WHO study on the reliability and validity of the alcohol and drug use disorder instruments: methods and overall results

Ustun, B.; Compton, W.; Mager, D.; Babor, T.; Baiyewu, O.; Chatterji, S.; Cottlet, L.; Gogus, A.; Mavreas, V.; Peters, L.; Pull, C.; Saunders, J.; Smeets, R.M.W.; Stipec, M-R.; Vrasti, R.; Hasin, D.; Room, R.; van den Brink, W.; Regier, D.; Blaine, J.; Grant, B.; Sartorius, N.

DOI
10.1016/S0376-8716(97)00087-2

Publication date
1997

Published in
Drug and Alcohol Dependence

ELSEVIER

Drug and Alcohol Dependence 47 (1997) 161–169

DRUG and ALCOHOL DEPENDENCE

WHO Study on the reliability and validity of the alcohol and drug use disorder instruments: overview of methods and results

B. Üstün [a,*], W. Compton [b], D. Mager [b], T. Babor [c], O. Baiyewu [d], S. Chatterji [a], L. Cottler [b], A. Göğüş [e], V. Mavreas [f], L. Peters [g], C. Pull [h], J. Saunders [i], R. Smeets [j], M.-R. Stipec [k], R. Vrasti [l], D. Hasin [m], R. Room [n], W. Van den Brink [o], D. Regier [p], J. Blaine [q], B.F. Grant [r], N. Sartorius [s]

[a] *Unit on Epidemiology, Classification and Assessment, Division of Mental Health and Prevention of Substance Abuse, World Health Organization, 1211 Geneva 27, Switzerland*
[b] *Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA*
[c] *Department of Psychiatry, University of Connecticut Health Center, Farmington, CT, USA*
[d] *Department of Psychiatry, University College Hospital, Ibadan, Nigeria*
[e] *Department of Psychiatry, Hacettepe University, Ankara, Turkey*
[f] *Department of Psychiatry, University of Athens, Athens, Greece*
[g] *Clinical Research Unit for Anxiety Disorders, St Vincent's Hospital, Darlinghurst, NSW, Australia*
[h] *Service de Neuropsychiatrie, Centre Hospitalier de Luxembourg, Luxembourg, Luxembourg*
[i] *Department of Psychiatry, Royal Prince Alfred Hospital, Camperdown, NSW, Australia*
[j] *Inspectorate for Health Care, Ministry of Health, Welfare and Sports, Rijswijk, The Netherlands*
[k] *Office of the Dean for Academic Affairs, Behavioural Sciences Research Institute, San Juan, Puerto Rico*
[l] *Psychiatric Hospital, Timis, Romania*
[m] *Department of Psychiatry, College of Physicians and Surgeons of Columbia University, New York, USA*
[n] *Addiction Research Foundation, Toronto, Ontario, Canada*
[o] *Amsterdam Institute for Addiction Research, University of Amsterdam, Amsterdam, The Netherlands*
[p] *Division of Epidemiology and Services Research, NIMH, Rockville, MD, USA*
[q] *Division of Clinical Research, National Institute on Drug Abuse, Rockville, MD, USA*
[r] *Division of Biometry and Epidemiology, National Institute of Alcohol Abuse and Alcoholism, Rockville, MD, USA*
[s] *Department of Psychiatry, University of Geneva, Geneva, Switzerland*

Accepted 22 May 1997

Abstract

The WHO Study on the reliability and validity of the alcohol and drug use disorder instruments is an international study which has taken place in 12 centres in ten countries, aiming to test the reliability and validity of three diagnostic instruments for alcohol and drug use disorders: the Composite International Diagnostic Interview (CIDI), the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) and a special version of the Alcohol Use Disorder and Associated Disabilities Interview schedule-alcohol/drug-revised (AUDADIS-ADR). The purpose of the reliability and validity (R&V) study is to further develop the alcohol and drug sections of these instruments so that a range of substance-related diagnoses can be made in a systematic, consistent, and reliable way. The study focuses on new criteria proposed in the tenth revision of the International Classification of Diseases (ICD-10) and the fourth revision of the diagnostic and statistical manual of mental disorders (DSM-IV) for dependence, harmful use and abuse categories for alcohol and psychoactive substance use disorders. A systematic study including a scientifically rigorous measure of reliability (i.e. 1 week test-retest reliability) and validity (i.e. comparison between clinical and non-clinical measures) has been undertaken. Results have yielded useful information on reliability and validity of these instruments at diagnosis, criteria and question level. Overall the diagnostic concordance coefficients (kappa, κ) were very good for dependence disorders (0.7–0.9), but were somewhat lower for abuse and harmful use categories. The comparisons among instruments and independent clinical evaluations and debriefing interviews gave important information about possible sources of unreliability, and provided useful clues on the applicability and consistency of nosological concepts across cultures. © 1997 Elsevier Science Ireland Ltd.

