Psychiatric Research Interview for Substance and Mental Disorders (PRISM): Reliability for Substance Abusers

Deborah S. Hasin, Ph.D., Kristin D. Trautman, M.S.S.W., Gloria M. Miele, Ph.D., Sharon Samet, M.S.W., Melissa Smith, M.A., and Jean Endicott, Ph.D.

Objective: The purpose of this study was to investigate the reliability of a new semistructured diagnostic interview, the Psychiatric Research Interview for Substance and Mental Disorders (PRISM), for substance-abusing patients. The reliability of psychiatric diagnoses for individuals who drink heavily or use drugs has been shown to be problematic. The PRISM was designed to improve the reliability for such individuals. _Method:_ A test-retest reliability study of the PRISM was conducted with 172 patients being treated in dual-diagnosis or substance abuse settings. _Results:_ Good to excellent reliability was shown for many diagnoses, including affective disorders, substance use disorders, eating disorders, some anxiety disorders, and psychotic symptoms. The interview has recently been updated for DSM-IV diagnoses. _Conclusions:_ The PRISM offers a method of producing psychiatric diagnoses with improved reliability for patients and other research subjects who have problems with alcohol or drugs.
(Am J Psychiatry 1996; 153:1195–1201)

Understanding and treating comorbid alcohol, drug, and psychiatric disorders has become increasingly important in the last 10 years. Many clinical and epidemiologic studies have revealed a high degree of co-occurrence of substance abuse and psychiatric disorders. However, the diagnosis of psychiatric disorders in individuals with heavy alcohol or drug use has been problematic. According to DSM-III-R, a psychiatric diagnosis can be made if "it cannot be established that an organic factor initiated and maintained the disturbance" (p. 223), and organic disorders are diagnosed on the basis of "evidence from the history, physical examination, or laboratory tests of a specific organic factor (or factors) that is judged to be etiologically related to the disturbance" (p. 103). Rather than offering guidelines on taking a clinical history, DSM-III-R suggests using biologically oriented tests such as urine and blood tests and "challenges with a known intoxicant." These procedures are of limited use in the most common diagnostic situation, when the main source of information is a clinical interview and possibly an informant's report, and when the time frame of assessment extends more than a few days into the past. This situation has been problematic for research as well as clinical practice. A review of the methodological problems in diagnosing coexisting DSM-III-R psychiatric disorders in substance abusers (1) indicated that this lack of clarity resulted in the use of very different diagnostic procedures from study to study.

As is widely known, several general-purpose diagnostic interviews have been developed to assess the main adult axis I psychiatric disorders as defined by the different sets of diagnostic criteria such as DSM-III-R. Clinician-administered procedures (2–4) leave the differentiation of "organic" versus "nonorganic" to clinical judgment. Interviews designed for lay interviewers (5, 6) leave this differentiation to subjects' attribution on a symptom-by-symptom basis. Both of these situations are conducive to criterion variance: differences in formal inclusion and exclusion criteria used to summarize data into psychiatric diagnoses (7). Criterion variance is one of the principal contributors to diagnostic unreliability.

There is surprisingly little empirical evidence on the reliability or validity of psychiatric diagnosis among heavy drinkers or drug users. In a sample of substance abusers in treatment, the Schedule for Affective Disorders and Schizophrenia, Lifetime Version (SADS-L) was administered 1 week and 4 weeks after admission; the reliability of lifetime diagnoses was found to be poor (8). We have been unable to find a report of a standard test-retest study of psychiatric diagnoses made with the National Institute of Mental Health Diagnostic Interview Schedule (DIS) (5) or the Composite Inter-

FIGURE 1. Examples of DSM-IV Depression Items and Interviewer Instructions on the Psychiatric Research Interview for Substance and Mental Disorders (PRISM)[a]

Interviewer instruction: Use following guidelines for assessing symptoms of depression in relation to baseline state:

Potential Major Depression
- If baseline (usual) mood is *good/euthymic*, symptoms must represent clear change from baseline state.
- If baseline mood is *low/dysthymic*, symptoms must change (re: frequency, persistence, severity, or number) as mood changes (e.g., worsen when depressed mood worsens and improve when mood improves).

