



Published in final edited form as:

Drug Alcohol Depend. 2008 January 1; 92(1-3): 27–36. doi:10.1016/j.drugalcdep.2007.06.001.

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

W. June Ruan, Risë B. Goldstein, S. Patricia Chou, Sharon M. Smith, Tulshi D. Saha, Roger P. Pickering, Deborah A. Dawson, Boji Huang, Frederick S. Stinson, and Bridget F. Grant
Laboratory of Epidemiology and Biometry, Division of Intramural Clinical and Biological Research, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Department of Health and Human Services (Drs. Goldstein, Chou, Smith, Saha, Dawson, Huang, Stinson, and Grant, Ms. Ruan, and Mr. Pickering), Bethesda, MD 20852

Abstract

This study presents test-retest reliability statistics and information on internal consistency for new diagnostic modules and risk factor of alcohol, drug, and psychiatric disorders the Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (AUDADIS-IV). Test-retest statistics were derived from a random sample of 1,899 adults selected from 34,653 respondents who participated in the 2004–2005 Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). Internal consistency of continuous scales was assessed using the entire Wave 2 NESARC. Both test and retest interviews were conducted face-to-face. Test-retest and internal consistency results for diagnoses and symptom scales associated with posttraumatic stress disorder, attention-deficit/hyperactivity disorder, and borderline, narcissistic, and schizotypal personality disorders were predominantly good ($\kappa > 0.63$; ICC > 0.69 ; $\alpha > 0.75$) and reliability for risk factor measures fell within the good to excellent range (intraclass correlations = 0.50–0.94; $\alpha = 0.64$ –0.90). The high degree of reliability found in this study suggests that new AUDADIS-IV diagnostic measures can be useful tools in research settings. The availability of highly reliable measures of risk factors of alcohol, drug, and psychiatric disorders will contribute to the validity of conclusions drawn from future research in the domains of substance use disorder and psychiatric epidemiology.

Keywords

Reliability; alcohol and drug use disorders; risk factors; test-retest reliability; general population; internal consistency

Corresponding Author/Author to Whom Requests for Reprints Should Be Addressed: Bridget F. Grant, Ph.D., Ph.D., Laboratory of Epidemiology and Biometry, Room 3077, Division of Intramural Clinical and Biological Research, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, 5635 Fishers Ln., M.S. 9304, Bethesda, MD 20892-9304, Tel. 301-443-7370, Fax 301-443-1400, E-mail bgrant@willco.niaaa.nih.gov.

Disclaimer: The views and opinions expressed in this report 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.

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.

1. Introduction

The Alcohol Use Disorder and Associated Disabilities Schedule – Diagnostic and Statistical Manual of Mental Disorders – Fourth Edition (AUDADIS-IV) is a fully structured diagnostic interview designed to assess alcohol, drug, and mental disorders according to DSM-IV diagnostic criteria in both clinical and general populations (Grant et al. 2001). The central focus of the AUDADIS-IV is on the measurement of alcohol and drug use, and alcohol and drug use disorders. Over the past 15 years, additional modules assessing a wide range of other psychiatric disorders and risk factors have been incorporated into the AUDADIS-IV. Initially, in 1991–1992, the AUDADIS-IV was used in a large-scale survey of the U.S. population, the National Institute on Alcohol Abuse and Alcoholism's (NIAAA) National Longitudinal Alcohol Epidemiologic Survey (NLAES: Grant et al., 1992). Since then, the AUDADIS-IV has been used in the 2001–2002 NIAAA Wave 1 National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) and its three-year Wave 2 follow-up, conducted in 2004–2005 (Grant et al., 2004a, 2004b).

The reliability of various modules of the AUDADIS-IV has been extensively tested in U.S. clinical (Canino et al., 1999; Hasin et al., 1997) and general population (Grant et al., 1995, 2003) samples and in other countries as part of the National Institutes of Health/World Health Organization Reliability and Validity Study (Chatterji et al., 1997; Vrsti et al., 1998). In these U.S. general population studies, 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 associated consumption measures ($\kappa = 0.68\text{--}0.97$) and family history of psychopathology ($\kappa = 0.56\text{--}0.87$) measures were good to excellent (Grant et al., 1995, 2003), while reliability coefficients for the disorders and measures were extremely similar in substance abuse (Hasin et al., 1997) and primary care settings (Canino et al., 1999). Reliability studies conducted in Jebel, Romania, Bangalore, India, and Sydney, Australia also yielded kappas ranging between 0.57 and 0.96 for DSM-IV alcohol and drug use disorders (Chatterji et al., 1997). Corresponding reliability coefficients derived from U.S. general population samples for DSM-IV mood ($\kappa = 0.58\text{--}0.65$), anxiety ($\kappa = 0.40\text{--}0.52$) and personality ($\kappa = 0.40\text{--}0.67$) disorders were somewhat lower, but in the fair to good range.

The present study represents the latest in a long series of test-retest studies of the AUDADIS-IV. This test-retest study was conducted as part of the Wave 2 NESARC and included 1,899 respondents who participated in the Wave 2 NESARC proper. Test-retest reliability statistics are presented for five new DSM-IV psychiatric disorders introduced for the first time in the Wave 2 NESARC: posttraumatic stress disorder (PTSD), attention-deficit/hyperactivity disorder (ADHD), and borderline, narcissistic, and schizotypal PDs. The test-retest reliability of major risk factors for alcohol, drug, and mental disorders was also examined. These risk factors included discrimination, acculturation, race-ethnic orientation, childhood adverse experiences, stressful life events, social support and social networks, perceived stress, alcoholism stigma, intimate partner violence, and sexual orientation, attraction, and behavior. In addition, internal consistency of symptom and risk factor scales was assessed using the entire ($N = 34,653$) Wave 2 NESARC.

Understanding the reliability of new diagnostic measures introduced in the Wave 2 NESARC is critical for alcohol and drug use disorder epidemiology because PTSD, ADHD, and borderline, narcissistic, and schizotypal PDs have been demonstrated to be highly comorbid with substance use disorders in both clinical and general population surveys (Grant et al., 2004a; Kessler et al., 2005a, 2005b, 2005c; Lenzenweger et al., 2006). Further, major risk factors examined in this study have shown strong and significant associations with alcohol and drug use disorders and a broad range of other psychiatric disorders (Brisco-Smith and Hinshaw, 2006; Dube et al., 2007; Kessler et al., 2005a; Krause et al., 2006; Mays and Cochran, 2001;

Sandfort et al., 2001; Sartor et al., 2007; Vega et al., 2004; Yoshihama and Horrocks, 2003). The reliability of risk factors for alcohol, drug, and psychiatric disorders is just as important as the reliability of diagnoses, as both determine the validity of conclusions reached in any epidemiologic or clinical study in this domain.

