study and constrained by prevalence considerations, merits attention in future studies that can focus on individual disorders. Third, multiple additional factors may affect reliability, such as participant characteristics and interval between test and retest interviews. These factors were not examined in the present study, but warrant investigation in future studies.

Strengths of the study are also noted. The AUDADIS-5 is a fully structured diagnostic interview, administered in computer-assisted form, which removes the possibility of errors in skip patterns. Interviewers received extensive standardized training. The psychometric properties of the AUDADIS-5 were examined in the general population, the population for which the AUDADIS-5 was primarily designed. The study encompassed a large sample size, maintained complete blindness between test and retest interviews, and held mode of instrument administration and level of interviewer expertise constant in test and retest interviews. The study employed a test-retest administration interval that reduced confounding between diagnostic reliability and onset of new illness. All of these strengths contributed to a reliability study with a rigorous design for indicating the reliability of the AUDADIS-5 measures in the hands of multiple interviewers.

In view of the widespread importance and influence of psychiatric diagnoses and dimensional measures in national epidemiologic studies for research, policy and clinical

work, understanding the psychometric properties of these measures is crucial. While a series of studies was previously conducted with the DSM-IV version of the AUDADIS (AUDADIS-IV), the substantial changes in diagnostic definitions appearing in the DSM-5 necessitated a new study. Future psychometric analyses of the AUDADIS-5 will focus on multivariate analyses of factors affecting its reliability and validity (e.g. exploratory and confirmatory factor analyses, relationship of diagnoses and dimensional measures of psychopathology to external correlates). Meanwhile, the reliability of AUDADIS-5 diagnoses and dimensional scales found in this study suggests that the AUDADIS-5 can be a useful diagnostic tool in a variety of research settings.

Acknowledgments

Role of Funding Source

Funding sources: Support is acknowledged from the National Institutes of Health. Contract HHSN275200900007C, K05AA014223, and the New York State Psychiatric Institute (Hasin). The National Epidemiologic Survey on Alcohol Abuse and Related Conditions –III was sponsored by the National Institute on Alcohol and Alcoholism and funded, in part, by the intramural Program, NIAAA, National Institutes of Health. NIAAA had no role in the study design, in the collection, analyses or interpretation of the data, in the writing of the report or the decision to submit the article for publication.

References

- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorder. Third. American Psychiatric Association Publishing; Washington, DC: 1980.
- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorder. Fourth. American Psychiatric Association Publishing; Washington, DC: 1994.
- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorder. Fourth. American Psychiatric Association Publishing; Washington, DC: 2004. Revised
- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorder. Fifth. American Psychiatric Association Publishing; Washington, DC: 2013.
- Bromet EJ, Dunn LO, Connel MM, Dew MA, Schulber HC. Longterm reliability of diagnosing lifetime major depression in a community sample. *Arch Gen Psychiatry*. 1986; 43:435–440. [PubMed: 3964022]
- Canino GJ, Bravo M, Ramirez R, Febo V, Fernandez R, Hasin DS. The Spanish Alcohol Use Disorder and Associated Disabilities Interview Schedule (AUDADIS): reliability and concordance with clinical diagnoses in a Hispanic population. *J Stud Alcohol*. 1999; 60:790–799. [PubMed: 10606491]
- Center for Disease Control and Prevention. Unweighted response rates for the NHANES 2011–2012. Center for Disease Control and Prevention; Atlanta, GA: 2013.
- Chatterji S, Saunders JB, Vraiti R, Grant BF, Hasin D, Mager D. Reliability of the alcohol and drug modules of the Alcohol Use Disorder and Associated Disabilities Interview Schedule – Alcohol/ Drug –revised (AUDADIS-ADR): an international comparison. *Drug Alcohol Depend*. 1997; 47:171–185. [PubMed: 9306043]
- Clark LA. Assessment and diagnosis of personality disorder: perennial issues and an emerging reconceptualization. *Annu Rev Psychol*. 2007; 58:227–257. [PubMed: 16903806]
- Close RJ, Sachs CJ, Dyne PL. Reliability of bimanual pelvic examinations performed in emergency departments. *West J Med*. 2001; 175:240–244. [PubMed: 11577050]
- Davis M, Fleiss JL. Measuring agreement for multinomial data. *Biometrics*. 1982; 38:1047–1051.
- Easton C, Meza E, Mager D, Ulüg B, Kilic C, Gö ü A, Babor TF. Test-retest reliability of the alcohol and drug use disorder sections of the schedules for clinical assessment in neuropsychiatry (SCAN). *Drug Alcohol Depend*. 1997; 47:187–94. [PubMed: 9306044]
- Fleiss, JL. *Statistical Methods For Rates And Proportions*. Wiley; New York: 1981.

