

Twelve-Month Test-Retest Reliability of the Structured Clinical Interview for DSM-III-R Personality Disorders in Cocaine-Dependent Patients

Roger D. Weiss, Lisa M. Najavits, Larry R. Muenz, and Cathryn Hufford

This study examined 12-month test-retest reliability of the Structured Clinical Interview for DSM-III-R Personality Disorders (SCID-II) in cocaine-dependent patients. Thirty-one patients completed the SCID-II during the second week of hospitalization for cocaine dependence, and again 12 months later. In both interviews, patients were asked to answer questions about their personality during the several years preceding admission to the hospital. Test-retest reliability, as

measured by κ , was relatively poor at .46. However, reliability of negative diagnoses (the absence of a disorder at both time points) was higher than reliability of positive diagnoses (the presence of a disorder at both time points). Reasons for the difficulty in attaining long-term test-retest reliability of axis II diagnoses in cocaine-dependent patients are discussed.

Copyright © 1995 by W.B. Saunders Company

PATIENTS WITH substance use disorders and comorbid personality disorders represent an important clinical subgroup. The presence of axis II disorders in patients with substance use disorders has been associated with longer and more severe drug use histories, lower scores on measures of life satisfaction, higher rates of additional psychiatric disorders, diminished treatment retention, and poorer outcome.¹⁻³ Cocaine-dependent patients in particular have been reported to have a high frequency of axis II diagnoses, with prevalence rates ranging from 31% to 74%.⁴⁻⁸

Unfortunately, accurately assessing personality disorders in the context of substance dependence can pose great difficulties. Certain personality traits that may otherwise be viewed as problematic, such as manipulativeness, may be an integral and even adaptive part of the lifestyle of an individual using illicit drugs. Moreover, if someone is recurrently using or withdrawing from drugs of abuse, it can be difficult to determine whether certain behaviors or psychiatric symptoms are a result of these processes, or whether they indicate a separate

co-occurring disorder. When drug use begins at an early age, differentiating drug-related behavior from personality problems is particularly complicated.

A number of structured interviews have been developed to standardize and improve reliability of the process of diagnosing personality disorders.⁹ One indication of the utility of any such instrument is its test-retest reliability, i.e., agreement on the diagnosis of the same patient at two different time points.⁹ This is particularly relevant for axis II disorders, since they are by definition seen as lifetime disorders that are highly resistant to change.¹⁰ If a disorder that is conceptualized as enduring cannot be reliably diagnosed at different times, one can question either the diagnosis itself, the instrument used to measure it, or both.

There has been little research on test-retest reliability of axis II diagnoses in patients with substance use disorders. Malow et al.⁴ used sections of the Structured Clinical Interview for DSM-III-R Personality Disorders (SCID-II)¹¹ to examine test-retest reliability of borderline and antisocial personality disorder diagnoses. Two different interviewers administered the antisocial and borderline modules of the SCID-II 48 hours apart; κ values for diagnostic agreement were .87 and .84, respectively, indicating excellent reliability. First et al.¹² conducted a test-retest reliability study of the SCID-II on 284 subjects in five different sites (three psychiatric facilities and two nonpatient sites); 46 subjects were recruited from a substance abuse treatment unit. Assessments were performed by two different interviewers, who evaluated patients 1 to 14 days apart. The overall weighted κ

From the Alcohol and Drug Abuse Program, McLean Hospital, Belmont, MA; Department of Psychiatry, Harvard Medical School, Boston, MA; and Department of Biostatistics in Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA.

Supported by Grants No. DA-05944, DA-07693, and DA-08631 from the National Institute on Drug Abuse, and the Dr. Ralph and Marian C. Falk Medical Research Trust.

Address reprint requests to Roger D. Weiss, M.D., Alcohol and Drug Abuse Program, McLean Hospital, 115 Mill St, Belmont, MA 02178-9106.

*Copyright © 1995 by W.B. Saunders Company
0010-440X/95/3605-0004\$03.00/0*

value for substance abuse patients was .39 (range, .11 to .92 for individual diagnoses), indicating poor agreement overall. Indeed, this represented the lowest level of agreement among three patient sites.

We are aware of no long-term test-retest reliability studies of the SCID-II in any population, including patients with substance use disorders. In this article, we report results of a study of 12-month test-retest reliability of the SCID-II in a population of patients who were initially hospitalized because of cocaine dependence.

