ELSEVIER

Drug and Alcohol Dependence 44 (1997) 133–141

DRUG and ALCOHOL DEPENDENCE

The alcohol use disorder and associated disabilities interview schedule (AUDADIS): reliability of alcohol and drug modules in a clinical sample

Deborah Hasin*, Kenneth M. Carpenter, Steven McCloud, Melissa Smith, Bridget F. Grant

Columbia University/NYS Psychiatric Institute, Box 123, 722 West 168th Street, New York, NY 10032, USA

Received 19 June 1996; accepted 7 December 1996

Abstract

The alcohol use disorder and associated disabilities interview schedule (AUDADIS), was designed for use in the general population, and was previously shown to have good reliability in a sample of household residents. However, measurement problems are different in clinical samples. Thus, a test-retest study was conducted of the AUDADIS in a clinical sample of 296 substance-using patients from substance- and psychiatrically-identified treatment settings. Reliability for current drug-specific AUDADIS dependence diagnoses was good to excellent for high-prevalence as well as low-prevalence drug categories. Reliability for abuse diagnoses was not as good, although this was due to the hierarchical nature of the abuse diagnosis itself, rather than its defining criteria. Demographic and other factors were investigated for their potential effects on the reliability of alcohol and cocaine diagnoses; low severity was the only consistent predictor of unreliability for both of these categories. Reliability of consumption variables was generally good, with a few notable exceptions. Results suggest that the AUDADIS can be used in research comparing treated to community samples of individuals with alcohol and drug diagnoses. © 1997 Elsevier Science Ireland Ltd.

Keywords: Diagnostic interview; Diagnostic reliability; Multinomial regression; Psychiatric assessment; Reliability

1. Introduction

The alcohol use disorders and associated disabilities schedule (AUDADIS; Grant et al., 1994a) is a fully structured diagnostic interview designed to assess alcohol and drug use as well as alcohol and drug use disorders in both clinical samples and the general population. The AUDADIS was initially used in a large-scale national survey conducted in the US in 1992 (Grant et al., 1994a). The reliability of the AUDADIS in the general population has been presented previously (Grant et al., 1995). These reliabilities were generally shown to be good to excellent, in a study whose design improved considerably on prior psychometric studies of diagnostic interviews in community samples.

As described in greater detail elsewhere (Grant et al., 1995), the AUDADIS is a fully-structured interview that can be given by either lay interviewers or clinicians. The AUDADIS was designed to make diagnoses of alcohol and drug use disorders in a manner that would overcome prior problems in the operationalization of these disorders and improve reliability. Thus, in contrast to earlier fully-structured diagnostic interviews such as the DIS (Robins et al., 1981), the clustering of symptoms within a 12-month period for present and past disorders was required. Additionally, the AUDADIS was designed to make these diagnoses in well-defined time frames, the last 12 months (current) and prior to the last 12 months (past). These time frames, unique to the AUDADIS among fully-structured interviews (although shared by a clinician-administered interview; Hasin et al., 1996) offer several benefits. First,

* Corresponding author. Tel.: +1 212 9605518; fax: +1 212 9602386.

PII S0376-8716(97)01332-X

they allow for the identification and study of fully- and partially-remitted cases, who may be of considerable interest in both treatment and epidemiologic research. Second, they allow for the specification of a current cluster of symptoms as well as symptoms experienced only in the past. This provides clarity about the disorder at the time when the symptoms are perhaps most relevant, as in baseline interviews for follow-up studies, and also investigation of disorders when recall is most recent and therefore best. In addition, the current and past time periods can be combined for the creation of the lifetime diagnosis, which is of most direct use in genetic studies. In the AUDADIS, a lifetime diagnosis is made in all subjects who ever met criteria for the disorder in the past, in the present, or both.

An important feature of the AUDADIS is that onset and recency were defined in terms of the clustered symptoms of each disorder, rather than the first- and last-occurring symptom of the disorder (as done in the DIS and CIDI (World Health Organization (1993). The information was designed for application of diagnostic computer programs that make diagnoses according to algorithms reflecting the diagnostic criteria (similar to the DIS and the CIDI).

Additional features of the AUDADIS are the alcohol and drug consumption sections. These ascertain quantity, frequency and patterning of alcohol and drug use as well as the overlap between alcohol and use of other drugs. Many other diagnostic interviews omit such information, since the information is not required for the diagnoses. However, the decision to omit quantity and frequency from the diagnostic criteria was originally made on the grounds that such information could not be obtained reliably from addicts (Guze et al., 1969; Guze, 1965). If amounts and patterns of alcohol and drug use can be obtained reliably, such information has the potential to add considerably to research on alcohol and drug use disorders.

