Ten-year stability and latent structure of the DSM-IV schizotypal, borderline, avoidant, and obsessive-compulsive personality disorders

Charles A. Sanislow
Todd D. Little
Carlos M. Grilo
Maria Daversa
John C. Markowitz, et al.

Available at: https://works.bepress.com/charles_sanislow/80/
Ten-Year Stability and Latent Structure of the DSM–IV Schizotypal, Borderline, Avoidant, and Obsessive-Compulsive Personality Disorders

Charles A. Sanislow  
Yale University School of Medicine

Todd D. Little  
University of Kansas and Schiefelbusch Institute for Life Span Studies

Emily B. Ansell and Carlos M. Grilo  
Yale University School of Medicine

Maria Daversa  
Harvard University Medical School and McLean Hospital

John C. Markowitz and Anthony Pinto  
Columbia University and New York State Psychiatric Institute

M. Tracie Shea  
Veterans Affairs Medical Center, Providence, Rhode Island, and Brown University Alpert Medical School

Shirley Yen  
Brown University

Andrew E. Skodol  
University of Arizona College of Medicine and the Sunbelt Collaborative

Leslie C. Morey  
Texas A & M University

John G. Gunderson and Mary C. Zanarini  
Harvard University Medical School and McLean Hospital

Thomas H. McGlashan  
Yale University School of Medicine

Evaluation of the validity of personality disorder (PD) diagnostic constructs is important for the impending revision of the Diagnostic and Statistical Manual of Mental Disorders. Prior factor analytic studies have tested these constructs in cross-sectional studies, and models have been replicated longitudinally, but no study has tested a constrained longitudinal model. The authors examined 4 PDs in the Collaborative Longitudinal Personality Disorders study (schizotypal, borderline, avoidant, and obsessive-compulsive) over 7 time points (baseline, 6 months, 1 year, 2 years, 4 years, 6 years, and 10 years). Data for 2-, 4-, 6- and 10-year assessments were obtained in semistructured interviews by raters blind to prior PD diagnoses at each assessment. The latent structure of the 4 constructs was differentiated during the initial time points but became less differentiated over time as the mean levels of the constructs dropped and stability increased. Obsessive-compulsive PD became more correlated with schizotypal and borderline PD than with avoidant PD. The higher correlation among the constructs in later years may reflect greater shared base of pathology for chronic personality disorders.

Keywords: schizotypal, borderline, avoidant, obsessive-compulsive, personality disorders

This article was approved by the publication committee of the Collaborative Longitudinal Personality Disorders Study. Funding for the Collaborative Longitudinal Personality Disorders Study is provided by National Institutes of Health Grants MH050837 to Brown University Department of Psychiatry and Human Behavior, MH050839 to Columbia University and New York State Psychiatric Institute, MH050840 to Harvard Medical School and McLean Hospital, MH050838 to Texas A&M University, and MH050850 to Yale University School of Medicine; further support for this study was provided by National Institutes of Health Grants MH079078 to John C. Markowitz, MH080221 to Anthony Pinto, MH073708 to Charles A. Sanislow, and MH069904 to Shirley Yen. Correspondence concerning this article should be addressed to Charles A. Sanislow, who is now at the Department of Psychology, Wesleyan University, 207 High Street, Middletown, CT 06459. E-mail: csanislow@wesleyan.edu
The Diagnostic and Statistical Manual of Mental Disorders (DSM) defines personality disorders (PDs) as stable and enduring, reflecting a persistent pattern of maladaptive personality throughout the life course. Approaches used to evaluate PD construct validity include testing this stability assumption by examining time to remission for PD diagnosis, stability of criteria within the diagnosis, and factor structure of the PD diagnostic constructs or clusters. Prospective tests of stability by several research groups, including our own Collaborative Longitudinal Study of Personality Disorders (CLPS), have shown that PDs tend to remit at rates higher than the DSM definition implies (e.g., Grilo et al., 2004; Shea et al., 2002; see also Laptook, Klein, & Dougherty, 2006; Zanarini, Frankenburg, Hennen, & Silk, 2003; Zanarini et al., 2007). Individual variability of PDs across time has also emerged from nonclinical samples (Lenzenweger, Johnson, & Willett, 2004).

A second approach to testing the validity of DSM PDs is to examine their latent structure. Here, results have been mixed. Evidence from other studies supports the DSM constructs and suggests that subdiagnostic levels of PD pathology have prognostic value. Using latent class analyses, Clifton and Pilkonis (2007) identified a group of individuals with subclinical borderline PD (BPD) diagnoses who more closely resembled individuals meeting full diagnostic criteria in their social-interpersonal and occupational functioning than they did non-BPD participants. Dimensional scoring of DSM criteria has been shown to more accurately predict psychosocial functioning than do PD categories (Skodol, Oldham, et al., 2005). In some studies, DSM PD diagnoses appear more stable when examined dimensionally than when examined categorically (e.g., Morey et al., 2007). Further, tests of the stability of the relative order of PD criteria suggest that individuals remain consistent in rank order of criteria over time, even when they fluctuate in severity or number of PD features (Grilo et al., 2004). Together, these findings modestly support the validity of the PD constructs.

