Acta Psychiatr Scand 2002: 105: 28–36
Printed in UK. All rights reserved

ACTA PSYCHIATRICA SCANDINAVICA
ISSN 0001-690X

Confirmatory factor analysis of DSM-IV borderline, schizotypal, avoidant and obsessive-compulsive personality disorders: findings from the Collaborative Longitudinal Personality Disorders Study

Sanislow CA, Morey LC, Grilo CM, Gunderson JG, Shea MT, Skodol AE, Stout RL, Zanarini MC, McGlashan TH. Confirmatory factor analysis of DSM-IV borderline, schizotypal, avoidant and obsessive-compulsive personality disorders: findings from the Collaborative Longitudinal Personality Disorders Study.
Acta Psychiatr Scand 2002: 105: 28–36. © Munksgaard 2002.

Objective: To test the diagnostic constructs implied by DSM-IV Axis-II personality disorders by examining relationships between different combinations of DSM-IV criteria.
Method: Confirmatory factor analysis was used to test the borderline, schizotypal, avoidant and obsessive-compulsive personality disorder constructs in a large treatment-seeking sample ($N = 668$) from a multisite study. A model based on the three DSM-IV Axis II clusters was also tested. Both models were tested against a unitary 'generic' model constructed from four criteria sets combined.
Results: Goodness-of-fit for both the three-cluster and four disorder models was significantly better than the unidimensional model, and the four-disorder model was significantly better than the three-cluster model. Results were replicated using data from 2-year follow-up obtained by interviewers blind to original Axis II diagnoses at baseline.
Conclusion: Support is provided for the DSM-IV disorder-level classification for schizotypal, borderline, avoidant and obsessive-compulsive personality disorders in a treatment-seeking sample.

Charles A. Sanislow[1], Leslie C. Morey[4], Carlos M. Grilo[1], John G. Gunderson[2], M. Tracie Shea[5], Andrew E. Skodol[3], Robert L. Stout[6], Mary C. Zanarini[2], Thomas H. McGlashan[1]

Collaborative Longitudinal Study of Personality Disorders (CLPS): [1]Yale University School of Medicine, [2]McLean Hospital and Harvard Medical School, [3]New York State Psychiatric Institute and Columbia University College of Physicians and Surgeons, [4]Texas A & M University, Department of Psychology, [5]Veterans Affairs Medical Center and Brown University Medical School, [6]Decision Sciences Institute and Brown University Medical School, USA

Key words: personality disorders; confirmatory factor analysis; borderline; schizotypal; avoidant; obsessive-compulsive

Charles Sanislow PhD, Yale University School of Medicine, PO Box 208098, New Haven, CT 06520–8098, USA

Accepted for publication June 1, 2001

Introduction

The personality disorder categories within Axis II of the DSM-III (1) were largely developed on the basis of committee consensus informed by theoretical judgments of personality pathology. The more recent DSM-IV (2) used some empirical data to refine personality disorder categories (3), although much of it was unpublished. Problems associated with disorder co-occurrence (4, 5) and diagnostic overlap (6) have persisted (7). While a variety of multivariate procedures including cluster analysis and the analysis of covariance structures such as exploratory and confirmatory factor analysis are available to test latent diagnostic constructs, studies employing these methods with personality disorders are limited and have produced mixed results.

Empirical validation studies of Axis II constructs

Prior to DSM-III (1), Tyrer and Alexander used cluster and factor analyses to study 24 personality attributes related to personality disorder obtained by structured diagnostic interview in a sample of psychiatric patients (8). They subdivided their sample into two groups: those whose primary diagnosis was personality disorder, and those with other diagnoses. Similar factor structures were obtained for

both populations, suggesting that personality disorder criteria differ mainly in degree of severity among psychiatric patients. This finding has been echoed in more recent studies, based on the DSM-III (9) and DSM-III-R (10), which have demonstrated similarities between the factor structure of clinical and community samples.

Kass and colleagues employed a four-point rating scale at the disorder level (i.e. clinicians rated the disorder, and not individual criteria) for the DSM-III personality disorders in a sample of psychiatric out-patients (11). Factor analysis of these ratings provided partial support for the three DSM clusters, with some suggestion of a fourth factor for compulsive personality disorder. Morey (12) used a Procrustean rotation approach using the Kass (11) data, as well as matrices derived from MMPI and MCMI personality disorder scales, and found that the three-factor solution could be identified, with varying success, in all three datasets. Lending further support to the three cluster (plus one disorder) model, Hyler and Lyons (13) replicated the Kass et al. (11) findings using data collected from ratings of patients made by a nationwide sample of psychiatrists. A limitation of these two studies is the data that were factor analyzed were based on ratings at the disorder level rather than the specific criteria. Both Hyler and Lyons (13) and Kass et al. (11) considered the possibility that the results may have reflected psychiatrists' notions about the clustering, rather than the nature of personality pathology itself. On the other hand, evidence for the homogeneity of the compulsive personality disorder diagnosis has also been demonstrated in a general population using criteria-level data (14).

Hyler and colleagues (15) also conducted a factor analysis of the entire DSM-III item set obtained by patient self-report, and then attempted to validate the resulting factors using ratings obtained from psychiatrists on a subsample of their study group. The results of their factor analysis yielded 11 factors that did not directly correspond to the DSM-III personality disorders. There was also considerable redundancy among the factors, as evidenced by follow-up regression analyses. They concluded by recommending a greater reliance on the three clusters, and hinted that a dimensional system might be helpful to differentiate more fine-grained differences beyond the clusters. On the other hand, more recent factor-analytical work carried out with DSM-IV data has not supported the three clusters, although those results were based on disorder-level analyses (16).

