

Drew Westen, Ph.D.
Jonathan Shedler, Ph.D.
Rebekah Bradley, Ph.D.

A Prototype Approach to Personality Disorder Diagnosis

Objective: Virtually no research has tested alternatives to the diagnostic method used since DSM-III, which requires decisions about the presence/absence of individual diagnostic criteria, followed by counting symptoms and applying cutoffs (the count/cutoff method). This study tested an alternative, prototype matching procedure designed to simplify diagnosis. The procedure was applied to personality disorders. **Method:** A random national sample of psychiatrists and clinical psychologists (N=291) described a randomly selected patient in their care. Clinician-provided diagnostic data were used to generate categorical and dimensional DSM-IV diagnoses (number of symptoms present per disorder). Clinicians also used one of two prototype matching systems to provide a diagnosis for the selected patient. **Results:** Prototype diagnosis led to reduced comorbidity relative to DSM-IV diagnosis, yielded similar estimates of validity in predicting criterion variables (adaptive functioning, treatment response, and etiology), and outperformed DSM-IV diagnosis in ratings of clinical utility and ease of use. Adding a personality health prototype further increased prediction. **Conclusions:** A simple prototype matching procedure provides a viable alternative for improving diagnosis of personality disorders in clinical practice. Prototype diagnosis has multiple advantages, including ease of use, minimization of artifactual comorbidity, compatibility with naturally occurring cognitive processes, and ready translation into both categorical and dimensional diagnosis.

(Reprinted with permission from *The American Journal of Psychiatry* 2006; 163:846–856)

As applied to psychiatry, classification has two components: 1) taxonomy (establishing diagnostic groupings) and 2) diagnosis (applying those groupings to individual cases) (1). Personality disorder researchers have focused considerable attention on taxonomy (i.e., refining categories and criteria); however, no one has systematically tested alternative methods of diagnosing cases since axis II first appeared in 1980. The approach used since DSM-III requires dichotomous (present/absent) decisions about roughly 80 diagnostic criteria followed by counting to determine whether the number of criteria exceeds cutoffs (hereafter, the *count/cutoff* approach).

The merits of a diagnostic system can be evaluated with respect to three classes of criteria. *Internal criteria* include characteristics internal to the system, such as coherence (Does the system describe conceptually meaningful syndromes?), comprehensiveness (Does it encompass the spectrum of pathology?),

and parsimony (Does it define distinct, non-redundant syndromes?). *External criteria* link the diagnostic constructs to conceptually relevant external criterion variables such as etiological factors, treatment response, level of adaptive functioning, and laboratory findings (2). *Clinical criteria* address the extent to which clinicians find the diagnostic system relevant and useful in real-world application. None of these classes of criteria is alone definitive. A diagnostic method that has high predictive validity but is not readily used by treating clinicians may not be diagnostically useful (3). This study applies these three classes of criteria to evaluate approaches to diagnosis of personality disorders.

WHY SHOULD AN ALTERNATIVE DIAGNOSTIC METHOD BE CONSIDERED?

The count/cutoff method emerged from the Research Diagnostic Criteria of the 1970s (4). It had

clear advantages over the subjective decision rules of DSM-II (5) and has facilitated tremendous progress in personality disorder research since that time. However, several limitations have become apparent. First, most personality traits are continuously, not dichotomously, distributed in nature (6). Second, comorbidity among personality disorder diagnoses is so high that researchers frequently report data at the level of the three axis II clusters rather than making specific personality disorder diagnoses. Third, for psychometric reasons, it is virtually impossible for criterion sets of only seven to nine criteria per disorder to describe complex, multifaceted personality disorder syndromes while simultaneously delineating distinct, nonoverlapping categories (7, 8). Fourth, the method does not take into account the cognitive processing parameters of human diagnosticians (cognitive economy). The diagnostic criteria were not selected or organized in a way that allows clinicians to form coherent mental representations of the disorders and are not linked by functional or causal relations important to human category judgment (9). Indeed, clinicians rarely follow the diagnostic procedures prescribed by DSM-IV, and when they do, the resulting diagnoses have limited reliability and validity (10, 11).

