

Short-Term Diagnostic Stability of Schizotypal, Borderline, Avoidant, and Obsessive-Compulsive Personality Disorders

M. Tracie Shea, Ph.D.

Robert Stout, Ph.D.

John Gunderson, M.D.

Leslie C. Morey, Ph.D.

Carlos M. Grilo, Ph.D.

Thomas McGlashan, M.D.

Andrew E. Skodol, M.D.

Regina Dolan-Sewell, Ph.D.

Ingrid Dyck, M.P.H.

Mary C. Zanarini, Ed.D.

Martin B. Keller, M.D.

Objective: Personality disorders are defined as enduring patterns of maladaptive behaviors and traits that are stable over time. This study prospectively examined the stability of four personality disorders (schizotypal, borderline, avoidant, and obsessive-compulsive) over a 1-year follow-up period.

Method: Subjects (N=668) were recruited from multiple clinical settings at four collaborating institutions. Subjects met criteria for one or more of the four personality disorders or were part of a comparison group of subjects with major depressive disorder and no personality disorder. Diagnoses were established by using semistructured interviews. Follow-up assessments, conducted 6 and 12 months after the baseline assessment, included monthly ratings of all criteria for the four personality disorders and weekly ratings of the course of major depressive disorder. The current report is based on 621 subjects

with complete data through 12 months of the follow-up period.

Results: Significantly more subjects in each personality disorder group remained at diagnostic threshold throughout the 12 months of the follow-up period than did those in the major depressive disorder group. A continuous measure of number of criteria met was highly correlated across the three assessments. The majority of personality disorder subjects, however, did not consistently remain at diagnostic threshold, and the mean number of criteria met decreased significantly for each group.

Conclusions: Individual differences in personality disorder features appear to be highly stable, although the number of criteria present decreases over time. Personality disorders may be characterized by stable trait constellations that fluctuate in degree of maladaptive expression.

(*Am J Psychiatry* 2002; 159:2036–2041)

Key to the definition of a personality disorder is the assumption of stability over time. According to DSM-IV, personality disorders are apparent by late adolescence or early adulthood and are characterized by a persisting pattern of maladaptive traits throughout adulthood. There are few data to support these assumptions, however (1–3). The majority of existing studies addressing this question have used DSM categorical diagnoses to define personality disorders, examining diagnostic or functional outcome over varying periods of time. A review of research on personality disorder diagnostic stability (3) found 11 studies (10 on borderline personality disorder), with follow-up periods ranging from weeks to 15 years. Diagnostic stability, defined as percent of study participants retaining their personality disorder diagnosis at the follow-up evaluation, ranged from 25% to 78%, with a mean of 56%. The few studies that have examined diagnostic stability for more than one personality disorder group have typically been limited in reporting rates of stability for individual personality disorders because of study group size (4–8). Many of the existing studies of diagnostic stability have also been limited by high attrition rates, absence of structured inter-

views for the initial diagnoses, lack of assessment of other axis II disorders, and lack of interrater reliability testing or reporting. The majority of studies are based on no more than two assessments, often with long intervals between them. Thus, it is difficult to draw firm conclusions regarding diagnostic stability of the DSM definition of personality disorders.

Limitations of the categorical model of personality disorders—including diagnostic overlap, heterogeneity within categories, and arbitrary thresholds for diagnoses—have led to greater attention to dimensional approaches to defining and assessing personality disorders. One approach is to use a continuous measure consisting of personality disorder criteria scores (9). In one study (8), stability of personality disorder features was assessed over a 4-year period in 250 subjects drawn from a nonclinical university population. There was a statistically significant decrease in mean level of personality disorder features over time, with most of the change occurring from the first to the second assessment and little change from the second to the third assessment. Correlations of the total personality disorder scores were moderate to high, suggesting stability of indi-

vidual differences in relative number of personality disorder features (8). Correlations for individual personality disorder scores were generally lower, but all were statistically significant. Another study examined the stability of personality disorder features over a 2-year interval in a community sample of 118 gay men (7). They reported that personality disorder diagnoses had low stability, whereas personality disorder symptom levels showed moderate stability.

