

Winter April, 2000

# The Collaborative Longitudinal Personality Disorders Study: Reliability of Axis I and II Diagnoses.

Mary C. Zanarini  
Andrew E. Skodol  
Donna S. Bender  
Regina T. Dolan  
Charles A. Sanislow, et al.

## **THE COLLABORATIVE LONGITUDINAL PERSONALITY DISORDERS STUDY: RELIABILITY OF AXIS I AND II DIAGNOSES**

Mary C. Zanarini, EdD, Andrew E. Skodol, MD,  
Donna Bender, PhD, Regina Dolan, PhD,  
Charles Sanislow, PhD, Elizabeth Schaefer, EdM,  
Leslie C. Morey, PhD, Carlos M. Grilo, PhD,  
M. Tracie Shea, PhD, Thomas H. McGlashan, MD,  
and John G. Gunderson, MD

Both the interrater and test-retest reliability of axis I and axis II disorders were assessed using the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) and the Diagnostic Interview for DSM-IV Personality Disorders (DIPD-IV). Fair-good median interrater  $\kappa$  (.40–.75) were found for all axis II disorders diagnosed five times or more, except antisocial personality disorder (1.0). All of the test-retest  $\kappa$  for axis II disorders, except for narcissistic personality disorder (1.0) and paranoid personality disorder (.39), were also found to be fair-good. Interrater and test-retest dimensional reliability figures for axis II were generally higher than those for their categorical counterparts; most were in the excellent range (>.75). In terms of axis I, excellent median interrater  $\kappa$  were found for six of the 10 disorders diagnosed five times or more, whereas fair-good median interrater  $\kappa$  were found for the other four axis I disorders. In general, test-retest reliability figures for axis I disorders were somewhat lower than the interrater reliability figures. Three test-retest  $\kappa$  were in the excellent range, six were in the fair-good range, and one (for dysthymia) was in the poor range (.35). Taken together, the results of this study suggest that both axis I and axis II disorders can be diagnosed reliably when using appropriate semistructured interviews.

---

From the Collaborative Longitudinal Personality Disorders Study: Brown University Department of Psychiatry and Human Behavior (Dolan, Shea), Columbia College of Physicians and Surgeons and New York State Psychiatric Institute (Bender, Skodol), Harvard Medical School and McLean Hospital (Gunderson, Schaeffer, Zanarini), Texas A & M University (Morey), and Yale University School of Medicine and Yale-New Haven Psychiatric Hospital (Grilo, McGlashan, Sanislow).

This study was supported by grants from the National Institute of Mental Health (R10 MH 50837, 50838, 50839, 50840, and 50850).

This publication has been reviewed and approved by the Publications Committee of the Collaborative Longitudinal Personality Disorders Study.

Address correspondence to Dr. Mary C. Zanarini, McLean Hospital, 115 Mill Street, Belmont, MA 02478; E-mail: zanarini@mclean.harvard.edu.

They also suggest that the reliability of axis II disorders is roughly equivalent to that reliability found for most axis I disorders.

Reliability is a key element to any study of diagnostic differentiation and stability because it sets upper limits on measures of validity, including diagnostic stability. Interrater reliability is a test of whether different raters process and score the same patient material in a similar manner. Test-retest reliability is more complicated and depends on consistency of patient self-report and interviewer differences in eliciting, understanding, and scoring clinical material.

Kappa ( $\kappa$ ) values have become the standard measure of reliability in psychiatry because they correct for chance agreements. According to Fleiss (1981), "for most purposes, values greater than .75 or so may be taken to represent excellent agreement beyond chance, values below .40 or so may be taken to represent poor agreement beyond chance, and values between .40 and .75 may be taken to represent fair to good agreement beyond chance." Using these figures as guidelines and following data described in a careful review article by Zimmerman (1994), most studies of the interrater reliability of the major axis II interviews designed to assess DSM-III/DSM-III-R personality disorders have found  $\kappa$  values in the good-excellent range (.62-.77), with an unweighted mean  $\kappa$  coefficient for any personality disorder being .75. Short interval (between one week and 10 days) test-retest reliability figures for these same instruments has been found to be somewhat lower. More specifically, most test-retest  $\kappa$  values were found to be in the fair-good range (.45-.69), with an unweighted mean  $\kappa$  coefficient for any personality disorder being .56. Only for antisocial personality disorder has an unweighted mean test-retest  $\kappa$  value in the excellent range (.77) been achieved.