The above-mentioned factor analytical studies relied on exploratory factor analysis (EFA), which has a limited capability to evaluate data with a hypothesized structure. Methodological advances have made confirmatory tests of covariance structures possible by using structural equation modeling (SEM) techniques to perform confirmatory factor analysis (CFA). In contrast to the more traditional EFA, CFA offers the advantage of statistical tests for goodness of fit when examining the resulting factor structures thus providing a more definitive evaluation of diagnostic constructs. Three studies have examined personality disorders using CFA (17–19).

Moldin and colleagues (17) examined the latent structure of DSM-III-R Axis II psychopathology in two normal control samples (parents and their offspring) from the New York High Risk Project. Using CFA, they found that a three-factor solution fit the data more effectively than one- or two-factor solutions. However, their three-factor model did not correspond with the three clusters of DSM-III-R. One limitation of this study was the restricted range of personality pathology (from the normal sample). Another, perhaps more important, issue concerns refitting of the model. Moldin and colleagues (17) tested successively revised models based on their initial empirical results. On one hand, this approach can lead to the development and refinement of a particular model. On the other hand, such refitting of a model based on the initial findings weakens the test of theory or model because initial results are fed back into the statistical testing procedures. In this sense, the test does not represent an a priori confirmatory test.

Bell and Jackson attempted to test directly the three clusters and 11 categories of DSM-III personality disorder in a psychiatric in-patient sample (18). Support for the categories was undermined by a high degree of correlation between criteria and non-corresponding disorders. Although their results indicated that the three-cluster model offered a less than optimal fit, they concluded that the three-cluster model best represented the data, and noted that some deviation from the DSM-specified clusters was evident from their cluster analysis results (schizotypal, borderline and histrionic loaded on Cluster C; dependent and compulsive on Cluster A). They recommended further study to understand better the overlap implied by the correlations of criteria across disorders.

The third study that used CFA to study to test personality disorders was more broadly based, in that it both compared and evaluated a number of various models of personality pathology in a range of samples (e.g. clinical and community) (19). The various models evaluated included (and were not limited to) the DSM Cluster model, the Five-Factor Model (20) and Cloninger and Svrakic's seven-

factor model (21). The results provided modest support for the DSM Clusters (Cluster B obtained the consistently highest factors across the various datasets, while the factors corresponding to Clusters A and C were less consistent). After evaluating alternative models, they concluded that the fit for the DSM (and other) models of personality pathology could be surpassed with the FFM or seven-factor models, but also noted that the addition of factors beyond four did not add significantly to the models.

Among other empirical approaches to the study of latent diagnostic constructs, Morey (22) cluster-analyzed DSM-III-R Axis II features obtained from clinician ratings of 291 patients who had been identified as having personality disorders, and the correlations among the personality disorder criteria generally supported the DSM-III-R model. The exception was that narcissistic and antisocial features grouped together, resembling the traditional psychopath (i.e. narcissistic-exploitative features might be viewed as a form of psychopathy). The results from this study may have partly been a result of the proximity of statistical procedure (i.e. cluster analysis) to the process used to develop the DSM categories. In other words, the results may have been more reflective of the conceptually based categories of the DSM than other methods that search for dimensional constructs. In contrast, dimensional models that have not corresponded well to the DSM framework have generally employed statistical methods that are less consonant with the procedures used to develop the DSM.

In sum, support for the DSM taxonomic system of personality disorders has not been demonstrated empirically. Alternatively, there has been some suggestion that a unitary conception of personality disorder may fit the data more effectively and, at the same time, be more parsimonious. Work by Dowson and Berrios using the Personality Disorder Questionnaire–Revised (PDQ-R (23)) suggested that the total number of personality disorder criteria was the most important derivative of the DSM-III-R Axis II classification (24). Similarly, Nestadt and colleagues demonstrated that a single explanatory factor best accounted for DSM-III Axis II variance except compulsive personality disorder (14). Further evidence for a unitary dimensional conception of personality disorder can be implied from the multiple studies reporting extensive co-occurrence of Axis II disorders (5, 7, 22). These high rates of reported diagnostic co-occurrence suggest that the DSM-IV system may not be the most parsimonious ordering of personality pathology, and the above studies provide conflicting results regarding various levels of support for the DSM clusters and disorders. One way to resolve these issues is by study of the latent diagnostic constructs using factor analysis.

Among methods based on the analysis of covariance structures, confirmatory factor analysis (CFA) (25) may be useful to evaluate the latent diagnostic constructs reflected in the DSM personality disorders as part of the construct validation process (26). With CFA, predictions are made about the relations among observed variables based on the hypothetical constructs they are purported to measure. If the model-implied covariances correspond with observed covariances, then the hypothetical constructs are supported (27). The purpose of the present study was to test various sets of relationships among diagnostic indicators as specified by the DSM-IV by conducting CFA on the DSM-IV criteria for four personality disorders (schizotypal, borderline, avoidant and obsessive-compulsive) in order to evaluate the hypothesized structure of these criteria implied by DSM-IV Axis II (i.e. personality disorder diagnoses). We evaluated the hypotheses that these criteria could be understood as (a) constituting four distinct disorders, or (b) constituting three distinct clusters of personality. To do so, we began by evaluating a unidimensional or generic model (i.e. assuming a unidimensional gradient of severity) of personality disorder for purposes of comparison, and then tested successively how much was gained (or lost) by comparing the degree of fit achieved with the three-cluster and four-disorder models.