The Collaborative Longitudinal Personality Disorders Study was designed to provide comprehensive data on several aspects of short and longer-term outcome of subjects meeting criteria for one (or more) of four DSM-IV axis II conditions: schizotypal, borderline, avoidant, and obsessive-compulsive personality disorders (10). The study includes an axis I comparison group of patients with major depressive disorder, selected because of the prototypic episodic course of this illness (i.e., remissions and relapses, which have been presumed to distinguish axis I from axis II) and because major depressive disorder is highly prevalent and has been well studied.

One of the goals of this multisite, prospective naturalistic longitudinal study is to examine the validity of the definition of personality disorders as enduring and stable. Here we address the question of short-term stability by using data from the first year of a prospective follow-up period. Categorical and continuous measures of stability are investigated. On the basis of the DSM definition of personality disorders, we hypothesized that each of the personality disorders would be more stable than an axis I comparison condition (major depressive disorder) and that the majority of each of the four personality disorders under study would continue to meet criteria over the 1-year follow-up period. The personality disorder groups were also compared for differences in stability.

Method

Subjects

A detailed description of the Collaborative Longitudinal Personality Disorders Study aims, background, design, methods, and study group characteristics has been reported separately (10, 11). Recruitment efforts were directed at obtaining a diverse and clinically representative sample. The majority of subjects were patients recruited from inpatient and outpatient clinical services affiliated with each of the four recruitment sites of the Collaborative Longitudinal Personality Disorders Study. Additional subjects with current or past psychiatric treatment were recruited by postings or advertising. Of 1,605 subjects screened, 668 (42%) were eligible and entered the study. The current report is based on 621 subjects (93% of the intake group) with complete data through 12 months of the follow-up period. The majority of the 621 subjects were female (64%) and Caucasian (77%), with a mean age of 32.8 years ($SD=8.1$). Less than half (40%) of the subjects were employed. Co-occurring axis I and axis II disorders were common (11). The mean number of axis I disorders for personality disorder subjects was 3.6 ($SD=1.7$). Most personality disorder subjects (64%) had more than one personality disorder diagnosis; the mean number of additional axis II disorders was 1.4 ($SD=1.6$). The

subjects followed did not differ significantly from the missing 47 subjects on any demographic or clinical variables examined.

Procedures

All participants signed written informed consent after a full explanation of study procedures. At the baseline evaluation, potential subjects were screened for possible personality disorder by completing a self-report Personality Screening Questionnaire, which consisted of items from the Personality Diagnostic Questionnaire (12) that pertained to the four targeted personality disorders. Subjects screening positive for one or more of the personality disorders were referred for further assessment. Subjects were also screened for the possible presence of current major depressive disorder with the self-report Depression Screening Questionnaire, which consists of items based on DSM-IV criteria. Subjects screening positive on the Depression Screening Questionnaire and negative for a personality disorder on the Personality Screening Questionnaire were referred for further assessment for the major depressive disorder comparison group.

Subjects were interviewed face-to-face by experienced interviewers. Interviewers with master's- or doctoral-level training underwent extensive training and continued reliability monitoring (10) in the administration of the axis I and axis II diagnostic measures—the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) (13) and the Diagnostic Interview for DSM-IV Personality Disorders (14). Diagnoses obtained from the Diagnostic Interview for DSM-IV Personality Disorders required convergent support from either the self-report Schedule for Nonadaptive and Adaptive Personality (15) or an independent clinician-rated Personality Assessment Form (16). When more than one target disorder was present, assignment to a primary personality disorder study group was determined by a severity-based algorithm (10).

Subjects screening positive on the Depression Screening Questionnaire were assessed with the SCID-I and Diagnostic Interview for DSM-IV Personality Disorders. Subjects with current major depressive disorder by DSM-IV criteria and no personality disorder were eligible for the axis I comparison group of subjects with major depressive disorder.

Study subjects were then interviewed 6 and 12 months after the baseline assessment. The four study personality disorders were assessed by using a modified version of the Diagnostic Interview for DSM-IV Personality Disorders, and all co-occurring axis I disorders were assessed by using the Longitudinal Interval Follow-Up Evaluation (17). These follow-up interviews were not blind and were conducted by the same (baseline) interviewer whenever possible.