Despite the acceptable levels of reliability that have been achieved in the studies reported upon above, controversies have continued in the area of the assessment of axis II pathology. The most relevant continuing controversy pertains to whether any of the several available interviews has superior psychometric properties to the others. There are no clear cut answers to this question, with all of the five main semistructured interviews (Structured Interview for the DSM Personality Disorders or SIDP [Stangl, Pfohl, Zimmerman, Bowers, & Corenthal, 1985], Personality Disorder Examination or PDE [Loranger, Susman, Oldham, & Russakoff, 1987], Diagnostic Interview for Personality Disorders or DIPD [Zanarini, Frankenburg, Chauncey, & Gunderson, 1987], Structured Clinical Interview for DSM Personality Disorders or SCID-II [First et al., 1995], and Personality Disorder Interview or PDI [Widiger, Trull, Hurt, Clarkin, & Frances, 1987]) attaining adequate levels of reliability, particularly if used by their developers (Zimmerman, 1994).

This controversy is not as salient in the realm of axis I assessment. Both the Diagnostic Interview Schedule (DIS) (Robins, Helzer, Croughan, & Ratcliff, 1981) and the Composite International Diagnostic Interview (CIDI) (Andrews & Peters, 1998) are reliable structured instruments designed to be used by lay interviewers. However, the Structured Clinical Interview for Axis I Disorders (SCID-I) developed by Spitzer and his colleagues in the

early 1980s (Spitzer, Williams, Gibbon, & First, 1984) has become the standard for assessing the presence of the major axis I disorders by clinically experienced raters.

Both the interrater and test-retest reliability of the DSM-III-R version of the SCID-I have been found to be adequate. In an interrater reliability study involving 54 patients rated by three interviewers (Skre, Onstad, Torgersen, & Kringlen, 1991), all axis I disorders diagnosed at least five times obtained a  $\kappa$  of .80 or higher, except simple phobia (.70) and social phobia (.72). In a multisite test-retest reliability study (Williams et al., 1992), lifetime diagnoses of bipolar disorder, drug abuse and dependence, anorexia, and bulimia attained mean weighted  $\kappa$  in the excellent range (>.75). The  $\kappa$  attained for the remaining mood disorders and substance use disorders in addition to the psychotic and anxiety disorders were in the fair-good range (.48-.73).

The current study is part of a larger study of the longitudinal course of four axis II disorders (avoidant personality disorder, borderline personality disorder, obsessive compulsive personality disorder, and schizotypal personality disorder) and a comparison group with major depression and no notable axis II pathology. The rationale, sample, and methodology of the Collaborative Longitudinal Personality Disorders Study (CLPS) have been described elsewhere (Gunderson et al., 2000).

The present study assesses the interrater and test-retest reliability of DSM-IV axis I and II disorders using the DSM-IV version of the SCID-I (First, Spitzer, Gibbon, Williams, 1996) and the Diagnostic Interview for DSM-IV Personality Disorders (DIPD-IV) (Zanarini, Frankenburg, Sickel, & Yong, 1996). We chose to use the DIPD-IV because of the relatively high interrater and test-retest reliability of its parent interview, its relative ease of administration, and the significant advantage of having its developer oversee reliability training. We also selected the DIPD-IV because it could easily be modularized for follow-along assessments.

Table 1 compares the reliability figures found for the original versions of the major axis II interviews in studies conducted by their developers. As can be seen, the interrater reliability of the DSM-III version of the DIPD compares favorably with that achieved by the PDE, PDI, and SIDP. The short-term test-retest reliability of the DSM-III version of the DIPD compares favorably with that achieved by the DSM-III version of the SIDP in addition to the DSM-III-R version of the SCID-II.