Potential Substance-Induced Depressions
- If identified period occurs during period of chronic substance use, include symptom if occurrence or severity is a clear change from substance-using baseline state.
- Exclude acute effects of intoxication/withdrawal caused by a change in level of substance use immediately before onset of symptom.
- Specific substances that can cause symptoms are listed at the last bullet (see "Relevant substance states") of each applicable item; use this information as a basis for probing temporal relationship of symptom to intoxication and/or withdrawal.

Insomnia	Insomnia Nearly Every Day	? 1 2 3
15. During that time, did you have trouble sleeping nearly every day? (Can you describe that to me? Was that a definite change from usual? Did you have trouble falling asleep? Did you wake up during the night and have trouble going back to sleep? Did you wake up much earlier than you wanted to? How much sleep were you actually getting during your usual sleeping hours? How much sleep do you usually get? How often did that happen? How long did that go on?)	• Include if has 25% estimated decrease in sleep 5+ out of 7 days over at least a 2-week period. • 25% decrease in sleep=about 2 hours less when 7–8 hours is usual, an hour and a half less when 6 hours is usual. • Include if has 25% initial, middle, or terminal insomnia but makes up for lost sleep during the day. • Exclude if asleep but had terrible dreams. • Exclude if asleep but exhausted in the morning. • Exclude if wakes up a few times and falls back to sleep in a few minutes. • Exclude loss of sleep due to external circumstances (e.g., construction noise outside) or hallucinations or delusions. • Relevant substance states: COC, STI intoxication; ALC, SED, COC, STI, HER, OPI withdrawal.[b]	Check all that apply: ☐ Initial ☐ Middle ☐ Terminal

Alcohol-Induced Syndrome	Substance-Induced Syndrome: Alcohol	? 1 2 3
31. Just before that time started when you were (depressed/SOE[c]), what was your drinking like? (How much were you drinking then?) Had you recently cut back on your drinking? Or started drinking more than you had been? Was there any change at all in your drinking habits? If yes to any: What was the change? When did that happen? Were you feeling (depressed/SOE[c]) before you (increased/decreased) your drinking?	• Full syndrome beginning and remitting *entirely* during chronic intoxication (i.e., intoxicated more than half the days for a month)=3 (steady state). • Full syndrome beginning during chronic intoxication or acute withdrawal and *lasting no longer than 4 weeks* after cessation of intoxication or withdrawal=3. • Acute alcohol withdrawal usually lasts up to 7 days. • Mixed or unclear etiology=2. • Be sure to justify fully any ratings of 2 or 3 in the narrative, as well as any ratings of 1 for episodes occurring during periods of heavy drinking.	☐ Steady state only

Specify drinking _____

1=Symptom not present   2=Subclinical   3=Symptom positive

[a]The left column contains the probes, the middle column gives descriptions of clinical features, and the right column is used for the ratings.
[b]COC=cocaine, STI=stimulant, ALC=alcohol, SED=sedative, HER=heroin, OPI=opiate.
[c]SOE=subject's own equivalent.

national Diagnostic Interview (6) for substance-abusing patients. In a large test-retest reliability study of the Structured Clinical Interview for DSM-III-R (SCID), diagnoses of affective disorders in patients from a substance abuse treatment program showed poor test-retest reliability (9), and the same was true for several psychiatric diagnoses among subjects at several sites who were current abusers of alcohol or drugs (10). Research on the SCID's differentiation between substance-induced and independent psychiatric diagnoses for patients with substance abuse showed little cross-sectional or predictive validity (11). No evidence of this type is yet available for the Schedules for Clinical Assessment in Neuropsychiatry (4). Poor reliability and/or validity of diagnoses reduces the possibility that a study (for

example, a treatment study) will be able to show significant differences between groups when diagnoses are an important grouping or outcome variable.

To address the lack of a diagnostic interview that is suitable for comorbidity research, we developed the Psychiatric Research Interview for Substance and Mental Disorders (PRISM). The original version of the PRISM was developed to evaluate subjects according to DSM-III-R criteria, although sufficient advance information on the DSM-IV criteria for substance use disorders was available to one of us (D.S.H.) that DSM-IV diagnoses of substance dependence and abuse could also be produced from this version. The PRISM shares some important general features with the SCID. For example, the PRISM uses a three-column format (figure 1

shows an example of PRISM items). This includes a middle column where the clinical feature to be rated is described (e.g., diagnostic symptoms or criteria), a left column with probes for interviewers to use in obtaining the needed information, and a right column where items are rated. Also, in the PRISM (as in the SCID and all semistructured diagnostic interviews), interviewers use the probes in the left column as a starting point but are expected to ask additional probing questions as needed to obtain the information for rating each item.

