Internal consistency, intercriterion overlap and diagnostic efficiency of criteria sets for DSM-IV schizotypal, borderline, avoidant and obsessive-compulsive personality disorders

Grilo CM, McGlashan TH, Morey LC, Gunderson JG, Skodol AE, Tracie SM, Sanislow CA, Zanarini MC, Bender D, Oldham JM, Dyck I, Stout RL. Internal consistency, intercriterion overlap and diagnostic efficiency of criteria sets for DSM-IV schizotypal, borderline, avoidant and obsessive-compulsive personality disorders.
Acta Psychiatr Scand 2001: 104: 264–272. © Munksgaard 2001.

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.

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

Collaborative Longitudinal Personality Study: [1]McLean Hospital and Harvard Medical School, [2]Brown University Department of Psychiatry and Human Behavior, [3]New York State Psychiatric Institute and Columbia University College of Physicians and Surgeons, [4]Yale University School of Medicine and [5]Texas A&M University Department of Psychology, USA

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-occur-

rence (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 inter-relatedness 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 retro-spective 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 inter-relatedness 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. caseness versus noncaseness) as evaluated from actuarial prediction (34). Since the early applications of Yerushalmy (35) to medical diagnoses, this work has evolved considerably (34, 36) and has contributed 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 diagnostic 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 cohesiveness of DSM-IV (2) PD diagnoses in adult patients assessed with semi-structured diagnostic interviews. In addition to evaluating *within*-category interrelatedness, we sought to use a similar psychometric approach to examine the degree of *between*-category criterion overlap. In addition, this represents 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 characteristics 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 supplemented by participants responding to postings and advertisements for an interview study of personality, 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 (schizotypal (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 participants 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-occurrence 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 variability 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 evaluation, 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

Table 1. Internal consistencies for DSM-IV personality disorders (N = 668)

Axis II DIPD-4 personality disorders	N	No. of items	Alpha	MeanIC	MedianIC
Cluster A					
Paranoid	81	7	0.70	0.25	0.22
Schizoid	18	7	0.47	0.11	0.11
Schizotypal	96	9	0.77	0.28	0.25
Cluster B					
Antisocial	49	22	0.87	0.24	0.23
Borderline	240	9	0.82	0.33	0.33
Histrionic	13	8	0.64	0.18	0.18
Narcissistic	36	9	0.76	0.26	0.25
Cluster C					
Avoidant	324	7	0.83	0.41	0.40
Dependent	49	9	0.71	0.24	0.22
Obsessive-compulsive	261	8	0.69	0.22	0.20
Research diagnoses					
Passive-aggressive	54	7	0.69	0.24	0.23
Depressive	177	7	0.75	0.30	0.31

Alpha = Cronbach's coefficient alpha (standardized variables); meanIC = mean intercriterion correlation; medianIC = median intercriterion correlation.

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 inter-criterion 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 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

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

	Paranoid	Schizoid	Schizotypal	Antisocial	Borderline	Histrionic	Narcissistic	Avoidant	Dependent	Obs-comp	Pass-Aggr
Schizoid	0.10										
Schizotypal	0.16	0.08									
Antisocial	0.08	0.03	0.08								
Borderline	0.11	0.05	0.06	0.11							
Histrionic	0.08	0.01	0.07	0.08	0.13						
Narcissistic	0.08	0.02	0.07	0.08	0.10	0.17					
Avoidant	0.13	0.10	0.09	0.05	0.11	0.04	0.04				
Dependent	0.07	0.05	0.01	0.03	0.15	0.11	0.06	0.15			
Obs-comp	0.06	0.03	0.01	0.00	0.00	0.03	0.07	0.02	0.02		
Pass–Aggr	0.13	0.06	0.09	0.12	0.17	0.15	0.17	0.11	0.12	0.05	
Depressive	0.11	0.06	0.06	0.04	0.12	0.06	0.04	0.19	0.12	0.05	0.14

ICMIC = inter-category median intercriterion correlation.

