

# Stability and Change in Personality Disorder Features

## *The Longitudinal Study of Personality Disorders*

Mark F. Lenzenweger, PhD

**Background:** There exists no empirical literature documenting the long-term longitudinal stability of personality pathology comparable to that available for normal personality. A number of test-retest studies have usefully established the short-term reliability of Axis II measures. However, the test-retest design is methodologically inadequate for resolving issues related to the long-term stability of personality disorder (PD). This prospective longitudinal study evaluated the stability of PD features in multiwave perspective.

**Methods:** Subjects (N = 250) drawn from a nonclinical university population were examined for PD features at 3 different time points using the International Personality Disorders Examination (IPDE) and the Millon Clinical

Multiaxial Inventory II (MCMI-II) during a study period of 4 years.

**Results:** Features of PD displayed considerable evidence of stability for individual differences and group means, at the dimensional level of analysis, on both the IPDE and the MCMI-II. Both measures revealed modest declines in PD features over time; however, the observed changes were associated with relatively small effect sizes.

**Conclusion:** Features of PD, viewed from a dimensional perspective, seem to be relatively stable in terms of individual differences and group means based on both clinical interview and self-administered PD assessments.

*Arch Gen Psychiatry.* 1999;56:1009-1015

**T**HERE EXISTS no empirical literature documenting the long-term longitudinal stability of the personality disorders (PD) comparable to that available for normal personality.<sup>1,2</sup> Although the *DSM*<sup>3-5</sup> asserts that PDs are enduring and stable over time and recognizable by adolescence or early adult life, supporting empirical data are virtually nonexistent. Test-retest studies<sup>6,7</sup> have been conducted on PDs, typically borderline personality disorder—however, they reveal noteworthy methodological shortcomings that limit their utility for resolving issues related to long-term stability of the PDs. The greatest limitation of all such studies lies in the research design itself and the fundamental inability of test-retest observations to adequately address stability, an established fact in lifespan research methodology.<sup>8-12</sup> Test-retest studies have usefully established the test-retest reliability of the primary Axis II assessment devices.<sup>13-15</sup> The methodological superiority, however, of the prospective multiwave longitudinal design for studying continuity and change in personality is well known.<sup>2,8,9,11,16</sup>

The Longitudinal Study of Personality Disorders (LSPD) was begun in 1990 as a National Institute of Mental Health-sponsored prospective multiwave longi-

tudinal study of personality pathology, normal personality, and temperament. A major goal of the project is the lifespan study of the stability of PD symptomatology. By using a nonclinical population, the LSPD avoids the complications attending the study of hospital/clinic PD cases (eg, treatment confounds, Berkson's bias). Finally, the LSPD began with subjects who were young enough to allow for an evaluation of the *DSM* assumption regarding age at onset for PD.

Stability of PD features can be assessed from 4 vantage points: rank order of individual differences, mean level (or group) stability, structural (factorial) stability, and ipsative (or intrapersonal) stability.<sup>17-19</sup> Additional approaches to stability evaluation involve structural equation modeling<sup>8,11,20</sup> and latent growth curve procedures.<sup>10,21</sup> This report from the LSPD focuses only on the stability of individual differences and mean levels for PD features.

## RESULTS

### AXIS I CHARACTERISTICS OF THE SAMPLE

The lifetime *DSM-III-R* Axis I diagnoses (**Table 2**) of the study subjects are for

From the Department of Psychology, Harvard University, Cambridge, Mass.

## SUBJECTS AND METHODS

### SUBJECTS

The 258 subjects in the LSPD were drawn from an initial sample that consisted of 2000 first-year undergraduate students at Cornell University, Ithaca, NY.<sup>22</sup> Subjects were assigned to either a possible personality disorder (PPD) or no personality disorder (NPD) group as determined by the International Personality Disorder Examination (IPDE) Screen (IPDE-S) (completed screens = 1684, response rate = 84.2%). The PPD subjects had to meet the diagnostic threshold for at least 1 specific *DSM-III-R* PD, whereas NPD subjects did not meet the *DSM-III-R*–defined threshold for diagnosis and had fewer than 10 PD features across all disorders. Extensive details concerning subject selection procedure and sampling are given elsewhere.<sup>22</sup> The 258 subjects consisted of 121 men (47%) and 137 women (53%); 134 (66 women) in the PPD group and 124 (71 women) in the NPD group. All subjects gave voluntary written informed consent and received an honorarium of \$50 at each wave. Of the initial 258 subjects, 250 completed all 3 assessment waves. Of the 8 subjects who did not complete all waves, 5 were in the PPD and 3 were in the NPD group; the proportions did not differ (goodness of fit  $\chi^2_1 = 0.50$ ,  $P > .47$ ). Of these 8 subjects, 6 transferred to other colleges and 2 died in motor vehicle crashes. This report concerns the 250 subjects completing all 3 assessment waves.