2. Method

2.1. Sample

The 2004–2005 Wave 2 NESARC (Grant et al., 2006) is the second wave that follows upon the Wave 1 NESARC, conducted in 2001–2002 and described in detail elsewhere (Grant et al., 2004a, 2004b). The Wave 1 NESARC was a representative sample of the adult population of the United States, including Alaska and Hawaii. The target population was the civilian population, 18 years and older, residing in households and group quarters, including military living off base, boarding houses, rooming houses, nontransient hotels and motels, shelters, facilities for housing workers, college quarters, and group homes. Face-to-face interviews were conducted with 43,093 respondents. The NESARC oversampled Blacks, Hispanics, and young adults aged 18 to 24 years. The overall response rate was 81.0%.

In the Wave 2 NESARC, attempts were made to conduct face-to-face reinterviews with all 43,093 respondents who had completed the Wave 1 interview. Excluding respondents not eligible for the Wave 2 interview (e.g., because they were deceased, had been deported, or were mentally or physically impaired), the Wave 2 response rate was 86.7%, reflecting 34,653 completed Wave 2 interviews. The cumulative response rate at Wave 2 was the product of this Wave 2 response rate and the response rate from Wave 1 (81.0%), or 70.2%. As in Wave 1, the Wave 2 NESARC data were weighted to reflect design characteristics of the NESARC survey and to account for oversampling. Adjustment for nonresponse across numerous variables, including age, race-ethnicity, sex, region, and presence of any Wave 1 NESARC substance use disorder or psychiatric disorder, was performed at the household and person levels. Weighted data were then adjusted to be representative of the civilian population of the United States on a variety of socioeconomic variables including region, age, race-ethnicity, and sex, based on the 2000 Decennial Census.

For the current reliability study, Wave 2 NESARC respondents who completed the entire AUDADIS-IV interview were randomly selected to participate in the retest interview at one of four Census Bureau regional offices. Each of the four regional offices conducted in-person reinterviews within 10 weeks after respondents had participated in the Wave 2 NESARC survey proper. The AUDADIS-IV sections tested, sample sizes ($n=450-552$), and response rates (87.9%–93.7%) for each test-retest site are shown in Table 1. Table 2 shows the sociodemographic characteristics of the study respondents at each regional office. No statistical differences were found between respondents and nonrespondents to the retest interview with regard to sex, race-ethnicity, education, or age, primarily because of the high response rates achieved in these test-retest studies. At sites that tested the DSM-IV diagnostic variables (Philadelphia and Detroit), respondents did not differ significantly from nonrespondents with respect to having been diagnosed with a DSM-IV psychiatric disorder based on the test interview.

Internal consistency was examined using the entire Wave 2 NESARC sample. Sociodemographic characteristics of this diverse sample are presented in Table 3.

2.2. Test-retest design

The test-retest design of the present study was identical to all prior test-retest studies of the AUDADIS-IV. Interviewer assignments during the initial test and subsequent retest were

randomized among the staff. There were approximately 82 to 114 interviewers who administered the test interview and between 14 and 36 interviewers who administered the retest interview at each study site. All interviewers had at least 5 years of field experience working on health-related surveys; each interviewer completed a week-long self-study and participated in a week-long in-class training session. No interviewer was allowed to interview the same respondent twice. Interviewers administering the retest interviews were also blind to the results of the initial interview. Any communication between the interviewers concerning information collected during interviews was strictly confidential, in accordance with Bureau of the Census and NIAAA standards and requirements. The test and retest intervals varied little by site, ranging from 2 to 10 weeks with averages of approximately 5.9 (SD = 0.58) weeks.

2.3. Measures tested

2.3.1. DSM-IV diagnostic modules—This test-retest study focused on five DSM-IV diagnoses, including two Axis I disorders, PTSD and ADHD. Test-retest statistics are presented for 12-month and lifetime diagnoses of PTSD and both childhood (before age 18) and adult (since age 18) ADHD. Reliability of DSM-IV lifetime diagnoses of Axis II disorders (borderline, schizotypal, and narcissistic PDs) are also reported. For each DSM-IV disorder, intraclass correlation coefficients were also computed on scales constructed from the associated diagnostic symptom items (i.e., symptom counts).

2.3.2. Acculturation and race-ethnic orientation—The 11-item acculturation scale was adopted from the Brief Acculturation Rating Scale-II for Mexican Americans (ARSM-A-II: Coronado et al., 2005; Cuellar et al., 1995, 2004; Deyo et al., 1985; Solis et al., 1990) and from a similar acculturation scale developed and evaluated among East Asian Americans, the East Asian Acculturation Measure (EAAM: Barry, 2001). Acculturation items focused strongly on language use, proficiency, and preference, as well as race-ethnic social preference. These questions were asked separately of Hispanics, Asians and Pacific Islanders, and all others, and rated using a 5-point scale (1 to 5). For the language items, the response categories were only Spanish (or Asian/Pacific Islander language or other non-English language), more Spanish than English, both equally, more English than Spanish, and only English. Social preference items (e.g., those with whom the respondent usually socialized) were associated with similar response categories, that is, from all Hispanic/Latino (or all Asian/Pacific Islander or all from my race-ethnic group) to all other race-ethnic group. Acculturation items were summed to form a scale with values ranging from 1 to 55, with higher scores indicating lower acculturation.

The race-ethnic identification scale, designed to measure the part of a person's self-concept that derives from his or her knowledge of, or membership in, a social group, was adopted from earlier scales that were assessed among diverse race-ethnic groups (Barry, 2002; Phinney, 1992; Rahim-Williams et al., 2007). Items conceptualized race-ethnic identification, race-ethnic pride, importance of race-ethnic heritage, role of race-ethnic background in respondents' interactions with others, and shared race-ethnic values, attitudes, and behaviors. The 8-item race-ethnic identification scale was scored on a 6-point Likert scale (1=strongly agree, 2=agree, 3=somewhat agree, 4=somewhat disagree, 5=disagree, 6=strongly disagree), with a range of values from 1 to 48. After appropriate items were reverse coded, higher scores indicated higher degrees of race-ethnic identification.

2.3.3. Discrimination—Six separate discrimination scales appearing in the AUDADIS-IV were modeled after the Experiences with Discrimination (EOD) scales developed by Krieger and colleagues (Krieger, 1990, 1998, 2003; Krieger and Sidney, 1997; Krieger et al., 2005). The original EOD scales were expanded to include discrimination based on overweight, religion, and physical disability, in addition to race-ethnicity, gender, and sexual orientation (gay, lesbian, or bisexual), as well as to accommodate two time periods: the past 12 months,

and prior to the past 12 months. Although these discrimination scales are conceptualized as measuring experiences of discrimination, not perceived discrimination, it is not clear whether perceptions and experiences with discrimination can be differentiated. The number and types of questions asked for each of the discrimination scales differed (Table 4), depending on the nature of the specific discrimination experiences being assessed. Each discrimination question assigned a value of 0 = “never,” 1 = “almost never,” 2 = “sometimes,” 3 = “fairly often,” or 4 = “very often,” and values were summed across each time period.