Assessment Instruments

The Diagnostic Interview for DSM-IV Personality Disorders (14) is a semistructured interview for assessment of DSM-IV axis II disorders. One or more questions are asked for each of the criteria, rated on a 3-point scale (0=not present; 1=present but clinically insignificant; 2=definitely present). The time frame covered is the prior 2 years, but traits must be characteristic of the person for most of his or her adult life in order to be counted toward a diagnosis. Interrater and test-retest reliabilities in the current study were comparable to published reports of reliability for other semistructured interviews for personality disorders (18). Interrater kappas for the four personality disorders ranged from 0.68 (borderline personality disorder) to 0.73 (avoidant personality disorder); test-retest kappas ranged from 0.69 (borderline personality disorder) to 0.74 (obsessive-compulsive personality disorder). Median reliability correlations for criteria scores ranged from 0.79 to 0.91 (interrater) and 0.65 to 0.84 (test-retest) for the four personality disorders (18).

To assess the longitudinal course of the study personality disorders, the Diagnostic Interview for DSM-IV Personality Disorders

ders was modified to record the presence of each criterion for the four personality disorders for each month of the follow-up interval. Interviewers asked the standard probes for presence of each criterion; if present at all during the interval, the subject was then queried about any change over the interval to determine whether or when the criterion was absent. Ratings were then made for each month of the interval for each criterion by using the aforementioned 3-point scale.

An additional reliability study was conducted to estimate the reliability of retrospective reporting by month on the modified Diagnostic Interview for DSM-IV Personality Disorders. At the 12-month assessment, interviewers assessed and rated month 6 in addition to months 7–12. Hence, month 6 was rated twice, first at the 6-month interview, then again 6 months later at the 12-month interview. Based on 453 cases with overlap data, the kappas for diagnoses at the two time points were 0.78 (schizotypal personality disorder), 0.70 (borderline personality disorder), 0.73 (avoidant personality disorder), and 0.68 (obsessive-compulsive personality disorder).

The Longitudinal Interval Follow-Up Evaluation (17, 19) is a semistructured interview rating system with demonstrated reliability for assessing the longitudinal course of mental disorders. From information obtained at the interview covering the interval followed, weekly psychiatric status ratings are made for each axis I disorder present. Psychiatric status ratings are based on either a 6-point or 3-point scale, indicating whether the individual meets full criteria, is in partial remission, or is in full remission from the given disorder.

Mental health treatment is also assessed by the Longitudinal Interval Follow-Up Evaluation, which includes detailed ratings of psychosocial and pharmacological treatments for all mental health contacts, frequency of sessions, length of treatment, and number of days of inpatient and partial hospitalization. Types and doses of all psychotropic medications are recorded on a weekly basis.

Data Analyses

Three indicators of personality disorder stability were investigated. First, the proportion of subjects remaining at diagnostic threshold all months of follow-up was examined as an indicator of the stability of DSM-IV categorical diagnoses. Second, change in mean number of personality disorder criteria from baseline to the 6- and 12-month assessments was examined for each personality disorder group. Third, correlations of number of criteria met at each of the three assessment points for each personality disorder group were examined. Number of criteria at baseline was based on the Diagnostic Interview for DSM-IV Personality Disorders; for the 6- and 12-month assessments, the number of criteria was based on the previous month rating of the modified Diagnostic Interview for DSM-IV Personality Disorders. The mean number of criteria is an indicator of the extent to which the personality disorder group on average retains the same level of personality disorder psychopathology. The correlations provide a measure of relative stability, i.e., the extent to which individuals retain their relative position within the subject pool in the type and level of personality features.

For the categorical measure, the combined personality disorder groups were compared with the major depressive disorder group by using chi-square analyses. Significant omnibus tests were followed by individual comparisons among the study groups. Repeated-measures analysis of variance (ANOVA) as per the general linear model procedure was used to examine the stability of the mean level of number of criteria met. The model included three levels of time (baseline, 6 months, and 12 months). Terms for the interaction of time with study site as well as five variables that showed significant differences among the study groups were also included in the model. The five variables were

gender, race, number of axis I disorders, number of axis II disorders, and treatment intensity over the 12 months of the follow-up period.

