Borderline Personality Disorder: Interrater Reliability of the Diagnostic Interview for Borderlines

by Jerome Kroll, Richard Pyle, Janet Zander, Kenneth Martin, Steven Lari, and Lloyd Sines

Abstract

Three interviewers, working in pairs and using the Diagnostic Interview for Borderlines (DIB), interviewed three samples of 10 adult psychiatric inpatients. The two members of each interviewer pair alternated in the role of interviewer. Each member of the interviewer pairs completed the interview schedule and scored the protocol according to the instructions provided by the authors of the DIB. Interrater reliability was determined for each interviewer pair using the Kappa statistic. Reliability values were at about the level reported recently by Spitzer, Forman, and Nee (1979) for DSM-III axis I diagnoses in general. Additional statistical analyses were performed to determine whether there were patient sex differences on DIB scores, whether the two male interviewers differed from the female interviewer in their ratings of male or female patients, and whether experience differences among the three interviewers were reflected in their rating of patients. Among the latter analyses, the only significant finding was that in the interviewer pair consisting of a very experienced male psychiatrist and a male first year psychology graduate student, the former rated male patients as less borderline than did the latter. No other difference was found among raters, or between sex of patient and interviewer sex or amount of clinical experience. It was concluded that the DIB is sufficiently reliable for use in clinical research.

Several recent articles on borderline conditions have pointed out that more data and fewer reviews and anecdotal reports are necessary at this time (Perry and Klerman 1978; Liebowitz 1979; Spitzer, Endicott, and Gibbon 1979; Perry and Klerman 1980). While this is undoubtedly true, it should be kept in mind that the reviews and reports remain equally necessary because the conceptual and methodological bases around which the data are gathered are themselves unclear. What is clear is that a sizable number of patients appear to share certain pathological traits and behaviors, but nevertheless defy easy categorization; they exist as individuals, but it is a puzzle as to whether the various nosological constructs by which we group these individuals accurately reflect either specific disease entities (including variations thereof) or even specific syndromes. Thus, both conceptual clarification and empirical studies must proceed in parallel.

Gunderson and Kolb (1976, 1978) have developed the most specific methodology to date for the diagnosis of borderline conditions. Items included in their semi-structured interview were initially derived from clinical and research reports and personal clinical experience, and were subsequently tested for reliability and concurrent validity with hospitalized young adults, ages 16 to 35. Although demonstrating reliability of such an instrument does not necessarily establish the instrument’s validity, it is a necessary first step in demonstrating the feasibility of identifying that subgroup of patients for more inten-
sive study. Gunderson and Kolb (1976) have reported interrater reliability data for their Diagnostic Interview for Borderlines (DIB); they found a mean intraclass r coefficient of .77 from the five sections (Social Adaptation, Impulse/Action, Affect, Psychosis, and Interpersonal Relations) of the DIB, and r's of .80 to .95 for total Scaled Scores, which are the basis of classification of patients as borderline or not borderline. However, the latter reliability coefficients were computed on samples of only four patients each.

The present article deals with interrater reliability of the DIB. It takes into account (1) sex of patient, (2) sex of interviewer, and (3) differences between interviewers in amount of clinical training and experience.

Methods

Interviewers. Three interviewers, working in pairs, examined three samples of 10 adult psychiatric inpatients. The three interviewers were selected to provide a range of interviewer experience, gender, and profession. Interviewer A was a male attending psychiatrist with 15 years' postresidency clinical experience in a variety of psychiatric settings. Interviewer B was a female chief resident psychiatrist in her third year of psychiatric training. Interviewer C was a male first-year graduate student in a doctoral program in clinical psychology. The three interviewers had no clinical or teaching assignments to the inpatient ward and had no knowledge of the patients they interviewed, other than the fact that they were hospitalized patients. No background information about the patients was provided to the interviewers. The DIB scores reflect solely the information developed during the DIB interviews. The interviewers were not trained by Gunderson but had access to his DIB procedure manual. The interviewers had no experience in the use of the DIB before the study.

Patients. The three samples of 10 patients each represented consecutively admitted patients on a 24-bed closed psychiatric ward in a 600-bed general hospital. The only patients excluded from the study were several patients with dementia or chronic schizophrenia who were not amenable to interviewing. Of the 30 patients interviewed, 14 were male and 16 were female. The age range for male patients was from 18 to 66, with a mean of 30.3, while in the female sample, the age range was from 26 to 58, with a mean of 37.4. The DSM-III diagnoses of the patients interviewed were as follows: Major Depressive Disorder—8; Bipolar Affective Disorder—5; Schizophrenia—4; Mixed Personality Disorder—4; Borderline Personality Disorder—3; Antisocial Personality Disorder—1; Alcohol Abuse—2; Bulimia—1; Intermittent Explosive Disorder—1; Hallucinogen Delusional Disorder—1. In a larger study of 117 consecutively admitted patients on the same ward, 15 percent of males and 21 percent of females were DIB-positive (DIB scores≥7) borderline; DSM-III diagnoses of borderline accounted for 4 percent of males and 13 percent of females (Kroll et al., in press).