- Grant, BF., Dawson, DA., Hasin, DS. The Alcohol Use Disorder and Associated Disabilities Interview Schedule Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition Version (AUDADIS-IV). National Institute on Alcohol Abuse and Alcoholism; Bethesda, MD: 2001.
- Grant BF, Dawson DA, Chou SP, Dufour MC, Pickering RP. The 12-month prevalence and trends in DSM-IV alcohol abuse and dependence, United States, 1991–1992 and 2001–2002. *Drug Alcohol Depend.* 2004; 74:223–234. [PubMed: 15194200]
- Grant BF, Dawson DA, Stinson FS, Chou SP, Kay W, Pickering RP. The Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (AUDADIS-IV): reliability of alcohol consumption, tobacco use, family history of depression and psychiatric diagnostic modules in a general population sample. *Drug Alcohol Depend.* 2003a; 71:7–16. [PubMed: 12821201]
- Grant, BF., Goldstein, RB., Chou, SP., Saha, TD., Ruan, WJ., Huang, B., Smith, SM., Zhang, H., Jung, J., Pickering, RP., Aivadyan, C., Greenstein, E., Hasin, DS. The Alcohol Use Disorder and Associated Disabilities Interview Schedule-Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition Version (AUDADIS-5). National Institute on Alcohol Abuse and Alcoholism; Rockville, MD: 2011.
- Grant BF, Harford TC, Dawson DA, Chou PS, Pickering RP. The Alcohol Use Disorder and Associated Disabilities Interview Schedule (AUDADIS: reliability of alcohol and drug modules in a general population sample. *Drug Alcohol Depend.* 1995; 39:37–44. [PubMed: 7587973]
- Grant, BF., Moore, TC., Shepard, J., Kaplan, K. Source And Accuracy Statement: Wave 1 National Epidemiologic Survey On Alcohol And Related Conditions (NESARC). National Institute On Alcohol Abuse And Alcoholism; Bethesda, MD: 2003b.
- Grant, BF., Moore, TC., Shepard, J., Kaplan, K. Source And Accuracy Statement: Wave 2 National Epidemiologic Survey On Alcohol And Related Conditions (NESARC). National Institute On Alcohol Abuse And Alcoholism; Rockville, MD: 2004.
- Grant, BF., Amsbary, M., Chu, A., Sigman, R., Kali, J., Sugawana, Y., Jiao, R., Goldstein, RB., Jung, J., Zhang, H., Chou, PS., Saha, TD., Huang, B., Ruan, WJ., Pickering, RP., Smith, SM. Source and Accuracy Statement: National Epidemiologic Survey on Alcohol and Related Conditions-III (NESARC-III). National Institute on Alcohol Abuse and Alcoholism; Rockville, MD:
- Hasin, DS., Aivadyan, C., Greenstein, E. Psychiatric Research Interview for Substance Use and Mental Disorders, Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (PRISM-5) Version. Columbia University, Department of Psychiatry; New York, NY: 2011.
- Hasin DS, 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]
- Helzer JF, Robins LN, Croughan JL, Welner A. Renard Diagnostic Interview (RDI): its reliability and procedural validity with physicians and lay interviewers. *Arch Gen Psychiatry.* 1981; 38:393–398. [PubMed: 7212969]
- Helzer JF, van der Brink W, Guth SE. Should there be both categorical and dimensional criteria for the substance use disorders in DSM-5? *Addiction.* 2006; 101(Suppl 1):17–22. [PubMed: 16930157]
- Krueger RF, Watson D, Barlow DH. Introduction to the special section: toward a dimensionally based taxonomy of psychopathology. *J Abnorm Psychol.* 2005; 114:491–493. [PubMed: 16351372]
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977; 33:159. [PubMed: 843571]
- Marin JR, Bilker W, Lautenbach E, Alpern ER. Reliability of clinical examinations for pediatric skin and soft-tissue infections. *Pediatrics.* 2010; 126:925–930. [PubMed: 20974788]
- Meyers JL, Shmulewitz D, Aharonovich E, Waxman R, Frisch A, Weizman A, Spivak B, Edenberg HJ, Gelernter J, Hasin DS. Alcohol metabolizing genes and alcohol phenotypes in an Israeli household sample. *Alcohol Clin Exp Res.* 2013; 37:1872–81. [PubMed: 23895337]
- Meyers JL, Shmulewitz D, Wall MM, Keyes KM, Aharonovich E, Spivak B, Weizman A, Frisch A, Edenberg HJ, Gelernter J, Grant BF, Hasin DS. Childhood adversity moderates the effect of ADH1B on risk for alcohol-related phenotypes in Jewish Israeli drinkers. *Addict Biol.* 2013; 17:105–118.
- Narrow WE, Kuhl EA. Dimensional approaches to psychiatric diagnosis in DSM-5. *J Ment Health Policy Econ.* 2011; 14:197–200. [PubMed: 22345361]