A subsequent repeated-measures ANOVA was conducted to test for differences among personality disorder groups in criteria change over the follow-up period. Since the number of possible criteria differs for the four personality disorders, change in proportion of criteria met was used as the dependent variable in this analysis. The model included the same set of interaction terms as the aforementioned within-group analyses. Pearson correlation coefficients were calculated to examine the stability of number of criteria met over the three assessment points for each of the four personality disorder criteria sets across all subjects.

Although adjusting for treatment effects in naturalistic studies is complex because of the well-known bias for those patients with the most severe problems to receive the most treatment (20), the possibility that our stability findings were confounded by differences in amount of treatment received was explored. A measure of treatment intensity was developed that used weights assigned to levels of care (inpatient, day hospital, or outpatient); these weights were multiplied by the amount of treatment received at each level during the 12 months of the follow-up period. The resulting scores were then included as an interaction term in the repeated-measures ANOVA to test for possible influences of amount of treatment received on stability.

All statistical analyses were conducted by using SAS version 6.12. An overall p value of <0.05 was set for determining statistical significance for the primary analyses.

Results

A significantly larger proportion of personality disorder subjects (44%) remained at or above diagnostic threshold for all 12 months compared with the major depressive disorder group (4%) ($\chi^2=53.7$, $df=4$, $p<0.001$). Relative to subjects in the major depressive disorder group, significantly higher proportions of subjects remained at full criteria in each of the personality disorder groups (schizotypal: $\chi^2=26.9$, $df=1$, $p<0.001$; borderline: $\chi^2=39.2$, $df=1$, $p<0.001$; avoidant: $\chi^2=66.8$, $df=1$, $p<0.001$; obsessive-compulsive: $\chi^2=41.7$, $df=1$, $p<0.001$). There was a significant difference among the personality disorder groups in the percent of subjects remaining at diagnostic threshold all 12 months ($\chi^2=13.3$, $df=3$, $p<0.004$). As seen in Figure 1, more subjects initially diagnosed with avoidant personality disorder remained at full criteria (56%, $N=82$) than did subjects with schizotypal personality disorder (34%, $N=28$) ($\chi^2=10.5$, $df=1$, $p<0.002$), borderline personality disorder (41%, $N=65$) ($\chi^2=7.8$, $df=1$, $p<0.006$), and obsessive-compulsive personality disorder (42%, $N=61$) ($\chi^2=5.7$, $df=1$, $p<0.02$).

Results of the repeated-measures ANOVAs (Table 1) showed a significant effect of time for each of the four personality disorder groups, reflecting the decrease in mean number of criteria met over time. Subsequent analyses showed that all significant change occurred between baseline and month 6, with no significant change between 6 and 12 months for any of the personality disorder groups. There were no significant interactions between time and site, number of axis I or axis II diagnoses, or gender. There were significant interactions of time and treatment intensity for schizotypal personality disorder ($F=5.02$, $df=2$, 146,

$p=0.008$) and for borderline personality disorder ($F=3.66$, $df=2$, 286 , $p<0.03$). Examination of the pattern of change for subjects receiving treatment of high and low intensity (determined by median split) revealed that subjects in the high treatment intensity groups showed less change. There was also a significant interaction of time and race in the obsessive-compulsive personality disorder group ($F=3.32$, $df=2$, 266 , $p=0.04$), which was due to a steeper time trend for the minority subjects.

Results of the repeated-measures ANOVA examining the interaction of time with personality disorder group in change in proportion of criteria met showed a significant group-by-time interaction ($F=2.53$, $df=6$, 1024 , $p<0.03$). Pairwise tests of the amount of change from baseline to the average of 6 and 12 months showed avoidant personality disorder as being significantly more stable than schizotypal personality disorder ($F=25.88$, $df=1$, 216 , $p\leq 0.0001$), borderline personality disorder ($F=7.27$, $df=1$, 292 , $p=0.007$), and obsessive-compulsive personality disorder ($F=12.34$, $df=1$, 280 , $p=0.0005$). In addition, schizotypal personality disorder changed less than did borderline personality disorder ($F=4.33$, $df=1$, 227 , $p<0.04$).