METHOD

Data were collected from 31 patients hospitalized for treatment of cocaine dependence at the Alcohol and Drug Abuse Treatment Center of McLean Hospital, Belmont, MA. Patients were recruited for this study as part of a larger investigation of predictors of outcome and the process of recovery and relapse in cocaine dependence. All patients provided written informed consent before entering the study.

Thirty-one patients for whom we were able to collect repeat SCID-II data at 12 months represented 49% of our original cohort of 63 patients who initially entered our larger follow-up study during this period. There were no significant differences between patients with and without 12-month repeat SCID-II evaluations on total number of in-hospital SCID-II diagnoses, sociodemographic characteristics (sex, race, age, or employment status), or cocaine use history (length, frequency, or amount).

Patients were identified as meeting DSM-III-R criteria for current cocaine dependence using the Structured Clinical Interview for DSM-III-R (SCID).¹³ Axis II diagnoses were initially obtained during the second week of hospitalization using the SCID-II,¹¹ which consists of two parts. The patient initially completes a self-report Personality Questionnaire consisting of 120 "yes-no" questions according to "the kind of person you generally are, that is, how you usually have felt or behaved over the past several years." The interviewer then conducts the SCID-II interview, reviewing affirmative responses on the Personality Questionnaire in more detail. For example, if a patient answers "yes" to a question such as "Do you often agree with people even when you think they are wrong?", the interviewer probes the answer to see if in fact this is a frequent rather than occasional occurrence.

During the year after admission, patients were evaluated with monthly administration of the Addiction Severity Index,¹⁴ urine toxicologic screens, and weekly substance use questionnaires. Moreover, a repeat Personality Questionnaire and SCID-II were administered at 12 months by the initial interviewer. In this second interview, the patient was asked to answer according to the time frame of the several years preceding entrance to the hospital 12 months earlier. SCID and SCID-II interviews were performed by a bachelor's-level research assistant (C.H.) who had several years of mental health experience and had previous training and

experience with the Schedule for Affective Disorders and Schizophrenia.¹⁵ She was trained in the use of SCID and SCID-II by the senior author (R.D.W.). After the interviewer had conducted three diagnostic assessments in the presence of the senior author in which independent diagnoses showed full agreement, she then conducted the interviews alone and reviewed the findings with the senior author.

Initial data analyses to determine concordance between baseline and 12-month retrospective axis II diagnoses were conducted using κ and percent agreement. κ represents the chance-corrected degree of agreement and ranges in value from -1 (representing complete disagreement), through 0 (chance level), to $+1$ (perfect agreement).¹⁶ Although standards for interpreting κ values vary somewhat, we used the methodology reported by First et al.,¹² since their study most closely resembled ours. κ values greater than .7 were thus considered good, with values between .5 and .7 rated as fair, and values less than .5 viewed as poor. Since κ values are likely to be unstable with very low and very high prevalence rates and since the prevalence of each axis II disorder was less than 30% in our sample (presented later), we also calculated percent agreement in our results. This statistic represents the number of cases with diagnostic concordance divided by the total number of cases. Moreover, since κ is equally reflective of agreement about presence or absence of diagnoses, we followed the recommendation of Cicchetti and Feinstein¹⁷ and further categorized agreement into "positive" and "negative" agreement; the former refers to agreement at both times that a diagnosis was present, and the latter indicates absence of a diagnosis at both times. These were then calculated separately as ratios in the following manner. For positive agreement, the numerator was the number of cases with a diagnosis at both times and the denominator was the combined number of cases with a diagnosis at admission, 12 months, or both. Negative agreement was calculated by dividing the number of cases with no diagnosis at both times by the combined number of cases with no diagnosis at admission, 12 months, or both.

RESULTS

The patient sample was predominantly white (94%) and employed (71%), with a slight preponderance of men (58%); the age (mean \pm SD) of the sample was 31.6 ± 6.3 years. They had used cocaine for 8.0 ± 5.1 years; most were intranasal users (61%), and the rest were cocaine smokers. The patients used 16 ± 36 g/wk and spent $\$5,846 \pm 7,436$ on cocaine during the previous 6 months.

At the 12-month assessment, all but two patients had been abstinent for at least the previous month: one had been drug-free for 2 weeks, and one was actively using cocaine (although he had no axis II diagnosis at either time).

Twenty-five of 31 patients (81%) had at least

one axis II disorder during hospitalization: eight patients (26%) had only one diagnosis, four (13%) had two, four (13%) had three, and nine (29%) had four or more. Twenty-nine (39%) of 75 personality disorders diagnosed during hospitalization met the minimum number of criteria for that disorder, and 46 (61%) exceeded the minimum.