## METHOD

A total of 12 master's or doctoral level raters participated in this study: three from each of the overall study's four patient recruitment sites (Brown, Columbia, Harvard, and Yale). Each rater had been trained in the use of the SCID-I and the DIPD-IV by one of us who is also the developer of the DIPD-IV (MCZ). Each rater first participated in a week long centralized training session and then returned to his or her home site to complete training. Each rater observed a minimum of three interviews and was observed conducting a minimum of three interviews. Training was considered complete when the principal investigator for each site, after observing

**TABLE 1. Interrater and Test-Retest Reliability of the Original Version of Major Axis II Interviews as Assessed by Developer and Colleagues**

Instrument	Sample	Antisocial	Avoidant	Borderline	Compulsive	Interrater Reliability						Schizoid	Schizotypal
						Dependent	Histrionic	Narcissistic	Paranoid	Passive Ag			
DIPD*	43 inpts	1.0	1.0	.94	.92	.92	.87	.87	.52	.95	—	—	.86
PDE**	60 inpts	.70	—	.96	.88	—	.77	—	—	—	—	—	.80
PDI†	67 inpts	.83	.69	.75	.67	.52	.71	.77	.71	.52	.71	.71	.64
SIDP†	43 inpts	—	.45	.95	.36	.88	.84	.85	—	—	—	—	.55
Short Interval Test-Retest Reliability (1-2 weeks)													
DIPD*	54 inpts	.84	.67	.85	.56	.77	.68	.78	.54	.46	—	—	.61
SCID-II <sup>‡</sup>	103 pts	.71	.71	.40	.24	.42	.74	.43	.68	.59	—	—	—
SIDP†	20 inpts	—	—	.58	—	—	.43	—	—	—	—	—	—

\*Zanarini et al., 1987; \*\*Loranger et al., 1987; †Widiger et al., 1985; ‡Stangl et al., 1995. All κ apply to DSM-III diagnoses, except for SCID-II test-retest reliability figures. All κ also pertain to diagnoses made five times or more.

a full axis I and II evaluation, certified that the rater was sufficiently prepared to begin to interview subjects independently.

Interrater reliability was assessed using 27 videotapes and median  $\kappa$  are presented for 84 rater pairs. Test-retest reliability was assessed using two direct interviews of 52 subjects. Approximately equal numbers of interviews were conducted at each site (Brown = 13, Columbia = 17, Harvard = 10, and Yale = 12) and as the reliability figures were very similar at all of the sites, only the overall test-retest results are presented below. The first interview was conducted as part of the study's intake procedure and the second interview, which was conducted by a separate blind interviewer, occurred seven to 10 days after the initial diagnostic assessment.

Subjects were recruited for the reliability studies at each site by the project coordinator for that site. Subjects were selected more for availability than prototypicality and thus, our reliability subjects represent a sample of convenience more than one of design. Dimensional reliability figures were also calculated for each axis II disorder (i.e., the number of criteria rated as present and of clinical significance for each disorder were summed for each interview and then compared with one another).

The  $\kappa$  statistic was used to calculate both the interrater and test-retest reliability figures for axis I diagnoses and categorical axis II diagnoses.  $\kappa$  were only computed for disorders diagnosed five times or more. This cutoff was selected to ensure that our results would be comparable with those of other research groups who used this cutoff as avoid the highly unstable results that can occur when lower cutoffs are used. Because not every interrater reliability tape was viewed by all 12 raters, median pairwise comparisons, which were computed separately for each disorder by taking the fiftieth percentile  $\kappa$  for all pairs of raters, were used rather than mean  $\kappa$ . For the same reason, the Pearson  $r$  statistic was used to calculate both the interrater and test-retest dimensional reliability figures rather than intraclass correlational techniques.

## RESULTS

Table 2 details the reliability results for categorical axis II diagnoses. As can be seen, fair-good median interrater  $\kappa$  were found for all disorders but antisocial. The median  $\kappa$  for antisocial personality disorder was 1.0. All of the test-retest  $\kappa$ , except for those for narcissistic personality disorder and paranoid personality disorder, were also in the fair-good range. Although the  $\kappa$  for narcissistic personality disorder (1.0) was extremely high, the  $\kappa$  for paranoid personality disorder (.39) was in the poor range as outlined by Fleiss. Interestingly, the test-retest  $\kappa$  for three of our four target disorders (avoidant, borderline, and obsessive compulsive personality disorder) were marginally higher than the interrater  $\kappa$  for these disorders.

