Internal consistency and intercriterion overlap within and between DSM-IV Axis II Personality Disorders: Findings from the Collaborative Longitudinal Study of Personality Disorders

Carlos M. Grilo
Thomas H. McGlashan
Leslie C. Morey, Texas A & M University - College Station
John G. Gunderson
Andrew E. Skodol, et al.

Available at: http://works.bepress.com/charles_sanislow/21/
Internal consistency, intercriterion overlap and diagnostic efficiency of criteria sets for DSM-IV schizotypal, borderline, avoidant and obsessive-compulsive personality disorders


Objective: To evaluate performance characteristics of DSM-IV Personality Disorders (PDs) criteria.

Method: Six hundred and sixty-eight adults recruited for the Collaborative Longitudinal Personality Disorders Study (CLPS) were assessed with diagnostic interviews.

Results: Within-category inter-relatedness was evaluated by Cronbach’s alpha and median intercriterion correlations (MIC). Cronbach’s alpha ranged from 0.47 to 0.87 (median = 0.71); seven of the 10 PDs had alphas greater than 0.70. Between-category criterion overlap was evaluated by ‘inter-category’ intercriterion correlations between all PD pairs (ICMIC). ICMIC values (median = 0.08) were lower than MIC values (median = 0.23). Diagnostic efficiency statistics (sensitivity, specificity, positive predictive power and negative predictive power) were calculated for schizotypal, borderline, avoidant and obsessive-compulsive PDs.

Conclusion: DSM-IV PD criteria sets have some convergent validity and discriminant validity: criteria for individual PDs correlate better with each other than with criteria for other PDs. Diagnostic efficiency statistics provide guidance regarding usefulness of criteria for inclusion or exclusion.

Introduction

Personality disorders (PD) have been included in the American Psychiatric Association’s Diagnostic and Statistical Manual (DSM) from the first edition (DSM-I) published in 1952 (1) to the current (4th) edition: DSM-IV (2). Particularly noteworthy was the placement of PD diagnoses on a separate Axis II in the multi-axial system of the DSM-III (3) (see Skodol (4) for a detailed ontogeny of the classification of PD). The past decade has witnessed a burgeoning of clinical and research attention to the PDs, but important and fundamental questions regarding different aspects of their validity remain (5–10). Some concerns regarding the validity of PDs, which include, for example, questionable reliability (see 10) and high diagnostic co-occurrence...
ence (11), naturally beg for basic psychometric work on the PD criteria and diagnoses (9, 12–16).

One aspect of construct validity of PD diagnoses is the internal consistency of their component symptoms. Internal consistency reflects the interrelatedness of a set of criteria and represents one approach to examining reliability (17). Relatively little work has been done on the internal consistency of DSM-defined PD diagnoses. Overall, studies of relatively large samples have generally not employed standardized interviews and most studies that have utilized diagnostic interviews have ascertained relatively small samples.

Morey (18) used data from a PD symptom checklist completed by raters performing a retrospective chart review of 37 psychiatric in-patients to examine internal consistency of three DSM-III (3) PD diagnoses. Cronbach’s (19) coefficient alpha was 0.36 for borderline PD, 0.39 for narcissistic PD and 0.54 for schizotypal PD. Livesley and Jackson (20), using data generated from clinicians asked to apply behavior sets judged to be associated with DSM-III (3) PDs to ‘classical’ cases, reported alpha coefficients ranging from 0.74 (for narcissistic PD) to 0.94 (for avoidant PD). Pfohl and colleagues (21) examined diagnostic cohesiveness of DSM-III (3) PDs using criterion data generated from structured diagnostic interviews administered to 131 patients. Low positive predictive values for specific criteria within several PDs and considerable diagnostic overlap between PDs were found.

Morey (22) used a national sample of patients and a symptom checklist for DSM-III (3) and DSM-III-R (23) PDs to examine differences between the two systems. To minimize the effects of differences in criterion list length, Morey (22) used intercriterion correlations. Using DSM-III (3) criteria, median intercriterion correlations ranged from 0.10 (for compulsive PD) to 0.34 (for dependent PD). Using DSM-III-R (23), values ranged from 0.10 (for obsessive compulsive PD) to 0.29 (for paranoid PD). Morey (22) noted that the comparable internal consistencies for the DSM-III and DSM-III-R showed that the DSM revisions increased coverage without decreasing internal consistency. Morey (23) also reported considerable diagnostic overlap among DSM-III-R (23) PDs, suggesting that the revisions increased the amount of diagnostic overlap between diagnoses.