Listening to audiotapes of discordant SCID test-retest pairs of interviews (provided to us by the Biometrics Research staff of the New York State Psychiatric Institute), we found that often interviewers conscientiously probed unclear points with subjects but remained unclear (and inconsistent) about how to rate the information. This occurred because the SCID does not provide sufficiently specific rating guidelines, allowing for item-level criterion variance (7). Many features of the PRISM were designed to overcome this problem. Some of these are specific to comorbidity issues, while others are more general.

New features of the PRISM that are specific to comorbidity include 1) relocation of the alcohol and drug sections from the middle of the interview to the beginning (following the introductory section); thus, PRISM interviewers have a thorough knowledge of the subject's patterns of drinking and drug use and history of alcohol and drug use disorders when they start covering psychiatric disorders; 2) an increase in the level of structure and completeness in taking an alcohol and drug use history before the alcohol and drug diagnostic sections; and 3) the addition of separate organic exclusion items for alcohol, drugs, and prescribed medications, as well as interviewer instructions and middle-column guidelines to assist in differentiating organic from nonorganic disorders. As we have noted, DSM-III-R is not very specific about making this differentiation on the basis of the clinical history, and there is controversy in the field about the most valid definitions. However, reliance solely on clinical judgment in previous semistructured interviews had not produced reliable diagnoses, so we believed that some guidelines were better than none.

New features of the PRISM that were designed to increase reliability in general, not just specific to comorbidity issues, included the following. 1) Specific rating guidelines were introduced throughout the interview, including frequency and duration requirements, exclusion specifications, and decision rules for frequently encountered sources of uncertainty. Since the PRISM was first used in a community sample (albeit of heavy drinkers [12]), great care was taken to clarify the distinctions between subthreshold and threshold ratings of symptoms. 2) Dissimilar components of individual criteria were separated into different items, to ensure that interviewers assess all of the components and to allow for simplified rating guidelines. 3) Many probes were added to the left column to standardize probing to some extent and to indicate the questions usually needed to

explore a response that does not fully answer an interviewer's question. Through repeated testing by experienced diagnostic interviewers, these probes were adjusted for suitability with a variety of subjects (black, white, English-speaking Hispanic, male, and female). 4) The level of structure in the introductory section was increased. Items were added to obtain indirect indicators of alcohol, drug, or psychiatric problems (for example, an interrupted education or employment history or a treatment history). 5) Computer programs were developed to produce diagnoses from the interview data. These programs reduce clerical error and also reduce the possibility that clinicians will diagnose by opinion rather than by the DSM criteria. 6) A training package consisting of a manual, videotapes, prescripted role plays, and other training aids was developed. This training standardizes semistructured interviewing techniques and comprehension of the interview material. The training assumes education of at least a master's degree level in a field such as social work or psychology.

At an intermediate point in the development of the PRISM, the instrument was subjected to a test-retest study with 75 community residents, most of whom were prescreened for heavier than average drinking (a description of the screening and recruitment procedures is provided elsewhere [12]). This preliminary study indicated reliabilities of 0.86 and 0.66 for current and past major depressive disorder, respectively, and 0.65 for dysthymia (our unpublished data). Reliability for alcohol, drug, and eating disorders was also generally good, but low prevalences in some areas prevented complete testing of the instrument, and problems in some sections indicated the need for further development. The PRISM was extensively revised after that first trial and subjected to a larger test-retest study in a clinical sample; this study is described below.

METHOD

The subjects were 172 patients being treated in one of two sites, an inpatient dual-diagnosis psychiatric setting and an outpatient drug counseling setting. The 82 female subjects were consecutively admitted patients, and the 90 male subjects were a random subgroup of consecutively admitted patients, so as to have a study group as evenly balanced for gender as possible. About 52% of the subjects were male, and about 20% had not completed high school. About 59% of the subjects were white, 29% black, 11% Hispanic, and the remainder "other." The mean age of the subjects was 34.97 years (SD=9.7).