* Corresponding author.

PII S0376-8716(97)00087-2

Keywords: WHO study; Alcohol; Drug use; Reliability; Validity; Dependence; Harmful use

1. Introduction

A reliable and valid diagnostic process is an essential prerequisite to almost all research and clinical practice in the area of alcohol, drug use and mental (ADM) disorders. These disorders are defined within classificatory systems such as the International Classification of Diseases (ICD-10: WHO, 1993a) and the Diagnostic and Statistical Manual of mental disorders (DSM-IV: American Psychiatric Association, 1994). Since these disorders are defined by their manifestations rather than direct disease markers, the reliability and validity of their assessment has been a major need as well as a challenge. The World Health Organization (WHO), in collaboration with the three US National Institutes on Health (NIH): the National Institute on Mental Health (NIMH), the National Institute of Alcohol Abuse and Alcoholism (NIAAA) and the National Institute of Drug Abuse (NIDA), began a joint project on diagnosis and classification of mental disorders, alcohol- and drug-related problems in 1979 (WHO, 1985). One of the primary aims of the project was to develop instruments that could be used cross-culturally for the assessment of ADM disorders. These instruments translated the 'operational criteria' of the ICD-10 and DSM-IV classification systems into questions and compiled the responses into diagnosis. The Composite International Diagnostic Interview (CIDI: WHO, 1993b) and the Schedules for Clinical Assessment in Neuropsychiatry (SCAN: WHO, 1993c) are two of the most prominent instruments developed for this purpose. While these two instruments have been extensively field-tested in a variety of cultural settings for mental disorders, adequate information about their cross-cultural applicability, usefulness, reliability and validity was not available for the alcohol and drug use disorder sections.

The WHO/NIH joint project, therefore, commenced a substantial programme of research on the assessment instruments for the alcohol and drug use disorders (Üstün and Wittchen, 1992). The programme had two phases. The first one aimed at studying the meanings and interpretation of alcohol and drug use and problems in different cultures and of their implications for creating uniform diagnostic standards and international assessment instruments. This phase, commonly referred to as the cross-cultural applicability research (CAR) study, was conducted at nine centres (Room et al., 1996). The second phase of the project focused mainly on the reliability and validity of the substance abuse instruments, namely the SCAN, the CIDI and a third instrument which used a different questioning style, yet the same operational criteria: a special version of the Alcohol Use Disorder and Associated Disabilities Interview schedule (AUDADIS: Grant and Hasin, 1992; Grant and Towle, 1990; Grant et al., 1995; Hasin et al., 1997) which was specifically modified for the R&V study as the AUDADIS alcohol/drug-revised (AUDADIS-ADR: WHO, 1992). (see also Chatterji et al., 1997).

The broad aim of this study was to systematically test the reliability and validity of three international diagnostic instruments (CIDI, SCAN and AUDADIS-ADR) for alcohol and drug use disorders. This was expected to help in the further development of the alcohol and drug sections of the these instruments towards making them more reliable, valid and suitable for cross-cultural use. The study also aimed to determine the reliability of the recently developed ICD-10 and DSM-IV criteria for substance-related disorders using the three instruments. The major aims of the study can be summarized as follows:

1. To determine the reliability of the assessment measures at item, criterion and diagnostic level using a test-retest design.
2. To determine the concordance between different diagnostic interviews (structured and semi-structured) in widely different cultural settings.

This study as a whole allows comparisons across 12 sites, three different instruments, various substances including alcohol, three diagnostic systems in which the level of reliability and validity could be studied at question, criteria and diagnostic levels. Table 1 summarizes the basic domains of the variables studied. The scope of this introductory report is much more narrow, highlighting the major results of the WHO collaborative study on reliability and validity of the alcohol and drug use disorder instruments. Major reliability and validity findings are primarily presented at the lifetime diagnostic level for alcohol and other drugs while data are also presented for alcohol at the item and criteria level.

2. Methods

2.1. Centres

This study followed a collaborative multicentre design, with the coordinating centre located at WHO, (Geneva) and 12 participating centres located at Amsterdam (The Netherlands), Ankara (Turkey), Athens (Greece), Bangalore (India), Farmington (Connecticut), Ibadan, (Nigeria), Jebel (Romania), Luxembourg (Luxembourg), St Louis (USA), San Juan (Puerto Rico) and two separate centres at Sydney (Australia). Besides covering major geographical regions of the world, these centres also represented different diagnostic traditions, classification approaches, drinking and drug use patterns, cultures, religions, languages and availability of heath care services (Table 2).