Material and methods

Subjects

Study subjects aged 18–45 years were evaluated as part of a prospective, repeated-measures project to examine the longitudinal course of personality disorders (28). Primarily treatment-seeking subjects at both in-patient and out-patient facilities (the sample was supplemented by subjects responding to postings or media advertising for an interview study of personality, who were currently seeking or receiving psychiatric treatment or psychotherapy, or who had recently been in psychiatric treatment or psychotherapy) were sampled for four representative personality disorders (borderline, schizotypal, avoidant and obsessive-compulsive) along with a control group meeting criteria for major depressive disorder but with no personality disorder. Potential subjects 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. schizo-

phrenia, schizophreniform or schizoaffective disorders) or organicity. Three of the disorders were chosen to represent each of the DSM-IV Axis II Clusters A, B and C (i.e. borderline, schizotypal and avoidant). The fourth disorder, obsessive-compulsive, was included because there has been some indication suggesting that this disorder stands apart from the three clusters (11, 13). The five targeted diagnostic groupings, drawn from a variety of treatment settings, ensured a full representation of Axis II pathology, and their treatment-seeking status provided an ecologically valid study group. A detailed description of selection procedure and rationale is provided elsewhere (28). All eligible subjects who began the assessment signed written informed consent after the research procedures had been fully explained. The final cohort was comprised of 668 subjects. Sixty-four per cent of the sample was female; 76% were Caucasian, 11% African American, 9% Hispanic, and the remaining percentage of other ethnic backgrounds (28). On average, subjects received 1.4 Axis II diagnoses, a rate that is comparable to previous studies (29–31). The diagnostic composition of the sample is described in detail elsewhere (32).

Assessment

Extensively trained research interviewers with master's or doctoral degrees assessed all subjects. The Structured Clinical Interview for DSM-IV Axis I Disorders–Patient Version (SCID-I/P) (33) and the Diagnostic Interview for DSM-IV Personality Disorders (DIPD-IV) (34) were among the assessments conducted. The DIPD-IV is a semi-structured diagnostic interview with several questions pertaining to each DSM-IV Axis II criterion. Each criterion is scored as either 0 for 'absent', 1 for 'present but of uncertain clinical significance', or 2 for 'present and clinically significant'. In our sample, median kappa coefficients (35) ranged from 0.58 to 1.0 for all Axis II disorders (36). For the sample overall, the mean number of positive criteria (i.e. '2') for the four personality disorders were as follows: for borderline, 3.4 (SD = 2.7; for schizotypal, 1.7 (SD = 2.0); for avoidant, 3.29 (SD = 2.4); and for obsessive-compulsive, 2.7 (SD = 2.1).

Approach to modeling

Confirmatory factor analytical methods (CFA) (25) were utilized to test latent diagnostic constructs implied by DSM-IV Axis II. To carry out these analyses, we utilized version 4.0 of the *Analysis of Moment Structures* software (37). Multiple fit indices were utilized to evaluate the various models more fully. Included were the comparative fit index (CFI) (38) the normed fit index (NFI) (38) and the root mean square error of approximation (RMSEA) (39). The CFI and the NFI both measure the fit of the model relative the null model, but the CFI is less affected by sample size (27). The RMSEA was included because it is a measure of fit that takes model parsimony into account (i.e. goodness-of-fit values can sometimes be inflated artificially as the number of parameters in the model are increased). The NFI and the CFI range from 0 (poor fit) to 1 (good fit). For the RMSEA, values less than 0.08 indicate an acceptable fit (40). Finally, chi-square indices of fit were also included to facilitate comparisons among models.

In general, structural equation modeling methods allow one to fit and refine models by changing parameters. In this context, statistical significance testing can become less conservative when the specified model is based more on the empirical aspects of the dataset rather than a priori theoretically based specifications that are being tested by the observations in the dataset. When the goal is to confirm or disconfirm a model, it is important to have a theoretical framework driving the analyses so that one does not simply proceed until the covariances are maximally explained. Thus, for *confirmatory* analyses, it is important to take precautions not to over-fit a model (i.e. continue with successive revisions until fit indices reach a compelling magnitude). Precise theoretical specifications about the model to be tested should ideally be made and tested without substantial revisions to avoid forcing, in a Procrustean manner, the data into preconceived notions. Our goal was to test explicitly the DSM-IV Axis II disorder and cluster levels of classification, and not to successively refit the models that we tested in order to optimize fit indices.

We tested all models using the diagnostic criteria from the four study disorders obtained from the DIPD-IV as observed variables that corresponded to the respective sets of latent constructs (i.e. the three clusters or four disorders). As a competing model, a one-factor model was specified based on a view of personality disorder as a singular construct that varies in severity. Independence of error terms was specified for these models. For the three-cluster and four-disorder models, we allowed the latent variables to be correlated.