Each discrimination scale was accompanied by 2 questions assessing reactions to unfair treatment or discrimination questions, proposed by Stancil et al. (2000), and shown in Table 3. These items were collectively scored as engaged (“do something/talk to others,” coded 2; moderate, i.e., “do something/keep to self,” coded 1; and passive, i.e., “accept it/keep to self,” coded 0).

2.3.4. Stressful life events and perceived stress—The AUDADIS-IV contained measures of objective stress and perceived stress. The objective stress measure included the occurrence of 14 stressful life events during the 12 months preceding the Wave 2 interview: (1) respondent moving or having someone new come to live with him or her; (2) being fired or laid off from a job; (3) being unemployed and looking for a job for more than a month; (4) trouble with a boss or coworker; (5) changes in job, job responsibilities, or work hours; (6) marital separation or divorce or breakup of a steady relationship; (7) serious problems with a neighbor, friend, or relative; (8) financial crisis, declaration of bankruptcy, or being unable to pay bills more than once; (9) serious trouble with the police or law; (10) being the victim of theft; (11) intentional damage to the respondent’s property or the property of someone who lived in the respondent’s home; (12) death of a family member or close friend; (13) one or more assaults, attacks, or muggings perpetrated against family members or close friends; and (14) family members or close friends having serious trouble with the police or law. The stressful life event scale was defined as the sum of all experienced life events, each coded 1 (range 0 to 14).

The Perceived Stress Scale-4 (PSS4) appearing on the AUDADIS-IV has been conceptualized as an assessment of cognitively mediated emotional responses to objectively stressful life events and not the objective life events themselves (Cohen and Williamson 1988; Cohen et al., 1983). Items on this scale included frequencies (0=never, 1=almost never, 2=sometimes, 3=fairly often, and 4=very often) during the past 12 months when respondents felt: (1) able to control important things in their lives; (2) confident about their abilities to handle personal problems; (3) that things were going their way; and (4) that difficulties were piling up so high that they could not overcome them. Items 2 and 3 were reverse coded so that higher scores indicated greater perceived stress.

2.3.5. Interpersonal support social networks—The 12-item general population version of the Interpersonal Support Evaluation List (ISEL12) was used to measure respondents’ perceptions of the current availability to them of potential social resources (Cohen and Hoberman, 1983; Cohen et al., 1985). The items are counterbalanced for social desirability; that is, half of the items are positive statements about social relationships (e.g., “If I were sick, I know I would find someone to help me with my daily chores”), while half are negative statements (e.g., “I feel that there is no one I can share my most private worries and fears with”). All items were coded 1=definitely false, 2=probably false, 3=probably true, 4=definitely true (range 1–16), with positive items negatively coded so that lower scores on the ISEL12 corresponded to lower levels of perceived availability of social resources. Detailed descriptions of these items appear in Cohen et al. (1985).

A detailed description of the Social Network Index (SNI) developed by Cohen and colleagues is given elsewhere (Cohen et al., 1997). The SNI assesses participation in 12 types of social relationships including relationships with spouse or partner, parents, parents-in-law, children, other relatives, close friends, workmates, schoolmates or teachers, neighbors, fellow volunteers (e.g., charity, community service work), members of groups without religious affiliation (e.g., social, recreational, professional), and members of religious groups. The total numbers of persons in each of these relationships whom respondents saw or talked with on the phone or Internet at least once every 2 weeks formed a continuous scale of the number of network members.

2.3.6. Alcoholism – stigma—The 12-item alcoholism stigma scale measures perceived devaluation of current or former alcoholics. The scale assesses respondents' agreement (1=strongly agree to 6=strongly disagree) with statements indicating that most people devalue alcoholics by perceiving them as failures, as less intelligent than other persons, and as individuals whose opinions need not be taken seriously. The alcoholism stigma scale was adapted from Link and colleagues' Perceived Devaluation Scale that was developed for general psychiatric patients as described in detail elsewhere (Link, 1982; Link et al., 1991, 2001). The scale is balanced with a high level of perceived devaluation indicated by agreement with 6 of the items and by disagreement with the remaining 6 items. Items are summed to produce a scale with scores ranging from 1 to 72, high scores reflecting a strong perception of devaluation.

2.3.7. Adverse childhood events—All questions about adverse childhood events (ACEs) related to respondents' first 17 years of life. Questions were adapted from the Adverse Childhood Events study (Dong et al., 2003; Dube et al., 2003) and were originally part of an extensive battery of questions appearing on the Conflict Tactics Scale (CTS; Straus, 1979; Straus and Gelles, 1990) and the Childhood Trauma Questionnaire (CTQ; Bernstein et al., 1994; Wyatt, 1985). Response categories for most scale items were 1=never, 2=almost never, 3=sometimes, 4=fairly often, and 5=very often. Response category values were summed across items to produce scales. The physical neglect items all required reverse coding and were associated with the response categories of 1=never true, 2=rarely true, 3=sometimes true, 4=often true, and 5=very often true.

Emotional abuse and physical abuse were defined by 3 and 2 questions from the CTS, respectively. For emotional abuse, questions asked how often respondents' parents or caregivers living in their home: (1) swore at, insulted, or said very hurtful things to respondents; (2) threatened to hit or throw something at respondents but didn't; and (3) acted in any other way that made respondents afraid that they would be physically hurt or injured. For physical abuse, the frequency of pushing, grabbing, shoving, slapping or hitting, and hitting so hard that respondents had marks or bruises or were injured were ascertained.

For both emotional and physical neglect, sets of 5 CTQ items were used. Items assessing physical neglect included the frequency with which respondents: (1) were made to do chores too difficult or dangerous for someone their age; (2) were left alone or unsupervised when they were too young to be alone; (3) went without things they needed like clothing, shoes, or school supplies; (4) went hungry or were not being provided with regular meals; and (5) had parents or caregivers fail to get them medical treatment when respondents were sick or hurt. Items assessing emotional neglect included the following: (1) there was someone in the respondent's family who wanted him or her to be a success; (2) there was someone in the family who helped the respondent feel important or special; (3) the respondent's family was a source of strength and support; (4) the respondent felt that he or she was part of a close-knit family; and (5) someone in the respondent's family believed in him or her.

Childhood sexual abuse was defined by 4 questions developed by Wyatt (1985). All sexual abuse questions asked about sexual experiences with an adult or any other person and were restricted to behaviors that respondents did not want and were experienced when respondents were too young to know what was happening. The sexual abuse scale included questions about touching and fondling, touching in a sexual way, and attempting and actually having sexual intercourse.