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

	BR	SEN	SPE	PPP	NPP	Item-total
Schizotypal PD						
Social anxiety	0.16	0.71	0.93	0.62	0.95	0.49
No close friends	0.25	0.58	0.81	0.34	0.92	0.30
Odd beliefs	0.20	0.78	0.90	0.56	0.96	0.50
Unusual experiences	0.25	0.81	0.85	0.48	0.96	0.51
Paranoid ideation	0.28	0.73	0.79	0.37	0.95	0.46
Ideas of reference	0.34	0.84	0.75	0.36	0.97	0.42
Odd behavior	0.06	0.38	0.99	0.86	0.91	0.44
Odd thinking/speech	0.09	0.46	0.98	0.77	0.92	0.49
Constricted affect	0.07	0.39	0.98	0.77	0.90	0.43
Mean	0.19	0.63	0.89	0.57	0.94	0.45
Borderline PD						
Intense anger	0.46	0.85	0.76	0.67	0.90	0.56
Affective instability	0.52	0.94	0.72	0.65	0.90	0.60
Chronic emptiness	0.42	0.71	0.74	0.61	0.82	0.42
Identity disturbance	0.32	0.62	0.85	0.70	0.80	0.47
Stress-related paranoia	0.32	0.68	0.87	0.75	0.83	0.53
Avoid abandonment	0.28	0.60	0.89	0.76	0.80	0.49
Self-injury, suicidality	0.26	0.59	0.93	0.82	0.81	0.52
Impulsivity	0.46	0.82	0.75	0.64	0.88	0.48
Unstable relationships	0.39	0.82	0.84	0.75	0.89	0.60
Mean	0.38	0.74	0.82	0.71	0.85	0.52
Avoidant PD						
Socially inept	0.59	0.91	0.71	0.75	0.90	0.60
Being rejected	0.52	0.89	0.83	0.83	0.89	0.66
Needs to be liked first	0.52	0.82	0.77	0.77	0.82	0.53
Feels inadequate	0.56	0.94	0.79	0.81	0.93	0.67
Fears being ridiculed	0.38	0.63	0.85	0.80	0.71	0.46
Avoids interpersonal contact	0.36	0.68	0.95	0.93	0.76	0.61
Fears embarrassment	0.37	0.66	0.91	0.88	0.74	0.56
Mean	0.47	0.79	0.83	0.82	0.82	0.58
Obsessive-Compulsive PD						
Rigid and stubborn	0.45	0.79	0.77	0.69	0.85	0.43
Miserly	0.14	0.29	0.96	0.82	0.68	0.27
Pack rat	0.40	0.64	0.75	0.62	0.76	0.29
Perfectionism	0.43	0.78	0.80	0.72	0.85	0.45
Details, rules, lists, order	0.29	0.60	0.92	0.83	0.78	0.43
Reluctant to delegate tasks	0.46	0.82	0.78	0.71	0.87	0.52
Inflexibility about morality	0.31	0.58	0.86	0.73	0.76	0.37
Workaholic	0.27	0.51	0.88	0.73	0.73	0.31
Mean	0.34	0.62	0.84	0.73	0.79	0.38

BR = base rate (participants with the individual criterion divided by total number participants).
SEN = sensitivity (participants with dx who have criterion divided by participants with dx).
SPE = specificity (participants without dx who do not have criterion divided by participants without dx).
PPP = positive predictive power (participants with criterion who have dx divided by participants with criterion).
NPP = negative predictive power (participants without criterion who do not have dx divided by participants without criterion).
Item-total = correlation of criterion and total number of criteria.

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).

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 provide 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 inter-relatedness 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