Dimensional reliability figures for axis II are detailed in Table 3. As can be seen, they were generally found to be substantially higher than their categorical counterparts. All of the median Pearson  $r_s$  for interrater reliability were in the excellent range except for that for schizoid personality disorder (.69). Seven test-retest reliability figures were also in the excellent range. However, the Pearson  $r_s$  for all three of the odd cluster dimensions

**TABLE 2. Interrater and Test-retest Reliability of Categorical Axis II Diagnoses\***

Axis II Disorder	Median Interrater $\kappa$ ( <i>N</i> = 84 Pairs of Raters)	Test-Retest $\kappa$ ( <i>N</i> = 52 Cases)
Schizotypal personality	— ( <i>N</i> = 3)	.64 ( <i>N</i> = 5)
Schizoid personality	— ( <i>N</i> = 2)	— ( <i>N</i> = 2)
Paranoid personality	.58 ( <i>N</i> = 7)	.39 ( <i>N</i> = 6)
Avoidant personality	.68 ( <i>N</i> = 15)	.73 ( <i>N</i> = 26)
Dependent personality	— ( <i>N</i> = 3)	— ( <i>N</i> = 1)
Obsessive Compulsive personality	.71 ( <i>N</i> = 18)	.74 ( <i>N</i> = 15)
Depressive personality	.67 ( <i>N</i> = 12)	.62 ( <i>N</i> = 14)
Passive Aggressive personality	.64 ( <i>N</i> = 7)	.51 ( <i>N</i> = 7)
Borderline personality	.68 ( <i>N</i> = 13)	.69 ( <i>N</i> = 25)
Histrionic personality	— ( <i>N</i> = 3)	— ( <i>N</i> = 2)
Narcissistic personality	— ( <i>N</i> = 4)	1.0 ( <i>N</i> = 5)
Antisocial personality	1.00 ( <i>N</i> = 8)	.70 ( <i>N</i> = 9)

\* $\kappa$  were only calculated for disorders diagnosed five times or more.

and for depressive and histrionic personality traits or features were in the fair-good range (.52–.72).

Table 4 details the  $\kappa$  obtained for axis I disorders. As can be seen, excellent median interrater  $\kappa$  were found for six of the 10 disorders diagnosed five times or more, whereas fair-good median interrater  $\kappa$  were found for the other four axis I disorders. In general, the test-retest reliability figures for axis I disorders were somewhat lower than the interrater

**TABLE 3. Interrater and Test-retest Reliability of Dimensional Axis II Diagnoses**

Axis II Disorder	Median Pearson $r_s$ ( <i>N</i> = 84 Pairs of Raters)	Test-Retest Pearson $r_s$ ( <i>N</i> = 52 Cases)
Schizotypal personality	.91	.65
Schizoid personality	.69	.52
Paranoid personality	.86	.71
Avoidant personality	.79	.83
Dependent personality	.87	.78
Obsessive Compulsive personality	.85	.82
Depressive personality	.92	.72
Passive Aggressive personality	.93	.76
Borderline personality	.90	.84
Histrionic personality	.83	.65
Narcissistic personality	.88	.82
Antisocial personality	.97	.92

**TABLE 4. Interrater and Test-Retest Reliability of Axis I Diagnoses\***

Axis I Disorder	Median Interrater $\kappa$ (N = 84 Pairs of Raters)	Test-Retest $\kappa$ (N = 52 Cases)
Major Depression	.80 (N = 19)	.61 (N = 38)
Dysthymia	.76 (N = 5)	.35 (N = 7)
Alcohol Abuse/Dependence	1.0 (N = 13)	.77 (N = 25)
Drug Abuse/Dependence	1.0 (N = 18)	.76 (N = 22)
Panic Disorder	.65 (N = 5)	.65 (N = 16)
Social Phobia	.63 (N = 5)	.59 (N = 14)
Obsessive-Compulsive disorder	.57 (N = 5)	.60 (N = 9)
Generalized Anxiety disorder	.63 (N = 5)	.44 (N = 14)
Post-Traumatic Stress disorder	.88 (N = 8)	.78 (N = 17)
Any eating disorder	.77 (N = 5)	.64 (N = 15)

\* $\kappa$  were only calculated for disorders diagnosed five times or more.

reliability figures. The test-retest  $\kappa$  for three of the disorders were in the excellent range. Of the  $\kappa$  for the remaining seven disorders, six were in the fair-good range, whereas one for dysthymia was in the poor range (.35).