Interviews were conducted shortly after admission and after withdrawal was complete, if applicable. A complete description of the study was given to the subjects, and then written informed consent was obtained. The response rate among those approached was 93%. The median interval between interviews was 7 days, and 75% of the subjects were reinterviewed within 12 days, although a few of the second interviews took place several weeks later. The second interviewer was always blind to the results of the first interview.

All interviewers had a master's degree in social work or psychology and at least some clinical experience with patients with substance abuse or psychiatric patients. Before being hired, all potential interviewers were required in their job interviews to show ability to do PRISM-type interviewing through role playing of brief segments of the PRISM. Some of the interviewers who worked on this study had

TABLE 1. Test-Retest Reliability (Kappa) of DSM-IV Diagnoses of Substance Use Disorders With the Psychiatric Research Interview for Substance and Mental Disorders (PRISM) (N=172)

| | Current | | | | Past | | | | Lifetime | | | |
| | | | Prevalence | | | | Prevalence | | | | Prevalence | |
Disorder	Kappa	SE	Time 1	Time 2	Kappa	SE	Time 1	Time 2	Kappa	SE	Time 1	Time 2
Dependence												
Alcohol	0.81	0.04	0.51	0.49	0.64	0.06	0.65	0.65	0.69	0.06	0.70	0.69
Anxiolytics	0.81	0.11	0.05	0.05	0.59	0.11	0.11	0.08	0.69	0.09	0.14	0.12
Cannabis	0.80	0.09	0.09	0.06	0.50	0.09	0.17	0.11	0.63	0.07	0.26	0.17
Cocaine	0.92	0.03	0.43	0.43	0.78	0.06	0.18	0.17	0.88	0.04	0.60	0.59
Hallucinogens	0.49	0.31	0.01	0.01	0.55	0.16	0.04	0.04	0.53	0.15	0.05	0.05
Heroin	0.94	0.04	0.10	0.12	0.94	0.04	0.11	0.11	0.95	0.03	0.21	0.22
Licit opiates	0.85	0.10	0.04	0.04	0.87	0.09	0.05	0.05	0.85	0.07	0.09	0.09
Stimulants	0.66	0.32	0.01	0.01	0.69	0.10	0.10	0.06	0.68	0.10	0.10	0.08
Abuse												
Alcohol	0.32	0.15	0.04	0.06	0.10	0.10	0.09	0.10	0.18	0.10	0.09	0.12
Anxiolytics	0.20	0.19	0.04	0.02	0.38	0.12	0.08	0.10	0.36	0.11	0.11	0.11
Cannabis	0.51	0.11	0.10	0.09	0.43	0.08	0.24	0.25	0.50	0.07	0.32	0.33
Cocaine	−0.01	0.01	0.01	0.01	0.43	0.20	0.02	0.03	0.28	0.17	0.04	0.04
Hallucinogens	−0.01	0.01	0.01	0.01	0.46	0.13	0.07	0.07	0.40	0.13	0.08	0.08
Heroin	1.00	0.00	0.01	0.01	0.32	0.25	0.02	0.02	0.49	0.22	0.02	0.02
Licit opiates	−0.01	0.01	0.01	0.01	0.14	0.15	0.03	0.03	0.26	0.16	0.04	0.04
Stimulants	—	—	—	—	0.42	0.14	0.06	0.07	0.39	0.14	0.06	0.07

previous experience in using the PRISM, while others did not. (Examination of the study results by interviewer did not show interviewer effects.) Training included use of the PRISM training manual, viewing videotapes of the DSM-III-R version of the PRISM, and role playing. Each interviewer was then observed several times by a supervisor to ensure that the interview was being administered correctly. Once the study began, regular staff meetings were held to prevent interviewer drift and to identify problem areas in the PRISM as the study progressed. As interviews were completed, they were checked by a supervisor and the data were entered into computer files and cleaned. The diagnostic programs were run on these data to produce the diagnoses.

In the statistical analysis, the kappa statistic was used to indicate the degree of interrater reliability beyond agreement that would be expected due to chance. Kappa values can range from 1.00 (perfect agreement) to −1.00 (total disagreement). Kappas of 0.75 and above indicate excellent agreement, from 0.40 to 0.74 indicate fair to good agreement, and below 0.39 indicate poor agreement (13). A kappa of 0 indicates agreement no better than chance. We include some kappas for relatively low-prevalence disorders because they offer additional information about certain disorders (particularly substance abuse) and about the psychometric performance of the interview.