A diagnostic interview designed for use in general population as well as clinical samples should show good reliability in both types of samples, but measurement problems in the two types of samples are different. In a general population sample, disorders are often milder, putting a greater test on the thresholds for distinguishing positive diagnoses from those of sub-threshold severity. Also, the prevalence of substance use disorders (especially current disorders) is lower in the general population. Very low prevalences provide a stringent test on the reliability of an instrument due to the low amount of true variance with which error variance can be compared. Low prevalence also limits investigation of detailed aspects of an instrument's measurement capabilities. In many clinical samples, of course, substance use disorders are more common, providing higher prevalences and more variance. Disorders and symptoms also tend to be more severe. However, polysubstance use and comorbidity among substance use disorders may blur distinctions between the symptoms of dependence and abuse in different drug classes, reducing reliability for different reasons. Also, the length of detailed questions on drug abuse and dependence symptoms in individuals who have used many different drug classes must answer may cause subjects to under-report drugs that are covered later in the interview, or under-report drugs in all categories in the second interview of a test-retest pair. Thus, empirical demonstration of the reliability of a diagnostic interview is required in clinical samples even if it has already been shown to work well in the general population. Below, we report on a test-retest study of the AUDADIS in 296 patients in different types of substance abuse treatment.

2. Methods

2.1. Subjects

Subjects were patients in two sites: a general substance abuse outpatient clinic and an inpatient dual-diagnosis psychiatric unit. Subjects were consecutive admissions to these two facilities when admission numbers were low, and randomly selected admissions when admission numbers were high enough that we could not interview all admissions. Of the 296 subjects, 58.4% were male and 41.6% were female. About 34.5% were black, 38.5% white and the rest English-speaking Hispanic. The mean age of the sample was 37.8 (S.D., 7.96; range, 18–59), with 6.4% of the sample aged 18–24 years, 78% of the sample aged 25–44 years, and the remainder 45 years and older. In this sample, 72.6% had earned a high school diploma or higher degree. Nineteen percent of the sample was married or living with someone as if married, 35.1% were divorced, separated or widowed, and 45.9% were never married.

2.2. Procedures

Subjects were recruited and interviewed as soon as possible after their entrance into treatment, with a few day's delay for acute withdrawal to subside if necessary. Initially, all female consecutive admissions and a random subsample of males were selected. However, about midway through the study, this procedure was modified to include all consecutive admissions. This was done to increase the speed of intake into the study.

At both sites, a site coordinator from the research staff met with the subjects individually, explained the study to them, and obtained informed consent. Subjects were interviewed and re-interviewed a mean of 10.4 days apart (S.D., 16.77), with a range of 1–159 days. A total of 30 subjects (10.2%) were re-interviewed more

than 14 days after their first interview. Different interviewers conducted the two interviews for each patient. The second interviewer was always blind to the results of the first interview. Subjects were paid for their participation. Of those asked to participate, 296 agreed and completed two interviews, yielding a final response rate of 93%.

2.3. Measures

We used the same version of the alcohol and drug sections of the AUDADIS that were used in the 1992 national survey (Grant et al., 1994b) and that were tested in the general population (Grant et al., 1995). We report on alcohol and drug abuse and dependence diagnoses derived from the Diagnostic Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV; American Psychiatric Association, 1994). Drug classes include anxiolytics, cannabis, cocaine, heroin, licit opiates, and stimulants.

2.4. Interviewers

A total of 15 interviewers conducted AUDADIS interviews for this study. Since the AUDADIS was designed for non-clinicians, this was the type of interviewer used in the study. Prospective interviewers were screened for negative attitudes concerning alcohol and drug use, or for other strong assumptions regarding alcohol or drugs that might affect their interviewing. Prospective interviewers were also asked to role-play a portion of the AUDADIS interview after only a few minutes of preparation as part of their employment interview. Prior experience with the AUDADIS (Hasin et al., 1997a,b) and with other diagnostic interviews (Hasin et al., 1997c) shows that serious problems with this early task predict unsatisfactory interviewing after training.

Interviewer training consisted of work with self-study materials including an interviewer's manual, didactic presentation of the structure and methods of the interview by an experienced trainer, role-playing of the AUDADIS sections according to pre-scripted role-plays, and specific training modules concerning alcohol and drug classes and effects (Grant et al., 1995). An additional part of interviewer training for this study was instruction and supervision on personal conduct in a clinical setting and on patient management during the interview. The site coordinators monitored these aspects of the study throughout, giving supervision when necessary.

2.5. Analyses

For dichotomous variables, kappa was used as a measure of reliability, that is, inter-rater agreement corrected for chance (Cohen, 1960). We used the form of kappa designed for random raters, although in practical terms, there were only minute differences between κ using the formulae for random or fixed raters. For continuous measures, we used the intraclass correlation coefficient (ICC; Shrout and Fleiss, 1979). We also used the formulation for random raters for ICC. Kappa and ICC produce equivalent results when the number of possible values of a continuous variable equals two. The stability of both kappa and ICC are indicated with 95% confidence intervals.

The values of ICC and kappa are interpreted similarly. Kappa and ICC both have a range of -1 to $+1$. A kappa or ICC of 0.0 indicates agreement exactly at the chance level. A kappa or ICC of 0.75 and above indicates excellent agreement, from 0.40 to 0.74 fair to good agreement, below 0.39 poor agreement (Fleiss, 1981). Negative values of kappa and ICC reflect systematic disagreement. All reliability statistics for the diagnostic criteria and classifications were computed for the full sample.