References

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. Washington, DC: American Psychiatric Association, 1952.
2. American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 4th edn. Washington, DC: American Psychiatric Association, 1994.
3. American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 3rd edn. Washington, DC: American Psychiatric Association, 1980.
4. Skodol AE. Classification, assessment, and differential diagnosis of personality disorders. J Pract Psychiatry Behav Health 1997;**3**:261–274.
5. Grilo CM, McGlashan TH, Oldham JM. The course and stability of personality disorders. J Pract Psychiatry Behav Health 1998;**4**:61–75.
6. Perry JC. Problems and considerations in the valid assessment of personality disorders. Am J Psychiatry 1992;**149**:1645–1652.
7. Westen D. Divergences between clinical and research methods for assessing personality disorders: implications for research and the evolution of Axis II. Am J Psychiaty 1997;**154**:895–903.
8. Widiger T. DSM-IV reviews of the personality disorders: introduction to special series. J Pers Disord 1991;**5**:122–134.
9. Widiger T, Trull TJ. Performance characteristics of the DSM-III-R personality disorder criteria sets. In: Widiger TA, Frances AJ, Pincus HA et al., eds. DSM-IV source book, vol. 4. Washington, DC: American Psychiatric Association, 1998.
10. Zimmerman M. Diagnosing personality disorders: a review of issues and research methods. Arch Gen Psychiatry 1994;**51**:225–245.
11. Oldham JM, Skodol AE, Kellman HD, Hyler SE, Rosnick L, Davies M. Diagnosis of DSM-III-R personality disorders by two structured interviews: patterns of comorbidity. Am J Psychiatry 1992;**149**:213–220.
12. Blashfield R, Draguns J. Evaluative criteria for psychiatric classification. J Abnorm Psychol 1976;**85**:140–150.
13. Gunderson JG. DSM-IV personality disorders: final overview. In: Widiger TA, Frances AJ, Pincus HA et al., eds. DSM-IV source book, vol. 4. Washington, DC: American Psychiatric Association, 1998.
14. Livesey WJ. Theoretical and empirical issues in the selection of criteria to diagnose personality disorders. J Pers Disord 1987;**1**:88–94.
15. Livesey WJ, ed. The DSM-IV personality disorders. New York: Guilford Press, 1995.
16. Clark LA, Livesley WJ, Morey L. Personality disorder assessment: the challenge of construct validity. J Pers Disord 1997;**11**:205–231.
17. Cortina JM. What is coefficient alpha? An examination of theory and applications. J Appl Psychol 1993;**78**:98–104.
18. Morey L. A psychometric analysis of five DSM-III-R categories. Pers Ind Diff 1985;**6**:323–329.
19. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;**16**:297–334.
20. 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.
21. Pfohl B, Coryell W, Zimmerman M, Stangl D. DSM-III personality disorders: diagnostic overlap and internal consistency of individual DSM-III criteria. Compr Psychiatry 1986;**27**:22–34.
22. Morey LC. Personality disorders in DSM-III and DSM-III-R: convergence, coverage, and internal consistency. Am J Psychiatry 1988;**145**:573–577.
23. American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 3rd edn, revised. Washington, DC: American Psychiatric Association, 1987.
24. Blais MA, Benedict KB, Norman DK. Establishing the psychometric properties of the DSM-III-R personality disorders: implications for DSM-V. J Clin Psychol 1998;**5**:795–802.
25. Nunnally J, Bernstein I. Psychometric theory, 3rd edn. New York: McGraw-Hill, 1994.
26. Blais MA, Hilsenroth MJ, Castlebury FD. Psychometric characteristics of the Cluster B personality disorders under DSM-III-R and DSM-IV. J Pers Disord 1997;**11**:270–278.
27. Becker DF, Grilo CM, Morey LC, Walker ML, Edell WS, McGlashan TH. Applicability of personality disorder criteria to hospitalized adolescents: evaluation of internal consistency and criterion overlap. J Am Acad Child Adolesc Psychiatry 1999;**38**:200–205.
28. Loranger AW. Personality disorder examination (PDE) manual. Yonkers, New York: DV Communications, 1988.
29. Blais MA, Norman DK. A psychometric evaluation of the DSM-IV personality disorder criteria sets. J Pers Disord 1997;**11**:168–176.
30. Maffei C, Fossati A, Agostoni I et al. Interrater reliability and internal consistency of the structured clinical interview for DSM-IV Axis II personality disorders (SCID-II), version 2.0. J Pers Disord 1997;**11**:279–284.
31. Grilo CM, McGlashan TH. Convergent and discriminant validity of DSM-IV axis II personality disorder criteria in adult outpatients with binge eating disorder. Compr Psychiatry 2000;**41**:159–162.
32. First MB, Spitzer RL, Gibbon M, Williams JBW. Structured clinical interview for DSM-IV Axis II personality disorders (SCID-II). Version 2.0. New York: New York State Psychiatric Institute, 1994.
33. Zanarini MC. The Diagnostic Interview for DSM-IV Personality Disorders (DIPD-IV). USA: McLean Hospital and Harvard Medical School, 1996.
34. Widiger TA, Hurt SW, Frances A, Clarkin JF, Gilmore M. Diagnostic efficiency and DSM-III. Arch Gen Psychiatry 1984;**41**:1005–1012.
35. Yerushalmy J. Statistical problems in assessing methods of medical diagnosis, with special references to X-ray techniques. Public Health Rep 1947;**62**:1142–1439.
36. Baldessarini RJ, Finkelstein S, Arana G. The predictive power of diagnostic tests and the effect of prevalence of illness. Arch Gen Psychiatry 1983;**40**:569–573.
37. Faraone SV, Biederman J, Sprich-Buckminster S, Chen W, Tsuang MT. Efficiency of diagnostic criteria for attention deficit disorder: toward an empirical approach to designing and validating diagnostic algorithms. J Am Acad Child Adolesc Psychiatry 1998;**32**:166–174.