## DISCUSSION

Three main findings have emerged from this study. The first finding is that the levels of both interrater and test-retest reliability that were found for categorical axis II disorders are quite good. Almost all of these  $\kappa$  were in the fair-good range. This is consistent with the results of numerous single site DSM-III and DSM-III-R studies reviewed by Zimmerman (1994). These results are also consistent with the results of a multisite study of the test-retest reliability of the DSM-III-R version of the SCID-II (First et al., 1995). In addition, they are consistent with the results of a single site study that used the SCID-II to assess both the interrater and test-retest reliability of axis II disorders according to DSM-IV criteria (Maffei et al., 1997).

As noted in Table 1, the interrater reliability of the original (DSM-III) version of the DIPD was previously assessed in a sample of 43 inpatients at McLean Hospital in Belmont, MA, whereas the test-retest reliability of this version of the instrument was assessed in a sample of 52 McLean inpatients (Zanarini, Frankenburg, Chauncey, & Gunderson, 1987). Interrater  $\kappa$  ranged from .52 to 1.0, with a median  $\kappa$  of .92. The test-retest  $\kappa$  ranged from .46 to .85, with a median  $\kappa$  of .68. In the current study, interrater  $\kappa$  ranged from .58 to 1.0, with a median  $\kappa$  of .68. The test-



retest  $\kappa$  ranged from .39 to 1.0, with a median  $\kappa$  of .69. In general, the test-retest reliability figures found in both studies are quite similar. However, the interrater reliability figures obtained in the current study are somewhat lower than in the earlier DIPD reliability study. This discrepancy is not surprising since the original DIPD reliability study involved only three raters (one of whom had developed the DIPD) and all interrater reliability interviews were conducted conjointly. This study design contrasts with that of the current study, which had 12 raters from four sites and where interrater reliability interviews were taped by one rater and then viewed later by the other 11 raters. It may be that videotaped interviews do not convey as much clinical information as in-person interviews. In our experience, it is easier to maintain hours of careful attention in a conjoint interview than to maintain attention while viewing a videotaped interview.

It should also be noted that the reliability figures for our four main axis II disorders (avoidant, borderline, obsessive compulsive, and schizotypal personality disorders) were marginally higher than those for the other axis II disorders. It is probable that this difference reflects both the greater frequency with which these disorders were found in our reliability samples and the greater emphasis that they received in training.

The second main finding in this study is that the dimensional levels of axis II reliability are substantially higher than those we found for categorical diagnoses. This finding is consistent with those of earlier studies that assessed axis II pathology dimensionally (Dreessen & Arntz, 1998; Lorange et al., 1994; Maffei et al., 1997). It also suggests that raters tended to agree on whether a subject displayed character pathology on a certain dimension but disagreed on the exact number of diagnostic criteria that the subject met. Of course, if axis I disorders were assessed dimensionally, the resulting criteria counts would also probably be shown to have higher reliability than their categorical counterparts.

The third main finding is that the  $\kappa$  we achieved for axis I disorders, both interrater and test-retest, are quite good. They are also as consistent with those achieved by earlier studies using DSM-III-R criteria for these disorders; both a single site study (Skre et al., 1991) that used raters of varying educational backgrounds and a multisite study that relied mainly on fully trained psychologists or psychiatrists as raters (Williams et al., 1992). This finding is important for a longitudinal study of personality disorders because axis I comorbidity is a crucial covariate in addition to being a key descriptive area of investigation.

Taken together, the results of this study suggest that both axis I and axis II disorders can be diagnosed reliably when using appropriate semistructured interviews. They also suggest that the reliability of axis II disorders is roughly equivalent to that found for most axis I disorders. This latter finding is notable because it is commonly assumed that axis II disorders are far less reliably diagnosed than axis I disorders. This finding is also notable given the "skip-out" structure of the SCID-I, which probably enhances reliability and the lack of this feature in the DIPD-IV. When viewed together, the results of this study should be particularly encouraging to all researchers who wish to assess axis II psychopathology.