In analyzing subject and other effects on agreement, bivariate tests were initially used to compare concordant with discordant cases, including χ^2 tests and t-tests. Multivariate investigation of subject and other effects on concordance were carried out with multinomial logistic regressions (Hosmer and Lemeshow, 1989). This type of regression analysis is designed for polytomous outcome categories. In multinomial regression, one level of the outcome variable is taken as the reference category, and the other levels are contrasted to it. In the analyses below, the discrepant cases were considered the reference cases. The concordant positive and concordant negative cases constituted the contrast groups.

3. Results

3.1. Reliability of diagnoses

Table 1 shows the results for categories analogous to those in the previous AUDADIS reliability paper (Grant et al., 1995). These are for DSM-IV dependence and abuse diagnoses combined. As Table 1 shows, the results for current (past 12 months) diagnoses were generally good to excellent, ranging from.62 for cannabis to.83 for heroin. For past disorders, κ ranged from 0.54 for alcohol disorders to 0.67 for heroin disorders. Table 2 shows more detailed results on reliability for alcohol and drug dependence diagnoses, showing these diagnoses by drug class and time frame. As shown, the κ for current dependence diagnoses for alcohol, heroin and stimulants were excellent, with cocaine only slightly below this level. This was true despite the widely varying prevalences of diagnoses in

Table 1
Reliability of AUDADIS DSM-IV alcohol and drug use diagnoses[a]

Diagnostic classification	κ Coefficients (S.E.) (prevalence first interview/prevalence second interview)	
	Current	Past
Alcohol abuse and dependence	0.74 (0.04)	0.54 (0.04)
	(60.47/56.42)	(57.77/50.33)
Any drug abuse and dependence	0.68 (0.04)	0.57 (0.04)
	(58.44/53.04)	(63.17/56.41)
Cannabis abuse and dependence[b]	0.62 (0.09)	0.57 (0.07)
	(6.42/5.74)	(13.18/9.80)
Cocaine abuse and dependence	0.71 (0.04)	0.60 (0.04)
	(45.27/39.53)	(51.69/44.25)
Heroin abuse and dependence[b]	0.83 (0.05)	0.67 (0.04)
	(12.50/11.62)	(17.91/15.21)

[a] All diagnoses coded: 0, no diagnosis; 1, abuse only: and 2, dependence.
[b] κ for no diagnosis and dependence only.

different drug classes. Reliability for current cannabis, anxiolytic and licit opiate dependence were in the fair to good range, despite extremely low prevalences in this sample. The κ for past diagnoses ranged from good (alcohol, cannabis, cocaine, heroin and stimulants) to fair (anxiolytics, licit opiates). The κ for lifetime diagnoses (current, past or both) were in ranges similar to those found for current and past disorders.

Table 3 shows the reliability of diagnoses of abuse. The top half of Table 3 shows abuse diagnoses made according to DSM-IV criteria, that is, abuse diagnoses were excluded if the criteria for dependence were met. In contrast to the results for dependence, κ for abuse were fair at best, and more often poor, with the exception of heroin and lifetime diagnoses of alcohol abuse. Recall that in DSM-IV, a diagnosis of abuse cannot be made if the subject has ever met criteria for dependence on a lifetime basis. Thus, reliability of abuse depends not only on the reliability of its own criteria but also on the reliability of the dependence category on which it is conditional. Where dependence is highly prevalent, the potential prevalence and variance for abuse decrease substantially. To further investigate the reliability of abuse, we created variables corresponding to non-hierarchical abuse, that is, abuse diagnoses made independently of whether the subject met criteria for dependence in that drug class. The bottom half of Table 3 also shows these 'non-hierarchical' abuse diagnoses. Relative to the hierarchical abuse diagnoses, all non-hierarchical abuse diagnoses are distinctively

Table 2
Test-retest reliability of DSM-IV AUDADIS diagnoses of substance dependence

Drug category	κ (S.E.) (prevalence first interview/prevalence second interview)		
	Current	Past	Lifetime
Alcohol	0.76 (0.04)	0.60 (0.05)	0.75 (0.04)
	(0.58/0.52)	(0.52/0.44)	(0.63/0.59)
Anxiolytics	0.66 (0.16)	0.50 (0.13)	0.57 (0.12)
	(0.02/0.02)	(0.04/0.04)	(0.04/0.05)
Cannabis	0.63 (0.11)	0.65 (0.10)	0.66 (0.09)
	(0.05/0.05)	(0.07/0.05)	(0.07/0.06)
Cocaine	0.72 (0.04)	0.64 (0.05)	0.72 (0.04)
	(0.40/0.33)	(0.44/0.32)	(0.51/0.42)
Heroin	0.86 (0.05)	0.69 (0.06)	0.8 (0.05)
	(0.12/0.10)	(0.15/0.11)	(0.18/0.14)
Licit opiates	0.59 (0.19)	0.46 (0.15)	0.59 (0.13)
	(0.02/0.01)	(0.04/0.02)	(0.04/0.03)
Stimulants	0.80 (0.20)	0.72 (0.11)	0.77 (0.10)
	(0.01/0.01)	(0.04/0.03)	(0.04/0.03)