Compared with the US population, the sample overrepresents Asian/Pacific Islanders and underrepresents those of Hispanic, African American, and working-class origins (**Table 1**).

### PD ASSESSMENTS

The LSPD has a prospective multiwave panel design<sup>8</sup> in which subjects are evaluated at 3 points in time (ie, freshman, sophomore, and senior years in college). All interview assessments were conducted by experienced clinical psychologists (PhD level) or psychiatric social workers. The masters-level social workers had an average of 16 years of clinical experience. Subjects also completed a large battery of self-administered assessments of normal personality, temperament, and other psychological factors. Finally, as the LSPD is a naturalistic prospective study, subjects were free to seek psychological treatment of their own accord.

#### IPDE *DSM-III-R* Screen

This is a 250-item self-administered true-false PD screening inventory developed by Armand W. Loranger, PhD. The diagnostic efficiency and psychometric properties of the IPDE-S in a 2-stage screen application were described in a prior report from the LSPD.<sup>22</sup>

#### International Personality Disorder Examination

The IPDE is the well-known semistructured interview designed for use by experienced clinicians for the assessment of both *DSM* and *International Classification of Diseases, 10th Revision (ICD-10)* PD features.<sup>13,23,24</sup> The IPDE<sup>13</sup> allows for both dimensional and categorical scoring of the *DSM* PDs. The *DSM-III-R* criteria were assessed in this study. Interviewers received training in IPDE administration and scoring by Loranger and were supervised throughout the

project by the author. Supervision by the author was done blind to the subjects' identity, putative PD status, and all prior assessment information. The interrater reliability for IPDE assessments was generally excellent at all 3 waves (intraclass correlations range, 0.84-0.92). Interviewers were blind to the screening results and putative PD status of the subjects and the same interviewer never saw the same subject more than once during the 3 study waves.

#### Structured Clinical Interview for *DSM-III-R*: Non-Patient Version

This is a well-known semistructured *DSM-III-R* Axis I clinical interview for use with nonpatients.<sup>25</sup> It was administered prior to the IPDE.

#### Millon Clinical Multiaxial Inventory II

The Millon Clinical Multiaxial Inventory II (MCMI-II)<sup>26,27</sup> is the well-known 175-item, true-false, self-administered inventory designed to coordinate with the *DSM-III-R* Axis II PDs. The MCMI-II possesses excellent psychometric properties<sup>26,27</sup> and yields separate dimensional scores for each of the 11 PDs.

### DATA ANALYSES

The primary analyses focused on the individual difference and mean level (or group) stability of PD features. These 2 forms of stability are conceptually independent and have different interpretations. Individual difference (ie, "rank order"<sup>17</sup> or "normative"<sup>18</sup>) stability concerns the extent to which individuals maintain their relative position within a group ranking over time. Level, or mean level, stability concerns the extent to which group means on a variable (or disorder) of interest remain invariant over time. All data analyses were conducted at the level of dimensional scores on the IPDE and the MCMI-II.

Individual difference stability is assessed using the Pearson correlation coefficient ( $r$ ) applied to actual symptom scores (ie, not ranks). The dimensional, cluster, and total scores for the PDs (IPDE and MCMI-II) were examined for rank order for wave 1 to wave 2, wave 2 to wave 3, and wave 1 to wave 3 comparisons. The differences between wave 1 to wave 2 vs wave 1 to wave 3 stabilities were tested to determine if the longer interval for the latter resulted in significantly lower stabilities; these correlated correlation coefficients were tested as per Meng et al.<sup>28</sup>

The stability of mean levels of the 11 Axis II disorders as well as the total number of PD features was evaluated for both measures using a repeated-measures analysis of variance (ANOVA) as implemented by MANOVA.<sup>29</sup> The ANOVAs were all of a group  $\times$  sex  $\times$  time design, with 2 between-subjects factors (group, sex) and 1 within-subjects factor (time). All F tests involving the within-subjects factor (time) were corrected using the procedure of Greenhouse and Geisser.<sup>30</sup> Two sets of 12 ANOVAs were conducted, one each for the IPDE and MCMI-II, for a total of 24 tests. A Bonferroni correction was used to maintain a nominal significance level of .05 for each effect repeatedly tested within each set of ANOVAs and only ANOVA results with  $P < .004$  (ie,  $.05/12$ ) are discussed as significant. Effect size estimates (partial  $\eta^2$ ) were calculated for significant group, time, and group  $\times$  time effects. The  $\chi^2$  test was used to contrast proportions and the kappa ( $\kappa$ ) coefficient was used to evaluate the stability of categorical PD diagnoses over time. All tests were 2-tailed.