- Pedersen G, Karterud S, Hummelen B, Wilberg T. The impact of extended longitudinal observation on the assessment of personality disorders. *Personal Ment Health*. 2013; 7:277–287. [PubMed: 24343977]
- Regier D. Merging categorical and dimensional diagnoses of mental disorders. *Epidemiol Psych Sci*. 2012; 21:267–269.
- Regier DA. Dimensional approaches to psychiatric classification: refining the research agenda for DSM-IV: an introduction. *Int J Methods Psychiatry Res*. 2007; 16:S1–S5.
- Ruan WJ, Goldstein RB, Chou SP, Smith SM, Saha RP, Pickering RP, Dawson DA, Huang B, Stinson FS, Grant BF. The Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (AUDADIS-IV): reliability of new psychiatric diagnostic modules and risk factors in a general population sample. *Drug Alcohol Depend*. 2008; 92:27–36. [PubMed: 17706375]
- Rueter MA, Chao W, Conger RD. The effect of systematic variation in retrospective conduct disorder reports on antisocial personality disorder diagnoses. *J Consult Clin Psychol*. 2000; 68:307–312. [PubMed: 10780131]
- Semler, G., Wittchen, HU. Das diagnostic interview schedule. Erst ergebnisse zur reliabilitat and differentiellen validitat der deutschen passing. In: Kommer, D., Roehric, B., editors. *Geneinde – Psychologische Perspektivin*. DGVt GWG; Koln, Germany: 1983. p. 109-117.
- Shouten H. Measuring pairwise agreement among many observers. *Biomol J*. 1980; 6:497–504.
- Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979; 86:420–428. [PubMed: 18839484]
- Substance Abuse and Mental Health Services Administration. Results from the 2012 National Survey on Drug Use and Health. Substance Abuse and Mental Health Services Administration; Rockville, MD: 2013.
- Vrasti R, Grant BF, Chatterji S, Ustun BT, Mager D, Olteanu I, Bodoi M. The reliability of the Romanian version of the alcohol module of the WHO Alcohol Use Disorder and Associated Disabilities Interview Schedule – Alcohol/Drug-Revised (AUDADIS-ADR). *Eur Addict Res*. 1998; 4:144–149. [PubMed: 9852366]
- Wacker, H.Bettegay, R.Mullejans, R.Schlossor, C.Stephanis, C.Ravavilas, A., Soldatos, C., editors. Using the CIDI-C in a General Population. Elsevier Science Publishers; New York: 1991. p. 27-36.
- Wallace DE, McGreal GT, O’Toole G, Holloway P, Wallace M, McDermott EW, Blake J. The influence of experience and specialization on the reliability of a common clinical sign. *Ann R Coll Surg Engl*. 2000; 82:336–338.
- Williams JB, Gibbon M, First MB, Spitzer RL, Davies M, Borus J, Howes MJ, Kane J, Pope HG Jr, Rounsaville B. The Structured Clinical Interview for DSM-III-R (SCID). II. Multisite test-retest reliability. *Arch Gen Psychiatry*. 1992; 49(8):630–6. [PubMed: 1637253]
- Wittchen HU. Reliability and validity studies of the WHO Composite International Diagnostic Interview (CIDI): a critical review. *J Psychiatry Res*. 1994; 28:57–84.
- Wittchen HU, Kessler R, Zhao S, Abelson J. Reliability and clinical validity of the UN-CIDI DSM-III-R generalized anxiety disorder. *J Psychiatry Res*. 1995; 29:95–110.
- Wittchen HU, Abelson JL, Kessler RC. Reliability and procedural validity of UM-CIDI DSM-III-R phobic disorders. *Psychol Med*. 1996; 26:1169–1177. [PubMed: 8931163]
- Zimmerman M. Diagnosing personality disorders: a review of issues and research methods. *Arch Gen Psychiatry*. 1994; 51:225–245. [PubMed: 8122959]