38. Milich R, Widiger TA, Landau S. Differential diagnosis of attention deficit and conduct disorders using conditional probabilities. J Cons Clin Psychol 1987;**55**:762–767.
39. Waldman ID, Lilienfeld SO. Diagnostic efficiency of symptoms for oppositional defiant disorder and attention-deficit hyperactivity disorder. J Consult Clin Psychol 1991;**59**:732–738.
40. Zanarini MC, Gunderson JG, Frankenburg FR, Chauncey DL. Discriminating borderline personality criteria from other axis II disorders. Am J Psychiatry 1990;**147**:161–167.
41. Fossati A, Maffei C, Bagnato M et al. A psychometric study of DSM-IV passive- aggressive (negativistic) personality disorder criteria. J Pers Disord 2000;**14**:72–83.
42. Gunderson JG, Shea MT, Skodol AE et al. The collaborative longitudinal personality disorders study: development, aims, design, and sample characteristics. J Pers Disord 2000;**14**:300–315.
43. Skodol AE, Stout RL, McGlashan TH et al. Co-occurrence of mood and personality disorders: a report from the collaborative longitudinal personality disorders study. Depress Anxiety 1999;**10**:175–182.
44. McGlashan TH, Grilo CM, Skodol AE et al. The collaborative longitudinal personality disorders study II: baseline patterns of DSM-IV Axis I/II and II/II diagnostic co- occurrence. Acta Psychiatr Scand 2000;**102**:256–264.
45. Stuart S, Pfohl B, Bataglia M, Bellodi L, Grove W, Cadoret R. The co-occurrence of DSM-III-R personality disorders. J Pers Disord 1998;**12**:302–315.
46. Shea MT, Pilkonis PA, Beckham E et al. Personality disorders and treatment outcome in the NIMH Treatment of Depression Collaborative Research Program. Am J Psychiatry 1990;**147**:711–718.
47. Clark LA. Schedule for Nonadaptive and Adaptive Personality (SNAP). Minneapolis, MN: University of Minnesota Press, 1993.
48. Zanarini MC, Skodol AE, Bender D et al. The collaborative longitudinal personality disorders study: reliability of Axis I and II diagnoses. J Pers Disord 2000;**14**:291–299.
49. Hattie J. Methodology review: assessing unidimensionality of tests and items. Appl Psychol Measure 1985;**9**:139–164.
50. Dawes RM. A note on base rates and psychometric efficiency. J Consult Psychol 1962;**26**:422–424.
51. Finn SE. Base rates, utilities, and DSM-III: shortcomings of fixed-rule systems of psychodiagnostics. J Abnorm Psychol 1982;**91**:294–302.
52. Meehl PE, Rosen A. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. Psychol Bull 1955;**52**:194–216.
53. Mataix-Cols D, Baer L, Rauch SL, Jenike MA. Relation of factor-analyzed symptom dimensions of obsessive-compulsive disorder to personality disorders. Acta Psychiatr Scand 2000;**102**:199–202.
54. Liotti G, Pasquini P. Predictive factors for borderline personality disorder: patients' early traumatic experiences and losses suffered by the attachment figure. Acta Psychiatr Scand 2000;**102**:282–289.