Table 3
Test-retest reliability of DSM-IV AUDADIS diagnoses of substance abuse

Drug categories	κ (S.E.) (prevalence first interview/prevalence second interview)		
	Current	Past	Lifetime
Hierarchical			
Alcohol	0.27 (0.16)	0.36 (0.11)	0.43 (0.10)
	(0.03/0.02)	(0.06/0.02)	(0.07/0.06)
Anxiolytics	0.00 (0.00)	0.41 (0.13)	0.41 (0.13)
	(0.00/0.00)	(0.04/0.03)	(0.04/0.03)
Cannabis	0.24 (0.20)	0.27 (0.11)	0.25 (0.11)
	(0.02/0.01)	(0.06/0.04)	(0.07/0.05)
Cocaine	0.10 (0.12)	0.24 (0.08)	0.23 (0.08)
	(0.04/0.02)	(0.12/0.08)	(0.13/0.09)
Heroin	0.01 (0.00)	0.16 (0.12)	0.16 (0.12)
	(0.01/0.00)	(0.04/0.03)	(0.04/0.03)
Licit opiates	0.00 (0.00)	0.17 (0.16)	0.14 (0.14)
	(0.01/0.00)	(0.03/0.01)	(0.03/0.01)
Stimulants	0.00 (0.00)	0.33 (0.25)	0.33 (0.25)
	(0.00/0.00)	(0.02/0.00)	(0.02/0.00)
Non-hierarchical			
Alcohol	0.79 (0.04)	0.36 (0.11)	0.79 (0.04)
	(0.53/0.52)	(0.06/0.06)	(0.55/0.53)
Anxiolytics	0.42 (0.17)	0.61 (0.09)	0.67 (0.09)
	(0.02/0.03)	(0.07/0.06)	(0.03/0.02)
Cannabis	0.70 (0.09)	0.67 (0.07)	0.68 (0.07)
	(0.06/0.06)	(0.13/0.10)	(0.14/0.10)
Cocaine	0.76 (0.04)	0.68 (0.04)	0.76 (0.04)
	(0.43/0.38)	(0.51/0.43)	(0.45/0.40)
Heroin	0.83 (0.05)	0.71 (0.06)	0.79 (0.05)
	(0.11/0.10)	(0.18/0.14)	(0.19/0.15)
Licit opiates	0.62 (0.13)	0.63 (0.11)	0.56 (0.11)
	(0.04/0.03)	(0.05/0.05)	(0.05/0.05)
Stimulants	0.80 (0.20)	0.78 (0.09)	0.78 (0.09)
	(0.01/0.01)	(0.05/0.05)	(0.05/0.05)

Table 4
Intraclass correlation coefficients for AUDADIS alcohol and drug abuse and dependence items by DSM-IV diagnosis

	Intraclass correlation coefficients (95% CI)	
	Abuse	Dependence
Past year		
Alcohol	0.82 (0.79, 0.85)	0.86 (0.83, 0.88)
Cannabis	0.64 (0.58, 0.69)	0.74 (0.69, 0.78)
Heroin	0.85 (0.82, 0.87)	0.89 (0.87, 0.91)
Cocaine	0.77 (0.73, 0.81)	0.81 (0.77, 0.84)
Prior to past year		
Alcohol	0.82 (0.79, 0.85)	0.77 (0.73, 0.81)
Cannabis	0.75 (0.71, 0.79)	0.79 (0.75, 0.82)
Heroin	0.80 (0.76, 0.83)	0.89 (0.87, 0.91)
Cocaine	0.74 (0.69, 0.78)	0.75 (0.71, 0.79)

higher, even though prevalences ranged from moderate to very low. Thus, the abuse criteria themselves (or the AUDADIS items representing them) did not seem to be the problem, but rather, their conditional reliance on dependence. This pattern of results is not unique to the AUDADIS. Very similar results have been found with a semi-structured diagnostic interview in a similar clinical sample (Hasin et al., 1997c) and in reliability studies of other instruments where this aspect of the findings remained either unanalyzed or unpublished.

Table 4 shows the ICCs for continuous measures of dependence severity created by summing symptom items within drug classes. As shown, these measures had excellent test-retest reliability for both current and past diagnoses for the entire sample.