Having a battered mother or female caregiver was defined by 4 questions from the CTS that assessed the frequency with which each respondent's father, stepfather, foster or adoptive father, or mother's boyfriend engaged in any of the following behaviors toward the respondent's mother, stepmother, foster or adoptive mother, or father's girlfriend: (1) pushing, grabbing, slapping, or throwing something at her; (2) kicking, biting, hitting her with a fist, or hitting her with something hard; (3) repeatedly hitting her for at least a few minutes; or (4) threatening her with a knife or gun, or using a knife or gun to hurt her.

The other ACE scale consisted of 6 items that measured global household dysfunction. All questions were coded 1=yes and 0=no, and summed across items to yield a scale score ranging from 0 to 6. The items included the following experiences before respondents were 18 years of age: having a parent or other adult with whom they lived who had an alcohol or drug problem, went to jail or prison, was treated or hospitalized for a mental illness, or attempted or committed suicide.

2.3.8 Intimate partner violence—Intimate partner violence (IPV) scales were adapted from other studies (Cunradi et al., 1999; Lipsky et al., 2006; White et al., 2002) for both respondents and their partners as perpetrators of abuse and each asked about abusive behaviors in the 12 months preceding the interview. The frequencies (0=never, 1=once, 2=2 to 3 times, 3=once per month, and 4=more than once per month) of 6 abusive behaviors were ascertained and values were summed across items to yield a scale with values ranging from 0 to 24, higher scores indicating greater IPV. IPV questions asked about the frequencies of: (1) pushing, grabbing, or shoving; (2) slapping, kicking, biting, or hitting; (3) threatening with a weapon like a knife or gun; (4) cutting or bruising; (5) having forced sex; and (6) inflicting an injury that required medical care.

2.3.9. Sexual orientation, attraction, and behavior—The 3 sexual questions were developed with Dr. Randall L. Sell of Columbia University and were based on questions appearing in the 1999 National Youth Risk Behavior Surveillance System (YRBSS: Centers for Disease Control and Prevention, 2001) and 2001 National Behavioral Risk Factor Surveillance System (BRFSS: Centers for Disease Control and Prevention, 2003) conducted by the Centers for Disease Control and Prevention, as well as the 1988 National Health Interview Survey (Centers for Disease Control and Prevention, 1988), among other national surveys. The first question measured sexual orientation by asking respondents which of three categories (heterosexual, gay or lesbian, or bisexual) best describes them. Respondents were also asked if they were sexually attracted only to males, mostly to males, equally to males and females, mostly to females, or only to females. Using the response categories associated with sexual attraction, respondents were also asked to describe the partners with whom they had sex.

2.4. Statistical analysis

The present study used the same statistical methodology as all prior test-retest studies of the AUDADIS-IV to derive reliability coefficients. That is, for dichotomous data elements, kappa was used as a measure of reliability and is defined as a measure of pairwise agreement corrected for chance (Fleiss, 1981; Shouten, 1980). For continuous measures derived from the test-retest

study, intraclass correlation coefficients (ICC) are presented as measures of reliability. Both measures assess stability reliability. Since our reliability design assumed that interviewers were randomly drawn from a larger population of interviewers, we used a one-way random effects ANOVA model to derive intraclass correlation coefficients (Shrout and Fleiss, 1979). Kappa and ICC 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 no better than 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, 1997).

Reliability of the composite measures that consisted of multiple items was assessed in the entire Wave 2 NESARC using Cronbach's alpha statistic (Cronbach, 1951). Alpha reflects the internal consistency of items within a scale and is a measure of the squared correlation between observed and true scores. Alpha ranges from a negative 1.0 to a positive 1.0, and usually values of 0.70 and above are considered acceptable (Nunnally, 1978). Other psychometric researchers (George and Mallery, 2003) provide the following standards in interpreting alpha values: ≥ 0.90 – excellent; ≥ 0.80 and < 0.90 – good; ≥ 0.70 and < 0.80 – acceptable; ≥ 0.60 and < 0.70 questionable; ≥ 0.50 and < 0.60 – poor; and ≥ 0.50 unacceptable. Because all individual items within each scale were comparably scaled, raw alpha coefficients were reported.

For the purposes of the present analyses, reliability statistics were calculated on all major variables with prevalences of 0.01 or greater. All analyses were conducted using the total sample from each site. Analyses by age, gender, and race-ethnicity were precluded by small sample sizes.

3. Results

3.1. Reliability of AUDADIS-IV DSM-IV diagnoses

Table 5 shows the reliability of the new DSM-IV diagnostic modules appearing in the Wave 2 NESARC. Test-retest reliability coefficients for PTSD, ADHD, and borderline, narcissistic, and schizotypal PDs were in the fair to good range for diagnoses (kappa values=0.63–0.77) and associated symptom scales (ICCs=0.69–0.75). Internal consistency of symptom scales associated with PTSD, ADHD, and borderline, narcissistic, and schizotypal PDs fell within the good range (alpha = 0.75–0.89).

3.2. Reliability of AUDADIS-IV risk factors

Table 6 presents the reliability results for AUDADIS-IV measures of risk factors for alcohol, drug, and psychiatric disorders. Test-retest reliabilities of the acculturation and race-ethnic identification scales were in the excellent range, as were the perceived stress and stressful life event scales (ICCs=0.78–0.94). All discrimination scales demonstrated at least good reliabilities (ICCs ≥ 0.50), with excellent reliability observed for the sexual orientation discrimination scales in both time periods (ICCs=0.78, 0.82). Kappa coefficients for measures of reaction to discrimination proposed by Stancil et al. (2000) for each type of discrimination were similar, ranging from 0.58 to 0.63. Reliability coefficients for interpersonal support evaluation (kappa=0.63), social networks (ICC=0.70), and alcoholism stigma (ICC=0.93) indicated good to excellent reliability. The childhood adverse experience scales were also associated with good to excellent reliability (ICCs=0.69–0.88). Kappa values for sexual orientation, sexual attraction, and sexual behavior fell in the good range (0.60–0.66). With few exceptions, continuous measures achieved better than acceptable levels (≥ 0.70) of internal consistency.

4. Discussion

The AUDADIS-IV demonstrated good test-retest and internal consistency reliability for the new diagnostic modules introduced in Wave 2 of the NESARC. Reliabilities of PTSD and ADHD dichotomous diagnoses were good, with slightly higher ICC and alpha coefficients generally observed for their dimensional scale counterparts. The reliability of lifetime PTSD was lower than that for the corresponding 12-month diagnosis. This result, found in past research (Grant et al., 2003), implicates recall bias as one factor that may lead to diminutions in reliability for diagnostic measures in general population test-retest studies. By contrast, reliability of childhood ADHD was lower than that for the manifestation of the disorder in adulthood. The prevalence of childhood ADHD in the U.S. general population has recently been estimated at 8.1%, with only 36.3% of individuals who had ADHD as children also experiencing the disorder as adults (Kessler et al., 2005). Thus, in this case, it may be that the lower prevalence of the disorder in adulthood compared with childhood resulted in the diminution of reliability observed in this study.