Table 1
WHO R&V study: basic domains of the variables studied

A. Sites	B. Substances	C. Instruments	D. Systems	E. Level of agreement	F. Level of validity and reliability
1. Amsterdam	1. Alcohol	1. AUDADIS-ADR	1. ICD-10	1. Item	1. Independent Clinic Evaluation (ICE)
2. Ankara	2. Cannabis	2. SCAN	• Harmful use	2. Criteria	2. Clinical diagnosis (charts)
3. Athens	3. Amphetamines	3. CIDI	• Dependence	3. Diagnosis	3. Urine analysis (ONTRAK)
4. Bangalore	4. Sedatives		2. DSM-III-R	4. Discrepancy interview protocol (DIP)	4. Clinical versus ICE
5. Farmington	5. Cocaine		• Abuse		
6. Ibadan	6. Opiates		• Dependence		
7. Jebel	7. PCP		3. DSM-IV		
8. Luxembourg	8. Hallucinogens		• Abuse		
9. St. Louis	9. Inhalants		• Dependence without physiological dependence		
10. San Juan	10. Other		• Dependence with physiological dependence		
11. Sydney (1)			A. Any diagnosis versus no diagnosis		
12. Sydney (2)			B. Specific diagnosis groups		

2.2. Instruments

This study used three diagnostic instruments; the CIDI, the SCAN and the AUDADIS-ADR. The CIDI is a fully structured interview schedule based on the operational diagnostic criteria of the ICD-10, DSM-III R and now DSM-IV. SCAN is a semi-structured interview conducted by a clinically experienced interviewer with the help of a comprehensive glossary. Symptoms are explored individually and then diagnoses are generated using computerized algorithms. In previous studies conducted under the auspices of the WHO/NIH joint project acceptable reliability was found at the diagnostic level for earlier versions of the CIDI and the SCAN (Wittchen et al., 1991; Cottler et al., 1991; Wing et al., 1990). The AUDADIS-ADR was added as the third instrument for use in this study because it used different phrasing of questions and operationalizations—e.g. confirmation of the clustering of the symptoms by the respondents rather than the algorithms—and had been found reliable in large general population surveys (Grant et al., 1995).

The translations of all instruments were independently made and reviewed including a comparison with the back-translation by experts using WHO guidelines. They were all field tested before use. The CIDI, AUDADIS-ADR and SCAN were extensively tested for their cross-cultural applicability in the CAR Study (Room et al., 1996).

The interviewers were trained in standard training courses lasting one week at a centralized location, separately for each instrument. Various quality assurance steps were taken including exchange of interviewers between sites and a consultant visiting all the sites, and observing and co-rating the interviews. Interviewer characteristics varied for each site and are described in more detail in the accompanying five studies presented in this series.

The diagnoses for all three instruments were derived by computer algorithms to assign diagnosis according to the logic of the diagnostic criteria. The algorithms were checked extensively for their adequate representation of the criteria by independent expert committees and were used as standard algorithms for WHO instruments (these are available from the first author at WHO on request).

2.3. Design

Each centre carried out the study with subjects who used alcohol and/or drugs, but focused specially on one or two principal drugs (see Table 2). The centre also selected one or more sub-studies (test-retest on CIDI, SCAN or AUDADIS-ADR, comparison studies on CIDI, SCAN and AUDADIS-ADR). Table 2 gives the sub-studies carried out at each of the centres, along with the main substance(s) studied at the centres. For example, the Amsterdam centre compared CIDI with SCAN and focused primarily on THC and opiates, while the Farmington centre did a test-retest comparison of the SCAN and focused mainly on cocaine as the principal drug.