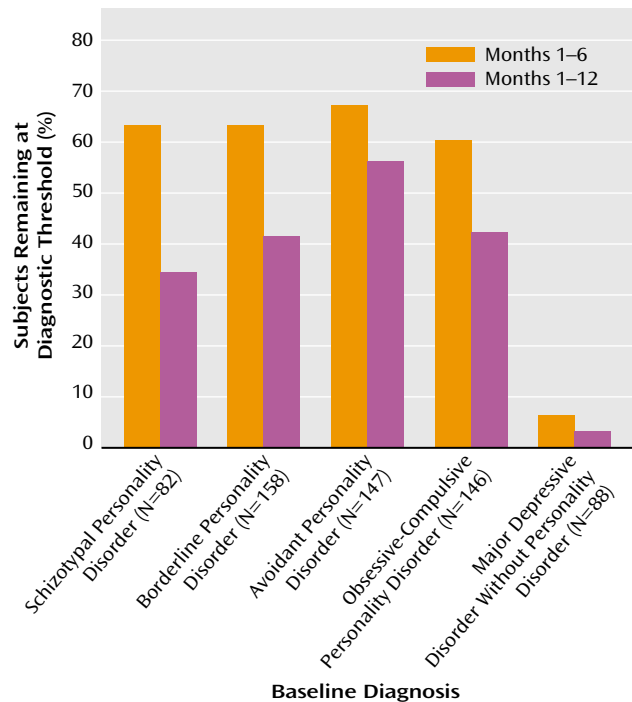
Correlation coefficients of number of criteria met over the three assessment points were uniformly large (ranging from 0.86 to 0.92) and highly significant ($p<0.0001$) across all three time points and for each of the personality disorder criteria sets (Table 2).

Discussion

Our findings suggest that whether personality disorders appear stable depends upon how stability is defined. When the traditional DSM categorical model is used, each of the personality disorder groups was found to have significantly higher rates of diagnostic stability than the major depressive disorder comparison group. This suggests that the constellation of behaviors and traits that comprise personality disorders are, as expected, more persistent than symptomatic episodes of major depressive disorder. It is important to note that findings for major depressive disorder cannot be interpreted as reflective of other axis I disorders or as providing a generalized support for a distinction between axis I and axis II on the basis of diagnostic stability. Many axis I disorders are now known to be more chronic than episodic, including anxiety disorders such as panic disorder (21). Furthermore, although more diagnostically stable than major depressive disorder, the majority of personality disorder subjects did not consistently remain at DSM-IV criteria thresholds when followed closely over time. Similarly, the significant decrease in mean number of criteria for each of the groups suggests decreases in severity of personality disorders over time. In contrast, when the relative stability of individual differences was examined, we found a high level of consistency.

Previous studies of the longitudinal diagnostic stability of borderline personality disorder have reported similar or

FIGURE 1. Diagnostic Stability for Months 1–6 and Months 1–12 Among Subjects Diagnosed at Baseline With Either a Personality Disorder (Schizotypal, Borderline, Avoidant, or Obsessive-Compulsive) or Major Depressive Disorder With No Personality Disorder



even higher rates of diagnostic stability despite much longer follow-up intervals (3). For example, the mean percentage of diagnostic stability for borderline personality disorder over 10 studies with follow-up lengths that extended up to 15 years is 57% (3), compared with 41% in the current study. This difference is likely due to a difference in the method of assessing diagnostic stability. Our follow-up assessment included monthly ratings for each criterion, and our indicators of diagnostic stability were based on presence of a sufficient number of criteria at a clinically significant level for every month of follow-up. Although more precise, this is a more stringent measure of stability than the usual method of basing ratings on evidence of the criteria at some unspecified frequency over the time period followed.