To our knowledge, this is the first study to examine the reliability of DSM-IV borderline, narcissistic, and schizotypal PD diagnoses in the general population. The reliabilities of these PDs were as good as or better than the corresponding reliabilities found in short-term test-retest studies of the same diagnoses in clinical samples (Zimmerman, 1994). However, test-retest reliability should be greater in clinical compared with general population surveys, since more severe cases of PDs are found in treatment, whereas milder cases tend to be found among community respondents. These results suggest that our efforts to reduce sources of unreliability often related to fully structured diagnostic interviews appear to have been successful. Specifically, the high degree of standardization of the AUDADIS-IV and its training module appears to have reduced three major sources of unreliability in instruments of this type: (1) the questions that assess psychiatric symptoms; (2) the symptom information provided by the respondent; and (3) the interpretation of the information provided by the respondent. The AUDADIS-IV is also unique in its requirement to assess stringently the DSM-IV clinical significance criterion, i.e., the requirement that each PD lead to distress and/or social or occupational impairment. Requiring individuals classified as having a particular PD to endorse the DSM-specified number of symptoms of the disorder in addition to meeting the clinical significance criterion may, in part, have been responsible for the good reliability observed for borderline, narcissistic, and schizotypal PDs in this study.

For most diagnoses assessed in this study, an order effect was observed. That is, prevalences generally decreased from test to retest interview. Although the present study cannot offer a definitive explanation for the observed order effect, the decline in prevalences of alcohol, drug, and psychiatric disorders from test to retest has been observed in tests of a variety of other psychiatric assessment instruments (Helzer, 1981; Bromet et al., 1986). The decline in prevalences has been attributed to either a reduction in reporting of symptoms, or inconsistency in positive responses to screening questions that were often used to route respondents past sections of the interview that were not relevant. However, the AUDADIS-IV does not have screening questions associated with ADHD or the three PDs assessed in this study. That is, all respondents answered symptom item and associated diagnostic questions in these modules. Although the AUDADIS-IV PTSD module does have screening questions, there was little discrepancy in responses to these questions at test and retest. The gross error rate for these screening questions was 2.0%, a rate that was too small to account for the observed order effect. Taken together, these findings suggest that the decreases in prevalences of the disorders between test and retest is more likely the result of a decline in symptom reporting rather than an increase in negative responses to screening questions. Further methodologic research should focus on this important but insufficiently understood phenomenon.

Reliability was also found to be somewhat greater when dimensional scales were examined, compared with their categorical counterparts. Consistent with the internal consistency results, ICC values were good to excellent for all DSM-IV psychiatric diagnoses assessed in this study. These results are consistent with prior research (Grant et al., 2003) and were expected. Continuous measures are more statistically informative than categorical measures and therefore should be more reliable. Further, less severe cases of disorder will necessarily have a greater adverse impact on the reliability of categorical measures than on their continuous counterparts.

The psychometric evaluation of major risk factors associated with alcohol, drug, and other psychiatric disorders is rare in the substance use disorder or psychiatric epidemiology literature. This study showed fair to good test-retest and internal consistency reliability for most risk factors, with acculturation, race-ethnic identification, sexual orientation discrimination, perceived stress, stressful life events, alcoholism stigma, and adverse childhood experience scales demonstrating reliabilities in the good to excellent range. These reliability results were not surprising since only measures of risk factors that had been found in prior psychometric studies to demonstrate satisfactory reliability were selected for inclusion in the AUDADIS-IV. However, our slight modifications to those measures necessitated a re-evaluation of their reliabilities, especially in general population samples, where many of these measures had not been assessed.

The excellent reliability of the ACE scales was not expected due to the sensitivity of these measures. It is possible that the objective and behavioral nature of these scales outweighed their sensitivity to yield excellent reliability. Alternatively, ACEs are extremely memorable, albeit painful, a situation that may have reduced recall biases related to the events and consequently increased their reliability. Another nonmutually exclusive possibility is that the rapport established between respondents and interviewers highly trained to ask sensitive questions may have contributed to the high reliability observed for these measures.

The high level of reliability found in this study for PTSD, ADHD, and borderline, narcissistic, and schizotypal PDs suggests that the AUDADIS-IV can be a useful diagnostic tool in research settings. The Wave 2 NESARC survey, from which the data from this study were derived, queries a wide range of clinical symptomatology that cuts across numerous DSM-IV Axis I and II disorders. The finding that dimensional symptom scales for the diagnoses examined in this study were highly reliable supports the need for continued and sustained research to construct and evaluate dimensionally based assessment instruments to improve upon the purely categorical approach to diagnoses that underlies the DSM-IV. Incorporating dimensional components in future DSM revisions promises to address concerns commonly cited with respect to the categorical model of diagnosis, that is, excessive comorbidity, heterogeneity among persons with the same disorder, and inconsistent, unstable, and arbitrary diagnostic boundaries between disordered and normal functioning (Oldham et al., 2005). Further, studies (e.g., Markon and Krueger, 2005; Krueger et al., 2006) that explicitly compare continuous and dichotomous models of DSM-IV disorders using sophisticated latent class and trait models appear warranted in helping to define better phenotypic targets for etiologic and treatment research.

Importantly, this study has also provided the research community with a reliable battery of risk factors for use in future epidemiologic research on alcohol, drug, and psychiatric disorders. The availability of such a broad range of reliable risk factor assessments promises to contribute to the reliability of future research and the conclusions drawn from it, and to sharpen the direction of inquiry it defines.