More recently, DSM-III-R (23) defined PD diagnostic criteria have been tested. Blais et al. (24) used data from PD symptom checklist ratings from 320 patients identified as having a PD. Only three of the 11 DSM-III-R (23) defined PDs had an alpha coefficient above 0.70, which was selected to represent an acceptable level (see 25). Blais and colleagues (26) generated criterion data for Cluster B PDs based on a retrospective review of archival medical records for 94 patients. Coefficient alphas ranged from 0.35 to 0.75 for the four DSM-III-R (23) defined cluster B PDs and from 0.57 to 0.74 using DSM-IV (2) criteria, suggesting some improvement in the progression from DSM-III-R to DSM-IV. Becker and colleagues (27) examined within-category interrelatedness as well as between category overlap for DSM-III-R (23) PD diagnoses in 38 adult psychiatric in-patients using criterion data generated by the Personality Disorder Examination (28). Alpha coefficients ranged from 0.52 (schizotypal PD) to 0.88 (antisocial PD). Seven of the 11 PDs had alphas greater than 0.70. Becker et al. (27) reported MIC values that were generally comparable or slightly higher than those reported by Morey (22). The degree of criterion overlap within PDs was generally higher than that observed for criteria between PD, suggesting some degree of PD diagnostic homogeneity.

Three recent reports have extended these analyses to DSM-IV-defined PDs (29–31). Blais and Norman (29) evaluated DSM-IV (2) PD criteria using data generated by a national sample of mental health professionals who rated 280 patients known to them using a symptom checklist. A median coefficient alpha of 0.73 was reported; seven of the 10 PDs had alphas greater than 0.70. However, analyses modeled after Morey (22) suggested that while convergent validity was generally adequate, discriminant validity was problematic, i.e. a high degree of correlations between criteria for different PDs was observed. Maffei and colleagues (30) reported adequate internal consistency for PD diagnoses (Armor’s theta ranged from 0.71 to 0.94) using data from 231 patients obtained with the Structured Clinical Interview for DSM-IV–Axis II, version 2.0 (32). Between-diagnosis criterion overlap was not reported. Grilo and McGlashan (31) examined within category inter-relatedness and between category overlap for DSM-IV (2) PDs in 70 out-patients using data generated by the Diagnostic Interview for DSM-IV Personality Disorders (DIPD-IV; 33). Cronbach’s alpha ranged from 0.64 to 0.93 (mean = 0.77); six of the 10 PD diagnoses had alphas greater than 0.70. ICMIC values (mean = 0.28) were slightly lower than MIC values (mean = 0.34), suggesting some degree of discriminant validity.

The above psychometric analyses provide a general sense of convergent and discriminant validity of PD criteria. Diagnostic efficiency analyses are required to inform decisions regarding continued refinement of the criteria sets (34). Diagnostic efficiency refers to the degree to which
criteria correctly discriminate diagnoses (i.e. case-
ness versus noncaseness) as evaluated from actu-
arial prediction (34). Since the early applications of 
Yerushalmy (35) to medical diagnoses, this work 
has evolved considerably (34, 36) and has con-
tributed to the ‘validation’ of certain Axis I 
disorders to refinements of diagnostic algorithms 
(37–39). Noteworthy is Widiger and colleagues, (34) 
cogent demonstration of the utility of positive and 
negative predictive power — in addition to 
sensitivity and specificity — for evaluating diag-
nostic efficiency of DSM-III (3) borderline PD 
criteria. Unfortunately, there exist few published 
diagnostic efficiency statistics for PD diagnoses, 
with reports for DSM-III PD diagnoses (21), for 
DSM-III borderline PD (34), for DSM-III-R 
borderline PD (40), and DSM-IV passive-aggressive 
PD (41) representing notable contributions. 
The purpose of the present study was to examine 
selected aspects of construct validity of PDs (i.e. the 
convergent and discriminant validity of the criteria 
sets). We evaluated the within-category cohesive-
ness of DSM-IV (2) PD diagnoses in adult patients 
assessed with semi-structured diagnostic interviews. 
In addition to evaluating within-category inter-
relatedness, we sought to use a similar psychometric 
approach to examine the degree of between-
category criterion overlap. In addition, this repre-
sents the first report of diagnostic efficiency analyses 
of the criteria for four selected DSM-IV (2) PD 
diagnoses: schizotypal, borderline, avoidant and 
obsessive-compulsive PDs.