The significant decrease in number of criteria met from baseline to 6 months may be due in part to this methodological difference, since our baseline assessment used the usual method of assessment (obtaining examples of clinical significance and not monthly ratings). Another consideration is the recruitment of subjects from clinical settings, which may have increased the likelihood of capturing subjects at their most impaired point. On the other hand, our findings for mean level change are remarkably similar to those reported by Lenzenweger (8) for a group of subjects from a nonclinical (university) setting that also used the same methodology for personality disorder assessment at both time points. The similarity in these find-

TABLE 1. Number of Diagnostic Criteria Met Over Time by Subjects Diagnosed at Baseline With Schizotypal, Borderline, Avoidant, or Obsessive-Compulsive Personality Disorder

Baseline Personality Disorder Diagnosis	Required for Diagnosis	Number of DSM-IV Criteria						Main Effect of Time (repeated-measures ANOVA)		
		Endorsed at Baseline		Endorsed at 6 Months		Endorsed at 12 Months		F	df	p
Schizotypal (N=82)	5	Mean	SD	Mean	SD	Mean	SD	12.31	2, 146	<0.0001
Borderline (N=158)	5	5.8	1.0	4.2	1.9	3.8	1.9	8.66	2, 286	0.0003
Avoidant (N=147)	4	6.8	1.4	4.9	2.2	4.2	2.4	6.22	2, 268	<0.003
Obsessive-compulsive (N=146)	4	5.5	1.1	4.2	2.1	4.0	2.2	10.65	2, 266	<0.0001
		5.2	1.1	4.0	2.0	3.4	1.8			

TABLE 2. Correlations Between Assessment Points for Number of Personality Disorder Criteria Met by Subjects Diagnosed at Baseline With Schizotypal, Borderline, Avoidant, or Obsessive-Compulsive Personality Disorder

Baseline Personality Disorder Diagnosis	Correlation (r) Between Assessment Points for Number of Criteria Met ^a		
	Baseline and 6 Months	Baseline and 12 Months	6 Months and 12 Months
Schizotypal (N=82)	0.91	0.88	0.90
Borderline (N=158)	0.90	0.84	0.90
Avoidant (N=147)	0.88	0.87	0.92
Obsessive-compulsive (N=146)	0.87	0.86	0.90

^a All correlations are significant at $p < 0.0001$.

ings (significant decreases in continuous criteria scores occurring between the first and second assessments, with minimal change between the second and third assessment) suggests that other factors, such as the effects of repeated assessments, may play a role in these findings.

In terms of individual differences, the findings show that individuals retain their relative position in their group. That is, despite a decrease in average in number of criteria, the amount and type of personality disorder features present, relative to other subjects in the group, remains consistent. Prior studies that used continuous measures of personality disorder features (7, 8) have similarly reported significant correlations across assessments, although the correlations were notably higher in the current study (ranging from 0.84 to 0.92 compared with 0.40–0.69 for the four personality disorder criteria scores [8]). This is likely due to differences in the subject groups. Because both prior studies were drawn from nonclinical settings, they likely had a narrower range of severity in personality disorder scores than the present study.

Avoidant personality disorder subjects were significantly more likely to remain at disorder threshold over the 12 months than subjects in the other groups and showed significantly less change on mean proportion of criteria met. This higher stability for avoidant personality disorder may be due to differences among the personality disorders in criteria. Some criteria are clearly more trait-like, in contrast to others that are more behaviorally anchored. For example, impulsive self-damaging behaviors, suicidal gestures, and threats (borderline personality disorder) or magical thinking that influences behavior (schizotypal personality disorder) are likely to be less persistent when measured on a monthly basis than more trait-like criteria

such as pervasive feelings of inadequacy and fear of rejection (avoidant personality disorder). Some of the more behaviorally anchored criteria, although perhaps clinically important expressions of underlying psychopathology when they occur, will be evident less frequently than more trait-like criteria, which may be expressed in a variety of ways. Hence the meaning of stability is quite different for the different types of criteria, highlighting the need for a clearer understanding and identification of the important dimensions that may underlie the various criteria of axis II (22). Use of more trait-based criteria for personality disorders with examples of behavioral indicators of the traits would likely increase diagnostic stability.