References

- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4. American Psychiatric Press; Washington, DC: 1994.
- Barry DT. Development of a new scale for measuring acculturation: the East Asian acculturation measure (EAAM). *J Immigr Health* 2001;3:193–197. [PubMed: 16228786]
- Barry DT. An ethnic identity scale for East Asian immigrants. *J Immigr Health* 2002;4:87–94. [PubMed: 16228764]
- Bernstein DP, Fink L, Hondelsman L, Foote J, Lovejoy M. Initial reliability and validity of a new retrospective measure of child abuse and neglect. *Am J Psychiatry* 1994;151:1132–1136. [PubMed: 8037246]
- Briscoe-Smith AM, Hinshaw SP. Linkages between child abuse and attention-deficit/hyperactivity disorder in girls: behavioral and social correlates. *Child Abuse Negl* 2006;30(11):1239–1255. [PubMed: 17097140]
- 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]
- Centers for Disease Control. The 1988 Health Interview Survey. Centers for Disease Control; Hyattsville, MD: 1989.
- Centers for Disease Control. The 1999 National Youth Risk Behavior Surveillance System. Centers for Disease Control; Hyattsville, MD: 2001.
- Centers for Disease Control. The 2001 National Behavioral Risk Factor Surveillance System. Hyattsville, MD: 2003.
- 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]
- Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav* 1983;24:385–396. [PubMed: 6668417]
- Cohen S, Hoberman HM. Positive events and social supports as buffers of life changes stress. *J Appl Soc Psychol* 1983;13:99–125.
- Cohen, S.; Mermelstein, R.; Kamarck, T.; Hoberman, HM. Measuring the functional components of social support. In: Sarason, IG.; Sarason, BR., editors. *Social Support: Theory, Research and Applications*. Sage; Newbury, NJ: 1985. p. 73-94.
- Cohen, S.; Williamson, G. Perceived stress in a probability sample of the United States. In: Spacek, S.; Oskamp, S., editors. *The Social Psychology of Health: Claremont Symposium on Applied Social Psychology*. Sage; Newbury Park, NJ: 1988. p. 31-67.
- Cohen S, Doyle WJ, Skoner DP, Rabin BS, Gwaltney JM. Social ties and susceptibility to the common cold. *JAMA* 1997;277:1940–1944. [PubMed: 9200634]
- Coronado GD, Thompson B, McLerran D, Schwartz SM, Koepsell TD. A short acculturation scale for Mexican-American populations. *Ethn Dis* 2005;15:53–62. [PubMed: 15720049]
- Cottler LB, Grant BF, Blaine J, Mavreas V, Pull C, Hasin D, Compton WM, Rubio-Stipec M, Mage D. Concordance of DSM-IV alcohol and drug use disorder criteria and diagnoses as measured by AUDADIS-ADR, CIDI and SCAN. *Drug Alcohol Depend* 1997;47:195–205. [PubMed: 9306045]
- Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297–334.
- Cuellar I, Arnold B, Maldonado R. Acculturation rating scale for Mexican-Americans – II: a revision of the original ARSMA scale. *Hispanic J Behav Sci* 1995;17:275–304.
- Cuellar I, Bastida E, Braccio SM. Residency in the United States, subjective well-being, and depression in an older Mexican-origin sample. *J Aging Health* 2004;16:447–466. [PubMed: 15271265]
- Cunradi CB, Caetano R, Clark CL, Schafer J. Alcohol-related problems and intimate partner violence among White, Black, and Hispanic couples in the U.S. *Alcohol Clin Exp Res* 1999;23:1492–1501. [PubMed: 10512315]
- Davies M, Fleiss JL. Measuring agreement for multinomial data. *Biometrics* 1982;38:1047–1051.

- Deyo RA, Diehl AK, Hazoda H, Stern MP. A simple language-based acculturation scale for Mexican Americans: validation and application to health care research. *Am J Public Health* 1985;75:51–55. [PubMed: 3966599]
- Dong M, Anda RF, Dube SR, Giles WH, Felitti VJ. The relationship of exposure to childhood sexual abuse to other forms of abuse, neglect, and household dysfunction during childhood. *Child Abuse Negl* 2003;27:625–639. [PubMed: 12818611]
- Dube SR, Felitti VJ, Dong M, Chapman DP, Giles MD, Anda RF. Childhood abuse, neglect and household dysfunction and the risk of illicit drug use: the adverse childhood experiences study. *Pediatrics* 2003;111:564–572. [PubMed: 12612237]
- Fleiss, JL. *Statistical methods for rates and proportions*. Wiley; New York: 1981.
- George, D.; Mallery, P. *SPSS for Windows Step by Step: A Simple Guide and Reference*. 11.0 update. 4. Allyn & Bacon; Boston: 2003. p. 231
- 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.; Dawson, DA.; Hasin, DS. *The Alcohol Use Disorder and Associated Disabilities Interview Schedule-DSM-IV Version*. National Institute on Alcohol Abuse and Alcoholism; Bethesda, MD: 2001.
- 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.; 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. Available at <http://www.niaaa.nih.gov>
- Grant BF, Hasin DS, Stinson FS, Dawson DA, Chou SP, Ruan WJ, Pickering RP. Prevalence, correlates, and disability of personality disorders in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *J Clin Psychiatry* 2004a;65:948–958. [PubMed: 15291684]
- Grant BF, Stinson FS, Dawson DA, Chou SP, Dufour MC, Compton W, Pickering RP, Kaplan K. Prevalence and co-occurrence of substance use disorders and independent mood and anxiety disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Arch Gen Psychiatry* 2004b;61:807–816. [PubMed: 15289279]
- 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]
- Kessler RC, Adler LA, Barkley R, Biederman J, Conners CK, Faraone SV, Greenhill LL, Jaeger S, Secnik K, Spenser T, Ustun TB, Zaslavsky AM. Patterns and predictors of attention-deficit/hyperactivity disorder persistence into adulthood: results from the National Comorbidity Survey Replication. *Biol Psychiatry* 2005a;57:1442–1451. [PubMed: 15950019]
- Kessler RC, Berglund PA, Demler O, Jin R, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* 2005b;62:593–602. [PubMed: 15939837]
- Kessler RC, Chiu WT, Demler O, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* 2005c;62:617–627. [PubMed: 15939839]
- Krause ED, Kaltman S, Goodman L, Dutton MA. Role of distinct PTSD symptoms in intimate partner reabuse: a prospective study. *J Trauma Stress* 2006;19:507–516. [PubMed: 16929505]
- Krieger N. Racial and gender discrimination: risk factors for high blood pressure? *Soc Sci Med* 1990;30:1273–1281. [PubMed: 2367873]
- Krieger N, Sidney S. Prevalence and health implications of anti-gay discrimination: a study of Black and White women and men in the CARDIA study. *Int J Health Serv* 1997;27:157–176. [PubMed: 9031018]