Results

Fit indices for the one-factor model uniformly indicated a poor fit to the data (χ^2 (495) = 4802.1, NFI = 0.39, CFI = 0.41 and RMSEA = 0.114). Next,

the three-factor model based on the three clusters was tested. Because avoidant and obsessive-compulsive personality disorders are considered part of cluster C, these criteria were combined for the observed variable set corresponding to the latent variable representing cluster C. The fit indices for the three-factor model suggested an appreciable improvement upon the one-factor model (χ^2 (492) = 2689.5; NFI = 0.66, CFI 0.70 and RMSEA = 0.082). Finally, the fit indices for the four-factor model suggested that this model provided a more acceptable fit to the data (χ^2 (489) = 1756.8; CFI = 0.83, NFI = 0.78 and RMSEA = 0.062). The model based on the four disorders is shown in Fig. 1.

Based on inspection of the results of the three models, the four-factor solution appeared to offer the better fit. While the NFI and CFI fell somewhat short of an ideal fit, the RMSEA value of 0.062 was within the 0.08 cut-off. To examine our hypotheses of these competing models more directly, we performed chi-square tests for the differences (χ^2_{diff}) to test the significance of the improved degree of fit offered by each successive model over the unidimensional model (because all three models shared identical observed variables, they qualified as 'nested' allowing statistical testing directly comparing the goodness of fit between the models). These results revealed that the four-factor model offered a significantly improved fit over the three-factor model (χ^2_{diff} (3) = 932.7, $P < 0.001$) and over the one-factor model (χ^2_{diff} (6) = 3045.3, $P < 0.001$). Note also that the three-factor model offered a significantly improved fit over the one-factor model (χ^2_{diff} (3) = 2112.6, $P < 0.001$).

To evaluate the results more completely we obtained testing the above progression of models, we repeated these analyses using data obtained from our 2-year follow-up. Recall that the diagnostic assessments that were conducted at the 2-year follow-up were done so by interviewers blind to original baseline Axis II diagnoses. The results for the one-factor model again indicated a uniformly poor fit (χ^2 (495) = 3017.8, NFI = 0.43, CFI = 0.47 and RMSEA = 0.101). For the three-factor model based on the DSM-IV clusters, some improved degree of fit was indicated (χ^2 (492) = 1980.3; NFI = 0.63, CFI 0.69 and RMSEA = 0.078). Finally, the four-factor 'disorder' model appeared to offer the most acceptable fit (χ^2 (489) = 1299.9; CFI = 0.83, NFI = 0.76 and RMSEA = 0.058); see Fig. 2. Regarding tests comparing the three models, the results again revealed that the four-factor model offered a significantly improved fit over the three-factor model (χ^2_{diff} (3) = 680.4, $P < 0.001$) and also over the one-factor model (χ^2_{diff} (6) = 1717.9, $P < 0.001$). Again, note that the three-factor model

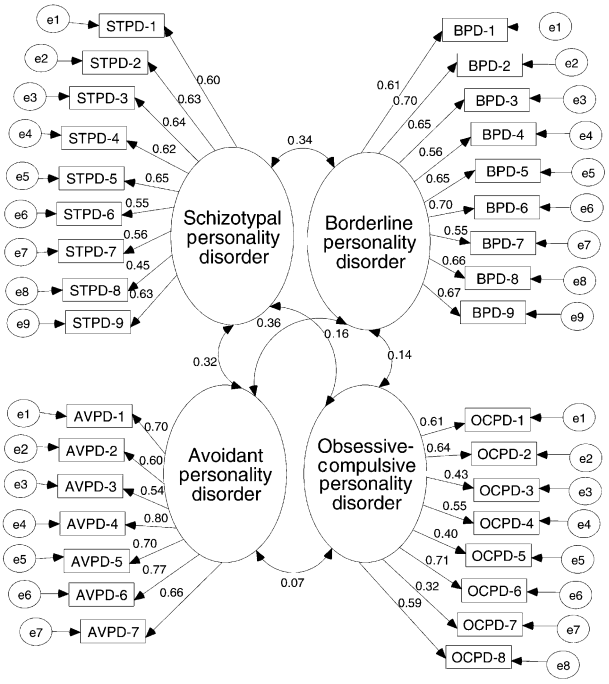

Fig. 1. Disorder-level model testing the four personality disorders studied in the CLPS (schizotypal, borderline, avoidant and obsessive-compulsive). Note: Observed variables are enclosed in rectangles, latent constructs in ellipses and error terms in circles. Reported factor score weights are standardized estimates. Numbering corresponds to DSM-IV criterion ordering for each personality disorder. For schizotypal, STPD-1 = 'ideas of reference', STPD-2 = 'odd beliefs, superstitious, magical thinking', STPD-3 = 'unusual experiences', STPD-4 = 'odd thinking/speech', STPD-5 = 'paranoid ideation', STPD-6 = 'constricted affect', STPD-7 = 'odd behavior', STPD-8 = 'no close friends', and STPD-9 = 'excessive social anxiety'. For borderline, BPD-1 = 'avoid abandonment', BPD-2 = 'unstable relationship', BPD-3 = 'identity disturbance', BPD-4 = 'impulsivity in two areas', BPD-5 = 'self-injury/suicide attempt', BPD-6 = 'affective instability', BPD-7 = 'chronic emptiness', BPD-8 = 'intense anger', and BPD-9 = 'stress-related paranoid ideation, dissociation'. For avoidant, AVPD-1 = 'avoids jobs with interpersonal contact', AVPD-2 = 'needs to be liked first', AVPD-3 = 'fears being ridiculed or shamed', AVPD-4 = 'preoccupied with being rejected', AVPD-5 = 'views self as socially inept', AVPD-6 = 'feels inadequate', and AVPD-7 = 'does not take risks because may be embarrassing'. For obsessive-compulsive, OCPD-1 = 'details, rules, lists, order', OCPD-2 = 'perfectionism interferes with task completion', OCPD-3 = 'workaholic', OCPD-4 = 'inflexible about morality, ethnics', OCPD-5 = 'pack rat', OCPD-6 = 'reluctant to delegate tasks', OCPD-7 = 'miserly', and OCPD-8 = 'rigid and stubborn'.

offered a significantly improved fit over the one-factor model (χ^2_{diff} (3) = 1037.5, $P < 0.001$).