Table 2
WHO R&V study: characteristics of participating sites

Location	Instrument comparisons	Main substance	*N*	Mean age	Mean years education	% Male	Language	Samples		Cultural description
								Non-treatment	Treatment	
Amsterdam	CIDI-SCAN	THC, opiates	150	37	12.1	66	Dutch	Two general practices and addicts' union 48%	Inpatients, 52%	All SES[b], Dutch and Surinamese, Catholic and Protestant
Ankara	SCAN-SCAN	Opiates, THC	150	34.7	9.7	89	Turkish	General population, 17% ex-patients, 17%	Inpatients, 33%, outpatients, 33%	Middle SES[b], Turkish, Moslem
Athens	SCAN-CIDI-AUDADIS	Opiates	155	33.7	11.7	67	Greek	Bars, prisons, 60%	Outpatients and inpatients, 40%	All SES, Greek, Greek Orthodox
Bangalore	AUDADIS-AUDADIS-SCAN[a]	THC, opiates	197	43.8	3.8	75	Kannada	General population from villages, 73%	Outpatients and inpatients, 27%	Lower SES[b], Dravidians and Indonarayans, Hindu
Farmington	SCAN-SCAN	Cocaine, opiates, THC	137	37	14.3	72	English	General population, and general practice, 69%	Inpatients 31%	Lower and middle SES[b], African American, Hispanic, and Caucasian, Catholic and Protestant
Ibadan	CIDI-SCAN	Cocaine, THC	150	33.1	9	80	Yoruba	General population, prisons, inpatients, 77%	Inpatients 23%	Lower and middle SES[b], Yoruba, Christian and Moslem
Jebel	AUDADIS-AUDADIS	Sedatives	149	36.9	12.5	68	Romanian	General practice, 68%	Inpatients, 32%	Lower and middle SES[b], Romanian, Orthodox, Catholic, Protestant
Luxemberg	CIDI-SCAN-AUDADIS	Opiates	131	39	11.3	72	French		Outpatients and inpatients 100%	All SES, caucasian, Catholic
San Juan	CIDI-CIDI (computerized)	THC	148	32.4	14.1	64	Spanish	General population, 26%	Outpatients (general and gastroenterology)74%	Lower SES[b], Hispanic, Catholic
St.Louis	CIDI-SCAN-AUDADIS	Cocaine, THC	150	36.5	12.2	67	English	General population, 69%	Inpatients, 31%	All SES, African American, caucasian, Christian, Protestant
Sydney (1)	CIDI-CIDI (computerized)	THC	140	31.1	13.4	35	English	General practice, methodone users, 69%[c]	Inpatients and outpatients, 31%	All SES, caucasian, Christian
Sydney (2)	AUDADIS-AUDADIS	Sedatives, amphetamines	152	30.9	13.4	54	English	Needle exchange program, 76%	Methadone users and rehab clinics, 24%	All SES, caucasians and other, Christian

[a] AUDADIS = AUDADIS – ADR.
[b] SES: socio-economic status.
[c] Methodone users of Sydney (1) site were not in treatment.

Table 3
WHO reliability and validity study sample

Study	N	Mean age	% Male	% Tx	% Unemployed	Mean years education
SCAN test-retest	287	35.9	81	33	30	12
CIDI test-retest	288	31.7	50	53	16	13.8
AUDADIS-ADR test-retest	500	37.2	66	60	14	9.9
CIDI vs. SCAN comparison	306	35.1	73	27	[a]	10.5
CIDI vs. SCAN vs. AUDADIS-ADR Comparison	436	36.4	69	57	[a]	11.7
SCAN vs. AUDADIS-ADR comparison	185	43.8	75	73	22	3.8

[a] Data not available.

The interval between any two interviews varied from 3 days to 3 weeks, the median being 1 week. All retest interviews were conducted by a different interviewer than the test interview, and retest interviewers were uninformed about prior responses. In comparison study protocols, the interviews were administered in a randomly alternating order.

2.4. Subjects

The study design required each centre to interview about 150 subjects. Approximately 75 subjects from each centre were selected on the basis of alcohol use while the other 75 were selected on the basis of drug use. Because these instruments were designed for epidemiological surveys, the reliability tests were aimed at enriched general population samples which therefore included not more than 50 subjects from special ADM treatment settings and at least 100 from the general population or primary care settings (that is from first level health care facilities such as general practice or outpatient clinics for family or internal medicine) which were non-substance abuse treatment settings. As can be seen in Table 2, most sites approached, but did not completely achieve, the study goal of having at least 75% of their samples from non-ADM treatment settings.

Table 2 summarizes the samples in different sites. Within the general sampling guidelines stated above, wide differences in sampling across the sites occurred. The treatment facilities from which respondents were selected included both inpatient and outpatient programs for the treatment of alcohol and other substance addiction at each centre. The particular substances abused varied across the sites considerably as expected as well as the sociodemographic characteristics of the samples. More importantly, the ways in which the non-treatment respondents were recruited varied significantly, although all met the criterion of being out-of-treatment and reporting at least a minimal amount of substance use (i.e. non-abstainers).