- Krieger N. Racial discrimination and skin color in the CARDIA study: implications for public health research. *Am J Public Health* 1998;88:1308–1313. [PubMed: 9736868]
- Krieger N. Does racism harm health? Did child abuse exist before 1962? On explicit questions, critical science, and current controversies: an ecological perspective. *Am J Public Health* 2003;93:194–199. [PubMed: 12554569]
- Krieger N, Smith K, Naishadham D, Hartman C, Barbeau EM. Experiences of discrimination: validity and reliability of a self-report measure for population health research on racism and health. *Soc Sci Med* 2005;61:1576–1596. [PubMed: 16005789]
- Krueger RF, Markon KE, Patric CJ, Iacono WG. Externalizing psychopathology in adulthood: a dimensional spectrum conceptualization and its implications for DSM-IV. *J Abnormal Psychol* 2006;115:85–95.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159. [PubMed: 843571]
- Lenzenweger MF, Lane MC, Loranger AW, Kessler RC. DSM-IV personality disorders in the National Comorbidity Survey Replication. *Biol Psychiatry*. 2007In Press
- Link BG. Mental patient status, work and income: an examination of the effects of a psychiatric label. *Am Sociol Rev* 1982;47:202–215. [PubMed: 7091929]
- Link BG, Mirotznik J, Cullen T. The effectiveness of stigma coping orientations: can negative consequences of mental illness labeling be avoided? *J Health Soc Behav* 1991;32:302–330. [PubMed: 1940212]
- Link BG, Stuenkel EL, Neese-Todd S, Asmussen S, Phelan JC. The consequences of stigma for the self-esteem of people with mental illness. *Psychiatr Serv* 2001;52:1621–1626. [PubMed: 11726753]
- Lipsky S, Caetano R, Field CA, Larkin GL. The role of intimate partner violence, race, and ethnicity in the help seeking behavior. *Ethn Health* 2006;11:81–100. [PubMed: 16338756]
- Markon KE, Krueger RF. Categorical and continuous models of liability to externalized disorders: a direct comparison. *Arch Gen Psychiatry* 2005;62:1352–1359.
- Mays VM, Cochran SD. Mental health correlates of perceived discrimination among lesbian, gay and bisexual adults in the United States. *Am J Public Health* 2001;91:1869–1876. [PubMed: 11684618]
- Oldham, JM.; Skodal, AE.; Bender, DS. *Textbook of personality disorders*. American Psychiatric Publishing, Inc; New York: 2005.
- Phinney JS. The multigroup ethnic identity measure: a new scale for use with diverse groups. *J Adolesc Res* 1992;7:156–176.
- Rahim-Williams FB, Riley JL, Herrera D, Campbell CM, Hastic BA, Fillingim RB. Ethnic identity predicts experimental pain sensitivity in African Americans and Hispanics. *Pain*. 2007In Press
- Research Triangle Institute. *Software for Survey Data Analysis (SUDAAN) Version 9.0*. Research Triangle Institute; Research Triangle Park, NC: 2004.
- Sandfort TGM, de Graf R, Bijl RV, Schnabel P. Same-sex sexual behavior and psychiatric disorders: findings from the Netherlands Mental Health and Incidence Study (NEMESIS). *Arch Gen Psychiatry* 2001;58:85–91. [PubMed: 11146762]
- Sartor CE, Lynskey MT, Heath AC, Jacob T, True W. The role of childhood risk factors in initiation of alcohol use and progression to alcohol dependence. *Addiction* 2007;102:216–225. [PubMed: 1722275]
- Shouten H. Measuring pairwise agreement among many observers. *Biometry* 1980;6:497–504.
- Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420–428. [PubMed: 18839484]
- Solis JM, Marks G, Garcia M, Shelton D. Acculturation, access to care, and use of preventive services by Hispanics: findings from HHANES 1982–1984. *Am J Public Health* 1990;80(Suppl):11–19. [PubMed: 9187576]
- Stancil TR, Hertz-Picciotto I, Schramm M, Watt-Morse M. Stress and pregnancy among African-American women. *Paediatr Perinat Epidemiol* 2000;14:127–135. [PubMed: 10791655]
- Straus M. Measuring intrafamily conflict and violence: the conflict tactics (CT) scales. *J Marriage Fam* 1979;41:75–88.

- Straus, M.; Gelles, R.J. Physical violence in American families: risk factors and adaptations to violence in 8,145 Families. Transaction Press; New Brunswick, NJ: 1990.
- Ustun B, Compton W, Mager D, Babor T, Baiyewu O, Chatterji S, Cottler L, Gogus A, Mavreas V, Peters L, Pull C, Saunders J, Smeets R, Stipek MR, Vrsti R, Hasin D, Room R, Van den Brink W, Regier D, Blaine J, Grant BF, Sartorius N. WHO Study on the reliability and validity of the alcohol and drug use disorder instruments: overview of methods and results. *Drug Alcohol Depend* 1997;47:161–169. [PubMed: 9306042]
- Vega WA, Sribney WM, Aguilar-Gaxiola S, Kolody B. 12-Month Prevalence of DSM-III-R psychiatric disorders among Mexican Americans: nativity, social assimilation, and age determinants. *J Nerv Ment Dis* 2004;192:532–541. [PubMed: 15387155]
- Vrsti 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* 1997;4:144–149. [PubMed: 9852366]
- White HR, Chen PH. Problem drinking and intimate partner violence. *J Stud Alcohol* 2002;63:205–214. [PubMed: 12033697]
- Wyatt GE. The sexual abuse of Afro-American and White American women in childhood. *Child Abuse Negl* 1985;9:507–519. [PubMed: 4084830]
- Yoshihama M, Horrocks J. The relationship between intimate partner violence and PTSD: an application of Cox regression with time-varying covariates. *J Trauma Stress* 2003;16:371–380. [PubMed: 12895020]
- Zimmerman M. Diagnosing personality disorders: a review of issues and research methods. *Arch Gen Psychiatry* 1994;51:225–245. [PubMed: 8122959]

Table 1
AUDADIS-IV Section, Sample Size and Response Rate by Regional Office

Regional Office	AUDADIS-IV Section	Initial Sample Size (n=2,076)	Interviewed Sample Size (n=1,899)	Response Rate (%)
Boston	Acculturation, race-ethnic orientation, race-ethnic discrimination, gender discrimination, perceived stress, social network, social support, adverse childhood experiences, physical, emotional and sexual abuse, and psychological and emotional neglect	518	476	91.9
Philadelphia	DSM-IV borderline, schizotypal and narcissistic personality disorders	552	485	87.9
Detroit	DSM-IV posttraumatic stress disorder, disability discrimination, religious discrimination, weight discrimination, alcoholism stigma	526	488	92.8
Denver	Attention-deficit/hyperactivity disorder, sexual orientation, attraction and behavior items, intimate partner violence, sexual orientation discrimination	480	450	93.7

Table 2
Sociodemographic Characteristics of Respondents by Regional Office

Regional Office	Gender		Ethnicity			Education			Age			
	Male		Black	White	Other	Less than High School	High School	Some College and Beyond	18-24	25-44	45-64	65+
Boston	41.8		11.0	85.9	3.1	12.4	24.0	63.6	4.6	43.1	34.2	18.1
Philadelphia	39.2		28.7	68.3	3.0	15.0	29.9	55.1	5.3	39.2	31.1	24.4
Detroit	41.8		21.6	75.7	2.7	13.3	30.7	56.0	6.8	41.2	34.0	18.0
Denver	41.9		4.9	91.3	3.8	16.0	23.5	60.5	6.9	39.5	35.0	18.6