3.2. Factors potentially influencing reliability of alcohol diagnoses

For these analyses, we classified subjects as being (a) concordant-positive, i.e. received a positive diagnosis at test and retest; (b) concordant-negative, i.e. received a negative diagnosis at test and retest; or (c) discordant, i.e. received one positive diagnosis and one negative diagnosis at test and retest. For alcohol dependence, subject characteristics potentially affecting reliability included sex, age, race, education, drug use defined as any lifetime use, number of dependence symptoms reported in the first interview (severity), and lifetime AA attendance. As noted above, we began with univariate tests of these factors. When comparing concordant-positive with discordant cases, no significant differences were found for sex. A significant difference was found for education ($\chi^2 = 8.33$, df = 2, $P = 0.02$), age ($t = 8.33$, df = 181, $P = 0.04$), AA attendance ($\chi^2 = 14.56$, df = 1, $P = 0.001$), and race ($\chi^2 = 7.98$, df = 2, $P = 0.02$). A significant effect was found for number of dependence symptoms ($t = 16.6$, df = 181, $P < 0.001$), with a higher number of symptoms predicting concordance. For concordant-negative vs. discordant cases, only severity of dependence ($t = 18.5$, df = 146, $P < 0.001$) and AA attendance ($\chi^2 = 5.35$, df = 1, $P = 0.02$) were significant differentiators. The number of days between the test and retest interview had no effect for either comparison.

Multivariate analyses of these factors allowed for simultaneous consideration of these effects. The discordant category was used as the reference category. All subject and other characteristics were entered into the model simultaneously. Table 5 shows the results of

Table 5
Multinomial logistic regression analyses for alcohol dependence discordant group versus negative concordant group (Model 1) and discordant group versus positive concordant group (Model 2)

Variable	Model 1		Model 2	
	Coefficient (S.E.)	Odds ratio (*P* value)	Coefficient (S.E.)	Odds ratio (*P* value)
Days between interview	0.038(0.05)	1.04 (0.45)	−0.017 (0.07)	0.98 (0.80)
Age	−0.053(0.07)	0.95 (0.47)	0.171 (0.08)	0.84 (0.09)
Gender	0.315 (0.91)	1.37 (0.73)	1.47 (1.11)	4.37 (0.19)
Race				
Caucasian	—	—	—	—
African-american	1.26 (1.28)	3.52 (0.32)	1.37 (1.22)	3.39 (0.26)
Other	1.63 (1.21)	5.12 (0.18)	1.72 (1.41)	5.60 (0.22)
Education				
Some college	—	—	—	—
High school	−0.146 (1.16)	0.86 (0.90)	2.98 (1.44)	19.72 (0.04)
<High school	0.165 (1.19)	1.19 (0.89)	2.44 (1.46)	11.57 (0.09)
Professional Tx	−1.27 (2.10)	0.28 (0.54)	0.085 (1.30)	1.09 (0.95)
AA meetings	0.959 (0.97)	2.63 (0.33)	−0.829 (1.27)	0.43 (0.52)
Dependence criteria	4.00 (0.91)	54.74 (<0.001)	−3.81 (3.63)	0.02 (<0.001)

Females served as the baseline group for gender analyses.
Dependence criteria entered as the number criteria met.

Table 6
Reliability of AUDADIS alcohol consumption measures

Consumption measure	Intraclass correlation coefficient (95% CI)			
	Beer	Wine	Liquor	All beverages combined
Days per year drank usual quantity of all beverage types combined (past year)	—	—	—	0.49 (0.41,0.56)
Days per year drank usual quantity of all beverage types combined (period of heaviest consumption)	—	—	—	0.72 (0.67,0.77)
Quantity consumed per occasion when drinking usual quantity (past year)	0.82 (0.78,0.85)	0.32 (0.23,0.40)	0.62 (0.55,0.68)	—
Quantity consumed per occasion when drinking heaviest quantity (past year)	0.80 (0.76,0.83)	0.30 (0.21,0.38)	0.56 (0.49,0.63)	—
Typical size, in ounces, of beverage consumed when drinking usual quantity (past year)	0.75 (0.70,0.79)	0.44 (0.35,0.52)	0.57 (0.50,0.64)	—
Typical size, in ounces, of beverage consumed when drinking heaviest quantity (past year)	0.66 (0.60,0.71)	0.40 (0.31,0.48)	0.69 (0.63,0.74)	—
Average daily ethanol intake (oz) from all beverage types when drinking usual quantities[a] (past year)	—		—	0.92 (0.90,0.93)
Average daily ethanol intake (oz) taking Into account usual and heaviest quantities of all beverage types (past year)	—	—	—	0.92 (0.90,0.93)
Average daily ethanol intake (oz) from all beverage types when drinking usual quantities (period of heaviest consumption)	—	—	—	0.64 (0.58,0.70)
Average daily ethanol intake (oz) taking into account usual and heaviest quantities of all beverage types (past year)	—		—	0.66 (0.60,0.71)

[a] Ordinal classification in ounces as: 1, less then 0.10; 2, 0.10–0.24; 3, 0.25–0.49; 4, 0.50–0.74; 5, 0.75–0.99; 6, 1.00–1.49; 7, 1.50–1.99;8, 2.00–2.49; 9, 2.50 or more.

these analyses for alcohol dependence. The comparison between concordant-positive and discordant cases showed that only number of dependence criteria and educational status continued to have significant effects in the multivariate analyses. Fewer alcohol dependence symptoms and lower educational status were associated with inter-rater discordance. When concordant-negative cases were compared with discordant cases, only number of dependence criteria had a significant effect. Parallel analyses conducted for cocaine dependence yielded similar results to those of alcohol dependence (not shown). Overall, the results indicated that borderline severity status accounted for the majority of discordant alcohol dependence and cocaine dependence classifications, controlling for other characteristics. Multivariate analyses for other drug classes were precluded by low prevalence.