**Material and methods**

**Subjects**

The study group for this report included all 668 
patients enrolled in the Collaborative Longitudinal 
Personality Disorders Study (CLPS). The aims, 
design, methodology and demographic character-
istics are described in Gunderson et al. (42) and the 
overall characterization of psychiatric (Axis I) and 
personality (Axis II) diagnoses are detailed in 
Skodol et al. (43) and McGlashan et al. (44).

Briefly, the CLPS is a longitudinal study that 
examines whether the separation of PDs to Axis II 
based on their putative stability, relative to the 
episodic unstable course of Axis I psychiatric 
disorders, is valid. Participants were recruited 
primarily from clinical services affiliated with the 
four CLPS recruitment sites located in somewhat 
diverse settings in the United States: Brown 
University (Providence, RI; N = 155), Columbia 
(New York City, NY; N = 161), Harvard/McLean 
Hospital (Belmont, MA; N = 190), and Yale (New 
Haven, CT; N = 162). Recruitment was supplemen-
ted by participants responding to postings and 
advertisements for an interview study of personal-
ity, who had previously or currently receiving 
some form of psychiatric treatment.

The overall CLPS study group of 668 participants 
is comprised participants recruited to meet criteria 
for at least one of four targeted PDs regardless of 
the presence or absence of additional PD (schizo-
typal (STPD; N = 86); borderline (BPD; N = 175); 
avoidant (AVPD; N = 157); obsessive-compulsive 
(OCPD; N = 153)) or for the control group 
comprised of major depressive disorder without 
PD (MDD; N = 97). Participants were prescreened 
to determine age eligibility (18–45 years) and 
treatment status and to exclude people with 
psychotic illnesses or acute confusional states.

Participants ranged in age from 18 to 45 
(M = 32.8, SD = 8.1). Sixty-four per cent of partici-
pants were female and 76% were Caucasian. 
Participants were generally distributed across the 
social classes, with relatively small representation 
from the lowest socioeconomic class. Forty-five per 
cent were out-patients in a variety of mental health 
settings, 11% were psychiatric in-patients, 5% were 
from medical settings and 39% were self-referred.

In our clinical study group, diagnostic co-
ocurrence between Axis I psychiatric disorders 
and Axis II personality disorders, as well as between 
Axis II disorders, was common (44). Thus, although 
participants were recruited for the five study groups, 
a wide distribution of all PDs characterized the 
participants assigned to the four specific PD groups. 
Participants with PDs had a mean of 1.4 (SD = 1.6) 
additional PD diagnoses (i.e. in addition to the 
‘primary’ PD diagnosis assigned). This overall rate 
of PD co-occurrence is similar to the mean = 1.8 
previously reported by Oldham and colleagues (11) 
and mean = 1.7 by Stuart and colleagues (45). 
Table 1 shows the frequency (column 1) of all the 
PD diagnoses in our overall study group. In order 
to meet the important goal of maximizing vari-
bility in the criteria, PD criteria data from the 
control group of major depressive disorder without 
PD were included in the present analyses.

**Procedures**

All participants signed written informed consent. 
Participants received a systematic diagnostic eval-
uation, which included the Diagnostic Interview for 
DSM-IV Personality Disorders (DIPD-IV; 33). The 
DIPD-IV (33) is a semi-structured diagnostic 
interview that assesses the presence of all of the 
10 recognized and the two ‘research criteria’ 
DSM-IV (2) PDs. The DIPD-IV (33) stipulates 
that criteria must be present and pervasive for a 
minimum of 2 years and characteristic of the person
for most of his or her adult life. DIPD-IV diagnoses for the five primary study groups required convergent support (42) from at least one of two assessment methods: (a) the Personality Assessment Form (PAF; 46) or (b) the Schedule for Non-adaptive and Adaptive Personality (SNAP; 47).

Interviews were conducted by doctoral and master’s level interviewers with at least two years of clinical experience with PD patients. The interviewers received intensive training from the developer of the DIPD-IV (Mary C. Zanarini) and received ongoing training, supervision and monitoring by the investigators. Inter-rater reliability and test–retest reliability have been established within- and across-sites and described in detail elsewhere (48). Briefly, test–retest reliability for the four PD diagnostic subgroups of primary focus in the CLPS ranged from kappa = 0.64 (STPD) to 0.74 (OCPD). Inter-rater reliability for the four PD subgroups were: kappa = 0.68 (AVPD and BPD), 0.71 (OCPD) and 1.0 (STPD); inter-rater reliability for the other PDs ranged from kappa = 0.58 to 1.0. Overall, these estimates of reliability compare favorably to those reported in the literature (see 5, 10).