Other methodological issues deserve comment. As noted, the follow-up interviews were not blind. While the Collaborative Longitudinal Personality Disorders Study design includes a full assessment with the Diagnostic Interview for DSM-IV Personality Disorders after 2 years, conducted by an interviewer blind to all previous data, the 6- and 12-month evaluations were conducted by the same interviewer in an effort to retain subjects. While lack of blindness allows the possibility of a bias toward more stability, use of the same interviewer provides the advantage of repeated contacts with the subject, which may increase the validity of ratings and diminishes error due to rater variance. It is possible that the lack of blindness has influenced the findings; if so, the current findings would presumably overestimate the diagnostic stability of these personality disorders.

A concern often expressed about longitudinal studies of clinically ascertained subjects is the potential confounding by treatment. In our preliminary examination of this issue, we found that for borderline personality disorder and schizotypal personality disorder, higher amounts of treatment were associated with less change in number of criteria met. This suggests that the amount of treatment received is driven in part by the severity of the disorder, a typical finding in naturalistic studies because of the selective bias in treatment seeking. Although it is possible that the personality disorders as a group might show less change if they had an untreated course, the similarity of findings in studies of nonclinical subjects (8) provides some reassurance that our findings are not biased by treatment. It is important to note that the current study was designed to address the clinically relevant question of personality disorder course in real-world clinical settings

and not the interesting, but distinct, question of the untreated course of personality disorders.

Another important aspect of stability concerns functional impairment. It is likely that when individuals lose criteria and drop below diagnostic threshold, impairment in functioning remains. The extent to which such impairment remains at a clinically significant level is an important piece of the stability picture, to be examined in future reports.

In conclusion, this initial report on short-term stability of schizotypal, borderline, avoidant, and obsessive-compulsive personality disorders shows each to be more diagnostically stable than an axis I comparison group of subjects with major depressive disorder, with a high degree of consistency in terms of individual differences in the number and type of personality disorder criteria met. At the same time, we found that the majority of subjects did not remain at personality disorder thresholds and that the mean level of criteria present for each of the four personality disorder criteria sets decreases significantly. Thus, while individuals are very consistent in terms of their rank order of personality disorder features, they may fluctuate in the severity or amount of personality disorder features present at any given point. While this may be indicative of a waxing and waning course of personality disorder pathology, the findings also reflect the limitations of the DSM criteria sets, particularly for assessment of change. Future reports will examine the course of the personality disorders over multiple subsequent assessments, the stability of individual criteria, the associated course of psychosocial functioning, and the influence of such factors as stressful life events and the course of co-occurring axis I disorders on personality disorder stability.

Received Oct. 11, 2001; revision received May 10, 2002; accepted June 11, 2002. From the Collaborative Longitudinal Personality Disorders Study. Address reprint requests to Dr. Shea, Department of Psychiatry and Human Behavior, Brown University Medical School, Duncan Bldg., 700 Butler Dr., Providence, RI 02906.

The Collaborative Longitudinal Personality Disorders Study is an ongoing, longitudinal, multisite, follow-along study of personality disorders that is funded by NIMH. Award sites are Brown University Department of Psychiatry and Human Behavior, Providence, R.I. (MH-50837); Columbia University and New York State Psychiatric Institute, New York (MH-50839); Harvard Medical School and McLean Hospital, Boston (MH-50840); Texas A&M University, College Station (MH-50838); and Yale University School of Medicine, New Haven, Conn. (MH-50850). This work has also been supported in part by NIMH grant MH-01654 to Dr. McGlashan.

This manuscript has been reviewed and approved by the publications committee of the Collaborative Longitudinal Personality Disorders Study.