3.3. Reliability of alcohol consumption measures

Table 6 shows the reliability of AUDADIS alcohol consumption measures in this clinical sample. As shown, reliability of usual drinking and drinking during heaviest consumption, days per year ranged from fair to good. Beverage-specific quantities for beer were all excellent. Beverage-specific quantities for liquor ranged from fair to excellent. Surprisingly, beverage-specific reliabilities for wine ranged from fair to poor. Past-year average ethanol consumption for usual and usual-plus-heaviest drinking was excellent, and in the fair range for period of heaviest consumption.

3.4. Reliability of drug use measures

Table 7 shows the κ indicating use of drugs in excess of an experimental level on a lifetime basis, and whether the drug was used at all in the last 12 months. As shown, reliabilities for the lifetime use indicator were good to excellent. Any use in the past 12 months was measured with excellent reliability for heroin, cannabis and cocaine, with the other drugs ranging from fair to good. Table 8 shows ICCs for age at first use of a drug for those who used it, and number of times used prior to the last 12 months and within the last 12 months. As shown, the reliability of age at first use was very high for all drug categories. Use during the last 12 months ranged from fair to excellent, with the sole exception of stimulants. The number of times a drug was used prior to the last 12 months also ranged from fair to excellent, with the sole exception of marijuana.

Table 7
Reliability of AUDADIS drug use measures[a]

Drug category	κ coefficient (S.E.) (prevalence first interview/ prevalence second interview)	
	Ever used 12+ times in lifetime	Used in the past 12 months
Sedatives	0.74 (0.04) (28.04/24.32)	0.55 (0.10) (8.11/4.73)
Tranquilizers	0.79 (0.04) (30.74/29.39)	0.71 (0.06) (12.84/12.50)
Cannabis	0.64 (0.05) (70.61/65.54)	0.77 (0.04) (35.14/31.08)
Cocaine	0.86 (0.04) (77.70/75.34)	0.83 (0.03) (57.43/55.07)
Heroin	0.91 (0.03) (34.8/32.8)	0.92 (0.03) (20.95/18.92)
Licit opiates	0.64 (0.04) (25.0/22.64)	0.59 (0.08) (11.49/8.78)
Stimulants	0.79 (0.04) (27.03/23.99)	0.54 (0.18) (2.36/1.35)
Methadone	0.76 (0.06) (10.81/9.80)	0.78 (0.09) (5.07/4.39)

[a] Measures coded; 0, no; 1, yes.

4. Discussion

The above results show that on the whole, in a clinical sample where measurement issues are considerably different than in the general population, the AUDADIS produced current diagnoses of drug and alcohol dependence with good to excellent reliability, and past and lifetime diagnoses that were also mainly good to excellent, although some were in the fair range. The reliability achieved was the case not only for high-variance diagnoses such as alcohol and cocaine, but also for very rare diagnoses of dependence such as stimulants and anxiolytics. A severity measure based on number of positive items also showed high reliability. As has been the case before (Hasin et al., 1997c), the reliability of abuse diagnoses was not as high as dependence, although the cause of this problem lay more with the hierarchical nature of the diagnosis than with the unreliability of its generally behavioral criteria.

In general, reliability of alcohol and drug use measures was good to excellent also. Most beverage-specific quantity measures for beer and liquor were excellent, as was average daily ethanol intake for usual and usual-plus-heaviest drinking in the past 12 months. Good to excellent reliability was obtained for drug-specific use past an experimental level and age at first use. Number of times each drug was used in the past generally ranged from fair to good, with the exception of marijuana.

The low reliability figures for wine prompted some post hoc investigation of the data. We found that wine was used less frequently in this sample than the other two beverage types, and that a few outliers who reported extremely high consumption in their first interview reported no wine consumption in their second interview. Also, heavy wine drinkers appeared less able to estimate their quantities of wine consumption consistently than heavy drinkers of other types of beverages, producing discrepancies in magnitudes of amounts between the test and retest interview that had a sharply attenuating effect on the reliability. The concept of 'a can' of beer is easily understood by most people, but most people are much less familiar with the quantities of wine consumed, either by the glass or by the bottle. Since fewer wine drinkers in the general population were extremely heavy drinkers, this issue did not have such a pronounced effect on the reliability of wine consumption in the general population (Grant et al., 1995).

AA attendence was included in the multivariate models as a factor potentially affecting concordance. This was done because participating in a forum that encourages talking about one's difficulties and discussing one's history on a fairly consistent basis (either in the meeting or with one's friends and sponsor) could influence the stability of reporting. Also, such participation would provide a framework in which individuals could conceptualize their problems which may in turn influence their reporting tendencies.