All participants had complete symptom criteria data available for all the DSM-IV (2) PDs, including the two ‘research’ diagnoses (passive-aggressive and depressive PDs). Internal consistency of symptom sets was determined by Cronbach’s alpha coefficient (19) and inter-criterion correlation analyses. Cronbach’s coefficient alpha (19) reflects the intercorrelation between items in a set (i.e. it is an estimate of the correlation between all possible split-half combinations of items), but alpha also increases as the number of items increases (17). For this reason, we additionally determined intercriterion correlation (MIC) to facilitate comparison between criterion lists of different lengths (22, 27; cf. 49). Two MICs were calculated: the mean and the median of the correlations for each possible pair of criteria within a given disorder. These coefficients (alpha and MIC) — calculated for each PD diagnosis — reflect the degree of within-category cohesiveness.

To examine criterion overlap between categories, we calculated an inter-category median intercriterion correlation (ICMIC) between all PDs. The ICMIC was calculated in a manner similar to the MIC: correlation coefficients were determined for all possible inter-category criterion pairs — each pair consisting of one criterion from each of the two categories being compared. We report ICMIC using the median of these correlation coefficients. The ICMIC reflects intercorrelation between categories and thus reveals the extent of criterion overlap.

Diagnostic efficiency statistics were calculated for the criteria sets for the four primary CLPS PD study groups (STPD, BPD, AVPD, and OCPD). Sensitivity (SEN) represents the proportion of participants with the diagnosis who have the criterion divided by participants with the diagnosis. Specificity (SPE) represents the participants without the diagnosis who do not have the criterion divided by the participants without the diagnosis. Positive predictive power (PPP) represents the participants with the criterion who have the diagnosis divided by participants with the criterion. Negative predictive power (NPP) represents the participants without the criterion who do not have the diagnosis divided by the participants without the criterion. We also calculated item-total correlations representing the correlation of the criterion and the total number of criterion. We
also report base rates for each individual criteria (i.e. participants with the particular criterion divided by total number of participants) (see 39). Lastly, we note (see 34,50,51) that PPP and NPP are affected by base rates of the diagnoses. Thus, PPP and NPP can fluctuate considerably across samples or study groups (36, 52).

Results

Table 1 shows alpha and both the mean and median intercorrelation values for each PD. Alpha coefficients ranged from 0.47 (schizoid PD) to 0.87 (antisocial PD) with a median alpha value of 0.71. Seven of the 10 DSM-IV PDs had alpha coefficients greater than 0.70 (schizoid, histrionic and obsessive compulsive PDs did not) and one of the two research categories (depressive PD) had an alpha greater than 0.70. Mean IC values ranged from 0.11 (schizoid PD) to 0.41 (avoidant PD) with a median value of 0.24. Median IC values ranged from 0.11 to 0.40 with a median of 0.23.

Given the CLPS recruitment procedure that targeted four specific PDs, we inspected the alpha and MIC values for those four PD groups separately from the other eight DSM-IV PD diagnoses. The mean (0.78) and median (0.80) alpha coefficients for the four PD groups were slightly higher than those for the other eight PD diagnoses (0.70 and 0.71, respectively). Similarly, the mean (0.31) and median (0.31) MIC values for the four PD groups were higher than for the other eight PD (0.23 and 0.24, respectively).

Table 2 provides the ICMIC values (median) for all personality disorder pairs. ICMIC values ranged from 0.00 to 0.19 with a median ICMIC value of 0.08. Thus, the overall ICMIC values were substantially lower than MIC values (from Table 1). This suggests that criteria for specific PD diagnoses are more inter-related with each other than they are with those for other PD diagnoses. Again, given our recruitment targeting four specific PD groups, we inspected the ICMIC values for these four PD groups separately with those for the other PD diagnoses. The mean ICMIC among the four PD diagnoses was 0.04, suggesting a high degree of discriminant validity. More central is the question of whether our targeting four PD artificially influenced this index of discriminant validity. We note that the mean ICMIC between the four CLPS PD study groups and other PD diagnoses (0.08) was identical to that between the criteria for the eight other PD diagnoses amongst themselves (0.08).

In our study characterized by base rates reflecting in large part our recruitment procedures targeting these four PD groups, we inspected the ICMIC values for these four PD groups separately with those for the other PD diagnoses. The mean ICMIC among the four PD diagnoses was 0.04, suggesting a high degree of discriminant validity. More central is the question of whether our targeting four PD artificially influenced this index of discriminant validity. We note that the mean ICMIC between the four CLPS PD study groups and other PD diagnoses (0.08) was identical to that between the criteria for the eight other PD diagnoses amongst themselves (0.08).