In sum, the results of the tests examining the differences in goodness-of-fit indicated that the least imperfect fit was achieved with the four-factor, or disorder-level model when compared to the three-cluster or unidimensional factor. This pattern of findings was nearly identical when our tests for modeling were applied to the 2-year follow-up data

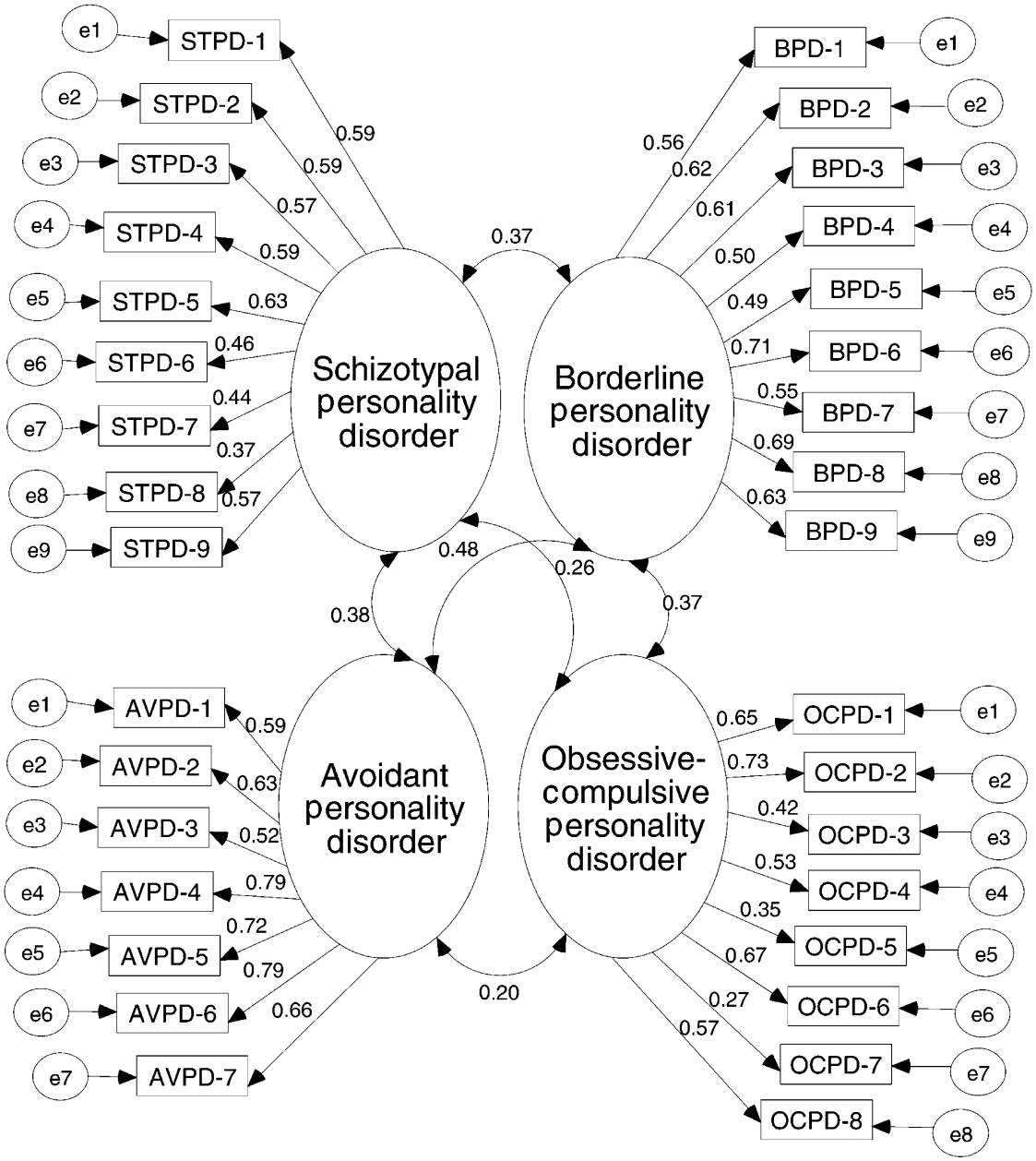

Fig. 2. Disorder-level model testing the four personality disorders studied in the CLPS (schizotypal, borderline, avoidant and obsessive-compulsive) using data from the 2-year follow-up. Note: Observed variables are enclosed in rectangles, latent constructs in ellipses, and error terms in circles. Reported factor score weights are standardized estimates. Numbering corresponds to DSM-IV criterion ordering for each personality disorder. For schizotypal STPD-1 = 'ideas of reference', STPD-2 = 'odd beliefs, superstitious magical thinking', STPD-3 = 'unusual experiences', STPD-4 = 'odd thinking/speech', STPD-5 = 'paranoid ideation', STPD-6 = 'constricted affect', STPD-7 = 'odd behavior', STPD-8 = 'no close friends', and STPD-9 = 'excessive social anxiety'. For borderline, BPD-1 = 'avoid abandonment', BPD-2 = 'unstable relationship', BPD-3 = 'identity disturbance', BPD-4 = 'impulsivity in two areas', BPD-5 = 'self-injury/suicide attempt', BPD-6 = 'affective instability', BPD-7 = 'chronic emptiness', BPD-8 = 'intense anger', and BPD-9 = 'stress-related paranoid ideation, dissociation'. For avoidant, AVPD-1 = 'avoids jobs with interpersonal contact', AVPD-2 = 'needs to be liked first', AVPD-3 = 'fears being ridiculed or shamed', AVPD-4 = 'preoccupied with being rejected', AVPD-5 = 'views self as socially inept', AVPD-6 = 'feels inadequate', and AVPD-7 = 'does not take risks because may be embarrassing'. For obsessive-compulsive, OCPD-1 = 'details, rules, lists, order', OCPD-2 = 'perfectionism interferes with task completion', OCPD-3 = 'workaholic', OCPD-4 = 'inflexible about morality, ethnics', OCPD-5 = 'pack rat', OCPD-6 = 'reluctant to delegate tasks', OCPD-7 = 'miserly', and OCPD-8 = 'rigid and stubborn'.

obtained by interviewers blind to the original Axis II diagnoses.

Discussion

We selected treatment-seeking individuals carefully to obtain a large sample representing four prototypic personality disorders: borderline, schizotypal, avoidant and obsessive-compulsive. The standardized, reliable semi-structured interview data obtained with the DIPD-IV (34, 36) from the baseline assessment was used to test the diagnostic constructs of these four disorders and the three clusters specified by DSM-IV Axis II using confirmatory factor analysis. The four-factor model corresponded to each of the four personality disorders in our study, and the three-factor model to each of the three DSM-IV clusters. A competing model, based on a unidimensional notion of personality disturbance, was used as a benchmark to test if the three cluster or four disorder models offered any advantage over a unidimensional conceptualization of personality disorder based on a gradient of severity. Overall, results from this study support the division of DSM-IV personality disorders into at least four disorders over a unidimensional model of personality disorder (based on the numerous studies documenting high co-occurrence among DSM-IV personality disorders), and over a three-factor model of personality disorder (based on the DSM-IV Axis II clusters).