In several factor analytic studies with exploratory or confirmatory factor approaches, researchers have examined the DSM PD constructs. These studies have mainly addressed the three PD clusters: Cluster A is odd–eccentric (paranoid, schizoid, and schizotypal), Cluster B is dramatic–emotional–erratic (antisocial, borderline, histrionic, and narcissistic), and Cluster C is anxious–fearful (avoidant, dependent, and obsessive compulsive; American Psychiatric Association [APA], 1996). One notable inconsistency is whether loadings of constructs for the individual disorders empirically conform to the three clusters specified by the DSM. Although support for a three-factor solution was found in a nonclinical population, loadings did not correspond to the three DSM clusters: Cluster A is odd–eccentric (paranoid, schizoid, and schizotypal), Cluster B is dramatic–emotional–erratic (antisocial, borderline, histrionic, and narcissistic), and Cluster C is anxious–fearful (avoidant, dependent, and obsessive compulsive; American Psychiatric Association [APA], 1996). Using data obtained from clinician ratings of PDs in adolescents, Durrett and Westen (2005) found support for the individual PD diagnostic constructs but not for the three-cluster organization. Bell and Jackson (1992), attempting to fit a three-factor solution to an inpatient clinical sample, found that although fit statistics were less than optimal, the data best corresponded to the three DSM clusters. O’Connor and Dyce (1998) did a comparative analysis using self-report measures of personality as well as DSM diagnoses in a clinical sample carefully selected for a broad range of pathology. They concluded that five- and seven-factor models fit the data better than did the three-cluster DSM model. However, there was no incremental gain in model fit beyond four factors. In sum, there is little empirical consensus for the optimal number of higher order factors for personality pathology.

Several studies raised a second, more specific question, namely whether obsessive–compulsive PD (OCPD) stands apart from the three clusters (e.g., Hylar & Lyons, 1988; Kass, Skodol, Charles, Spitzer, & Williams, 1985; Livesley, Jackson, & Schroeder, 1992; Livesley, Jang, & Vernon, 1998; Nestadt et al., 1994; Tyrer & Alexander, 1979). Morey (1986), however, demonstrated that a three-cluster solution including OCPD could be forced with Procrustean procedures with the Kass et al. (1985) data. In an Italian patient sample, Fossati et al. (2000) found results supporting three factors, but only one (odd–eccentric) aligned with the Diagnostic and Statistical Manual of Mental Disorders (4th ed.; DSM–IV; APA, 1996) clusters. Yang, Bagby, Costa, Ryder, and Herbst (2002) tested the DSM–IV cluster structure in a sample of Chinese patients, and their results did not support the three clusters. Rodebaugh, Chambless, Renneberg, and Fydrich (2005) analyzed a combination of archival data sets, and their results revealed better support for a three-factor model than for one-factor model with confirmatory tests. Finally, O’Connor (2005) reported good fit for a three-factor model with data based on the five-factor model approach (Neuroticism, low Agreeableness, Extroversion/Introversion) but found that adding a fourth factor capturing Conscientiousness and obsessive–compulsive features better accounted for the data. In sum, patient studies testing the DSM three-cluster model have produced mixed results.

The CLPS research group has also studied the covariance structures of PD constructs and found some support for the DSM constructs (Sanislow et al., 2002). Specifically, we tested a four-factor structural model corresponding to the DSM–IV diagnostic constructs schizotypal PD (STPD), BPD, avoidant PD (AVPD), and OCPD. Results showed good model fit at baseline (Year 0) and again at 2-year follow-up, based on diagnoses made by raters blind to baseline diagnoses (Sanislow et al., 2002). Although these findings supported the structure implied by the DSM diagnostic constructs, they did not uphold the relative weighting implicit in the ordering of symptom criteria within each diagnostic construct (APA, 1996). More recent work has also demonstrated inconsistencies in the criterion hierarchy for BPD (Karterud, Pedersen, Gude, & Falkum, 2004). However, variation in DSM criteria hierarchies across studies is not necessarily surprising, given that some criteria serve different functions for the constructs (e.g., predictive of the construct versus evidencing stability). Further, it seems reasonable to expect these functions, as well as the relation of the criteria to the constructs, to vary depending on the population sampled. For instance, suicidal behavior may predict a poorer outcome in a clinical sample than in a non–treatment-seeking sample. In a review of PD factor studies, Sheets and Craighead (2007) concluded that studies testing DSM structure with nonpatient community samples generally showed less support for the DSM than did those with patient populations.

Evaluating PD validity by testing stability at the level of the diagnostic construct can supersede these influences and fluctuations. Here, we examine the stability of four PD constructs longitudinally as well as their overlap with the other PD constructs. We extend prior work (Sanislow et al., 2002) in two ways. First, we test the stability of the four CLPS PDs (STPD, BPD, AVPD, and OCPD) over a longer, 10-year interval. Second, we examine the
PD constructs across seven assessment points in a single longitudinal model. That is, rather than testing separate models at each time point (cf. Sanislow et al., 2002), we tested a single panel model using the entire 10-year CLPS sample to directly evaluate the stability of the constructs. PDs were modeled in a large treatment-seeking sample at seven time points: Year 0 (baseline at study entry), Month 6, and Year 1, then Year 2, Year 4, Year 6, and Year 10. Participants entered the study with a primary PD diagnosis of STPD, BPD, AVPD, or OCPD or with a diagnosis of major depressive disorder with no PD. Participants targeted for one of the four PDs were not excluded for the presence of comorbid PDs. To examine the joint characteristics of change across time, growth curves were estimated. The panel model allowed estimation of the stability of the individual constructs, whereas the growth curve model allowed estimation of the nature of the specific characteristics of change over time. Models were controlled for demographic characteristics of age, sex, and race. On the basis of the DSM–IV premise of stability and distinctiveness, we hypothesized that compared with diagnostic approaches based on criterion cutoffs, the DSM constructs STPD, BPD, and AVPD would show stability within constructs and discriminant validity between constructs in the omnibus model. Given prior conflicting findings, we were uncertain whether the OCPD construct would demonstrate greater associations over time with the other Cluster C disorder, AVPD.