Table 3 provides the diagnostic efficiency statistics (SEN, SPE, PPP, NPP) along with base rates and item-total correlations for each criterion of the four PD study groups. Overall, mean base rates of individual criteria ranged from 0.19 (STPD) to 0.47 (AVPD). We comment briefly on these values descriptively, since adequate statistical analyses are not available to compare such conditional probability values (nor is our goal here to compare performance between diagnoses as much to produce data that speak to criterion usefulness as potential inclusion/exclusion markers).

In our study characterized by base rates reflecting in large part our recruitment procedures targeting these four PD diagnoses, mean PPP values were, in descending order: 0.82 (AVPD), 0.73 (OCPD), 0.71 (BPD) and 0.57 (STPD). The mean NPP values were, in descending order: 0.94 (STPD), 0.85 (BPD), 0.82 (AVPD), and 0.79 (OCPD). Variability was observed in the overall patterns of PPP and NPP within PD diagnoses. For example, in STPD the mean PPP (0.57) and NPP (0.94) varied greatly, whereas for AVPD the mean PPP and NPP were identical (0.82).

Our PPP and NPP values suggest the potential usefulness of certain criteria as inclusion and/or inclusion criteria (see 38,39). For STPD, whereas the absence of any criterion is generally predictive of absence of the diagnosis (i.e. high NPP values above 0.90), variability in PPP provide potential

Table 2. ICMIC values for DSM-IV personality disorders (N = 668)

<table>
<thead>
<tr>
<th></th>
<th>Paranoid</th>
<th>Schizoid</th>
<th>Schizotypal</th>
<th>Antisocial</th>
<th>Borderline</th>
<th>Histrionic</th>
<th>Narcissistic</th>
<th>Avoidant</th>
<th>Dependent</th>
<th>Obs-comp</th>
<th>Pass-Aggr</th>
<th>Depressive</th>
<th>ICMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizoid</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>Schizotypal</td>
<td>0.16</td>
<td>0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>Antisocial</td>
<td>0.08</td>
<td>0.03</td>
<td>0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>Borderline</td>
<td>0.11</td>
<td>0.05</td>
<td>0.06</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>Histrionic</td>
<td>0.08</td>
<td>0.01</td>
<td>0.07</td>
<td>0.08</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>Narcissistic</td>
<td>0.08</td>
<td>0.02</td>
<td>0.07</td>
<td>0.08</td>
<td>0.10</td>
<td>0.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>Avoidant</td>
<td>0.13</td>
<td>0.10</td>
<td>0.09</td>
<td>0.05</td>
<td>0.11</td>
<td>0.17</td>
<td>0.15</td>
<td>0.04</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>Dependent</td>
<td>0.07</td>
<td>0.05</td>
<td>0.01</td>
<td>0.03</td>
<td>0.15</td>
<td>0.17</td>
<td>0.15</td>
<td>0.11</td>
<td>0.06</td>
<td>0.15</td>
<td></td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>Obs-comp</td>
<td>0.06</td>
<td>0.03</td>
<td>0.01</td>
<td>0.00</td>
<td>0.09</td>
<td>0.17</td>
<td>0.15</td>
<td>0.03</td>
<td>0.07</td>
<td>0.02</td>
<td>0.02</td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>Pass–Aggr</td>
<td>0.13</td>
<td>0.06</td>
<td>0.09</td>
<td>0.12</td>
<td>0.17</td>
<td>0.17</td>
<td>0.15</td>
<td>0.11</td>
<td>0.12</td>
<td>0.05</td>
<td>0.05</td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>Depressive</td>
<td>0.11</td>
<td>0.06</td>
<td>0.06</td>
<td>0.04</td>
<td>0.12</td>
<td>0.06</td>
<td>0.04</td>
<td>0.19</td>
<td>0.12</td>
<td>0.05</td>
<td>0.05</td>
<td></td>
<td>0.14</td>
</tr>
</tbody>
</table>

ICMIC = inter-category median intercriterion correlation.
hints regarding usefulness of criterion as inclusion criteria. Three criteria (no close friends, paranoid ideation and ideas of reference) had PPP values <0.40 suggesting limited usefulness as inclusion criteria. The three observational criteria (odd behavior, odd thinking/speech, constricted affect) and social anxiety seemed useful as inclusion criteria given the high PPP values. Moreover, these four criteria occurred infrequently (base rates <0.16) but were highly specific to STPD (SPE >0.93).