Table 1 lists concordance rates between baseline and 12-month diagnoses. Our overall κ value for all axis II diagnoses was .46, with 81.8% agreement; κ values for individual diagnoses ranged from $-.15$ (avoidant) to $+.78$ (schizotypal). Percent agreement for individual disorders ranged from 51.7 (borderline) to 100 (schizoid).

Our results showed a discrepancy between percent agreement and κ values for some disorders, with the former measure appearing to show better concordance than the latter, e.g., with avoidant and obsessive-compulsive personality disorders. When we calculated positive and negative agreement separately (Table 1), negative agreement was higher than positive agreement for each individual diagnosis and for the sum total of all personality disorder diagnoses.

We wondered whether the low ratio of positive agreement for certain disorders could have been due to the unreliability of diagnosing a relatively small number of patients with a large

number of disorders. We thus calculated the ratio of positive agreement among patients who had three or more personality disorders during hospitalization, since poor test-retest reliability in this group could affect results of the entire sample (e.g., if several patients with multiple diagnoses during hospitalization had none at 12 months). However, we found that the total ratio of positive agreement for patients who had three or more personality disorders was .75 (compared with an overall ratio of positive agreement for the entire sample of .40). Moreover, the ratio of positive agreement in the multiple-disordered group was equal to or better than that found in the overall group for each individual diagnosis.

We collapsed 11 individual personality disorders into three clusters (A, B, and C) to see whether test-retest reliability for each cluster was better than that found for individual disorders. For this analysis, if a patient had a diagnosis from within a particular cluster during hospitalization and another diagnosis from within the same cluster at 12 months, we rated this as positive agreement. Test-retest reliability for personality disorder clusters was similar to that found for individual diagnoses: clusters A, B, and C had κ values of $+.47$, $+.27$, and $+.42$, respectively. The ratio of negative agreement exceeded that of positive agreement for clusters

Table 1. Concordance of Admission SCID-II Diagnoses and 12-Month Retrospective SCID-II Diagnoses in Cocaine-Dependent Patients (N = 31)

Disorder	Present at Admission and 12 Months	Absent at Admission and 12 Months	Present at Admission Only	Present at 12 Months Only	κ	Percent Agreement†	Ratio of Negative Agreement‡	Ratio of Positive Agreement§
Avoidant	0	22	6	3	$-.15$	71.0	.71	.00
Dependent	2	25	3	1	$+.43$	87.1	.86	.33
Obsessive-compulsive	1	26	2	2	$+.26$	87.1	.87	.20
Passive-aggressive	5	22	2	2	$+.71$	90.0	.85	.56
Paranoid	6	18	3	4	$+.47$	77.5	.72	.46
Schizotypal	2	28	1	0	$+.78$	96.8	.97	.67
Schizoid	0	31	0	0	*	100.0	1.00	*
Histrionic	9	16	4	2	$+.59$	80.6	.73	.60
Narcissistic	4	23	1	3	$+.59$	87.1	.85	.50
Borderline	6	10	8	7	$+.02$	51.7	.40	.29
Antisocial	6	17	4	4	$+.41$	74.2	.68	.43
Total	41	238	34	28	$+.45$	81.8	.79	.40

*Cannot be computed when prevalence of a disorder is 0.

†Number of cases with diagnostic agreement divided by total number of cases.

‡Ratio of cases with no diagnosis at both times divided by cases with no diagnosis at admission, 12 months, or both.

§Ratio of cases with a positive diagnosis at both times divided by cases with a diagnosis at admission, 12 months, or both.

||Not a conventional κ , since it refers to number of diagnoses rather than people.

A (.72 *v.* .46) and C (.59 *v.* .50), although positive agreement was higher in cluster B (.62 *v.* .33), perhaps reflecting overlap among these diagnoses.

DISCUSSION

In this study of 12-month test-retest reliability of the SCID-II in cocaine-dependent patients, we found that positive agreement (i.e., agreement at both times that a particular diagnosis was present) was relatively poor, whereas negative agreement (i.e., agreement at the two times regarding the absence of a diagnosis) was good. This discrepancy between positive and negative agreement, combined with a relatively low prevalence rate for many of the personality disorder diagnoses in this patient population, led to high percent agreement, and to less impressive agreement when measured by κ . Better reliability of the statistic measuring percent agreement is to be expected, since it does not correct for chance; indeed, this is the reason for using κ . However, our findings point out the importance of separating positive agreement and negative agreement in reliability studies of this nature, particularly when relatively uncommon disorders are being studied. Moreover, our data, in combination with those reported by First et al.,¹² point to the difficulty of attaining test-retest reliability for axis II diagnoses in patients with substance use disorders.