The specific non-treatment sampling frames were as follows: Amsterdam recruited general medical patients from a variety of outpatient practices. Ankara recruited opiate users from lists of persons treated between 1 and 5 years prior to the interview, while out-of- treatment alcohol users were recruited from general hospital outpatient gastrointestinal clinic and community volunteers (recruited from the friends and families of patients and medical students). Athens selected drug abusing respondents from a prison and alcohol users from taverns and bars. Bangalore selected respondents from local villages through key informants to identify groups of substance users who were then asked to recruit other users, through outreach to key locations where substance users were known to congregate, and through targeting a village with high rates of alcohol consumption by women. Farmington recruited respondents from inpatient medical facilities and from the community using a newspaper advertisement asking for volunteers for 'a study designed to evaluate the effectiveness of a research interview'. Ibadan selected subjects from inpatient medical and surgical units of a university hospital, from a local prison and volunteers from the community (recruited by discharged substance abuse treatment patients). Jebel selected non-substance abuse treatment respondents from general psychiatry outpatient clinics (non-psychotic patients only), community volunteers from a local factory and an unemployment agency, and from general practice patients. In Luxembourg non-substance abuse treatment respondents were recruited from attendees in the emergency department of a university hospital, from inpatient wards of a general medical hospital and from general practice outpatient practices. The San Juan site recruited patients using stratified random selection techniques from several medical clinics. The St. Louis site selected respondents receiving general medical services from a satellite community facility where respondents were recruited from street outreach. The Sydney CIDI site recruited non-substance abuse treatment subjects from several outpatient general medical practices. Finally, the Sydney AUDADIS-ADR site selected their out-of-treatment respondents from a street outreach prevention program. It is precisely because of these differences, which are typical of the broad inclusion criteria in clinical research, that the instruments could be said to have been tested in diverse populations with a broad range of alcohol and drug use disorder severity.

Table 4
Test-retest reliability of ICD-10 and DSM-IV harmful use, abuse and dependence diagnoses among users by instrument and substance

Substance	ICD-10 dependence	ICD-10 harmful use	DSM-IV dependence	DSM-IV abuse
SCAN				
Alcohol	0.76 (0.04)	0.35 (0.10)	0.73 (0.04)	0.60 (0.06)
Cannabis	0.75 (0.08)	0.40 (0.15)	0.86 (0.07)	0.56 (0.09)
Opiates	0.93 (0.07)	NA	0.76 (0.12)	NA
Sedatives	0.76 (0.08)	NA	0.82 (07)	0.44 (0.17)
Cocaine	0.71 (0.10)	NA	0.71 (0.10)	0.39 (0.22)
Hallucinogens	NA	0.44 (0.33)	NA	0.10 (0.06)
Amphetamines	0.63 (0.33)	NA	0.63 (0.33)	0.63 (0.33)
CIDI				
Alcohol	0.75 (0.04)	0.60 (0.07)	NA	NA
Cannabis	0.69 (0.07)	0.41 (0.11)		
Opiates	0.80 (0.08)	NA		
Sedatives	0.48 (0.11)	NA		
Cocaine	0.76 (0.07)	0.50 (0.15)		
Hallucinogens	0.79 (0.20)	0.72 (0.13)		
Amphetamines	0.76 (0.10)	0.62 (0.17)		
AUDADIS-ADR				
Alcohol	0.68 (0.04)	0.17 (0.12)	0.66 (0.04)	0.49 (0.06)
Cannabis	0.44 (0.06)	0.38 (0.11)	0.46 (0.06)	0.41 (0.06)
Opiates	0.73 (0.13)	NA	0.68 (0.13)	0.20 (0.18)
Sedatives	0.73 (06)	0.56 (0.15)	0.79 (0.06)	0.60 (0.09)
Cocaine	0.46 (0.23)	0.11 (0.18)	0.65 (0.23)	0.30 (0.15)
Hallucinogens	0.64 (0.11)	0.42 (0.14)	0.40 (0.15)	0.41 (0.11)
Amphetamines	0.63 (08)	NA	0.64 (0.07)	0.48 (0.10)

In summary, samples were enriched using various strategies to provide an adequate distribution across substance, gender, and non-treatment respondents. Generally, patients were recruited in usual ways (advertisements, contact by treatment facility doctors). Subjects who were within the first month of treatment were excluded. All respondents signed informed consent forms before being allowed to participate in the project and all study protocols were reviewed by local Institutional Review Boards. Subjects were paid culturally relevant amounts to cover their travel and time. Attrition (refusals and dropouts) was below 10% overall.

2.5. Statistical analysis

The chance corrected degree of agreement on alcohol and drug use disorder diagnoses and alcohol diagnostic criteria and symptom questions was calculated using the kappa statistic (Fleiss, 1981) both for test-retest agreement of the same instrument and concurrence between two instruments. κ's were calculated using a two by two comparison of positive versus negative substance users. For example, κ's were calculated on persons diagnosed with dependence at test compared with retest. Reliability was similarly calculated for each diagnostic criterion and for each interview question. Generally, values less than 0.40 are considered poor agreement; 0.40–0.64 values are fair; 0.65–0.74 values are good; and values greater than 0.75 are excellent.