For BPD, whereas the absence of any criterion is generally predictive of absence of the diagnosis (i.e. high NPP values above 0.80), variability in PPP provide potential hints regarding usefulness of criterion as inclusion criteria. Four criteria (intense anger, affective instability, chronic emptiness and impulsivity) had PPP <0.67 suggesting only moderate usefulness as inclusion criteria. The remaining criteria, most notably self-injury and suicidality, seemed useful as inclusion criteria given the PPP values (>0.70).

Table 3. Item analysis for four study group personality disorders: diagnostic efficiency statistics

<table>
<thead>
<tr>
<th>Category</th>
<th>BR</th>
<th>SEN</th>
<th>SPE</th>
<th>PPP</th>
<th>NPP</th>
<th>Item-total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizotypal PD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social anxiety</td>
<td>0.16</td>
<td>0.71</td>
<td>0.93</td>
<td>0.62</td>
<td>0.95</td>
<td>0.49</td>
</tr>
<tr>
<td>No close friends</td>
<td>0.25</td>
<td>0.58</td>
<td>0.81</td>
<td>0.34</td>
<td>0.92</td>
<td>0.30</td>
</tr>
<tr>
<td>Odd beliefs</td>
<td>0.20</td>
<td>0.78</td>
<td>0.90</td>
<td>0.56</td>
<td>0.96</td>
<td>0.50</td>
</tr>
<tr>
<td>Unusual experiences</td>
<td>0.25</td>
<td>0.81</td>
<td>0.85</td>
<td>0.48</td>
<td>0.96</td>
<td>0.51</td>
</tr>
<tr>
<td>Paranoid ideation</td>
<td>0.28</td>
<td>0.73</td>
<td>0.79</td>
<td>0.37</td>
<td>0.95</td>
<td>0.46</td>
</tr>
<tr>
<td>Ideas of reference</td>
<td>0.34</td>
<td>0.84</td>
<td>0.75</td>
<td>0.36</td>
<td>0.97</td>
<td>0.42</td>
</tr>
<tr>
<td>Odd behavior</td>
<td>0.06</td>
<td>0.38</td>
<td>0.99</td>
<td>0.86</td>
<td>0.91</td>
<td>0.44</td>
</tr>
<tr>
<td>Odd thinking/speech</td>
<td>0.09</td>
<td>0.46</td>
<td>0.98</td>
<td>0.77</td>
<td>0.92</td>
<td>0.49</td>
</tr>
<tr>
<td>Constricted affect</td>
<td>0.07</td>
<td>0.39</td>
<td>0.98</td>
<td>0.77</td>
<td>0.90</td>
<td>0.43</td>
</tr>
<tr>
<td>Mean</td>
<td>0.19</td>
<td>0.63</td>
<td>0.89</td>
<td>0.57</td>
<td>0.94</td>
<td>0.45</td>
</tr>
<tr>
<td>Borderline PD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intense anger</td>
<td>0.46</td>
<td>0.85</td>
<td>0.76</td>
<td>0.67</td>
<td>0.90</td>
<td>0.56</td>
</tr>
<tr>
<td>Affective instability</td>
<td>0.52</td>
<td>0.94</td>
<td>0.72</td>
<td>0.65</td>
<td>0.90</td>
<td>0.60</td>
</tr>
<tr>
<td>Chronic emptiness</td>
<td>0.42</td>
<td>0.71</td>
<td>0.74</td>
<td>0.61</td>
<td>0.82</td>
<td>0.42</td>
</tr>
<tr>
<td>Identity disturbance</td>
<td>0.32</td>
<td>0.62</td>
<td>0.85</td>
<td>0.70</td>
<td>0.80</td>
<td>0.47</td>
</tr>
<tr>
<td>Stress-related paranoia</td>
<td>0.32</td>
<td>0.68</td>
<td>0.87</td>
<td>0.75</td>
<td>0.83</td>
<td>0.53</td>
</tr>
<tr>
<td>Avoid abandonment</td>
<td>0.28</td>
<td>0.60</td>
<td>0.89</td>
<td>0.76</td>
<td>0.80</td>
<td>0.49</td>
</tr>
<tr>
<td>Self-injury, suicidality</td>
<td>0.26</td>
<td>0.59</td>
<td>0.93</td>
<td>0.82</td>
<td>0.81</td>
<td>0.52</td>
</tr>
<tr>
<td>Impulsivity</td>
<td>0.46</td>
<td>0.82</td>
<td>0.75</td>
<td>0.64</td>
<td>0.88</td>
<td>0.48</td>
</tr>
<tr>
<td>Unstable relationships</td>