There are several potential explanations for the latter finding. First, reliably diagnosing personality disorders is an inherently difficult process. However, recent development and use of standardized instruments to diagnose axis II disorders has clearly improved reliability of these diagnoses,⁹ so this reason is likely insufficient to explain our findings. Similarly, the SCID-II itself does not appear to have a specific design flaw that makes test-retest reliability particularly difficult to achieve, since joint-interview interrater reliability studies of the SCID-II have shown reliability as good as that reported with other standardized instruments.⁹

Another possible explanation relates to the length of time between interviews; in the study of test-retest reliability with the SCID-II that has thus far yielded the best results, the second interview occurred within 2 days of the first.⁴ Indeed, test-retest reliability in all personality

disorder assessments has been shown to decline as the interval between the two interviews increases.⁹ This may be due to poor memory (particularly in a substance-dependent population) or to a change over time in clinical status. Results found in our study are thus similar to those found in other longer-term (i.e., several weeks to 6 months) test-retest reliability studies of personality disorder diagnoses in patients with different coexisting axis I disorders.⁹ A change over time in the status of one's axis I disorder may affect informant bias, which may in turn alter patients' responses to questions about their personality traits.¹ As our group has discussed in previous reports,^{18,19} patients who enter treatment for substance dependence may possess certain attitudes and biases about the relationship between their personality traits and substance use; these attitudes may change substantially during the ensuing year, particularly in patients (like our sample) who do well in treatment. Indeed, certain aspects of the treatment of substance use disorders are designed to encourage patients to reconsider their beliefs about the relationship between substance use and behavior. In 12-step treatment, for example, patients may be encouraged to view some of their manipulative behaviors as part of their "disease," rather than as independent negative personality traits. Alternatively, they may be advised to acknowledge their "character defects" as a necessary part of recovery. Patients who are exposed to a year of this type of treatment and respond favorably may thus view their previous behavior differently at a reinterview than they had 12 months earlier. Depending on the specific nature of a patient's treatment, he or she could view previous behavior as either more or less indicative of an independent personality disorder.

It is unclear whether patients who continued to use cocaine regularly over time would have shown the same results our patients did. By virtue of our study design, we oversampled good-outcome patients, who were readily located and willing to return for follow-up interviews. Undersampling poor-outcome patients, who may have had more persistently disordered personalities (or who at least may have viewed themselves this way), could have negatively affected our reliability estimates. Similarly, axis

II state-trait artifact could have influenced our results, since patients' current personality traits may influence reporting of previous symptoms.²⁰

There were several other factors specific to our study that also may have affected our results. First, our small sample size limits generalizability of the study and suggests the need for further study of this subject in larger samples. Second, we used a bachelor's-level interviewer. Although she had mental health experience, had previously administered other structured diagnostic interviews, had established joint-interviewer reliability with the senior author for in-hospital assessments, and had reviewed all of her findings with the senior author, her lack of formal psychopathology training may have contributed to the unreliability of the assessments. However, our results were comparable to those reported by First et al.,¹² in which interviewers were all experienced mental health professionals with at least a master's degree. A third methodologic limitation was the fact that we used the same interviewer at both times.⁹ However, this should have biased the results toward higher reliability, not lower, since we did not have lack of interrater agreement as a potential source of unreliability. Finally, extreme variability in the amount of cocaine used by our patient population (as indicated by large standard deviations) suggests that it might be interesting, in a larger sample, to study the relationship between

amount of use and diagnostic reliability. Although these data are not sufficient to comment on this issue, one could hypothesize that heavier users either may have poorer memories or may have a more difficult time reliably distinguishing between drug-related behaviors and enduring personality traits.

Diagnosing axis I psychiatric disorders in patients with substance use disorders is hampered by some of the same methodologic difficulties that we have described earlier for personality disorders. For example, Rounsaville and Kleber,¹ in a study using the Schedule for Affective Disorders and Schizophrenia,¹⁵ found that diagnostic concordance for lifetime axis I disorders in opiate addicts was very good or better at a conjoint interview, moderate to acceptable at 6 months, and poor at 2.5 years. Hasin²¹ has recently tried to address some of the methodologic difficulties inherent in this diagnostic process by designing a structured interview (the Psychiatric Research Interview for Substance and Mental Disorders, formerly called the SCID for Alcohol and Drugs) that examines the relationship between substance use and psychiatric symptoms in a standardized and detailed manner. However, no analogous approach has been attempted with axis II disorders. Such an approach may be fruitful in helping to improve reliability (and thus clinical utility) of axis II diagnoses in patients with substance use disorders.