**Table 1. Demographic Characteristics of the Longitudinal Study of PD Sample (N = 250) for Subjects Available at 3 Waves of the Study\***

	Father, %	Mother, %
Parental education		
1-12 y	12.4	18.8
13-15 y	16.0	20.4
≥16 y	68.8	58.8
Not available	2.8	2
Parental occupation		
Clerical/sales/laborer/craftsman/operatives	12.0	22.0
Management/official	26.8	12.8
Professional/technical	52.4	42.0
Not available or homemaker	8.8	23.2
Race	<b>Percentage</b>	
African American	3.6	
Latin/Hispanic	4.8	
Caucasian/Anglo	72.0	
Asian/Pacific Islander	17.2	
Native American	0.8	
Other	1.6	
Sex, No. (%)	<b>Probable PD Group (n = 129)</b>	<b>No PD Group (n = 121)</b>
M	64 (49.6)	53 (43.8)
F	65 (50.4)	68 (56.2)

\*All data are presented as percentage unless otherwise indicated. PD indicates personality disorder.

definite and probable cases; a disorder was considered present if a subject met the criteria for the disorder at wave 1, wave 2, or wave 3 assessments. Eighty-one (62.8%) of the PPD subjects received an Axis I diagnosis compared with 32 (26.4%) NPD subjects ( $\chi^2_1 = 33.30$ ,  $P < .001$ ). Forty-one (31.8%) PPD subjects vs 21 (17.4%) NPD subjects reported a prior history of treatment by wave 3 ( $\chi^2_1 = 6.97$ ,  $P < .008$ ).

### INDIVIDUAL DIFFERENCES STABILITY

Stability coefficients ( $r$ ) for the total number of PD features present on the IPDE range from 0.61 (wave 1/wave 3) to 0.70 (wave 1/wave 2) (**Table 3**). For the Axis "clusters," stability coefficient ranges were: cluster A (0.48-0.61), cluster B (0.60-0.78), and cluster C (0.52-0.67). Differences in the stabilities for the wave 1/wave 3 interval vs the wave 1/wave 2 interval were tested to determine if lower stabilities were observed for the longer time span. Wave 1/wave 2 stability coefficients were significantly higher for schizoid, antisocial, histrionic, and narcissistic PDs, clusters A and B, and total PD features (all  $P < .05$ ). Overall, schizoid and antisocial PD features showed the highest level of individual differences stability on the IPDE.

For the overall mean MCMI-II PD dimensional score, stability coefficients range from 0.70 (wave 1/wave 3) to 0.77 (wave 1/wave 2) (**Table 4**). For clusters, stability coefficients ranges were: cluster A (0.64-0.73), cluster B (0.75-0.81), and cluster C (0.65-0.72). Differences in the MCMI-II stability coefficients were tested for the wave 1/wave 3 vs wave 1/wave 2 intervals. Wave 1/wave 2 stability coefficients were higher for paranoid, borderline, histrionic, nar-

**Table 2. Lifetime DSM-III-R Axis I SCID-NP Diagnoses, Definite and Probable (N = 250)\***

Disorder	Subjects, No. (%)†		$\chi^2$	P‡
	Controls (n = 121)	PPD Group (n = 129)		
Major depression	16 (13.2)	47 (36.4)	17.84	<.001
Bipolar/bipolar NOS	1 (0.8)	6 (4.7)	3.36	.07
Dysthymia (current only)	3 (2.5)	13 (10.1)	6.02	.01
Other affective disorder	10 (8.3)	32 (24.8)	12.22	<.001
Alcohol abuse	2 (1.7)	7 (5.4)	2.56	.11
Alcohol dependence	2 (1.7)	13 (10.1)	7.86	.005
Drug abuse	3 (2.5)	3 (2.3)	0.006	.94
Drug dependence	0 (0.0)	4 (3.1)	3.81	.05
Social phobia	6 (5.0)	22 (17.1)	9.18	.002
Simple phobia	8 (6.6)	15 (11.6)	1.88	.17
Panic	3 (2.5)	1 (0.8)	1.15	.28
Obsessive-compulsive disorder	4 (3.3)	8 (6.2)	1.14	.28
Anorexia	4 (3.3)	4 (3.1)	0.01	.93
Bulimia	2 (1.7)	11 (8.5)	5.99	.01
Eating disorder NOS	0 (0.0)	1 (0.8)	0.94	.33
Any Axis I diagnosis	32 (26.4)	81 (62.8)	33.30	<.001

\*SCID-NP indicates Structured Clinical Interview for DSM-III-R: Non-Patient Version; PPD, possible personality disorder; and NOS, not otherwise specified.