Method

Participants

Participants aged 18–45 years at study entry were evaluated as part of the CLPS. The CLPS is a prospective, repeated measures study that examined the course of PDs. For a more detailed description of the study design and aims, see Gunderson et al. (2000); for sample characteristics, see McGlashan et al. (2000). Primarily treatment-seeking participants at inpatient or outpatient facilities who were or had recently been in psychiatric treatment or psychotherapy were sampled for four representative PDs (borderline, schizotypal, avoidant, and obsessive-compulsive); a control group meeting criteria for major depressive disorder but no PD was also included. Media advertising and postings supplemented recruitment. Potential participants were prescreened to determine age eligibility and treatment status or history and to assist in excluding patients with active psychosis, acute substance intoxication or withdrawal, a history of schizophrenia-spectrum psychosis (i.e., schizophrenia, schizophreniform, or schizoaffective disorders), or organicity. The sample comprised 733 participants. The original cohort consisted of 668 participants followed through 10 years. The supplemental cohort of 65 additional minority participants, which was sampled to provide a more representative racial base, was not followed beyond Year 4 (Year 6 and Year 10 values were imputed as described below). The sample was 69% Caucasian, 15% African American, and 13% Hispanic, with the remainder from other ethnic backgrounds; 64% were women and 36% were men. All participants provided informed, written consent to study procedures prior to entry.

Three disorders were chosen to represent the DSM–IV Axis II Clusters A, B, and C (STPD, BPD, and AVPD, respectively). The fourth disorder, OCPD, was included because evidence suggested it might stand apart from the three clusters. The four targeted PD diagnoses and the treatment-seeking sample were drawn from varied settings that provided a spectrum of PD pathology, a distribution enhanced by the major depressive disorder contrast group. Presence of other PDs was not an exclusion criterion, and participants received 2.1 Axis II diagnoses on average, a rate comparable with other clinical studies (e.g., Blashfield, McElroy, Pfohl, & Blum, 1994; Oldham et al., 1995; Stuart et al., 1998). Further, participants’ treatment-seeking status provided an ecologically valid study group.

Assessment

Extensively trained research interviewers with master’s or doctoral degrees assessed all participants, and researchers were monitored for ongoing reliability. The Structured Clinical Interview for DSM–IV Axis I Disorders—Patient Version (SCID-I/P; First, Spitzer, Gibbon, & Williams, 1996) was used to assess Axis I disorders, and the Diagnostic Interview for DSM–IV Personality Disorders (DIPD-IV; Zanarini, Frankenburg, Sickel, & Yong, 1996) was used for Axis II disorders. The DIPD-IV is a semistructured diagnostic interview containing several questions pertaining to each DSM–IV Axis II criterion. Each criterion is scored 0 for absent, 1 for present but of uncertain clinical significance, or 2 for present and clinically significant. In our sample, median kappa coefficients (Cohen, 1960) ranged from .69 to .97 for all Axis II disorders (Zanarini et al., 2000). The DIPD-IV was administered at baseline (Year 0). The Month 6 and Year 1 assessments for the four PDs had a modified version of the DIPD-IV, the DIPD-Follow Along Version (FAV; Zanarini & Shea, 1996), in which ratings are made on a scale with 0, 1, or 2 for each month during the time period being queried. Reliability on the DIPD-IV-FAV based on the rating of two overlapping time points (Month 6 was rated twice for 453 cases) resulted in kappa coefficients of .78 for STPD, .70 for BPD, .73 for AVPD, and .68 for OCPD (see Shea et al., 2002). The Month 6 and Year 1 assessments were followed by blind assessments with the DIPD-IV; interviewers had no knowledge of participants’ PD diagnostic status from prior interviews at Years 2, 4, 6, and 10.

Analyses

Structural equation modeling (SEM; Hoyle & Smith, 1994) was used with LISREL 8.80 software (Jöreskog & Sörbom, 2008) with maximum likelihood estimation. Latent variables were computed for each of the four PDs—STPD, BPD, AVPD, and OCPD—to address our key question regarding the stability of the constructs these DSM–IV diagnoses represented. To represent the constructs, the indicators (i.e., individual PD criteria) were parceled following recommendations by Kishton and Widaman (1994; see also Little, Cunningham, Shahar, & Widaman, 2002). For the ratings obtained with the DIPD-IV-FAV, parcels were derived from the averaged values of each criterion over the 6 month assessment interval. parceling items offers several advantages over use of individual criteria as indicators that are pertinent to our goals (Little et al., 2002). Aggregate sets of items produce indicators that are more likely to have continuous properties and to have a more normal distribution that better fulfills the maximum likelihood assumptions than do nonparceled criteria. For model estimation, parcels
require fewer parameter estimates than do models that use the items individually. Finally, parcels are more reliable than items; hence, error variances of the parceled sets of items are smaller than are the items themselves.

To construct parcels for the present study, preliminary analyses of the items were carried out to determine the optimal balanced groupings of items. Three parcels were formed empirically for each of the four disorders, so that each grouping of averaged items evenly represented the common variance of the construct (i.e., all parcels exhibited an evenly distributed range of intercorrelations among the component criteria used to compose the parcels; see Little et al., 2002, for details of creating balanced parcels). The composition of the parcels, held constant across the assessment time points, is shown in Table 1.

### Missing Data

Across the 10 years of the study, only 16.4% of the overall data was missing, meeting acceptable standards of less than 20%, to use modern imputation procedures (Schafer & Graham, 2002). For this study, we used the SAS procedure Proc MI (Version 9.12) to address missing data, specified the expectation maximization (EM) algorithm to establish prior estimates, and used the Markov chain Monte Carlo procedure (MCMC) to impute missing values. The imputation process included sex, race, and age as well as all diagnostic variables and a participant’s membership in the original cohort or the supplemental minority sample (including all appropriate interaction terms). The imputation procedure was run 100 times to ensure maximal generalizability, given the presence of missing data (Enders, in press).