The rates of dependence and agreement were calculated only among persons who, on a lifetime basis, reached a certain threshold of use. In the CIDI and AUDADIS-ADR, for alcohol dependence, this was 12 drinks in their lifetime; for other drug dependence, the threshold was use 'greater than five times' during their lifetime. For the SCAN, this threshold was 'more than once or twice' during their lifetime for all substances. Analyses were limited to alcohol, opiate, cannabis, sedative, and cocaine because they represented a broad spectrum of use and because sample sizes were appropriately large for each.

3. Results

The total number of cases included in the study was 1825, each centre contributing between 131 and 197 cases. (Table 2). The mean age of the total sample was 37.2 ± 10.8 years. Of the sample, 68%, consisted of men and the rest were women. Overall, 23% of the cases came from specialized treatment settings, the rest being from primary care or general population. The numbers of cases interviewed with each of the instruments were 1118 with CIDI, 1244 with SCAN and 1058 with AUDADIS-ADR.

Table 3 presents demographic details of the samples used for each of the instrument at various centres. Test-retest studies were performed at seven sites: The

Table 5
Agreement on lifetime ICD-10 dependence diagnoses between the CIDI, SCAN and AUDADIS-ADR among users of selected substances

Diagnosis	CIDI vs. SCAN	AUDADIS-ADR vs. SCAN	CIDI vs. AUDADIS-ADR
Alcohol dependence	0.64 (0.05)	0.61 (0.05)	0.69 (0.08)
Opiate dependence	0.68 (0.08)	0.64 (0.04)	0.70 (0.07)
Cannabis dependence	0.44 (0.07)	0.36 (0.13)	0.48 (0.14)
Sedative dependence	0.48 (0.06)	0.50 (0.08)	0.39 (0.12)
Cocaine dependence	0.56 (0.11)	0.51 (0.12)	0.52 (0.13)

SCAN test-retest sites were Ankara and Farmington; CIDI test-retest sites were San Juan and Sydney, and the AUDADIS-ADR test-retest sites were Bangalore, Jebel and Sydney. CIDI vs. SCAN comparison were done at Amsterdam, Athens, Ibadan, Luxembourg and St. Louis. SCAN vs. AUDADIS-ADR comparisons were done at Athens, Bangalore, Luxembourg and St. Louis. CIDI vs. SCAN vs. AUDADIS-ADR comparison were done at Athens, Luxembourg and St. Louis.

Table 4 gives the kappas for the test-retest reliability of alcohol and drug use disorder diagnoses for each instrument for ICD-10 and DSM-IV formulations. For example, for ICD-10 alcohol dependence lifetime diagnoses, CIDI test-retest reliability is 0.75 while the corresponding value for SCAN is 0.76 and for AUDADIS-ADR is 0.68. Similarly the κ's associated with other drugs are shown in the same way. Generally all three instruments showed good to excellent chance-corrected agreement (κ) for dependence categories for alcohol 0.66–0.76, opiates 0.73–0.93, cannabis 0.44–0.86, sedatives 0.48–0.82; cocaine 0.46–0.76 and amphetamines 0.63–0.73, but reliability was lower for abuse and harmful use diagnoses. It is noteworthy that the reliability of diagnosis is lowest for lifetime diagnosis, and gets higher for past year and current diagnosis (Chatterji et al., 1997; Cottler et al., 1997). Similarly, we note the kappas among users only; if they are reported on the basis of the whole sample they are slightly higher. Hence the κ's reported here represent the strictest measure, indicating the lowest possible limits.

Table 5 gives the κ's reflecting the concordance between instruments for alcohol and selected drug dependence diagnoses. Concordance between all instrument pairs was good for alcohol and opiate dependence, but poor to fair for cannabis, sedative and cocaine dependence. Concordance between instrument pairs was generally in the poor range for harmful use and abuse diagnoses, a result discussed more fully in other reports of this series (Cottler et al., 1997; Pull et al., 1997).

4. Discussion

This study presents an international collaboration which enabled a large number of comparisons to be made in order to identify the issues of reliability and validity in alcohol and drug use disorders. The magnitude of this sample size and richness which allowed the systematic exploration of the reliability and concordance of instruments was only possible due to a joint effort of the collaborating sites. This study has completed many essential reliability and validity exercises on the alcohol and drug sections of these international diagnostic assessment instruments, and looked into possible ways to improve our classification and assessment systems.