Procedure. Interviewers A and B examined the first series of 10 patients, interviewers A and C a second sample, and interviewers B and C the final sample of 10 patients. Thus, each interviewer was involved in rating 20 patients. Within each sample of 10 patients, the two raters alternated in the role of the interviewer.

Results

Figure 1 shows the distribution of DIB Scaled Scores obtained in the present sample. The range of scores from male patients was from 1 to 9, with a mean of 4.9 and a standard deviation of 2.39, while the range of scores for female patients was from 2 to 8, with a mean of 4.7 and a standard deviation of 1.86. The observed difference between mean scores for male and female patients (.20) is not statistically reliable (t = 1.13, p > .20). According to Kolb and Gunderson (1980), a scaled score of 7 or higher (possible range 0–10) indicates that the patient is borderline. Four male and two female patients (20 percent of the total sample) were rated by both interviewers as borderline, while three additional female patients (10 percent of the sample) were rated as borderline by one (but not both) of the interviewers. Thus, 31 percent of the female patients and 29 percent of the male patients were identified as borderline by at least one of the interviewers. These incidence figures are similar to those reported by Gunderson and Kolb (1978) in a sample of hospitalized young adults.

Reliability. The data in table 1 indicate the Kappa coefficients for evaluating agreement between interviewers on the diagnosis of borderline or not borderline (a cutting score of 7 indicated border-
Figure 1. DIB scaled scores of 14 male and 16 female adult psychiatric inpatients

*2 ratings per patient

DIB Scaled Scores

Interviewer Differences. Individual interviewer Scaled Score means and standard deviations, by sex of patient, are shown in table 2. When each interviewer pair was considered separately, two-way analyses of variance of the data in table 2 failed to reveal significant differences between the means of male and female patients, or between the female interviewer and either of the two male interviewers. In interviewer pair AC, however, the interaction of interviewer with sex of patient was significant. Specifically, the male psychiatrist with 15 years of clinical experience rated male patients as less borderline, in terms of mean Scaled Scores, than did the male first-year graduate student in clinical psychology. The latter’s mean ratings for male and female patients were similar in magnitude to the former’s mean for female patients.

Discussion

The Diagnostic Interview for Borderlines (DIB) appears to be a reliable tool in the search for and study of psychiatric patients who manifest borderline personality characteristics. Its reliability appears to extend over a somewhat wider age range than originally studied by Gunderson and Kolb (1976). Beyond the narrow question of psychometric reliability, however, the present study addresses the question of possible sex differences among psychiatric patients in the extent to which they manifest characteristics of the borderline disorder—a phenomenon that has been reported in related conditions (Warner 1978; Kroll, Chamberlain, and Halpern 1979; Spalt 1980). The data failed to identify a sex difference in mean interview ratings, but the noticeably (though not statistically reliably) larger standard deviation in the male sample is of interest and may call for further study. Finally, while there was no difference in the ratings of male and female patients as a function of the sex of the interviewer, one significant disagreement between members of one interviewer pair was detected. Both interviewers in that pair were male, one a highly experienced psychiatrist and the other a clinical psychology student at the beginning of his professional training. Whether the observed difference represents a chance occurrence, or implicates a discipline and/or an experience factor as leading to disagreement, is unclear. However, the finding that the inexperienced
Table 2. DIB scaled score means and standard deviations by interviewer and sex of patient

<table>
<thead>
<tr>
<th>Interviewer pair (n = 10 patients per pair)</th>
<th>Male patients (n = 14)</th>
<th>Female patients (n = 16)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Mean</td>
</tr>
<tr>
<td>AB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interviewer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>5</td>
<td>4.20</td>
</tr>
<tr>
<td>B2</td>
<td>5</td>
<td>4.80</td>
</tr>
<tr>
<td>AC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interviewer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A3</td>
<td>6</td>
<td>3.33</td>
</tr>
<tr>
<td>C3</td>
<td>6</td>
<td>4.67</td>
</tr>
<tr>
<td>BC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interviewer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td>5.83</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>5.50</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>4.89</td>
</tr>
</tbody>
</table>

1 Male Board-certified psychiatrist.
2 Female third year psychiatric resident.
3 Male first year psychology graduate student.

psychology student and the advanced psychiatric resident showed good levels of agreement, as did the two psychiatrists with different levels of experience, somewhat confounds the issue, and more systematic study of these questions will be required.

References


Spitzer, R.L.; Endicott, J.; and Gibbon, M. Crossing the border into borderline personality and borderline schizophrenia. *Archives of General Psychiatry*, 36:17–24, 1979.


The Authors

Jerome Kroll, M.D., Richard Pyle, M.D., Janet Zander, M.D., Steven Lari, M.D., and Lloyd Sines, Ph.D., are at the Department of Psychiatry, University of Minnesota Medical School, and Kenneth Martin is at the Department of Psychology, University of Minnesota, Minneapolis, MN.