### Evaluation of Model Fit

Three fit indices were used to evaluate model fit, each offering certain advantages: the root-mean-square–error of approximation (RMSEA; Steiger, 1990), the nonnormed fit index (Bentler, 1990), and the comparative fit index (CFI; Bentler, 1990). The RMSEA accounts for model parsimony when evaluating model fit. Values less than .08 indicate good fit (Browne & Cudeck, 1993). The nonnormed fit index and CFI both measure fit relative to the appropriate longitudinal and/or multiple-group null model (i.e., assuming no relationships or zero correlations among model indicators, no changes or differences in indicator means, and no changes or differences in indicator variances; Widaman & Thompson, 2003) with values above .90 generally considered a good fit and those over .95 considered an excellent fit.

### Comparative Tests of Model Fit

To evaluate the comparative model fit, we used the maximum likelihood chi-square statistic to test for factorial similarities and differences across groups or across time in the form of nested-model comparisons. Because the chi-square difference test is overly sensitive to large sample sizes, we took appropriate, conservative measures. For instances in which the reliable structural components were being evaluated and the chi-square difference test is appropriate, concerns of excessive power were addressed by adopting a more stringent p value. For omnibus chi-square difference tests, we adopted a value of .005 (see Little & Slegers, 2005). For invariance tests, concerns arise over evaluating the invariance of the measurement parameters when a large number of parameter estimates are involved. Because these parameters reflect the fallible aspects of measurement (i.e., the loadings, residuals, intercepts), it is recommended in these instances that model invariance be evaluated with model-based information rather than an omnibus test (Cheung & Rensvold, 2002; Little, 1997). Therefore, when evaluating the tenability of the invariance constraints, we used two recommended criteria: (a) a change in CFI less than .01 and (b) the point estimate of the RMSEA falling within the confidence interval of the preceding model (see Cheung & Rensvold, 2002; Little, Bovaird, & Slegers, 2006).

### Progression of Modeling

Our primary goal was to evaluate the factorial structure, construct stability, and intraindividual change patterns among the four PDs (STPD, BPD, AVPD, and OCPD) spanning 10 years of longitudinal data. The first set of analyses tested factorial invariance...
ance across the seven time points (baseline or Year 0, Month 6, Year 1, Year 2, Year 4, Year 6, Year 10), with the expected factorial structure specified at each assessment point. In addition, the four-factor model was tested across the seven waves of data separately for male and female participants, to determine whether different models for each sex were warranted. In the next step, factorial invariance for the sample as a whole was tested to discern stability characteristics of the four PDs. This first set of analyses examined correlations among the disorders within each assessment as well as across the four time points. A second set of analyses specified growth curves to more clearly illustrate the relative intraindividual change patterns of each disorder.

Results

Part 1: Evaluation of Factorial Invariance

Table 2 shows model fit statistics for the progression of tests to examine factorial invariance. As indicated in Table 2, the progression of models showed excellent model fit (see Models 1–3). Moreover, inspection of the residuals and modification indices indicated that no further estimates would improve model fit. Specifically, the criteria for evaluating the steps of factorial invariance (i.e., a change in CFI less than .01 and the point estimate of the RMSEA falling within the confidence interval of the prior model) were well satisfied, indicating strong invariance across time and sex.

Next, we examined whether the correlations among the PD constructs were the same across men and women by testing whether the correlations among the PD constructs at each wave are the same across men and women (see Table 3). The chi-square difference test was significant, $\chi^2(378, N = 733) = 754.34, p < .0001$, indicating that there are sufficient differences among the correlations across the constructs for men and women to examine the longitudinal patterns separately by sex. In addition to the correlation differences, the omnibus test for any mean difference on any of these diagnostic constructs was significant, with a nested chi-square difference of $\chi^2(28, N = 733) = 109.74, p < .0001$ (see Table 3). Table 4 and Table 5 present the latent correlations, variances, and means for women and men, respectively.

Overall, the correlations among the constructs within each time point were quite discrete at baseline. For example, the highest correlation of .46 was between STPD and BPD in women, indicating that less than 20% of the variance was shared. These correlations do not change appreciably in women at 6 months or 1 year, but by Year 2, the OCPD diagnosis ratings begin to show modest correlations with the other diagnostic categories. In general, these correlations increase over time. The association between STPD and BPD also shows a steady increase for women, with correlations of around .45 during the first 2 years increasing to .73 (over 50% variance overlapping) by Year 10. The pattern for men was generally similar but with most correlations being somewhat smaller than for women. For example, the correlation between STPD and BPD is .35 at baseline (compared with .46 in women) and increases to .60 by Year 10. These longitudinal changes in the strength of the correlations were significant for both men and women ($p < .0001$; see Table 3).

Because the correlations among the constructs were statistically different for women and men, we examined the longitudinal stability relationships separately for men and women. As evidenced in Tables 4 and 5, the cross-time stability correlations of each construct were reasonably high. These correlations are generally .7 or greater when the time span between is around 2 years (and the 4-year span between Year 6 and Year 10 shows similar levels of stability). When estimated as predictive (autoregressive) relationships over time, the indirect stability coefficients (i.e., when one or more time points separate the measurement occasions) remain quite high for most time points (see Table 6). Note that the levels of stability when separated by the same amount of time are about the same, regardless of the specific time point of assessment. This pattern is consistent with a steady change process. Moreover, the indirect stability over the whole time span of the study (i.e., from baseline to Year 10) remains significant ($p < .0001$). In fact, all of the indirect effects are significant ($p < .0001$), indicating reliable stability in these indirect pathways over the 10 years span of the study.