Because our patient sample is large, demographically diverse and derived from a large number of different clinical sites, generalizability to many other clinical populations is expected. Nonetheless, our sample may not be generalizable to some clinical settings (e.g. non-urban, non-academic-affiliated facilities) or to non-clinical community samples, where the composition of the sample would be expected to vary. Our recruitment procedures ensured a full diagnostic range for the borderline, schizotypal, avoidant and obsessive-compulsive personality disorders selected carefully for our study purposes, and our major depressive disorder–no personality disorder control group provided non-personality disorder variability in criteria. Expected rates of diagnostic co-occurrence were found in our sample (32), providing further support of its representative nature.

The results of this study may assist in subsequent revisions of the DSM. For the DSM models that we tested, the best fit was the model based on four disorders. This suggests one of two things (and perhaps both). First, these results support prior studies that have indicated obsessive-compulsive personality disorder stands apart from the three clusters of DSM-IV (11, 13), as well as other studies that have not replicated the three clusters of the DSM (10, 18). Secondly, the disorder-level specification of the DSM system is modestly supported. The progression of results evidenced by our sequential testing of one, three and four-factor models was compelling both in terms of the goodness-of-fit and the theoretical rationale on which it was based. Thus, in the realm of these carefully selected personality categories, disorder-level classification of the DSM-IV Axis II is favored over three-cluster and unidimensional generic cataloging of DSM-IV personality disorder criteria.

Given high rates of diagnostic co-occurrence both in general (7, 29–31), and in our sample (32), as well as other studies that have demonstrated a large portion of variance explained by a single factor (14, 24), we found it surprising how poorly the 'generic' unidimensional model did. Others have suggested that personality disorders are not inherently unidimensional (41), and that measurement artifacts can account for the appearance that they are. Our findings support this idea, suggesting that there is more to personality pathology than a single dimension.

A more general test of the DSM-IV disorder-level model would ideally involve tests of all DSM personality disorders; because participants were recruited for membership into one of the four personality disorders, borderline, schizotypal, avoidant and obsessive-compulsive (or a major depressive disorder/no personality disorder control group), we are limited to examining only those four personality disorders in our confirmatory model. However, we stress that our selection process did not exclude subjects with other DSM-IV personality disorders. This provided us with the considerable variability necessary for our analyses. As we noted above, our patterning of co-occurrence among Axis II disorders is comparable to other studies. Finally, the replication of our results with the 2-year data where any potential selection biases would be expected to differ from the baseline assessment provides compelling support for the disorder level classification.