Highlights

- This study examined test-retest reliability of AUDADIS-5 DSM-5 substance use and psychiatric disorder diagnoses in a general population sample.
- Reliability of substance use disorder and selected psychiatric disorder diagnoses and criterion scales ranged from fair to excellent.
- The reliability of AUDADIS-5 DSM-5 diagnoses and criterion scales found in this study underscores the utility of the assessment instrument in general population samples.

Table 1

Sociodemographic characteristics of reliability study respondents (n=1,006)

Sociodemographic characteristics	N	% (SE)
Sex		
Male	437	43.4 (1.56)
Female	569	56.6 (1.56)
Age		
18–29	98	9.7 (0.93)
30–44	433	43.0 (1.56)
45–64	400	39.8 (1.54)
65 +	75	7.5 (0.82)
Race-ethnicity		
White	651	64.7 (1.51)
Black	154	15.3 (1.14)
American Indian/Alaska native	28	2.80 (0.52)
Asian/Pacific Islander	26	2.60 (0.52)
Hispanic	147	14.6 (1.11)
Education		
Less than high school	142	14.1 (1.09)
High school	271	27.0 (1.39)
Some college or higher	593	58.9 (1.55)
Employed		
Yes	706	70.2 (1.46)
No	300	29.8 (1.46)

Note: SE = Standard error of percentage.

Table 2

Reliability of AUDADIS-5 DSM-5 substance use disorder diagnoses

Disorder	Kappa (SE)	Prevalence % (Test/Retest)
Past year use disorder		
Tobacco	0.87 (0.02)	55.98/50.60
Alcohol	0.62 (0.03)	31.41/24.35
Cannabis	0.41 (0.06)	6.56/5.96
Cocaine	<i>_a</i>	<i>_a</i>
Stimulant	<i>_a</i>	<i>_a</i>
Opioids	0.40 (0.07)	3.48/3.18
Prior to the past year use disorder		
Tobacco	0.50 (0.04)	46.02/42.63
Alcohol	0.60 (0.03)	53.68/46.42
Cannabis	0.41 (0.04)	14.91/13.52
Cocaine	0.53 (0.05)	8.05/7.06
Stimulant	0.54 (0.05)	7.46/5.17
Opioids	0.47 (0.06)	4.87/5.86

Note: SE = Standard errors.

^aPrevalence at test and/or retest of <0.03.

Table 3

Reliability of dimensional AUDADIS-5 DSM-5 substance use disorder criteria scales

Disorder	ICC	95% CI
Past year use disorder		
Tobacco	0.84	(0.83–0.87)
Alcohol	0.85	(0.83–0.87)
Cannabis	0.70	(0.66–0.74)
Cocaine	0.75	(0.72–0.78)
Stimulant	0.80	(0.77–0.82)
Sedative/tranquilizer	0.80	(0.77–0.83)
Hallucinogen	0.59	(0.47–0.63)
Inhalant/solvent	<i>_a</i>	<i>_a</i>
Club drug	0.45	(0.37–0.52)
Opioids	0.71	(0.67–0.74)
Heroin	0.50	(0.42–0.56)
Prior to past year use disorder		
Tobacco	0.83	(0.81–0.85)
Alcohol	0.83	(0.81–0.84)
Cannabis	0.71	(0.67–0.74)
Cocaine	0.84	(0.82–0.86)
Stimulant	0.84	(0.81–0.85)
Sedative/tranquilizer	0.72	(0.68–0.75)
Hallucinogen	0.61	(0.56–0.65)
Inhalant/solvent	0.61	(0.56–0.65)
Club drug	0.73	(0.69–0.76)
Opioids	0.79	(0.76–0.81)
Heroin	0.84	(0.83–0.86)

Note: ICC = Intraclass correlation coefficient; CI = Confidence interval.

^aPrevalence at test and/or retest of <0.03.