Part 2: Evaluation of Change Relationships and Participant Characteristics

To examine the change relationships in mean levels, we fit simultaneous latent growth curve models to the four constructs. Fit statistics for the growth curves, which are shown in Table 2 (Model 4), were in the very good range. We then tested the similarities and the differences in the trajectories across men and

<table>
<thead>
<tr>
<th>Model number</th>
<th>Model description</th>
<th>Chi square</th>
<th>df</th>
<th>RMSEA</th>
<th>Lower 90% confidence interval</th>
<th>Upper 90% confidence interval</th>
<th>Nonnormed fit index</th>
<th>Comparative fit index</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Configural invariance (no constraints) over time or across sex</td>
<td>6,256.51</td>
<td>5,544</td>
<td>.019</td>
<td>.016</td>
<td>.021</td>
<td>.995</td>
<td>.996</td>
</tr>
<tr>
<td>2</td>
<td>Weak invariance (loadings invariant) across sex and time</td>
<td>6,422.41</td>
<td>5,648</td>
<td>.019</td>
<td>.017</td>
<td>.022</td>
<td>.995</td>
<td>.996</td>
</tr>
<tr>
<td>3</td>
<td>Strong invariance (loadings and intercepts) across sex and time</td>
<td>6,670.36</td>
<td>5,752</td>
<td>.021</td>
<td>.018</td>
<td>.023</td>
<td>.994</td>
<td>.995</td>
</tr>
<tr>
<td>4</td>
<td>Fit of the growth curve model for seven time points across men and women</td>
<td>1,535.55</td>
<td>760</td>
<td>.051</td>
<td>.047</td>
<td>.055</td>
<td>.968</td>
<td>.978</td>
</tr>
</tbody>
</table>

Note. RMSEA = root-mean-square error of approximation; df = degree of freedom.
women (see Table 2 for test results). We found that the proportions of change from time point to time point were functionally the same across men and women for all four constructs, \( \chi^2(20, N = 733) = 34.72, p = .043 \). When we tested whether the magnitude of these mean-level changes were similar or different for women and men, we found that the mean level at baseline of BPD and STPD were different and the mean change in STPD was different (see Table 3). The estimated trajectories for the growth curves are shown in Figure 1A (AVPD and OCPD) and Figure 1B (STPD and BPD, broken down by sex). Overall and as anticipated, the mean level of criteria dropped significantly over time (i.e., in the remitted direction) for all disorders (see Figure 1A and B, and Tables 4 and 5). This drop was most pronounced in the early years (e.g., Year 2 to Year 4) and then tended to level off. The STPD scores showed the least number of mean-level changes but did show sex differences in the change pattern. More pronounced drops in the means levels were found for BPD, AVPD, and OCPD, with the most pronounced drops seen between baseline and Year 2.

### Discussion

Our work extends prior efforts to evaluate PD constructs by testing their latent structure longitudinally. In contrast to prior work, the stability of four CLPS DSM–IV PD constructs (AVPD, BPD, STPD, and OCPD) was tested in a single longitudinal model at seven measurement points over a 10 year period. Thus, the latent structure of the constructs was examined in the context of longitudinal stability, a key component of the PDs as they are currently structured. The PD constructs best represent a personality pathology that is both enduring and distinct. Thus, if PDs are to retain the designation of “enduring patterns” (APA, 1996, p. 630) in the Diagnostic and Statistical Manual of the Mental Disorders (5th ed.; DSM–V), consideration of the increased correlation among the diagnoses is warranted.

Regarding stability, it is important to distinguish the significant statistical stability of the model from clinically meaningful stability. The general trajectories of the latent growth curve models showed a lessening of the constructs (i.e., in the direction of less pathology). The patterning of the indirect effects suggests that the majority of change in the constructs occurred early on, much of it during the first year. All four PD constructs exhibited increased stability in later years (Year 4 to Year 10). However, the mean levels of the constructs were much lower during these later time points and suggest a clinically significant reduction in pathology. Thus, only some aspect of each construct endures. However, due to the heterogeneity of the criteria as well as limitations imposed by the polythetic scoring system (i.e., different combinations of criteria can represent the diagnosis), it is not possible to tease this out with the present approach. We have suggested elsewhere, however, that some aspects of PDs may be more traitlike and enduring, whereas other aspects may be episodic in nature. For instance, affect-related criteria found in BPD are less likely to remit over time than are behavioral criteria (Zanarini et al., 2007) and are more frequently endorsed at later follow-up assessments (McGlashan et al., 2005; see also Sanislow & McGlashan, 1998).

The finding of lower mean levels on the constructs over the long term is interesting in light of the apparent disjunction between PD diagnoses, which appear to be less stable relative to their functional impairment (e.g., Skodol et al., 2002; Skodol, Pagnani et al., 2005). Comparing BPD with an Axis II contrast group, Zanarini and colleagues (2005) noted that some improvement in psychosocial functioning was associated with BPD remission status, although vocational deficits were still pronounced (Zanarini, Fran...
Table 4
Latent Correlations, Variances, and Means for Women