References

- Grilo CM, McGlashan TH: Stability and course of personality disorders. *Curr Opin Psychiatry* 1999; 12:157-162
- Grilo CM, McGlashan TH, Oldham JM: Course and stability of personality disorders. *J Practical Psychiatry and Behavioral Health* 1998; 4:61-75
- McDavid JD, Pilkonis PA: The stability of personality disorder diagnoses. *J Personal Disord* 1996; 10:1-15
- Barasch A, Frances A, Hurt S, Clarkin J, Cohen S: Stability and distinctness of borderline personality disorder. *Am J Psychiatry* 1985; 142:1484-1486
- Vaglum P, Friss S, Karterud S, Mehlum L, Vaglum S: Stability of the severe personality disorder diagnosis: a 2-5 year prospective study. *J Personal Disord* 1993; 7:348-353
- Ferro T, Klein DN, Schwartz JE, Kasch KL, Leader JB: 30-month stability of personality disorder diagnoses in depressed outpatients. *Am J Psychiatry* 1998; 155:653-659
- Johnson JG, Williams JBW, Goetz RR, Rabkin JG, Lipsitz JD, Remien RH: Stability and change in personality disorder symptomatology: findings from a longitudinal study of HIV+ and HIV- men. *J Abnorm Psychol* 1997; 106:154-158
- Lenzenweger MF: Stability and change in personality disorder features: findings from a longitudinal study of personality disorders. *Arch Gen Psychiatry* 1999; 56:1009-1015
- Oldham JM, Skodol AE: Charting the future of axis II. *J Personal Disord* 2000; 14:17-29
- Gunderson JG, Shea MT, Skodol AE, McGlashan TH, Morey LC, Stout RL, Zanarini MC, Grilo CM, Oldham JM, Keller MB: The Collaborative Longitudinal Personality Disorders Study: development, aims, design, and sample characteristics. *J Personal Disord* 2000; 14:300-315
- McGlashan TH, Grilo CM, Skodol AE, Gunderson JG, Shea MT, Morey LC, Zanarini MC, Stout RL: The Collaborative Longitudinal Personality Disorders Study: baseline axis I/II and II/II diagnostic co-occurrence. *Acta Psychiatr Scand* 2000; 102:256-264
- Hyler SE, Skodol AE, Kellman HD, Oldham JM, Rosnick L: Validity of the Personality Diagnostic Questionnaire—Revised: comparison with two structured interviews. *Am J Psychiatry* 1990; 147:1043-1048
- First MB, Spitzer RL, Gibbon M, Williams JBW: Structured Clinical Interview for DSM-IV Axis I Disorders Research Version (SCID-I). New York, New York State Psychiatric Institute, Biometrics Research, 1996
- Zanarini MC, Frankenburg FR, Sickel AE, Yong L: The Diagnostic Interview for DSM-IV Personality Disorders (DIPD-IV). Belmont, Mass, McLean Hospital, 1996
- Clark LA: Schedule for Nonadaptive and Adaptive Personality (SNAP). Minneapolis, University of Minnesota Press, 1993
- Shea MT, Glass DR, Pilkonis PA, Watkins J, Docherty JP: Frequency and implications of personality disorders in a sample of depressed inpatients. *J Personal Disord* 1987; 1:27-42
- 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
- Zanarini MC, Skodol AE, Bender D, Dolan R, Sanislow C, Schaeffer E, Morey LC, Grilo CM, Shea MT, McGlashan TH, Gunderson JG: The Collaborative Longitudinal Personality Disorders Study: reliability of axis I and II diagnoses. *J Personal Disord* 2000; 14:291-299
- Warshaw MG, Keller MB, Stout RL: Reliability and validity of the Longitudinal Interval Follow-Up Evaluation for assessing outcome of anxiety disorders. *J Psychiatr Res* 1994; 28:531-545
- Cochran W: Planning and Analysis of Observational Studies. New York, John Wiley & Sons, 1983
- Keller MB, Yonkers KA, Warshaw MG, Pratt LA, Gollan JK, Massion AO, White K, Swartz AR, Reich J, Lavori PW: Remission and relapse in subjects with panic disorder and panic with agoraphobia: a prospective short-interval naturalistic follow-up. *J Nerv Ment Dis* 1994; 82:290-296
- Shea MT: Some characteristics of the axis II criteria sets and their implications for assessment of personality disorders. *J Personal Disord* 1992; 6:377-381