†Values are the number of cases and percentages per diagnostic category observed in the 2 subject groups. Cases consist of both definite and probable lifetime Axis I diagnoses combined for the entire study period; some subjects have more than 1 Axis I diagnoses combined for the entire study period; some subjects have more than 1 Axis I diagnosis.

‡Pearson  $\chi^2$  test (2-tailed),  $df = 1$ .

**Table 3. Stability Coefficients for DSM-II-R PD Features Assessed With the IPDE Across 3 Study Waves (N = 250)\***

Disorder	Assessment Waves†		
	Wave 1/ Wave 2	Wave 2/ Wave 3	Wave 1/ Wave 3
Paranoid	0.47	0.38	0.39
Schizoid	0.74	0.70	0.59
Schizotypal	0.51	0.46	0.45
Cluster A	0.61	0.60	0.48
Antisocial	0.73	0.68	0.65
Borderline	0.59	0.43	0.50
Histrionic	0.55	0.54	0.44
Narcissistic	0.64	0.46	0.47
Cluster B	0.78	0.61	0.60
Avoidant	0.44	0.69	0.48
Passive-aggressive	0.46	0.47	0.44
Compulsive	0.46	0.50	0.40
Dependent	0.46	0.57	0.56
Cluster C	0.52	0.67	0.59
Total PD features	0.70	0.66	0.61

\*IPDE indicates International Personality Disorder Examination; PD, Personality disorder. The IPDE was scored using the dimensional method.

†Values are Pearson product-moment correlation coefficients based on 250 cases. All  $r$  values are significant at  $P < .001$  (2-tailed).

...cistic, and avoidant PDs; cluster A, B, and C; and total PD features (all  $P < .05$ ). Overall, antisocial and passive-aggressive PD dimensions showed the highest level of individual differences stability on the MCMI-II.

## MEAN LEVEL STABILITY

The repeated-measures ANOVA results for IPDE are summarized with the *F* statistics, *P* values, and effect size estimates for the group, time, and group × time interactions (**Table 5**). The results for the sex factor will not be discussed owing to a relative absence of significant effects for the factor. Only effects associated with *P* ≤ .004 are interpreted. For the IPDE, the PPD group consistently displayed greater PD symptomatology

relative to the NPD group (main effect). The total PD features on the IPDE showed a main effect for time, indicating a decline over time. A significant group × time interaction superseded the main effects of group and time, indicating a more rapid decline in PD features among the PPD (vs NPD) subjects over time. The mean (± SD) total PD scores for the PPD group over time were: wave 1 = 16.72 (±15.21), wave 2 = 8.91 (±10.95), and wave 3 = 8.68 (±10.58); and for the NPD group were: wave 1 = 4.90 (±5.91), wave 2 = 3.24 (±5.73), and wave 3 = 3.87 (±5.52).

On the IPDE, paranoid, borderline, narcissistic, histrionic, dependent, obsessive-compulsive, and avoidant PD features followed a longitudinal pattern similar to that seen for the total scores (Table 5). Generally, most symptom change (decline) occurred from wave 1 to wave 2 assessments, whereas symptom levels remained relatively constant from wave 2 to wave 3 in both groups. An effect for time (as either a main effect or in interaction with group) was detected for 9 of the 11 PDs assessed by the IPDE. The mean effect sizes, for the significant findings, were as follows: group (.08), time (.08), and group × time interaction (.04).

On the MCMI-II overall average base rate-adjusted dimensional score, the PPD group consistently displayed higher levels of PD symptomatology (main effect) (**Table 6**). A group × time interaction indicated a more rapid decline in PD features among the PPD (vs NPD) subjects over time. The mean (± SD) for the average base rate-adjusted PD dimensional score on the MCMI-II were: for the PPD subjects, wave 1 = 53.42 (±11.08), wave 2 = 48.49 (±12.24), and wave 3 = 46.30 (±12.03); for the NPD subjects: wave 1 = 37.82 (±7.44), wave 2 = 35.87 (±7.33), and wave 3 = 34.95 (±8.44).

**Table 4. Stability Coefficients for DSM-III-R PD Features Assessed With the MCMI-II Across 3 Study Waves (N = 250)\***

Disorder	Assessment Waves†		
	Wave 1/ Wave 2	Wave 2/ Wave 3	Wave 1/ Wave 3
	Paranoid	0.75	0.70
Schizoid	0.68	0.61	0.61
Schizotypal	0.63	0.55	0.55
Cluster A	0.73	0.64	0.64
Antisocial	0.74	0.70	0.70
Borderline	0.73	0.72	0.64
Histrionic	0.76	0.73	0.69
Narcissistic	0.76	0.74	0.68
Cluster B	0.81	0.78	0.75
Avoidant	0.76	0.67	0.67
Passive-aggressive	0.78	0.76	0.73
Compulsive	0.66	0.69	0.60
Dependent	0.63	0.61	0.59
Cluster C	0.72	0.71	0.65
Total PD features	0.77	0.74	0.70

\*MCMI-II indicates Millon Clinical Multiaxial Inventory II; PD, personality disorder. MCMI-II dimensions are base-rate transformed scores.