Table 4

Reliability of AUDADIS-5 DSM-5 mood, anxiety and trauma and stress-related and personality disorder criteria scales

Disorder	Kappa (SE)	Prevalence % (Test/Retest)
Past year		
Major depression	0.40 (0.04)	39.04/30.0
Manic	<i>a</i>	<i>a</i>
Dysthymia	0.39 (0.05)	15.34/13.26
Panic	0.49 (0.05)	16.07/14.09
Social phobia	0.43 (0.06)	13.29/13.21
Specific phobia	0.50 (0.05)	15.08/13.10
Generalized anxiety	0.51 (0.04)	32.94/29.98
Posttraumatic stress	0.41 (0.06)	30.16/25.16
Prior to past year		
Major depression	0.40 (0.04)	65.94/50.02
Manic	0.42 (0.04)	6.37/5.25
Dysthymia	0.48 (0.05)	22.91/17.93
Panic	0.48 (0.05)	22.02/19.05
Social phobia	0.46 (0.05)	15.87/16.27
Specific phobia	0.35 (0.06)	16.47/14.68
Generalized anxiety	0.39 (0.04)	43.06/37.14
Posttraumatic stress	0.44 (0.04)	35.52/29.92
Lifetime		
Conduct	0.46 (0.06)	14.29/12.70
Antisocial	0.46 (0.05)	13.69/11.90
Borderline	0.54 (0.04)	37.37/31.63
Schizotypal	0.49 (0.05)	22.51/18.53

^aPrevalence at test and/or retest of <0.03.

Table 5

Reliability of AUDADIS-5 DSM-5 dimensional mood, anxiety, trauma and stress-related and personality scales

Disorder	ICC	95% CI
Major depression	0.59	(0.50–0.65)
Mania	0.50	(0.38–0.56)
Dysthymia	0.65	(0.58–0.70)
Panic	0.79	(0.75–0.82)
Social phobia	0.74	(0.69–0.80)
Specific phobia	0.78	(0.73–0.81)
Generalized anxiety	0.79	(0.72–0.83)
Posttraumatic stress	0.69	(0.64–0.75)
Intrusion	0.74	(0.70–0.78)
Avoidance	0.66	(0.60–0.72)
Change in cognition/mood	0.72	(0.67–0.77)
Change in arousal/activity	0.71	(0.65–0.75)
Conduct	0.70	(0.63–0.75)
Antisocial	0.60	(0.51–0.68)
Borderline	0.66	(0.60–0.72)
Schizotypal	0.70	(0.63–0.75)

Note: ICC = Intraclass correlation coefficient; CI = Confidence interval.

Table 6

Reliability of age at first onset and age at most recent onset of AUDADIS-5 DSM-5 substance, mood, anxiety and trauma and stress-related disorders

Disorder	Age At First Onset		Age At Most Recent Onset	
	ICC	95% CI	ICC	95% CI
Substance use disorder				
Tobacco	0.61	(0.50–0.69)	0.76	(0.69–0.81)
Alcohol	0.74	(0.69–0.79)	0.80	(0.76–0.84)
Cannabis	0.76	(0.64–0.83)	0.70	(0.57–0.82)
Cocaine	0.85	(0.74–0.92)	0.86	(0.69–0.94)
Stimulant	0.91	(0.83–0.95)	0.67	(0.53–0.89)
Opioid	0.64	(0.55–0.73)	0.51	(0.40–0.61)
Mood, anxiety and trauma and stress-related disorders				
Major depression	0.92	(0.81–0.97)	0.86	(0.82–0.89)
Manic	<i>_a</i>	<i>_a</i>		
Dysthymia	0.75	(0.58–0.85)	0.87	(0.78–0.92)
Panic	0.79	(0.70–0.87)	0.62	(0.42–0.75)
Social phobia	0.69	(0.59–0.77)	0.48	(0.31–0.62)
Specific phobia	0.64	(0.55–0.71)	0.42	(0.28–0.54)
Generalized anxiety	0.82	(0.77–0.86)	0.73	(0.65–0.79)
Posttraumatic stress	0.79	(0.69–0.87)	0.62	(0.42–0.75)
Conduct/personality disorder				
Conduct	<i>_b</i>		<i>_b</i>	
Antisocial	<i>_b</i>		<i>_b</i>	
Borderline	0.61	(0.50–0.69)	0.86	(0.83–0.89)
Schizotypal	0.65	(0.54–0.74)	0.87	(0.83–0.90)

Note: ICC = Intraclass correlation coefficient; CI = Confidence interval.

^aPrevalence at test and/or retest of <0.03.

^bNot directly assessed.