Our results indicated that education was related to concordant-positive vs. discordant comparions, but not to concordant-negative vs. discordant comparisons.

Table 8
Intraclass correlation coefficients for AUDADIS drug use measures

Drug categories	Interclass correlation coefficients (95% CI)		
	Age at first use	Number of times used prior to past year	Number of times used in past 12 months
Cannabis	0.92 (0.90,0.94)	0.38 (0.27,0.48)	0.72 (0.67,0.76)
Cocaine	0.79 (0.73,0.83)	0.58 (0.50,0.65)	0.66 (0.60,0.71)
Heroin	0.93 (0.90,0.95)	0.64 (0.51,0.74)	0.79 (0.75,0.82)
Licit opiates	0.94 (0.91,0.96)	0.50 (0.32,0.65)	0.50 (0.42,0.57)
Stimulants	0.85 (0.78,0.90)	0.75 (0.64,0.83)	0.05 (0.00,0.14)
Tranquilizers	0.90 (0.86, 93)	0.61 (0.47,0.72)	0.62(0.56,0.68)

The effect of education on concordance may have been confounded by the relationship of education to the likelihood of getting a diagnosis. Those with lower education were more likely to receive an alcohol diagnosis, and all those in the concordant-positive versus discordant cases received the diagnosis in at least one interview. Thus, education was more likely to have affected this comparison than the comparison that included subjects whose education was better,on average: the concordant-negative versus discordant cases.

Note that the tables reflected reliability when all subjects evaluted at each site were included. This included subjects who had not used particular drugs, based on the assumption that their answers would be negative to all the drug diagnostic questions for the substances they did not use. In some instances, a research question for a particular project involves understanding the performance of an interview or a diagnostic system on one particular drug diagnosis among subjects who are all known to have used that drug. In such a case, the ability of the interview used to identify users is irrelevant since the sample would consist only of subjects previously known to be users of the drug. An appropriate reliability study for such a case would involve a series of subjects pre-screened as users of the particular substance of interest. Thus, no non-users would be included in the reliability sample. However, more often, the research question involves multiple alcohol or drug diagnoses in a clinical or community setting where histories of substance use are more diverse. In such a case, the agreement between measurement procedures on diagnoses is conditional on the instruments' identification of whether subjects are users of a drug or not, and part of the instrument's psychometric performance includes this aspect of the evaluation.

In response to the interest of some investigators, we note that we did create subsets of subjects by computer who were 'users' for each drug category and computed κ for the 'user-only' group. This mimics a situation in which only users of a particular drug were sampled at a particular type of setting. In general, κ for the user-only group were slightly lower than the κ reported in the above tables, generally by about 0.05. This is not enough of a difference to change the meaning of a result. In a few instances, including only users increased the kappa because the cause of discrepency between test and retest interviews involved whether subjects were considered users or not. In a few other instances, inclusion of users-only decreased the kappa enough to change the meaning of the result, for example, reliability of current heroin diagnoses ($\kappa = 0.66$, S.E. $= 0.10$) for users-only. It is not clear whether this was due to a true substantive effect, or whether it was a the result of a sample perturbation, given the number of κ produced when the tests were re-run for users only.

The reliability results of this study, in conjunction with the reliabilities demonstrated earlier in the general population suggest that the AUDADIS may be used with confidence in comparisons of clinical and household samples on issues pertaining to alcohol and drug diagnoses, as well as to alcohol and drug use. Thus, between treated and untreated subjects with dependence, comparisons of the course of their disorder, treatment history, and potential risk factors for dependence may be examined using the AUDADIS without the potential artifactual effect on the data of measurement problems pertaining to one type of sample but not the other. This type of research has the potential to reveal a great deal about the natural history of alcohol and drug dependence.

The alcohol and drug modules of the AUDADIS take an average of 20 min to administer in the general population. In the present study, we did not time individual modules. However, since the number of drugs is higher, on average, in clinical settings, the drug module takes somewhat longer in patient samples. However, both sections of the AUDADIS can still usually be completed in 30–40 min with patients.

Semi-structured diagnostic interviews in psychiatry have been subject to test-retest investigation with non-patient as well as patient samples (see, for example, Andreasen et al., 1981; Williams et al., 1992). However, the usual type of testing for fully structured interviews in both clinical and general population samples is procedural validity testing, e.g. Anthony et al., 1985; Hasin et al., 1987; Helzer et al., 1985; Hesselbrock et al., 1982; Robins et al., 1982. To our knowledge, such short-term reliability studies have not been previously conducted with fully structured interviews in clinical as well as general population samples, perhaps on the grounds that subjects would always answer identically to first and second administration of fully structured items. However, in the (likely) event that items in fully-structured diagnostic interviews are not perfect measures of the underlying concepts, short-term test-retest reliability has an important place in understanding the psychometric properties of fully structured as well as semi-structured interviews. The completion of this study marks the first instance in the development of fully-structured diagnostic interviews in which test-retest reliability statistics are available for both clinical and general population samples. Future psychometric studies of the AUDADIS include analyses of its reliability and validity in other cultures for comparative purposes. To achieve these goals, reliability studies in both clinical and population samples have recently been completed in India, Romania, and Australia, and validity studies have recently been conducted in India, St. Louis, Belgium and Greece. Once the results from all of the psychometric research is available, the AUDADIS will have undergone a thorough and extensive series of test-retest reliability evaluations.