†Values are Pearson product-moment correlation coefficients based on 250 cases. All *r* values are significant at *P* < .001 (2-tailed).

**Table 5. Repeated-Measures ANOVA for PD Features for DSM-III-R Axis II Disorders Assessed by the IPDE\***

PD	Effect								
	Group			Time†			Group × Time‡		
	F	P	ES‡	F	P	ES‡	F	P	ES‡
SZD	3.63	.06	.015	5.41	.005	.022	0.45	.63	.002
SZT	12.36	.001	.048	24.74	.001	.091	2.64	.07	.011
PAR	26.78	.001	.098	9.74	.001	.038	6.87	.002	.027
BDL	36.82	.001	.130	11.95	.001	.046	15.18	.001	.058
ANT	16.44	.001	.063	4.68	.01	.019	2.06	.13	.008
NAR	29.95	.001	.109	34.90	.001	.124	20.60	.001	.077
HST	23.13	.001	.086	35.64	.001	.127	7.04	.002	.028
DEP	19.25	.001	.073	11.14	.001	.043	13.20	.001	.051
CPS	34.03	.001	.122	29.64	.001	.108	7.73	.001	.030
AVD	9.15	.003	.036	15.86	.001	.061	10.84	.001	.042
PAG	20.29	.001	.076	11.17	.001	.043	1.55	.22	.006
Total PD	49.90	.001	.169	47.17	.001	.161	24.41	.001	.090

\*ANOVA indicates analysis of variance; IPDE, International Personality Disorder Examination; PD, personality disorder; SZD, schizoid; SZT, schizotypal; PAR, paranoid; BDL, borderline; ANT, antisocial; NAR, narcissistic; HST, histrionic; DEP, dependent; CPS, obsessive-compulsive; AVD, avoidant; and PAG, passive-aggressive. Total PD indicates total personality disorder feature dimensional score on the IPDE. Number of cases equals 250 (8 cases with incomplete data omitted from the analysis). The data were analyzed at the level of dimensional scores on the IPDE. Results for the between-subjects main effect of sex as well as all 2-way and 3-way interactions involving the sex factor are omitted given the relative absence of significant results (see text). Degrees of freedom for *F* tests: group (1, 246), time (2, 492), time × group (2, 492). Degrees of freedom reflect the between-subjects sex factor.

†Within-subjects *F* tests evaluated for significance as per Greenhouse and Geisser.<sup>30</sup>

‡ES indicates effect size (partial  $\eta^2$ ), a statistic that describes the proportion of variability in PD features attributable to a factor (group, time, or group × time).

**Table 6. Repeated-Measures ANOVA for PD Features for DSM-III-R Axis II Disorder Dimensions Assessed by the MCMI-II\***

Disorder	Effect								
	Group			Time†			Group × Time†		
	F	P	ES‡	F	P	ES‡	F	P	ES‡
SZD	1.76	.19	.007	4.79	.009	.019	0.96	.38	.004
SZT	58.66	.001	.193	16.18	.001	.062	4.03	.02	.016
PAR	100.36	.001	.290	19.81	.001	.075	3.06	.05	.012
BDL	141.25	.001	.365	15.49	.001	.059	11.35	.001	.044
ANT	83.30	.001	.253	9.93	.001	.039	1.74	.18	.007
NAR	33.28	.001	.119	4.38	.01	.017	0.43	.64	.002
HST	10.49	.001	.041	3.60	.03	.014	0.17	.84	.001
DEP	0.04	.84	.000	14.41	.001	.055	0.61	.54	.002
CPS	2.57	.11	.010	1.07	.34	.004	1.09	.34	.004
AVD	49.92	.001	.169	18.58	.001	.070	11.74	.001	.046
PAG	133.12	.001	.351	20.20	.001	.076	10.00	.001	.039
Total PD	143.23	.001	.362	42.42	.001	.146	8.20	.001	.032

\*ANOVA indicates analysis of variance; MCMI-II, Millon Clinical Multiaxial Inventory II; SZD, schizoid; SZT, schizotypal; PAR, paranoid; BDL, borderline; ANT, antisocial; NAR, narcissistic; HST, histrionic; DEP, dependent; CPS, obsessive-compulsive; AVD, avoidant; PAG, passive-aggressive; and PD, personality disorder. Total PD = overall average of the 11 MCMI-II personality disorder base rate-adjusted dimensions. Number of cases equals 250 (8 cases with incomplete data omitted from the analysis). The data were analyzed at the level of base rate-adjusted dimensional scores on the MCMI-II. Results for the between-subjects main effect of sex as well as all 2-way and 3-way interactions involving the sex factor are omitted given the relative absence of significant results (see text). Degrees of freedom for F tests: group (1, 246), time (2, 492), time × group (2, 492). Degrees of freedom reflect the between-subjects sex factor.