<table>
<thead>
<tr>
<th>Measure</th>
<th>Baseline</th>
<th>Month 6</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 4</th>
<th>Year 6</th>
<th>Year 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST</td>
<td>—</td>
<td>—</td>
<td>46</td>
<td>37</td>
<td>37</td>
<td>14</td>
<td>03</td>
</tr>
<tr>
<td>BP</td>
<td>46</td>
<td>31</td>
<td>25</td>
<td>12</td>
<td>04</td>
<td>22</td>
<td>13</td>
</tr>
<tr>
<td>AV</td>
<td>37</td>
<td>37</td>
<td>37</td>
<td>12</td>
<td>36</td>
<td>45</td>
<td>36</td>
</tr>
<tr>
<td>OC</td>
<td>14</td>
<td>09</td>
<td>03</td>
<td>02</td>
<td>04</td>
<td>02</td>
<td>02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measure</th>
<th>Baseline</th>
<th>Month 6</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 4</th>
<th>Year 6</th>
<th>Year 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST</td>
<td>85</td>
<td>31</td>
<td>25</td>
<td>12</td>
<td>04</td>
<td>22</td>
<td>13</td>
</tr>
<tr>
<td>BP</td>
<td>40</td>
<td>85</td>
<td>36</td>
<td>04</td>
<td>42</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>AV</td>
<td>33</td>
<td>36</td>
<td>82</td>
<td>05</td>
<td>36</td>
<td>45</td>
<td>36</td>
</tr>
<tr>
<td>OC</td>
<td>14</td>
<td>01</td>
<td>08</td>
<td>86</td>
<td>10</td>
<td>09</td>
<td>14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measure</th>
<th>Baseline</th>
<th>Month 6</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 4</th>
<th>Year 6</th>
<th>Year 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST</td>
<td>81</td>
<td>30</td>
<td>25</td>
<td>12</td>
<td>04</td>
<td>22</td>
<td>13</td>
</tr>
<tr>
<td>BP</td>
<td>43</td>
<td>77</td>
<td>33</td>
<td>04</td>
<td>41</td>
<td>43</td>
<td>09</td>
</tr>
<tr>
<td>AV</td>
<td>32</td>
<td>32</td>
<td>79</td>
<td>06</td>
<td>32</td>
<td>39</td>
<td>90</td>
</tr>
<tr>
<td>OC</td>
<td>08</td>
<td>03</td>
<td>08</td>
<td>86</td>
<td>10</td>
<td>09</td>
<td>14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measure</th>
<th>Baseline</th>
<th>Month 6</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 4</th>
<th>Year 6</th>
<th>Year 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST</td>
<td>73</td>
<td>23</td>
<td>26</td>
<td>12</td>
<td>04</td>
<td>22</td>
<td>13</td>
</tr>
<tr>
<td>BP</td>
<td>38</td>
<td>70</td>
<td>34</td>
<td>11</td>
<td>34</td>
<td>43</td>
<td>12</td>
</tr>
<tr>
<td>AV</td>
<td>30</td>
<td>28</td>
<td>70</td>
<td>06</td>
<td>32</td>
<td>40</td>
<td>82</td>
</tr>
<tr>
<td>OC</td>
<td>11</td>
<td>05</td>
<td>08</td>
<td>80</td>
<td>10</td>
<td>09</td>
<td>14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measure</th>
<th>Baseline</th>
<th>Month 6</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 4</th>
<th>Year 6</th>
<th>Year 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST</td>
<td>64</td>
<td>24</td>
<td>23</td>
<td>09</td>
<td>34</td>
<td>53</td>
<td>15</td>
</tr>
<tr>
<td>BP</td>
<td>38</td>
<td>56</td>
<td>23</td>
<td>08</td>
<td>38</td>
<td>44</td>
<td>23</td>
</tr>
<tr>
<td>AV</td>
<td>33</td>
<td>29</td>
<td>63</td>
<td>03</td>
<td>29</td>
<td>38</td>
<td>75</td>
</tr>
<tr>
<td>OC</td>
<td>18</td>
<td>16</td>
<td>13</td>
<td>56</td>
<td>23</td>
<td>14</td>
<td>64</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measure</th>
<th>Baseline</th>
<th>Month 6</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 4</th>
<th>Year 6</th>
<th>Year 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST</td>
<td>61</td>
<td>29</td>
<td>29</td>
<td>14</td>
<td>67</td>
<td>36</td>
<td>31</td>
</tr>
<tr>
<td>BP</td>
<td>32</td>
<td>46</td>
<td>23</td>
<td>14</td>
<td>32</td>
<td>36</td>
<td>31</td>
</tr>
<tr>
<td>AV</td>
<td>24</td>
<td>26</td>
<td>59</td>
<td>08</td>
<td>22</td>
<td>36</td>
<td>42</td>
</tr>
<tr>
<td>OC</td>
<td>12</td>
<td>13</td>
<td>06</td>
<td>52</td>
<td>17</td>
<td>19</td>
<td>11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measure</th>
<th>Baseline</th>
<th>Month 6</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 4</th>
<th>Year 6</th>
<th>Year 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST</td>
<td>55</td>
<td>32</td>
<td>31</td>
<td>15</td>
<td>64</td>
<td>39</td>
<td>35</td>
</tr>
<tr>
<td>BP</td>
<td>36</td>
<td>42</td>
<td>24</td>
<td>13</td>
<td>31</td>
<td>51</td>
<td>27</td>
</tr>
<tr>
<td>AV</td>
<td>22</td>
<td>26</td>
<td>47</td>
<td>07</td>
<td>23</td>
<td>33</td>
<td>53</td>
</tr>
<tr>
<td>OC</td>
<td>13</td>
<td>18</td>
<td>09</td>
<td>42</td>
<td>20</td>
<td>26</td>
<td>13</td>
</tr>
</tbody>
</table>

Means: 47, 1.0, 1.1, 89, 41, 79, 90, 72, 35, 71, 81, 63, 32, 73, 73, 60, 29, 58, 64, 49, 29, 50, 53, 45, 27, 39, 49, 38