A potential criticism of our work is that the fit indices obtained were not uniformly high. However, Gurtman and Pincus (42) have noted that over-reliance on the fit indices as the *sine qua non* for evaluating a model in the absence of other considerations (such as practical utility, or the theoretical basis of a model), rests on a mistaken assumption that multivariate procedures such as CFA provide conclusive evidence in and of themselves (43). Our procedure progressively testing two successive models exactly as specified by DSM-

IV without revision, and comparing them to the benchmark unidimensional model, is a very conservative approach. In light of our approach, and the additional support provided by the uniform patterning of factor scores in the four factor model (see Figs 1 and 2) we conclude that, for our sample, the results of the four-disorder model are favored over a unidimensional model of personality disorder, and over a three-cluster model. These results are made more compelling by the replication achieved using the 2-year follow-up data obtained by interviewers blind to baseline Axis II diagnoses.

Another potential limitation in our results concerns the possibility of a 'halo' effect. Because the DIPD-IV (34) queries about each disorder criteria in succession, and because we were recruiting for specific personality disorders, it is possible that these procedural elements introduced a bias where interviewers 'inflated' the subsequent ratings in a disorder where criteria were initially endorsed. On the other hand, this assessment procedure allows the interviewer evaluate the prototypical nature of the disorder more fully.

These results also inform the composition of the DSM-IV criteria groups at the item level (refer to Figs 1 and 2), a result consistent with those of Morey (22) using cluster analysis methodology. In the four-disorder model, it was noteworthy that for each of the four disorders that the criterion hypothesized to be most prototypic by the authors of DSM-IV, did not load most heavily on its corresponding factor. In other words, the rank ordering of criteria implied by the DSM-IV was not supported in our sample. For borderline, the two items that loaded most heavily on the corresponding factor for the baseline tests of the four-disorder model were 'affective instability' and 'unstable relationships'. At the 2-year follow-up, 'affective instability' again loaded highest, this time followed by 'intense anger'. For the schizotypal factor, 'paranoid ideation' was the most heavily loaded criterion for both the baseline and 2-year analyses for the four-factor model tests. For avoidant, 'preoccupied with being rejected' loaded most heavily on its correspondent factor for baseline and follow-up tests. For the obsessive-compulsive factor, it was the criterion 'reluctant to delegate tasks' at baseline and 'perfectionism' at follow-up that loaded most heavily. It is also worth noting that for borderline personality disorder, the newly added criterion 'transient, stress-related paranoid ideation' converged well with the items carried over from the original DSM-III-R criteria. The obsessive-compulsive personality disorder criterion 'miserly' did the most poorly of all items for both baseline and follow-up tests of the four-disorder model. Overall, the consistency of the loadings of specific criteria between the baseline and the 2-year follow-up tests of the four-disorder model were remarkable.

To conclude, our results indicate a great deal of heterogeneity among the four personality disorders that we studied (borderline, schizotypal, avoidant and obsessive-compulsive) that was multifactorial, and not captured simply along a dimension of severity. However, the poor performance of the one-factor model may be explained at least partially by our selection procedures including the clinical nature of our sample. These results demonstrate that, in a sample that is carefully diagnosed, the criteria can be separated into four relatively discrete groups. While the results of this study support the division of Axis II personality pathology into at least four groupings, we are unable to answer the question if more divisions are appropriate. It will also be interesting to see if similar results can be found in more naturalistically selected samples with varying rates of personality disorder criteria, diagnoses and severity.

Acknowledgements

Support for this work was provided by NIH/NIMH grants R10 50837, 50839, 50840, 50838, and 50850, from NIH/NIMH MH 01654. This manuscript was approved by the CLPS publication committee. We thank Todd Little for statistical advice and helpful comments on an earlier draft of this manuscript.

References

1. American Psychiatric Association (APA) Diagnostic and statistical manual of mental disorders, 3rd edn. Washington, DC: American Psychiatric Association, 1980.
2. American Psychiatric Association (APA) Diagnostic and statistical manual for mental disorders, 4th edn. Washington, DC: American Psychiatric Association, 1994.
3. Widiger T, Trull TJ. Performance characteristics of DSM-III-R personality disorders. In: DSM-IV sourcebook. Washington, DC: American Psychiatric Press, 1994:357–373.
4. Gunderson JG. DSM-IV personality disorders: final overview. In: DSM-IV sourcebook. Washington, DC: American Psychiatric Press, 1994:1123–1140.
5. Zimmerman M, Coryell W. The reliability of personality disorder diagnoses in a non-patient sample. J Pers Dis 1989;**3**:53–57.
6. Morey LC. Personality disorders in DSM-III and DSM-III-R: convergence, coverage, and internal consistency. Am J Psychiatry 1988;**145**:573–577.
7. Grilo CM, McGlashan TH, Oldham JM. The course and stability of personality disorders. J Pract Psychiatr Behav Health 1998;**4**:61–75.
8. Tyrer P, Alexander J. Classification of personality disorder. Br J Psychiatry 1979;**135**:163–167.
9. Livesley WJ, Jackson DN, Schroeder ML. Factorial structure of traits delineating personality disorders in clinical and general population samples. J Abnorm Psychol 1992; **101**:432–440.