Table 3
Sociodemographic Characteristics of Wave 2 NESARC Respondents

	n	%	(SE)
Sex			
Male	14,564	47.9	0.34
Female	20,089	52.1	0.34
Race-ethnicity			
White	20,161	70.9	1.54
Black	6,587	11.1	2.19
Hispanic	6,359	11.6	1.19
Asian	968	4.3	0.52
Native American	578	2.2	0.18
Age			
20–29	4,913	16.3	0.31
30–34	3,116	9.1	0.22
35–44	7,487	20.6	0.29
45–54	6,825	20.0	0.27
55–64	5,135	14.6	0.23
65+	7,177	19.3	0.34
Education			
Less than high school	5,514	14.0	0.45
High school graduate	9,452	27.5	0.53
Some college and beyond	19,687	58.5	0.63
Marital status			
Married/cohabiting	18,866	63.8	0.48
Widowed/separated/divorced	9,149	18.9	0.26
Never married	6,638	17.4	0.45

Table 4

Structure of AUDADIS-IV Discrimination Scales

	Race-ethnicity	Gender	Religion	Disability	Weight	Sexual orientation
A. Now I'd like to know about how often you have experienced discrimination, been prevented from doing something, or been hassled or made to feel inferior in any of the following situations: (1) because you are Hispanic or Latino/ of your race-ethnicity; (2) because you are male or female; (3) because of your religion; (4) because of your disability; (5) because of your weight; (6) because of your sexual orientation. During the last 12 months/before 12 months ago, about how often did you experience discrimination...						
(1) Ability to obtain health care/health insurance	✓	✓	✓	✓	✓	✓
(2) In how you were treated when you got care	✓	✓	✓	✓	✓	✓
(3) In public, like on the street, in stores or in restaurants	✓	✓				✓
(4) Obtaining a job, on the job, or getting admitted to school or training program, or in the courts or by the police, or obtaining housing	✓	✓	✓	✓	✓	✓
(5) Called (racist/sexist) name(s)	✓					✓
(6) Made fun of, picked on, pushed, shoved, hit or threatened with harm	✓					✓
(7) (In public settings/Access to public facilities) like bathrooms, restaurants, elevators or public transportation				✓	✓	
B. When you are treated unfairly because of your (race-ethnicity/gender/religion/disability/weight/sexual orientation):						
(1) Do you usually accept it as a fact or do you try to do something about it?						
(2) Do you usually talk to other people about it or do you keep it to yourself?						

Table 5
Reliability of DSM-IV Diagnoses and Associated Symptom Scales

Disorder	Diagnosis		Symptom Scale			
	Kappa	(SE)	Prevalence (time 1/ time 2)	ICC ^d	95% CI ^b	Alpha
Posttraumatic stress disorder	0.77	(0.09)	(0.22/0.19)	0.69	(0.64, 0.72) ^c	0.84
12-month Lifetime	0.64	(0.11)	(0.26/0.20)			
Attention deficit/hyperactivity disorder	0.71	(0.12)	(0.05/0.06)	0.75	(0.71, 0.78) ^c	0.89
Child	0.63	(0.09)	(0.03/0.02)	0.75	(0.74, 0.79)	0.83
Adult	0.71	(0.06)	(0.10/0.09)	0.72	(0.67, 0.75)	0.77
Borderline personality disorder (lifetime)	0.70	(0.09)	(0.11/0.07)	0.71	(0.67, 0.76)	0.75
Narcissistic personality disorder (lifetime)	0.67	(0.10)	(0.04/0.06)			
Schizotypal personality disorder (lifetime)						

Note: Number of items for symptom scales: posttraumatic stress disorder (n=18); attention-deficit/hyperactivity disorder (n=20); borderline (n=18), narcissistic (n=18), and schizotypal (n=22) personality disorders.

^a ICC = intraclass test-retest reliability coefficient.

^b CI = 95% confidence interval.

^c Symptom scales for post-traumatic stress disorder and attention deficit/hyperactivity disorder based on lifetime symptoms.

Table 6

Reliability of AUDADIS-IV Risk Factors

	ICC ^a	95% CI ^b	Alpha
Acculturation (33)	0.79	(0.76, 0.82)	0.85
Race-ethnic identification (24)	0.78	(0.74, 0.80)	0.79
Discrimination-race-ethnicity			
12-month (6)	0.68	(0.64, 0.72)	0.74
Prior to last 12 months (6)	0.64	(0.59, 0.68)	0.69
Discrimination-gender			
12-month (6)	0.62	(0.53, 0.72)	0.72
Prior to last 12 months (6)	0.59	(0.55, 0.64)	0.70
Discrimination-religion			
12-month (3)	0.64	(0.60, 0.68)	0.62
Prior to last 12 months (3)	0.50	(0.42, 0.58)	0.68
Discrimination-disability			
12-month (4)	0.59	(0.53, 0.65)	0.70
Prior to last 12 months (4)	0.63	(0.59, 0.68)	0.69
Discrimination-weight			
12-month (5)	0.79	(0.74, 0.80)	0.75
Prior to last 12 months (5)	0.77	(0.72, 0.78)	0.78
Discrimination-sexual orientation			
12-month (6)	0.82	(0.79, 0.85)	0.84
Prior to last 12 months (6)	0.78	(0.74, 0.80)	0.81
Perceived stress (4)	0.82	(0.76, 0.85)	0.84
Stressful life events (12-month) (14)	0.94	(0.93, 0.95)	0.86
Interpersonal support evaluation (12)	0.63	(0.58, 0.68)	0.82
Social networks (12)	0.70	(0.68, 0.76)	0.64
Alcoholism-stigma (12)	0.93	(0.92, 0.94)	0.82
Childhood physical abuse (3)	0.79	(0.76, 0.82)	0.78
Childhood sexual abuse (4)	0.86	(0.72, 0.78)	0.90
Childhood psychological abuse (3)	0.88	(0.82, 0.90)	0.85
Childhood emotional neglect (5)	0.82	(0.79, 0.85)	0.80
Childhood physical neglect (4)	0.80	(0.78, 0.84)	0.89
Battered mother/female caregiver (4)	0.76	(0.72, 0.79)	0.86
Other childhood adverse events (6)	0.69	(0.66, 0.73)	0.72
Intimate partner violence –respondent (6)	0.79	(0.76, 0.82)	0.70
Intimate partner violence-partner (6)	0.76	(0.73, 0.80)	0.73
Sexual orientation (1)	0.60 ^c (0.07) ^d	(0.03/0.03) ^e	<i>f</i>
Sexual attraction (1)	0.66 (0.06)	(0.06/0.04)	<i>f</i>
Sexual behavior (1)	0.62 (0.09)	(0.03/0.02)	<i>f</i>

^aICC = intraclass test-retest reliability coefficient.

^bCI = 95% confidence interval.

^cKappa value.

^dStandard error of Kappa in parentheses.

^ePrevalence of test/prevalence of retest.

^fAlpha coefficients not assessed for dichotomous measures.