Note. N = 467. Estimates are based on 100 imputations of the missing data. Note also that the square of the correlations indicates the shared variance among constructs and that these values are estimates of the reliable variance (i.e., with measurement error removed). For example, a correlation of .71, indicates that 50% of the reliable variance overlaps between constructs. ST = schizotypal personality disorder; BP = borderline personality disorder; AV = avoidant personality disorder; OC = obsessive-compulsive personality disorder.
Table 5
Latent Correlations, Variances, and Means for Men

<table>
<thead>
<tr>
<th>Measure</th>
<th>Baseline</th>
<th>Month 6</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 4</th>
<th>Year 6</th>
<th>Year 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>.35</td>
<td>.26</td>
<td>.30</td>
<td>.34</td>
<td>.27</td>
<td>.30</td>
<td>.25</td>
</tr>
<tr>
<td>AV</td>
<td>.29</td>
<td>.32</td>
<td>.33</td>
<td>.34</td>
<td>.33</td>
<td>.35</td>
<td>.34</td>
</tr>
<tr>
<td>OC</td>
<td>.19</td>
<td>.23</td>
<td>.09</td>
<td>.12</td>
<td>.15</td>
<td>.13</td>
<td>.14</td>
</tr>
<tr>
<td>ST</td>
<td>.90</td>
<td>.91</td>
<td>.84</td>
<td>.45</td>
<td>.70</td>
<td>.82</td>
<td>.61</td>
</tr>
<tr>
<td>BP</td>
<td>.36</td>
<td>.20</td>
<td>.14</td>
<td>.15</td>
<td>.43</td>
<td>.36</td>
<td>.44</td>
</tr>
<tr>
<td>AV</td>
<td>.23</td>
<td>.15</td>
<td>.25</td>
<td>.16</td>
<td>.66</td>
<td>.31</td>
<td>.39</td>
</tr>
<tr>
<td>OC</td>
<td>.10</td>
<td>.13</td>
<td>.17</td>
<td>.13</td>
<td>.15</td>
<td>.10</td>
<td>.13</td>
</tr>
<tr>
<td>ST</td>
<td>.84</td>
<td>.81</td>
<td>.76</td>
<td>.31</td>
<td>.65</td>
<td>.68</td>
<td>.57</td>
</tr>
<tr>
<td>BP</td>
<td>.24</td>
<td>.13</td>
<td>.15</td>
<td>.11</td>
<td>.55</td>
<td>.29</td>
<td>.66</td>
</tr>
<tr>
<td>AV</td>
<td>.15</td>
<td>.11</td>
<td>.17</td>
<td>.15</td>
<td>.51</td>
<td>.17</td>
<td>.29</td>
</tr>
<tr>
<td>OC</td>
<td>.06</td>
<td>.06</td>
<td>.07</td>
<td>.06</td>
<td>.11</td>
<td>.06</td>
<td>.11</td>
</tr>
<tr>
<td>ST</td>
<td>.82</td>
<td>.79</td>
<td>.70</td>
<td>.45</td>
<td>.47</td>
<td>.43</td>
<td>.35</td>
</tr>
<tr>
<td>BP</td>
<td>.33</td>
<td>.23</td>
<td>.19</td>
<td>.17</td>
<td>.44</td>
<td>.36</td>
<td>.39</td>
</tr>
<tr>
<td>OC</td>
<td>.08</td>
<td>.08</td>
<td>.09</td>
<td>.09</td>
<td>.14</td>
<td>.08</td>
<td>.14</td>
</tr>
</tbody>
</table>

Means 0.62 0.85 1.1 0.91 0.58 0.62 0.87 0.71 0.47 0.54 0.79 0.65 0.46 0.55 0.74 0.61 0.33 0.42 0.65 0.53 0.31 0.41 0.55 0.48 0.29 0.35 0.54 0.43
Variances 0.24 0.25 0.36 0.23 0.25 0.19 0.36 0.22 0.23 0.19 0.34 0.20 0.18 0.18 0.35 0.20 0.08 0.16 0.35 0.16 0.08 0.15 0.36 0.18 0.07 0.14 0.35 0.15

Note. Note that the square of the correlations indicates the shared variance among constructs and that these values are estimates of the reliable variance (i.e., with measurement error removed). For example, a correlation of .71 indicates that 50% of the reliable variance overlaps between constructs. ST = schizotypal personality disorder; BP = borderline personality disorder; AV = avoidant personality disorder; OC = obsessive-compulsive personality disorder.
kenburg, Hennen, Reich, & Silk, 2005). The persistent low-grade stability shown by the constructs in the present study is consistent with patterns of functional impairment.

Our results may be considered in relation to comorbidity. It has been suggested that high rates of comorbidity reflect core traits shared by different PDs (see Lynam & Widiger, 2001). Perhaps it is those who experience the greatest range of disturbance across constructs who also suffer most enduringly. Such an explanation is consistent with the higher levels of comorbid pathology typical of more disturbed populations and with findings from other studies showing nonremitting BPD cases had greater comorbidity with other Axis II disorders than did those that did remit (Zanarini et al., 2001, 2007, 2006). There is also the possibility that some participants were misdiagnosed (i.e., over diagnosed) at study entry. If that were the case, Widiger (2005) has suggested that the disorders would then appear to lack discriminant validity. However, we would also expect to see greater differences in indirect effects between baseline diagnoses to Year 10 diagnoses and Month 6 diagnoses to Year 10 diagnoses. Instead, we identified a phenomenon of decreased overall level that is relatively consistent over time.