REFERENCES

1. Rounsaville B, Kleber H. Psychiatric disorders and the course of opiate addiction: preliminary findings on predictive significance and diagnostic stability. In: Mirin S (ed): *Substance Abuse and Psychopathology*. Washington, DC: American Psychiatric, 1984:133-151.
2. Stark M, Campbell B. Personality, drug use, and early attrition from substance abuse treatment. *Am J Drug Alcohol Abuse* 1988;14:475-485.
3. Nace EP, Davis CW, Gaspari JP. Axis II comorbidity in substance abusers. *Am J Psychiatry* 1991;148:118-120.
4. Malow RM, West JA, Williams JL, Sutker PB. Personality disorders classification and symptoms in cocaine and opioid addicts. *J Consult Clin Psychol* 1989;57:765-767.
5. Bunt G, Galanter M, Lifshutz H, Castaneda R. Cocaine/"crack" dependence among psychiatric inpatients. *Am J Psychiatry* 1990;147:1542-1546.
6. Kleinman PH, Miller AB, Millman RD, Woody GE, Todd T, Kemp J, et al. Psychopathology among cocaine abusers entering treatment. *J Nerv Ment Dis* 1990;178:442-447.
7. Pettinati HM, Cabezas JJ, Evans BD, Meyers K. Incidence of personality disorders in cocaine vs. alcohol-dependent females. *Natl Inst Drug Abuse Res Monogr Ser* 1990;105:369-370.
8. Weiss RD, Mirin SM, Griffin ML, Gunderson JG, Hufford C. Personality disorders in cocaine dependence. *Compr Psychiatry* 1993;34:145-149.
9. Zimmerman M. Diagnosing personality disorders. *Arch Gen Psychiatry* 1994;51:225-245.
10. Loranger AW, Sartorius N, Andreoli A, Berger P, Buchheim P, Channabasavanna SM, et al. The International Personality Disorder Examination. *Arch Gen Psychiatry* 1994;51:215-224.
11. Spitzer RL, Williams JBW, Gibbon M, First MB. Structured Clinical Interview for DSM-III-R Personality Disorders (SCID-II). Interview and User's Guide. Washington, DC: American Psychiatric, 1990.
12. First MB, Spitzer RL, Gibbon M, Williams JBW, Davies M, Borus J, et al. The Structured Clinical Interview

for DSM-III-R Personality Disorders (SCID-II). II. Multi-site test-retest reliability study. *J Pers Disord*. In press.

13. Spitzer RL, Williams JBW, Gibbons M. Structured Clinical Interview for DSM-III-R. New York, NY: Biometrics Research Department, New York State Psychiatric Institute, 1987.

14. McLellan AT, Luborsky L, O'Brien CP, Woody GE. An improved diagnostic evaluation instrument for substance abuse patients: the Addiction Severity Index. *J Nerv Ment Dis* 1980;168:26-33.

15. Endicott J, Spitzer RL. A diagnostic interview: the Schedule for Affective Disorders and Schizophrenia. *Arch Gen Psychiatry* 1980;37:837-844.

16. Cohen JA. A coefficient of agreement for nominal scales. *Educ Psychol Measure* 1960;20:37-46.

17. Cicchetti DV, Feinstein AR. High agreement but low

kappa. II. Resolving the paradoxes. *J Clin Epidemiol* 1990;43:551-558.

18. Griffin ML, Weiss RD, Mirin SM, Wilson H, Bouchard-Voelk B. The use of the Diagnostic Interview Schedule in drug dependent patients. *Am J Drug Alcohol Abuse* 1987;13:281-291.

19. Weiss RD, Mirin SM, Griffin ML. Methodological considerations in the diagnosis of coexisting psychiatric disorders in substance abusers. *Br J Addict* 1992;87:179-187.

20. Rice JP, Rochberg N, Endicott J, Lavori PW, Miller C. Stability of psychiatric diagnoses: an application to the affective disorders. *Arch Gen Psychiatry* 1992;49:824-830.

21. Hasin D. Diagnostic interviews and assessment of alcohol use disorders: background, reliability, and validity. *Alcohol Health Res World* 1991;4:293-302.