RESULTS

Substance use disorders. As shown in table 1, the reliabilities for PRISM DSM-IV diagnoses of current alcohol and drug dependence, including licit opiates as well as heroin, were generally very high. The main exception was hallucinogens. Kappas were generally lower for past diagnoses of dependence but still excellent for cocaine, heroin, and licit opiates. The reliabilities for DSM-III-R diagnoses of dependence (not shown) were generally in a similar range. Kappas for the abuse diagnoses are also shown in table 1. If coded so that dependence hierarchically excluded abuse on a lifetime basis, as required in DSM-III-R and DSM-IV, abuse generally showed poor to chance-level agreement for current, past, and lifetime diagnoses. An alternative interpretation with respect to the DSM-IV criteria is

that an abuse diagnosis beginning before the onset of dependence should be considered positive. When such cases were recoded as positive, the reliabilities were slightly better but still low (not shown). Further investigation revealed that these low reliabilities were not due to the abuse criteria themselves, since abuse classified as positive regardless of dependence status showed kappas ranging from good to excellent for all drugs except hallucinogens (not shown).

Affective disorders. The reliability coefficients for DSM-III-R affective and other psychiatric disorders represent agreement on nonorganic disorders (termed "primary" in DSM-IV). As shown in table 2, kappas for current and past major depression were 0.81 and 0.64, respectively, while for current dysthymia kappa was 0.61, closely replicating the results of the earlier test-retest PRISM study of 75 cases that we have described. Kappas for current and past bipolar I disorder were 0.65 and 0.95, respectively. Kappas for bipolar II disorder were fair for current diagnoses and poor for past disorders.

Anxiety disorders. Reliabilities for current diagnoses of panic disorder, social phobia, and agoraphobia without panic were in the good-to-acceptable range (table 2). Current agoraphobia with panic achieved a kappa only in the low-to-fair range. There were no cases of current generalized anxiety disorder in the group. Kappas for diagnoses of past anxiety disorders ranged from 0.66 for agoraphobia without panic to 0.49 for generalized anxiety disorder and agoraphobia with panic. Obsessive-compulsive disorder was not assessed in the DSM-III-R version of the PRISM because of its expected low prevalence among alcohol and drug abusers.

Psychotic symptoms. The DSM-III-R version of the PRISM followed the SCID (nonpatient version) convention of providing ratings of psychotic symptoms but not of distinct psychotic disorders (8). As shown in table 2,

TABLE 2. Test-Retest Reliability (Kappa) of DSM-III-R Diagnoses of Psychiatric Disorders With the Psychiatric Research Interview for Substance and Mental Disorders (PRISM) (N=172)

Disorder	Current				Past				Lifetime			
			Prevalence				Prevalence				Prevalence	
	Kappa	SE	Time 1	Time 2	Kappa	SE	Time 1	Time 2	Kappa	SE	Time 1	Time 2
Affective												
Major depression	0.81	0.05	0.37	0.34	0.64	0.06	0.47	0.46	0.74	0.05	0.52	0.52
Dysthymia	0.61	0.13	0.07	0.04	—	—	—	—	—	—	—	—
Bipolar I	0.65	0.13	0.05	0.06	0.95	0.05	0.07	0.06	0.84	0.08	0.08	0.08
Bipolar II	0.56	0.23	0.02	0.02	0.27	0.23	0.02	0.02	0.49	0.22	0.02	0.02
Anxiety												
Panic	0.68	0.09	0.13	0.10	0.68	0.09	0.14	0.10	0.64	0.09	0.16	0.13
Agoraphobia with panic	0.48	0.18	0.05	0.03	0.45	0.15	0.06	0.04	0.45	0.15	0.06	0.04
Agoraphobia without panic	0.50	0.31	0.02	0.01	0.66	0.32	0.01	0.01	0.50	0.31	0.02	0.01
Generalized anxiety	—	—	—	—	0.49	0.31	0.01	0.01	0.49	0.31	0.01	0.01
Social phobia	0.61	0.13	0.06	0.07	0.52	0.14	0.10	0.10	0.52	0.14	0.10	0.10
Psychotic symptoms	0.63	0.10	0.09	0.10	0.76	0.08	0.13	0.12	0.79	0.07	0.15	0.14
Eating												
Anorexia	0.80	0.20	0.02	0.01	0.87	0.09	0.05	0.05	0.87	0.09	0.05	0.05
Bulimia	0.83	0.12	0.03	0.04	0.87	0.09	0.04	0.05	0.87	0.09	0.04	0.05
Binge eating	0.74	0.17	0.03	0.02	0.79	0.14	0.03	0.02	0.79	0.14	0.03	0.02
Personality												
Antisocial	—	—	—	—	—	—	—	—	0.60	0.07	0.25	0.20
Borderline[a]	—	—	—	—	—	—	—	—	0.83	0.07	0.23	0.22

[a]Assessed on 86 cases.

reliabilities for current, past, and lifetime psychotic symptoms were good to excellent.