<td>0.39</td>
<td>0.82</td>
<td>0.84</td>
<td>0.75</td>
<td>0.89</td>
<td>0.60</td>
</tr>
<tr>
<td>Mean</td>
<td>0.38</td>
<td>0.74</td>
<td>0.82</td>
<td>0.71</td>
<td>0.85</td>
<td>0.52</td>
</tr>
<tr>
<td>Avoidant PD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Socially inept</td>
<td>0.59</td>
<td>0.91</td>
<td>0.71</td>
<td>0.75</td>
<td>0.90</td>
<td>0.60</td>
</tr>
<tr>
<td>Being rejected</td>
<td>0.52</td>
<td>0.89</td>
<td>0.83</td>
<td>0.83</td>
<td>0.89</td>
<td>0.66</td>
</tr>
<tr>
<td>Needs to be liked first</td>
<td>0.52</td>
<td>0.82</td>
<td>0.77</td>
<td>0.77</td>
<td>0.82</td>
<td>0.53</td>
</tr>
<tr>
<td>Feels inadequate</td>
<td>0.56</td>
<td>0.94</td>
<td>0.79</td>
<td>0.81</td>
<td>0.93</td>
<td>0.67</td>
</tr>
<tr>
<td>Fears being ridiculed</td>
<td>0.38</td>
<td>0.63</td>
<td>0.85</td>
<td>0.80</td>
<td>0.71</td>
<td>0.46</td>
</tr>
<tr>
<td>Avoids interpersonal contact</td>
<td>0.36</td>
<td>0.68</td>
<td>0.95</td>
<td>0.93</td>
<td>0.76</td>
<td>0.61</td>
</tr>
<tr>
<td>Fears embarrassment</td>
<td>0.37</td>
<td>0.66</td>
<td>0.91</td>
<td>0.88</td>
<td>0.74</td>
<td>0.56</td>
</tr>
<tr>
<td>Mean</td>
<td>0.47</td>
<td>0.79</td>
<td>0.83</td>
<td>0.82</td>
<td>0.82</td>
<td>0.58</td>
</tr>
<tr>
<td>Obsessive-Compulsive PD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rigid and stubborn</td>
<td>0.45</td>
<td>0.79</td>
<td>0.77</td>
<td>0.69</td>
<td>0.85</td>
<td>0.43</td>
</tr>
<tr>
<td>Misery</td>
<td>0.14</td>
<td>0.29</td>
<td>0.96</td>
<td>0.82</td>
<td>0.68</td>
<td>0.27</td>
</tr>
<tr>
<td>Pack rat</td>
<td>0.40</td>
<td>0.64</td>
<td>0.75</td>
<td>0.62</td>
<td>0.76</td>
<td>0.29</td>
</tr>
<tr>
<td>Perfectionism</td>
<td>0.43</td>
<td>0.78</td>
<td>0.80</td>
<td>0.72</td>
<td>0.85</td>
<td>0.45</td>
</tr>
<tr>
<td>Details, rules, lists, order</td>
<td>0.29</td>
<td>0.60</td>
<td>0.92</td>
<td>0.83</td>
<td>0.78</td>
<td>0.43</td>
</tr>
<tr>
<td>Reluctant to delegate tasks</td>
<td>0.46</td>
<td>0.82</td>
<td>0.78</td>
<td>0.71</td>
<td>0.67</td>
<td>0.52</td>
</tr>
<tr>
<td>Inflexibility about morality</td>
<td>0.31</td>
<td>0.58</td>
<td>0.86</td>
<td>0.73</td>
<td>0.76</td>
<td>0.37</td>
</tr>
<tr>
<td>Workaholic</td>
<td>0.27</td>
<td>0.51</td>
<td>0.88</td>
<td>0.73</td>
<td>0.73</td>
<td>0.31</td>
</tr>
<tr>
<td>Mean</td>
<td>0.34</td>
<td>0.62</td>
<td>0.84</td>
<td>0.73</td>
<td>0.79</td>
<td>0.38</td>
</tr>
</tbody>
</table>
For AVPD, the absence of four criteria (socially inept, fear of being rejected, need to be liked first and feeling inadequate) is generally predictive of absence of the diagnosis (i.e. high NPP values above 0.82). Variability in PPP provides potential hints regarding usefulness of criterion as inclusion criteria. Two criteria (socially inept, needs to be liked first) had PPP <0.77, which while suggesting moderate usefulness as inclusion criteria did not seem as useful as the remaining five criteria as inclusion criteria. These two criteria (socially inept and need to be liked first) were characterized further by high base rates and relatively low specificity to the AVPD diagnosis.