By the later years (Year 6 to Year 10), it was noteworthy that the OCPD construct, postulated in DSM to reside in Cluster C (anxious–fearful), was more correlated with the STPD–Cluster A–based construct (odd–eccentric) and the BPD–Cluster B–based construct (erratic–emotional–dramatic), although showing little overlapping variance with the AVPD–Cluster C–based construct. This finding echoes other reports noting higher co-occurrence of OCPD with Cluster A PDs than with Cluster C PDs (e.g., Blais, McCann, Benedict, & Norman, 1997; Rossi, Marinangeli, Butti, Kalyvoka, & Petruzzi, 2000) and raises interesting possibilities. It may be that there is a persistent, maladaptive core of the OCPD construct that is less related to the anxious–fearful cluster (C) than to more severe clusters (A and B). Perhaps one component of OCPD is more associated with severe personality pathology, whereas other aspects reflect personality pathology in the anxious–fearful domain. This possibility is supported by recent factor analyses that identify two factors, perfectionism and rigidity, within the OCPD construct (Ansell, Pinto, Edelen, & Grilo,
Clarifying this would help to explain prior inconsistent findings (e.g., Fossati et al., 2000; Hyler & Lyons, 1988; Kass et al., 1985; Livesley et al., 1992, 1998; Morey, 1986; Nestadt et al., 1994; Tyrer & Alexander, 1979; Yang et al., 2002) and might further identify a core personality trait prognostic for more enduring personality pathology. It is interesting to note that other studies have described a loss of the interpersonal control associated with OCPD to be related to explosive outbursts of anger (Villemarette-Pittman, Stanford, Greve, Houston, & Mathias, 2004) and a greater incidence of impulsive aggression relative to normal and noncompulsive PD controls (Stein et al., 1996).

Interesting sex differences were found between the BPD and STPD constructs, but not the AVPD and OCPD constructs. The BPD and STPD sex differences are best illustrated in the growth curves plotted separately for men and women for these two disorders (see Figure 1B). The mean level of the STPD construct was significantly higher in men than in women at baseline. This difference declined through the 10 years of the study, with the gap...
narrowing to a negligible difference by Year 10. For the BPD construct, the mean level was higher for women compared with men, and this difference persisted through the 10 years of the study. The BPD findings reflect prior-reported findings from our studies testing sex bias in PD diagnosis (e.g., Boggs et al., 2005; Johnson et al., 2003) and are consistent with findings from other studies showing generally higher levels of BPD symptoms for women (e.g., Jane, Oltmanns, South, & Turkheimer, 2007).

This study has certain strengths and limitations. A cautionary note is that any revisions to the diagnostic system for personality pathology should take into account converging evidence from multiple sources and would ideally be informed by longitudinal studies of non–treatment-participating individuals with clinical levels of disturbance. As discussed above, many of the CLPS participants were receiving various forms of treatment, but because of the naturalistic design of the study, treatments were not controlled, and this precludes examination of treatment effects. Among the strengths of the present study is the large number of minority participants relative to other studies reported in the field. The focus on four PDs is a potential limitation. Participants were recruited with STPD, BPD, OCPD, and AVPD. Results may have been different if we had selected more broadly across all PDs. Even though those four disorders were targeted, participants typically met criteria for several PDs. Thus, this concern is moderated by the range of PD pathology evidenced in the CLPS sample, which was comparable with other clinical samples that used broader selection criteria (e.g., Blashfield et al., 1994; Oldham et al., 1995; Stuart et al., 1998). Nonetheless, results may not generalize to other clinical populations acquired with different selection criteria. Generalization to less disturbed symptomatic volunteers would also not be expected as different results in the latent structure of PDs have been found between clinical and community populations (Sheets & Craighhead, 2007).

Our use of parceling is both a strength and limitation. A decided strength is that it increased the reliability of the estimations for the PD constructs (e.g., Little et al., 2002). By reducing measurement error, the diagnostic constructs can be better captured than they would be by simply summing the criteria. This suggests that dimensional approaches in which psychometric properties are carefully considered may better serve to capture PD constructs (see Cuthbert, 2005). However, the parceling approach does preclude an examination of the strength and ordering of the relationships of individual PD criteria to their presumed constructs. This limits our ability to draw conclusions about matters such as potential differential stability of certain criteria as noted above.

It is also not possible to completely characterize the apparent overlap that we might term construct comorbidity in the context of the present study. The overlap, of course, could be due to a variety of factors, including criterion overlap, related traits, or undifferentiated pathology in chronically disturbed individuals. These questions are of interest for future work. From the present study, it is clear that some central core of personality pathology evident from DSM constructs does endure; yet, the distinctiveness of the diagnostic categories does not. Thus, there does appear to be a problem with the DSM PDs in their present framework in that the most stable and enduring personality pathology does not retain the distinct qualities of the PD constructs. Clarifying the enduring qualities of personality pathology is an important consideration for the DSM–V.

References


Received January 17, 2008
Revision received March 19, 2009
Accepted March 23, 2009

---

**Low Publication Prices for APA Members and Affiliates**

**Keeping you up-to-date.** All APA Fellows, Members, Associates, and Student Affiliates receive—as part of their annual dues—subscriptions to the *American Psychologist* and *APA Monitor*, High School Teacher and International Affiliates receive subscriptions to the *APA Monitor*, and they may subscribe to the *American Psychologist* at a significantly reduced rate. In addition, all Members and Student Affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

**Essential resources.** APA members and affiliates receive special rates for purchases of APA books, including the *Publication Manual of the American Psychological Association*, and on dozens of new topical books each year.

**Other benefits of membership.** Membership in APA also provides eligibility for competitive insurance plans, continuing education programs, reduced APA convention fees, and specialty divisions.

**More information.** Write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242.