Eating disorders. These conditions occurred rarely in this group but were assessed reliably nevertheless (table 2). Kappas for anorexia and bulimia were excellent. Binge-eating disorder is a condition analogous to bulimia but without the compensatory behaviors, such as purging to curtail weight gain (14, 15, and DSM-IV, p. 731). The reliability for current binge-eating disorder fell just short of the excellent range.

Personality disorders. Antisocial and borderline personality disorders were included in the PRISM because of their high prevalence in substance-abusing populations. Borderline personality disorder (added to the interview partway through the study) showed excellent reliability, while antisocial personality disorder showed reliability in the low end of the good range (table 2).

DISCUSSION

In a number of ways, this test-retest study showed that substantial progress was made toward the goal of achieving reliable diagnoses of psychiatric as well as substance use disorders for alcohol- and drug-abusing subjects. PRISM reliabilities for many affective disorders, anxiety disorders, and psychotic symptoms were appreciably better for alcohol and drug abusers than has been demonstrated with other semistructured diagnostic interviews. The findings are discussed in conjunction with their influence on the DSM-IV revision of the PRISM.

Diagnoses of DSM-IV substance dependence showed excellent reliability, but the reliability of diagnoses of hierarchically defined substance abuse was poor. The abuse criteria themselves do not have low reliability when considered nonhierarchically. This problem has also been shown in other instruments, largely in unpublished data, causing most reliability data on alcohol and drugs to be published only for dependence or in combined form as substance use disorder (abuse *or* dependence). While the DSM-IV version of the PRISM includes separate diagnoses of substance abuse and dependence to match the DSM-IV criteria, the general concept of substance abuse and its differentiation from dependence remains problematic. Some resolution of this problem will probably be required before any interview schedule can produce reliable abuse diagnoses.

The reliability of the bipolar II diagnosis was fair for current disorder and poor for past disorder. DSM-III-R does not include specific criteria for this diagnosis because of past difficulties with its reliability, so we used an algorithm analogous to the DSM-IV criteria. Clearly, a higher kappa would be preferable. However, the PRISM did no worse on this diagnosis than other instruments that were not tested specifically on substance abusers (9, 16, 17).

Investigation of the PRISM data on cases that had discrepant ratings for psychotic symptoms revealed that the training materials had allowed interviewers to maintain somewhat different concepts about how to rate the items. This information was used to clarify the DSM-IV versions of the PRISM training materials, middle column guidelines, and interviewer instructions for the psychotic symptoms section. The psychotic symptoms section has also been expanded to make diagnoses of specific psychotic disorders rather than the more global rating of "psychotic symptoms."

Generalized anxiety disorder occurred at a low prevalence in this group and did not show high reliability.

This diagnosis has shown reliability problems in other study groups as well (9, 18). The rating guidelines for generalized anxiety disorder items have been clarified as much as possible for the DSM-IV version of the PRISM.

PRISM DSM-III-R antisocial personality disorder had a lower kappa than has been found in other studies (19). Although the main focus of these other studies was personality disorder, and most involved some level of prescreening for this type of disorder, we feel that the kappa for PRISM diagnoses of antisocial personality disorder reflected two problems in the DSM-III-R version of the PRISM. First, PRISM items on conduct symptoms determined whether they had occurred at all and then whether they occurred before age 15, since age 15, or in both time frames. (This was done to investigate aspects of the antisocial personality disorder criteria.) Some of the discordance with respect to antisocial personality disorder occurred because subjects identified inconsistent time frames for conduct symptoms. This complexity has been removed from the DSM-IV version of the PRISM, in which conduct symptoms are now queried and coded only in the "before age 15" time frame. Second, we had attempted to increase the focus on lack of remorse to improve the validity of the antisocial personality disorder section, but the several PRISM remorse items were not very reliable. In the DSM-IV version of the PRISM, coverage of remorse has been made more standard and comparable to that in other diagnostic interviews, which should increase the reliability of the antisocial personality disorder diagnosis to customary levels.