The three instruments used in the study, SCAN, CIDI and AUDADIS-ADR prove to be instruments that yield reliable diagnosis for alcohol and drug dependence, but reliability was generally poor for corresponding harmful use and abuse diagnoses. These instruments provide basic information on the presence of the condition listed in the ICD-10, DSM-III-R and DSM- IV as well as additional information on their onset, recency and temporal clustering. There were some differences among the chance-corrected agreement (κ) levels between the instruments, but few of these were significant. Hence the results should not be seen as a competition but rather as a quality assurance that the translation of diagnostic criteria in the classification system allows operational assessments to be made in a reliable way. It was noteworthy that SCAN as a semi-structured instrument, which was anticipated as a potential source for unreliability, proved to be equally reliable in the area of alcohol and drug use disorders.

Reliability coefficients, however, allow us to make some inferences about the validity of the dependence and harmful use or abuse categorizations. Robust reliability figures for dependence categories in alcohol, opiates, cannabis, cocaine, sedatives, and amphetamines provide additional support for the validity of the 'dependence' syndrome concept in different cultures. This may indicate that we are dealing with a real construct, that is 'a disease' underlying the conceptualization and operationalization of dependence disorder.

Table 6
Reliability of ICD-10 alcohol dependence for interview items, criteria and diagnosis by instrument

Items	κ	Criteria	κ	Diagnosis
CIDI				
	0.59, 0.40	Compulsive use	0.65	
	0.61, 0.07, 0.69, 0.66	Impaired control	0.75	
	0.59, 0.71, 0.67, 0.56, 0.60, 0.67, 0.65, 0.50, 0.64, 0.76, 0.67	Withdrawal	0.74	
	0.69, 0.23	Tolerance	0.83	
	0.47, 0.62	Interest neglect	0.53	
0.75	0.48, 0.57, 0.56	Continued use	0.69	Dependence
SCAN				
	0.81	Compulsive use	0.81	
	0.60, 0.55	Impaired control	0.57	
	0.72, 0.74	Withdrawal	0.78	
	0.68	Tolerance	0.67	
	0.61, 0.68	Interest neglect	0.75	
0.76	0.62, 0.66, 0.61	Continued use	0.68	Dependence
AUDADIS-ADR				
	0.48, 0.55	Compulsive use	0.59	
	0.53, 0.50, 0.56, 0.56, 0.57	Impared control	0.60	
	0.57, 0.56, 0.65, 0.59, 0.59, 0.60, 0.66, 0.58, 0.57, 0.62, 0.51, 0.64	Withdrawal	0.57	
	0.59, 0.54	Tolerance	0.64	
	0.50, 0.50, 0.58, 0.55	Interest neglect	0.66	
0.68	0.61, 0.62, 0.59, 0.56, 0.60, 0.58, 0.64, 0.65, 0.79, 0.71, 0.68, 0.67, 0.79, 0.63	Continued use	0.67	Dependence

Chance-corrected test-retest agreement (κ) among users.

On the other hand, lower reliability puts a ceiling on the validity of the abuse harmful use categories. This indicates that even with the new diagnostic classifications in the ICD-10 and DSM-IV, conceptualization and operationalization has not yet reached a satisfactory level. Perhaps this is due to the nature of these categories which are not yet well-defined conditions but merely a transitional state between extended use and dependence. Additionally it may be due to the subjectivity of the 'social' criteria of 'harm' or 'hazard' to daily life which has different content than the more 'biological' criteria of withdrawal, tolerance and craving.

Table 6 shows a 'cascade of reliability' which represents the calculation of reliability at item and criterion level in addition to the diagnostic category level. The results of these analyses for a single substance (alcohol) and for a single diagnostic category (lifetime alcohol dependence in ICD-10) are shown using the data only among users. This analysis, currently being conducted for all substances and each classification system, shows the items and criteria that contribute strongly to the overall reliability for a diagnostic category using a particular instrument and also helps in identifying the 'bottleneck' items.

The item-criteria-diagnosis cascade gives us some clues how these operationalizations could be improved to give better reliability in our assessments. For example 'impaired control' or 'neglect of interests' questions were associated with the lowest test-retest agreement for alcohol. If these criteria could be better operationalized, we can better define our construct and improve the overall reliability of evaluation. Using this method, further modifications were made and continue to be made in the second revisions of CIDI, SCAN and AUDADIS-ADR in terms of questions, ratings, and their algorithms. Similarly, this data base will be useful in further refinement of diagnostic classifications such that it may guide the restructuring the diagnostic criteria and giving weights for their reliability. Additionally the study design will serve as a model to study the reliability and validity of other ADM disorders and disablements.