†Within-subjects F tests evaluated for significance as per Greenhouse and Geisser.<sup>30</sup>

‡ES indicates effect size (partial  $\eta^2$ ), a statistic that describes the proportion of variability in PD features attributable to a factor (group, time, or group × time).

The results for the MCMI-II (Table 6) for specific PDs were somewhat less uniform than those observed for the IPDE. Borderline, avoidant, and passive-aggressive MCMI-II PD dimensions followed a longitudinal pattern similar to that seen for the overall average MCMI-II dimensional score (ie, PPD subjects displaying a more rapid decline over time) indicated by a significant time × group interaction. Subjects in the PPD group had higher scores on the schizotypal, paranoid, antisocial, narcissistic, and histrionic PD dimensions (main effects). Time main effects were observed for the schizotypal, paranoid, antisocial, and dependent PD dimensions, indicating a decline for both subject groups over time, with the largest changes typically occurring from wave 1 to wave 2. An effect for time (as a main effect or in interaction with group) was detected for 7 of the 11 PD dimensions assessed by the MCMI-II. The mean effect sizes, for the significant findings, were as follows: group = 0.22, time = 0.06, and group × time interaction = .04.

These analyses were judged to be the most sensitive for evaluating stability because they were based on a dimensional approach to the PDs. In clinical practice, however, categorical methods of classification remain prominent and therefore the stability of the classification of “any definite PD” in these data was examined. The proportion of subjects meeting the criteria for a definite PD of any type (including not otherwise specified) on the IPDE at wave 1 was 7.6%, wave 2 was 5.6%, and wave 3 was 2.8%. The  $\kappa$  for wave 1/wave 2 = 0.45 ( $P < .001$ ). The  $\kappa$  coefficient could not be computed for other comparisons owing to the prevalences falling below 5%,<sup>31</sup> and such categorical analysis, therefore, remains inconclusive. The MCMI-II does not generate categorical diagnoses. Furthermore, the presence of an

Axis I disorder of any type at wave 1 did not significantly moderate PD feature changes over time for either the IPDE or MCMI-II.

#### COMMENT

These initial data from the LSPD suggest that PD features display relatively high levels of individual difference stability and appreciable mean level stability, with some change occurring over time. Categorical diagnoses could not be reliably evaluated for stability owing to a low prevalence, perhaps reflective of the conservative diagnostic thresholds embodied in the IPDE. To date, there are no other published reports or large-scale studies underway that address the full range of Axis II disorders from a prospective multiwave longitudinal perspective. The wave 1 to wave 2 findings of the LSPD are broadly consistent with the prior 2-occasion studies mentioned earlier<sup>6,7</sup>; however, the absence of other multiwave data sets necessarily precludes comparison beyond wave 2 assessments.

The stability of individual differences is relatively high for total PD features assessed by both IPDE and MCMI-II. Differences in the rank order stability of the total PD feature scores across the 2 methods were expected given that the format of the MCMI-II remains constant over time, whereas clinical interviewers on the IPDE change. All 11 disorders revealed substantial stability of individual differences across the 4-year time span. The rank order stability of PD features is comparable to that found for major normal personality trait dimensions, where median stability coefficients range from 0.46 to 0.75.<sup>(1p87)</sup> Finally, the rank order stability coefficients for these PD data would be higher were one to correct for

unreliability of measurement, which serves to underestimate true stability.

For mean level (ie, group) stability, the pattern of results reveals evidence of statistically significant changes in the mean levels of PD features, particularly within the PPD group, over time. However, despite being statistically significant, the changes were relatively small as revealed through the effect size estimates. Overall, the data suggest a relatively high level of stability for PD group means over time for both measures (IPDE, MCMI-II). The greatest changes in the mean levels observed for both measures tended to occur between the wave 1 and wave 2 assessments, a pattern of findings not unusual for prospective multiwave longitudinal studies.<sup>12</sup> Change in mean levels from wave 2 to wave 3 was slight and longitudinal methodologists<sup>12</sup> argue that this portion of the study period is the more accurate index of any real change. This is because changes from wave 1 to wave 2 might reflect some regression to the mean effect, even if relatively small, whereas change during the post-wave 2 assessments is more suggestive of genuine change free of regression to the mean effects.<sup>12</sup>