10. Schroeder ML, Livesley WJ. An evaluation of DSM-III-R personality disorders. Acta Psychiatr Scand 1991;**84**:512–519.
11. Kass F, Skodol AE, Charles E, Spitzer RL, Williams JB. Scaled ratings of DSM-III personality disorders. Am J Psychiatry 1985;**142**:627–630.
12. Morey LC. A comparison of three personality disorder approaches. J Psychopathol Behav Assess 1986;**8**:25–30.
13. Hyler SE, Lyons M. Factor analysis of the DSM-III personality disorder clusters: a replication. Compr Psychiatry 1988;**29**:304–308.
14. Nestadt G, Eaton WW, Romanoski AJ, et al. Assessment of DSM-III personality structure in a general-population survey. Compr Psychiatry 1994;**35**:54–63.
15. Hyler SE, Lyons M, Rieder RO, Young L, Williams JB, Spitzer RL. The factor structure of self-report DSM-III axis II symptoms and their relationship to clinicians' ratings. Am J Psychiatry 1990;**147**:751–757.
16. Fossati A, Maffei C, Bagnato M et al. Patterns of covariation of DSM-IV personality disorders in a mixed psychiatric sample. Compr Psychiatry 2000;**41**:206–215.
17. Moldin SO, Rice JP, Erlenmeyer-Kimling L, Squires-Wheeler E. Latent structure of DSM-III–R Axis II psychopathology in a normal sample. J Abnorm Psychol 1994; **103**:259–266.
18. Bell RC, Jackson HJ. The structure of personality disorders in DSM-III. Acta Psychiatr Scand 1992;**85**:279–287.
19. O'Connor BP, Dyce JA. A test of models of personality disorder configuration. J Abnorm Psychol 1998;**107**:3–16.
20. Widiger TA, Costa PT. Personality and personality disorders. J Abnorm Psychol 1994;**103**:78–91.
21. Cloninger CR, Svrakic DM. Differentiating normal and deviant personality by the seven-factor personality model. In: Strack S, Lorr M, eds. Differentiating normal and abnormal personality. New York: Springer, 1994:40–64.
22. Morey LC. The categorical representation of personality disorder: a cluster analysis of DSM-III-R personality features. J Abnorm Psychol 1988;**97**:314–321.
23. Hyler SE, Skodol AE, Kellman HD et al. Validity of the personality diagnostic questionnaire–revised: comparison with two structured interviews. Am J Psychiatry 1990;**147**: 1043–1048.
24. Dowson JH, Berrios GE. Factor structure of DSM-III-R personality disorders shown by self-report questionnaire: implications for classifying and assessing personality disorders. Acta Psychiatr Scand 1991;**84**:555–560.
25. Joreskog KG. A general approach to confirmatory maximum likelihood factor analysis with addendum. In: Joreskog KG, Sorbom D, eds. Advances in factor analysis and structural equation models. Cambridge, MA: ABT Books, 1979:21–43.
26. Cronbach LG, Meehl PE. Construct validity in the psychological literature. Psychol Bull 1955;**105**:317–327.
27. Kline RB. Principles and practice of structural equation modeling. New York: Guilford, 1998.
28. Gunderson JG, Shea MT, Skodol AE et al. The Collaborative Longitudinal Personality Disorders Study: development, aims, design, and sample characteristics. J Pers Dis 2000;**14**:300–315.
29. Oldham JM, Skodol AE, Kellman HD et al. Comorbidity of axis I and axis II disorders. Am J Psychiatry 1995;**152**:571–578.
30. Stuart S, Pfohl B, Battaglia M, Bellodi L, Grove W, Cadoret R. The cooccurrence of DSM-III-R personality disorders. J Pers Dis 1998;**12**:302–315.
31. Blashfield RK, McElroy RA, Pfohl B, Blum N. Comorbidity and the prototype model. Clin Psychol Sci Pract 1994; **1**:96–99.
32. McGlashan TH, Grilo CM, Skodol AE et al. The Collaborative Longitudinal Personality Disorders Study: baseline Axis I/II and II/II diagnostic co-occurrence. Acta Psychiatr Scand 2000;**102**:256–264.
33. First MB, Spitzer RL, Gibbon M, Williams JBW. Structured clinical interview for DSM-IV Axis I disorders. New York: New York State Psychiatric Institute, 1996.
34. Zanarini MC, Frankenburg FR, Sickel AE, Yong L. The diagnostic interview for DSM-IV personality disorders. McLean Hospital, Laboratory for the Study of Adult Development. Belmont, MA: 1996.
35. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Measure 1960;**20**:37–46.
36. Zanarini MC, Skodol AE, Bender D et al. The Collaborative Longitudinal Personality Disorders Study: reliability of Axis I and II diagnoses. J Pers Dis 2000;**14**:291–299.
37. Arbuckle JL, Wothke W. AMOS 4.0 User's guide. Chicago, IL: Small Waters, 1999.
38. Bentler PM. Comparative fit indices in structural models. Psychol Bull 1990;**107**:238–246.
39. Steiger JH. Structural model evaluation and modification: an interval estimation approach. Multivariate Behav Res 1990;**25**:173–180.
40. Browne MW, Cudeck R. Alternative ways of assessing model fit. In: Bollen KA, Long JS, eds. Testing structural equation models. Newbury Park, CA: Sage, 1993:136–162.
41. Livesley WJ, Jackson DN. The internal consistency and factorial structure of behaviors judged to be associated with DSM-III personality disorders. Am J Psychiatry 1986;**143**: 1473–1474.
42. Gurtman MB, Pincus AL. Interpersonal adjective scales: confirmation of circumplex structure from multiple perspectives. Pers Soc Psychol Bull 2000;**26**:374–384.
43. Roberts S, Pashler H. How persuasive is a good fit? A comment on theory testing. Psychol Rev 2000;**107**:358–367.