This study focused entirely on reliability. Validity is also an important dimension of an instrument's psychometric performance. In the absence of a "gold standard," validity research is more complex than reliability testing, since multiple measures are usually involved. Investigation of the validity of the PRISM is outside the scope of this report but is currently underway and will be reported when complete.

One may question the generalizability of this study because of the qualifications of the interviewers and supervisor, the level of care taken in the hiring, training, and supervision, and the quality checking of the data. All of these factors undoubtedly contributed to the reliability reported here. If we had used research assistants as interviewers (a common practice), the reliabilities might well have been lower. In our view, the methods used in this study would be applicable to any data collection involving complex diagnostic issues and a semistructured interview. Good research assistants can almost always ask the initial probes about symptoms and code the responses. However, without considerable extra training, such interviewers do not have the knowledge to formulate semistructured follow-up probes on the timing, intensity, or characteristics of symptoms. Perhaps, as empirical knowledge about the relationships of specific substance use patterns to different types of psychopathology accumulates, a future interview can be developed with probes on these relationships that are substance-specific but simple enough to be administered by research assistants or other lay interviewers.

DSM-IV provides considerably more information on making psychiatric diagnoses of substance-abusing patients than earlier nomenclatures and also provides (for the first time) some guidelines for using the clinical history to differentiate the two types of syndromes. As the PRISM was revised for DSM-IV, adjustments were made in the interview to take these changes into account, to improve further the reliability of PRISM diagnoses based on aspects of the test-retest data, and to shorten and simplify the interview wherever possible. The main additions across all sections of the PRISM consist of 1) guidelines in the middle column on the psychotomimetic effects of specific substances, 2) items providing for characterization of the primary/secondary distinction between substance use disorders and psychiatric disorders, and 3) items on onset, recency, and duration of substance-induced psychiatric disorders as well as primary psychiatric disorders. Figure 1 contains an example of material from the depression section of the DSM-IV version of the PRISM. The figure illustrates the specificity that enables interviewers to rate information consistently. For completeness, sections covering specific phobias, obsessive-compulsive disorder, and posttraumatic stress disorder have also been added. The updated DSM-IV version takes between 50 and 150 minutes to administer once interviewers have become familiar with the interview. The range depends on the complexity of the substance use and psychiatric history.

A longitudinal version of the PRISM, the PRISM-L, has also been prepared for follow-up studies. A notable feature of the PRISM-L is the inclusion of timeline grids modeled on the Longitudinal Interval Follow-up Evaluation (20) to rate remission, relapse, and other aspects of the clinical status of disorders as well as use of alcohol and drugs over time. This type of data is used in survival analyses of the course of disorders and interrelationships between disorders when patients are followed prospectively. The PRISM-L will be undergoing a test-retest reliability study shortly, which will be reported later.

Recall that a number of PRISM features were designed to enhance general reliability apart from comorbidity issues. Thus, we suspect that the PRISM would work well for subjects who do not abuse alcohol or drugs. Without the need for a detailed alcohol and drug history, the PRISM would also be shorter for such subjects. However, whether the PRISM would offer better reliability than other instruments for these subjects is speculative unless empirically tested.

In summary, the PRISM, a semistructured diagnostic interview, has demonstrated reliability at least as good for substance-abusing patients as other interviews have been shown to be for general samples. Since initial development of the PRISM took place with community residents and good preliminary results were also obtained in that setting, the PRISM should also be useful in groups where substance abuse is common even if the

subjects are not in treatment for such problems. Since a number of investigators are currently planning studies that include the PRISM as the diagnostic interview, presentation of these reliability results seems timely. The PRISM interview is a bit more demanding in terms of personnel and resources than some other widely used interviews. However, the findings suggest that when diagnostic issues are complex (as they are in substance-using populations), the improved reliability is worth the extra effort in terms of the ability to test hypotheses and investigate relationships between psychiatric diagnoses and other variables in subjects who abuse alcohol or drugs.