The overall theoretical implication of these mean level data suggests continuity consistent with the DSM definition of PD. Clinically, these data suggest also that change is possible in personality pathology, which may conceivably be catalyzed through therapeutic efforts. One must remember that these are changes at the group level and some persons within the group may have changed rather notably. What, however, is causing the mean level declines over time, though they are small effects? The observed change is not likely to be explained entirely by a regression to the mean effect because (1) highly reliable measures were used, (2) subjects were assessed on 2 clinical instruments (IPDE, MCMI-II) that were different from the selection instrument (IPDE-S), (3) the variances within each of the 2 groups were broadly comparable over time, and (4) the groups were not at the ceiling or floor on either measure. Moreover, regression to the mean as an effect per se loses much of its significance in a multiwave approach.<sup>12,32</sup> The first-year college experience might have heightened PD symptomatology at wave 1, especially in those subjects who were already somewhat affected. However, broadly comparable patterns of results (ie, small-magnitude declines in group mean levels over time) have been observed for normal personality and such change is viewed as substantively trivial.<sup>1</sup> Continued lifespan follow-up study of these individuals may help to shed further light on this issue of mean level stability and change.

Several caveats should be mentioned in considering these LSPD data. Compared with the US population at large, the LSPD sample is more homogeneous in age, educational achievement, and social class, which may differentially affect the study results. Also, the LSPD was begun when the subjects were first-year university students and some of the most severely affected individuals with PD might have never successfully enrolled in college and therefore would not be included in the study. However, one must recall the Axis I diagnostic data for the LSPD subjects (Table 2), diagnosed according to rigorous clinical thresholds, before ascribing undue levels of mental health to these subjects. Multiwave longitudi-

nal data from clinically based PD samples, mindful of their limitations for generalization, are needed for contrast with the LSPD results. Finally, one could argue that the stability observed for PD features in this study might be artifactual (ie, subjects portray themselves as more consistent than they really are); however, a large methodological literature in the personality assessment domain has rendered such an assertion untenable.<sup>1</sup> The present data will be further dissected using additional approaches to the analysis of stability, including analysis of latent growth curves and factorial invariance, as well as structural equation-based approaches to the analysis of stability and change.<sup>8,10,11,21</sup> William James<sup>33(p121)</sup> claimed that "by the age thirty, the character has set like plaster, and will never soften again," and that indeed seems to be true of normal personality.<sup>1</sup> However, clinical experience suggests that some PD features may diminish with age. Continued lifespan study of the LSPD subjects will offer the opportunity to determine if James' view also holds for PD. Finally, this study is best viewed as heuristic and hypothesis-generating, undertaken within a context of exploration rather than a context of justification or confirmation.

Accepted for publication July 28, 1999.

This research was supported in part by grant MH45448 from the National Institute of Mental Health, Washington, DC (Dr Lenzenweger).

Preliminary results of this study were presented at the Fifth Meeting of the International Congress on the Disorders of Personality, International Society for the Study of Personality Disorders, University of British Columbia, Vancouver, June 25, 1997; and at the 13th Annual Meeting of the Society for Research in Psychopathology, Harvard University, Cambridge, Mass, November 14, 1998.

I thank Margaret Dyer, MSW; Linda Foltz, PhD; Mickey Goldstein, MSW; Joan Lovejoy, MSW; Eileen Maxwell, MSW; Cynthia Neff, PhD; Judy Raabe, PhD; Carol Skinner, MSW; and Ba Stopha, MSW for their able assistance in conducting clinical assessments. I thank Stephen Cornelius, PhD, and Robert Rosenthal, PhD, for useful statistical consultations and Jerome Kagan, PhD, for discussions regarding stability concepts. I also thank Lauren Korfine, PhD; Armand W. Loranger, PhD; Theodore Millon, PhD; and Paul T. Costa, PhD, for their comments on an earlier version of this article.

Reprints: Mark F. Lenzenweger, PhD, Department of Psychology, Harvard University, 33 Kirkland St, Cambridge, MA 02138.

## REFERENCES

1. McCrae RR, Costa PT. *Personality in Adulthood*. New York, NY: Guilford Press; 1990.
2. Funder DC, Parke RD, Tomlinson-Keasey C, Widaman K. *Studying Lives Through Time: Personality and Development*. Washington, DC: American Psychological Association; 1993.
3. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*. Washington, DC: American Psychiatric Association; 1994.
4. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Third Edition*. Washington, DC: American Psychiatric Association; 1980.

5. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Revised Third Edition*. Washington, DC: American Psychiatric Association; 1987.
6. Perry JC. Longitudinal studies of personality disorders. *J Personal Disord*. 1993; 7(suppl):63-85.
7. McDavid JD, Pilkonis PA. The stability of personality disorder diagnoses. *J Personal Disord*. 1996;10:1-15.
8. Kessler RC, Greenberg DF. *Linear Panel Analysis: Models of Quantitative Change*. New York, NY: Academic Press Inc; 1981.
9. Rogosa D. Myths about longitudinal research. In: Schaie KW, Campbell RT, Meredith W, eds. *Methodological Issues in Aging Research*. New York, NY: Springer Publishing Co Inc; 1988.
10. Rogosa D, Brandt D, Zimowski M. A growth curve approach to the measurement of change. *Psychol Bull*. 1982;92:726-748.
11. Nesselroade JR, Baltes PB, eds. *Longitudinal Research in the Study of Behavior and Development*. New York, NY: Academic Press Inc; 1979.
12. Nesselroade JR, Stigler SM, Baltes PB. Regression toward the mean and the study of change. *Psychol Bull*. 1980;88:622-637.
13. Loranger AW, Sartorius N, Andreoli A, Berger P, Buchheim P, Channabasavanna SM, Coid B, Dahl A, Diekstra RFW, Ferguson B, Jacobsberg LB, Mombour W, Pull C, Ono Y, Regier D. The International Personality Disorder Examination (IPDE): the World Health Organization/Alcohol, Drug Abuse, and Mental Health Administration International Pilot Study of Personality Disorders. *Arch Gen Psychiatry*. 1994;51:215-224.
14. 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.
15. Trull TJ. Temporal stability and validity of two personality disorder inventories. *Psychol Assess*. 1993;5:11-18.
16. Menard S. *Longitudinal Research*. Newbury Park, Calif: Sage Publications Inc; 1991.
17. Kagan J. Perspectives on continuity. In: Brim O, Kagan J, eds. *Constancy and Change in Human Development*. Cambridge, Mass: Harvard University Press; 1980.
18. Alwin DF. Aging, personality, and social change: the stability of individual differences over the adult life span. In: Featherman DL, Lerner RM, Perlmutter M, eds. *Life-Span Development and Behavior*. Vol 12. Hillsdale, NJ: Lawrence A Erlbaum Associates; 1994.
19. Caspi A, Bem DJ. Personality continuity and change across the life course. In: Pervin L, ed. *Handbook of Personality: Theory and Research*. New York, NY: Guilford Press; 1990.
20. Collins LM, Horn JL, eds. *Best Methods for the Analysis of Change: Recent Advances, Unanswered Questions, Future Directions*. Washington, DC: American Psychological Association; 1991.
21. McArdle JJ, Epstein D. Latent growth curves within developmental structural equation models. *Child Dev*. 1987;58:110-133.
22. Lenzenweger MF, Loranger AW, Korfine L, Neff C. Detecting personality disorders in a nonclinical population: application of a 2-stage procedure for case identification. *Arch Gen Psychiatry*. 1997;54:345-351.
23. Loranger AW, Sartorius N, Janca A, eds. *Assessment and Diagnosis of Personality Disorders: The International Personality Disorder Examination (IPDE)*. New York, NY: Cambridge University Press; 1996.
24. World Health Organization. *International Classification of Diseases, 10th Revision (ICD-10)*. Geneva, Switzerland: World Health Organization; 1992.
25. Spitzer RL, Williams JBW, Gibbon M, First M. *User's Guide for the Structured Clinical Interview for DSM-III-R*. Washington, DC: American Psychiatric Press; 1990.
26. Millon T. *Millon Clinical Multiaxial Inventory II Manual*. Minneapolis, Minn: National Computer Systems Inc; 1987.
27. Choca JP, Van Denburg E. *Interpretive Guide to the Millon Clinical Multiaxial Inventory*. 2nd ed. Washington, DC: American Psychological Association; 1997.
28. Meng X-L, Rosenthal R, Rubin DB. Comparing correlated correlation coefficients. *Psychol Bull*. 1992;111:172-175.
29. *SPSS for Windows Advanced Statistics Release 6.1*. Chicago, Ill: SPSS Inc; 1994.
30. Greenhouse SW, Geisser S. On methods in the analysis of profile data. *Psychometrika*. 1959;24:95-112.
31. Grove WM, Andreasen NC, McDonald-Scott P, Keller MB, Shapiro RW. Reliability studies of psychiatric diagnosis: theory and practice. *Arch Gen Psychiatry*. 1981;38:408-413.
32. Baltes PB, Reese HW, Nesselroade JR. *Life-span Developmental Psychology: Introduction to Research Methods*. Monterey, Calif: Brooks/Cole Publishing Co; 1977.
33. James W. *Principles of Psychology*. Vol 1. New York, NY: Dover Publications Inc; 1950. Originally published 1890.