## REFERENCES

- Andrews, G., & Peters, L. (1998). The psychometric properties of the Composite International Diagnostic Interview. *Social Psychiatry and Psychiatric Epidemiology*, 33, 80-88.
- Dreessen, L., & Arntz, A. (1998). Short-interval test-retest interrater reliability of the Structured Clinical Interview for DSM-III-R Personality Disorders (SCID-II) in outpatients. *Journal of Personality Disorders*, 12, 138-148.
- First, M. B., Spitzer, R. L., Gibbon, M., Williams, J. B. W., Davies, M., Borus, J., Howes, M., Kane, J., Pope, H. G., & Rounsaville, B. (1995). The Structured Clinical Interview for DSM-III-R Personality Disorders (SCID-II). Part II: Multisite test-retest reliability study. *Journal of Personality Disorders*, 9, 92-104.
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1996). *Structured Clinical Interview for DSM-IV Axis I Disorders*. New York: New York State Psychiatric Institute, Biometrics Research Department.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: Wiley.
- Gunderson, J. G., Shea, M. T., Skodol, A. E., McGlashan, T. H., Morey, L. C., Stout, R. L., Zanarini, M. C., Grilo, C. M., Oldham, J. M., & Keller, M. B. (2000). *The Collaborative Longitudinal Personality Disorders Study: Development, Aims, design, and sample characteristics*.
- Loranger, A. W., Sartorius, N., Andreoli, A., Berger, P., Buchheim, P., Channabasavanna, S. M., Coid, B., Dahl, A., Diekstra, R. F. W., Ferguson, B., Jacobsberg, L. B., Mombour, W., Pull, C., Ono, Y., & Reiger, D. A. (1994). The International Personality Disorder Examination: The World Health Organization/Alcohol, Drug Abuse, and Mental Health Administration International Pilot Study of Personality Disorders. *Archives of General Psychiatry*, 51, 215-224.
- Loranger, A. W., Susman, V. L., Oldham, J. M., & Russakoff, L. M. (1987). The Personality Disorder Examination: A preliminary report. *Journal of Personality Disorders*, 1, 1-13.
- Maffei, C., Fossati, A., Agostoni, I., Barraco A., Bagnato, M., Deborah, D., Namia, C., Novella, L., & Petrachi, M. (1997). Interrater reliability and internal consistency of the Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II), version 2.0. *Journal of Personality Disorders*, 11, 279-284.
- Robins, L. N., Helzer, J. E., Croughan, J., & Ratcliff, K. S. (1981). National Institute of Mental Health Diagnostic Interview Schedule: Its history, characteristics, and validity. *Archives of General Psychiatry*, 38, 381-389.
- Skre, I., Onstad, S., Torgersen, S., & Kringle, E. (1991). High interrater reliability for the Structured Clinical Interview for DSM-III-R Axis I (SCID-I). *Acta Psychiatrica Scandinavica*, 84, 167-173.
- Spitzer, R. L., Williams, J. B. W., Gibbon, M., & First, M. B. (1984). *Structured Clinical Interview for DSM-III Axis I Disorders*. New York, New York State Psychiatric Institute, Biometrics Research Department.
- Stangl, D., Phofl, B., Zimmerman, M., Bowers, W., & Corenthal, C. (1985). A structured interview for the DSM-III personality disorders. *Archives of General Psychiatry*, 42, 591-596.
- Widiger, T. A., Trull, T. J., Hurt, S. W., Clarkin, J., & Frances, A. (1987). A multidimensional scoring of the DSM-III personality disorders. *Archives of General Psychiatry*, 44, 557-563.
- Williams, J. B. W., Gibbon, M., First, M. B., Spitzer, R. L., Davies, M., Borus, J., Howes, M., Kane, J., Pope, H. G., Rounsaville, B., & Wittchen, H. (1992). The Structured Clinical Interview for DSM-III-R (SCID). II. Multi-site test-retest reliability. *Archives of General Psychiatry*, 49, 624-629.
- Zanarini, M. C., Frankenburg, F. R., Chauncey, D. L., & Gunderson, J. G. (1987). The Diagnostic Interview for Personality Disorders: Interrater and test-retest reliability. *Comprehensive Psychiatry*, 28, 467-480.
- Zanarini, M. C., Frankenburg, F. R., Sickel, A. E., & Yong, L. (1996). *The Diagnostic Interview for DSM-IV Personality Disorders*. McLean Hospital, Laboratory for the Study of Adult Development, Belmont, MA.
- Zimmerman, M. (1994). Diagnosing personality disorders: A review of issues and research methods. *Archives of General Psychiatry*, 51, 225-245.