Acknowledgements

This paper is based on the data and the experience obtained in the WHO collaborative project on reliability and validity of the alcohol and drug use disorders instruments funded by the WHO/NIMH/NIAAA/NIDA joint project (UO1MH35883). The collaborating investigators in this study have been: WHO headquarters Geneva: T.B. Üstün (Principal Investigator), N. Sartorius (Former PI until 1995). NIAAA: B.F. Grant, L. Towle (until 1995); NIDA: J. Blaine; NIMH: D. Regier. In the field research sites: Amsterdam: R. Smeets; Ankara: A. Göğüş; Athens: V. Mavreas; Bangalore: S. Chatterji; Farmington: T. Babor; Ibadan: O. Baiyewu; Jebel: R. Vrasti; Luxembourg: C. Pull; St

Louis: L. Cottler; San Juan: M. R-Stipec; Sydney: (site 1) L. Peters, (site 2) J. Saunders. Consultants: W. van den Brink, D. Carlson, W. Compton, D. Hasin, D. Mager, R. Room. Shekhar Saxena made valuable contributions to the first draft of this paper.

References

American Psychiatric Association (1994) Diagnostic and Statistical Manual of Mental Disorders, 4 th ed. (DSM-IV). American Psychiatric Association, Washington DC.

Chatterji, S., Saunders, J., Vrasti, R, Grant, B., Hasin, D., Mager, D., 1997. 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. 47, 171–185 (this issue).

Cottler, L.B., Robins, L.N., Grant, B., Blaine, J., Towle, L.H., Wittchen, H.U., Satorius, N., 1991. The CIDI Core substance abuse and dependence questions: Cross-cultural and nosological issues. Br. J. Psychiatr. 159, 653–658.

Cottler, L.B., Göğüş, A., Ulüğ, B, Babor, T., 1997. Test-retest reliability of SCAN: Substance abuse sections.

Fleiss, J., 1981. Statistical Methods for Rates and Proportions, 2nd ed. Wiley, New York.

Grant, B.F., Harford, T.C., Dawson, D.A., Chou, P.S., Pickering, R.P., 1995. The alcohol use disorder and associated disabilities interview schedule (AUDADIS) Reliability of alcohol and drug modules in a general population sample. Drug Alcohol Depend. 39, 37–44.

Grant, B.F., Hasin D.S., 1992. The Alcohol Use Disorder and Associated Disabilities Interview Schedule. National Institute on Alcohol Abuse and Alcoholism, Rockville, MD.

Grant, B.F., Towle, L., 1990. Standardized diagnostic interviews for alcohol research. Alcohol Health Res. World 14, 340–348.

Hasin, D.S., Carpenter, K.M., McCloud, S., Smith, M., Grant, B.F., 1997. The alcohol use disorder and associated disabilities interview schedule (AUDADIS): reliability of alcohol and drug modules in a clinical sample. Drug Alcohol Depend. 44 (2,3) 133–141.

Pull, C., Saunders, J., Mavreas, V., Cottler, L., Grant, B., Hasin D., Blaine, J., Mager, D., Üstün, T.B., 1997. Concordance between ICD-10 Alcohol and Drug Use Disorder Criteria and Diagnoses as measured ba the AUDADIS-ADR, CIDI and SCAN: results of a cross-national study. Drug Alcohol Depend. 47, 207–216 (this issue).

Room, R., Janca, A., Bennett, L., Schmidt, L., Sartorius, N., 1997. Cross-cultural applicability research on diagnosis and assessment of substance use disorders: An overview of methods and selected results. Addiction 91, 199–220.

Üstün, T.B., Wittchen, H.-U., 1992. Instruments for the assessment of substance use disorders. Curr. Opin. Psychiatr. 5. 412–419.

Wing, J., Babor, T., Brugha, T., Burke, J., Cooper, J., Giel, R., Jablenski, A., Regier, D., Sartorius, N., 1990. SCAN. Schedules for clinical assesment in neurophychiatry. Arch. Gen. Psychiatr. 47, 589–593.

Wittchen, H.U., Robins, L.N., Cottler, L.B., Sartorius, N., et al., 1991. Cross-cultural feasibility, reliability and sources of variance of the Composite International Diagnostic Interview (CIDI). Crit. J. Psychiatr. 159, 645–653.

World Health Organization and Alcohol, Drug Abuse and Mental Health Administration, 1985. Mental disorders, alcohol- and drug-related problems: International perspectives on their diagnosis and classification. Excerpta Med. 1985.

World Health Organization, 1993b. Composite International Diagnostic Interview: Core Version 1.1. Geneva: World Health Organization, Geneva, Switzerland.

World Health Organization, 1993a. The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research. World Health Organization, Geneva, Switzerland.

World Health Organization, 1992. Alcohol Use Disorders and Associated Disabilities Interview Schedule—Alcohol/Drug-Revised. World Health Organization, Geneva, Switzerland.

World Health Organization, 1993c. Schedules for Clinical Assessment in Neuropsychiatry. American Psychiatric Press, Washington DC.