For OCPD, the absence of three criteria (rigid and stubborn, perfectionism, reluctant to delegate tasks) is generally predictive of absence of the diagnosis (i.e. high NPP values above 0.85). Variability in PPP provides hints regarding usefulness of criterion as inclusion criteria. Two criteria (miserly, details and rules) had PPP >0.92 showed usefulness as inclusion criteria.

Discussion

Our psychometric analysis used symptom data from a large treatment-seeking study group obtained through semi-structured diagnostic interviews performed by highly trained and carefully monitored research evaluators. The findings suggest that the PD criteria sets have some degree of convergent validity (as indicated by acceptable alphas) and that the PD criteria have some discriminant validity: the criteria for any given PD generally correlated better with each other (reflected in the MIC value) than with the criteria for other PDs (reflected in the ICMIC values). We also report diagnostic efficiency statistics for the four PD study groups (STPD, BPD, AVPD, OCPD).

It is important to note that our assessment interview did not involve randomized assessment of PD criteria. In other words, criteria for a given disorder were all assessed sequentially, in accordance with the structure of the DIPD-IV (33) interview. Our assessment methodology — while perhaps more clinically ecological — does have the possibility of a ‘halo’ effect (i.e. clinicians might be inclined to rate criteria similarly within a given disorder), which could artificially inflate internal consistency and minimize intercorrelations between categories. Moreover, given the modest convergence between diagnoses generated by available diagnostic interviews (e.g. 11) or perhaps for the individual criteria, our results could be somewhat different if we had utilized alternative instruments.

Since our patient study group is large, demographically diverse and derived from a large number of different clinical sites, generalizability to many other clinical populations is expected. None the less, our study group may not be generalizable to some clinical settings (e.g. non-urban, non-academic-affiliated facilities) or to community (non-clinical) populations. Different samples of varying composition may produce different patterns of associations among criteria within and between PDs. We note here that our recruitment procedures ensured a full diagnostic range for four PDs of primary interest (STPD, BPD, AVPD, OCPD) and our control group (MDD) contributed needed non-PD-level variability in criteria. While this strategy has its considerable strengths (i.e. good representation of these four PDs and representation of subjects with PDs across the three DSM-IV (2) clusters), it is possible that a comparable diagnostic range of criteria may not be available for the remaining PDs. However, we note the substantial frequency of all PDs present in this study group and that our rates of PD co-occurrence are similar to those reported for other samples (11, 45). Moreover, while we observed that the degree of criterion inter-relatedness was somewhat greater for the four PD study groups, that the degree of between-category overlap (ICMIC) did not differ between the four PD groups versus the other (six plus two research category) PD diagnoses.

We stress, however, that our recruitment procedures could have influenced base-rates and that the base-rates of the PD diagnoses in our clinical study groups may differ from those in other settings. Base rates influence PPP and NPP (34, 50, 51). Thus, PPP and NPP can fluctuate across samples (36, 52). It is crucial that other researchers publish similar data in order to allow ‘cross-validation’ of some of the performance characteristics of PD.

In conclusion, our findings suggest overall adequate, albeit modest, within-category interrelatedness for the DSM-IV PD criteria (convergent validity). ICMIC values were lower than MIC values (suggesting some discriminant validity). We also report base rates and diagnostic efficiency statistics for individual criteria for schizotypal, borderline, avoidant and obsessive-compulsive PDs. Variability in positive predictive power and negative predictive power provide some guidance regarding usefulness of specific criteria as inclusion or exclusion criteria. These findings, however, require replication and extension to other patient groups as well as to community samples characterized by different base rates of PD criteria and diagnoses. Such psychometric analyses of the PD components should also be considered in relation to
Axis I psychiatric disorders and states (e.g. 53) and developmental histories and putative risk factors (e.g. 54). Continued psychometric work is needed to complement expert clinical experience to further refine PD constructs.

Acknowledgement

This study was supported by grants: R10 MH 50837, 50838, 50839, 50840, 50850 and K05 MH 1645 from the National Institutes of Mental Health.

References


Performance characteristics of PDs


33. Zanarini MC. The Diagnostic Interview for DSM-IV Personality Disorders (DIPD-IV). USA: McLean Hospital and Harvard Medical School, 1996.