References

American Psychiatric Association (1994). Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition (DSM-IV), American Psychiatric Association, Washington, DC.

Andreasen, N.C., Grove, W.M., Shapiro, R.W., Keller, M.B., Hirshfeld, R.M.A. and McDonald-Scott, P. (1981) Reliability of Lifetime Diagnosis: a multicenter collaborative perspective. Arch. Gen. Psychiatry 38, 400–405.

Anthony, J.C., Folstein, M., Romanowski, A.J., Von Korff, M.R., Nestadt, G.R., Chahal R. (1985) Comparison of the lay DIS and a standardized psychiatric diagnosis. Arch. Gen. Psychiatry 42, 667–675.

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

Fleiss, J.L. (1981) Statistical Methods for Rates and Proportions. Wiley, New York.

Guze, 1965.

Guze, S.B., Goodwin, D.W. and Crane, J.B. (1969) Criminality and psychiatric disorders. Arch. Gen. Psychiatry 20, 583–591.

Grant, B.F., Hartford, T.C., Dawson, D.A., Chou, P.S. and Pickering, R.P. (1995). The Alcohol Disorder and Associated Disabilities Schedule (AUDADIS): Reliability of alcohol and drug modules in a general population sample. Drug Alcohol Depend. 39, 37–44.

Grant, B.F., Hartford, T., Dawson, D., Chou, P., Dufour, M. and Pickering, R. (1994) Prevalence of DSM-IV alcohol abuse and dependence: United States, 1992. Alcohol Health Res. World 18, 243–248.

Grant, B.F., Peterson, A., and Dawson, D.A. and Chou, S.P. (1994) Source and Accuracy Statement or the 1992 National Longitudinal Alcohol Epidemiological Survey (NLAES), National Institute on Alcohol Abuse and Alcoholism, Rockville, MD.

Hasin, D.S. and Grant, B.F. (1987) Diagnosing depressive disorders in patients with alcohol and drug problems: a comparison of the SADS-L and the DIS. J. Psychiatric Res., 21, 301–311.

Hasin et al., 1996.

Hasin et al., 1987.

Hasin, D., Li, Q., McCloud, S. and Endicott, J. (1997a) Agreement between DSM-III, DSM-III-R, DSM-IV, and ICD-10 alcohol diagnoses in U.S. community-sample heavy drinkers. Addiction. (in press).

Hasin, D. McCloud, S., Li, Q. and Endicott, J., (1997b) Cross-system agreement among demographic subgroups: DSM-III, DSM-III-r, DSM-IV, and ICD-10 diagnoses of alcohol use disorders. Drug Alcohol Depend. (in press).

Hasin, D.S., Trautman, K.D., Miele, G.M., Samet, S., Smith, M. and Endicott, J. (1997c). Psychiatric research interview for substance and mental disorders (PRISM): Reliability in substance abusers. Am. J. Psychiatr. (in press).

Helzer, J.E., Robins, L.N., McEvoy, M.A., Spitznagle, E.L., Stoltzman, R.K. and Farmer A et al. (1985) A comparison of clinical and diagnostic interview schedule diagnoses: physician reexamination of lay-interviewed cases in the general population. Arch. Gen. Psychiatry 42, 657–666.

Hesselbrock, V., Stabenau, J., Hesselbrock, M., Mirkin, P. and Meyer, R. (1982) A comparison of two interview schedules: the SADS-L and the NIMH DIS. Arch. Gen. Psychiatry 39, 674–677.

Hosmer, D. and Lemeshow, S. (1989) Applied Logistic Regression. Wiley, New York.

Robins et al., 1981.

Robins, L.N., Helzer, J.E., Ratcliff, K.S. and Seyfired, W. (1982) Validity of the Diagnostic Interview Schedule, version II: DSM-III diagnoses. Psychol. Med. 12, 855–870.

Shrout, P.E. and Fleiss, J.L. (1979) Intraclass correlations: Uses in assessing rater reliability. Psychol. Bull. 86, 420–428.

Williams, J., Gibbon,M., First, M., Spitzer, R., Davies, M., Borus, J., Howes, M., Kane, J., Pope, H., Rounsaville, B. and Wittchen, H. (1992) Structured Clinical Interview for DSM-III-R (SCID). II. Multisite test-retest reliability. Arch. Gen. Psychiatry 49, 198–205.

World Health Organization. (1993) Composite International Diagnostic Interview: Core Version 1.1. World Health Organization, Geneva.