REFERENCES

1. Weiss R, Mirin S, Griffin M: Methodological considerations in the diagnosis of coexisting psychiatric disorders in substance abusers. Br J Addiction 1994; 87:179–187
2. Endicott J, Spitzer RL: A diagnostic interview: the Schedule for Affective Disorders and Schizophrenia. Arch Gen Psychiatry 1978; 35:837–844
3. Spitzer RL, Williams JBW, Gibbon M, First MB: The Structured Clinical Interview for DSM-III-R (SCID), I: history, rationale, and description. Arch Gen Psychiatry 1992; 49:624–629
4. Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, Jablenski A, Regier D, Sartorius N: SCAN: Schedules for Clinical Assessment in Neuropsychiatry. Arch Gen Psychiatry 1990; 47: 589–593
5. Robins LN, Helzer JE, Croughan J, Ratcliff KS: The National Institute of Mental Health Diagnostic Interview Schedule: its history, characteristics, and validity. Arch Gen Psychiatry 1981; 38: 381–389
6. Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J, Farmer A, Jablenski A, Pickens R, Regier DA, Sartorius N, Towle LH: The Composite International Diagnostic Interview: an epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. Arch Gen Psychiatry 1988; 45:1069–1077
7. Spitzer RL, Endicott J, Robins E: Clinical criteria for diagnosis and DSM-III. Am J Psychiatry 1975; 132:1187–1192
8. Rounsaville B, Anton S, Carroll K, Budde D, Prusoff B, Gawin F: Psychiatric diagnoses of treatment-seeking cocaine abusers. Arch Gen Psychiatry 1991; 48:43–51
9. Williams JBW, Gibbon M, First MB, Spitzer RL, Davies M, Borus J, Howes MJ, Kane J, Pope HG, Rounsaville B, Wittchen H-U: The Structured Clinical Interview for DSM-III-R (SCID), II: multisite test-retest reliability. Arch Gen Psychiatry 1992; 49:630–636
10. Bryant K, Rounsaville B, Spitzer R, Williams J: Reliability of dual diagnosis: substance dependence and psychiatric disorders. J Nerv Ment Dis 1992; 189:251–257
11. Kadden R, Kranzler H, Rounsaville B: Validity of the distinction between "substance-induced" and "independent" depression and anxiety disorders. Am J Addictions 1995; 4:107–117
12. Hasin D, McCloud S, Li Q, Endicott J: Cross-system agreement among demographic groups: DSM-III, DSM-III-R, DSM-IV and ICD-10 diagnoses of alcohol use disorders. Drug Alcohol Depend 1996; 41:127–135
13. Fleiss J: Statistical Methods for Rates and Proportions, 2nd ed. New York, John Wiley & Sons, 1981
14. Spitzer RL, Devlin M, Walsh BT, Hasin D, Wing R, Marcus M, Stunkard A, Wadden T, Yanovski S, Agras S, Mitchell J, Nonas C: Binge eating disorder: a multisite field trial of the diagnostic criteria. Int J Eat Disord 1992; 11:191–203
15. Spitzer RL, Yanovski S, Wadden T, Wing R, Marcus MD, Stunkard A, Devlin M, Mitchell J, Hasin D, Horne RL: Binge eating disorder: its further validation in a multisite study. Int J Eat Disord 1992; 13:137–153
16. Spitzer RL, Endicott J, Robins E: Research Diagnostic Criteria: rationale and reliability. Arch Gen Psychiatry 1978; 35:773–782
17. Andreasen NC, Grove WM, Shapiro RW, Keller MB, Hirschfeld RM, McDonald-Scott P: Reliability of lifetime diagnosis: a multicenter collaborative perspective. Arch Gen Psychiatry 1981; 38:400–405
18. Manuzza S, Fyer AJ, Martin LY, Gallops MS, Endicott J, Gorman J, Liebowitz MR, Klein DF: Reliability of anxiety assessment, I: diagnostic agreement. Arch Gen Psychiatry 1989; 46: 1093–1101
19. Zimmerman M: Diagnosing personality disorders: a review of issues and research methods. Arch Gen Psychiatry 1994; 51: 225–245
20. Keller MB, Lavori PW, Friedman B, Nielsen E, Endicott J, McDonald-Scott P, Andreasen NC: The Longitudinal Interval Follow-up Evaluation: a comprehensive method for assessing outcome in prospective longitudinal studies. Arch Gen Psychiatry 1987; 44:540–548