A PROTOTYPE MATCHING APPROACH TO PERSONALITY DISORDER DIAGNOSIS

Elsewhere we have proposed a prototype matching approach to personality disorder diagnosis. This approach was designed to facilitate accurate classification while taking into consideration the cognitive requirements of the human diagnostician (12–14). Approaches to classification based on prototypes or exemplars have a long history in cognitive science and were first applied to psychiatric diagnosis 25 years ago (15–17). The proposed method presents clinicians with each personality disorder in its ideal or “pure” form. These prototype descriptions are presented in paragraph rather than list form and are psychologically richer and more detailed than DSM-IV criterion sets (which are limited to seven to nine features per disorder), allowing diagnosticians to form mental representations of psychologically coherent syndromes in which behavior and inner experience are linked by meaningful functional relations. (We thank Robert Spitzer and Michael First for convincing us of the superiority of the paragraph format for this purpose, as well as for helping design the questions we used to assess clinical criteria in this study.) To make a diagnosis, diagnosticians rate the overall similarity or “match” between a patient and the prototype using a 5-point rating scale, considering the prototype as

Figure 1. Prototype Description of Antisocial-Psychopathic Personality Disorder

Patients who match this prototype tend to be deceitful and to lie and mislead others. They take advantage of others, have minimal investment in moral values, and appear to experience no remorse for harm or injury caused to others. They tend to manipulate others' emotions to get what they want; to be unconcerned with the consequences of their actions, appearing to feel immune or invulnerable; and to show reckless disregard for the rights, property, or safety of others. They have little empathy and seem unable to understand or respond to others' needs and feelings unless they coincide with their own. Individuals who match this prototype tend to act impulsively, without regard for consequences; to be unreliable and irresponsible (e.g., failing to meet work obligations or honor financial commitments); to engage in unlawful or criminal behavior; and to abuse alcohol. They tend to be angry or hostile; to get into power struggles; and to gain pleasure or satisfaction by being sadistic or aggressive toward others. They tend to blame others for their own failures or shortcomings and believe that their problems are caused entirely by external factors. They have little insight into their own motives, behavior, etc. They may repeatedly convince others of their commitment to change but then revert to previous maladaptive behavior, often convincing others that “this time is really different.”

5	Very good match (patient <i>exemplifies</i> this disorder; prototypical case)	Diagnosis
4	Good match (patient <i>has</i> this disorder; diagnosis applies)	
3	Significant match (patient has <i>significant features</i> of this disorder)	Features
2	Slight match (patient has minor features of this disorder)	
1	Little or no match (description does not apply to this patient)	

a whole rather than counting individual symptoms (Figure 1).

This method generates both categorical and dimensional diagnoses. Ratings of 4 or 5 denote a categorical diagnosis (“caseness”), and a rating of 3 translates to the concept of “features” or subthreshold pathology. The method parallels diagnosis in many areas of medicine, where variables (e.g., blood pressure) are measured on a continuum but physicians by convention refer to certain ranges as “borderline” or “high.” The ready translation of dimensional into categorical diagnosis facilitates communication among professionals, overcoming a significant limitation of dimensional diagnosis.

This study compares four methods of personality disorder diagnosis, focusing on the cluster B disorders, because they are the most frequently studied,

have the best documented correlates, and are among the most prevalent of the personality disorders. The first method is *DSM-IV categorical diagnosis* (the count/cutoff approach). The second is a dimensionalized version of the count/cutoff approach in which the patient's score for each personality disorder equals the number of criteria met for the disorder (hereafter, *DSM-IV dimensional diagnosis*). (We included this second method, which is widely used in personality disorder research, to control for any effects that might be attributable simply to the differences between categorical and dimensional diagnosis.)

The third and fourth methods are alternative implementations of prototype matching, reflecting two ways of generating prototypes. The *clinician prototype* method reflects the shared understanding of experienced clinicians regarding the important features of each DSM-IV personality disorder. A national sample of experienced clinicians used a 200-item Q-sort instrument for assessing personality pathology (the Shedler-Westen Assessment Procedure-200 [SWAP-200]) to describe their mental prototype of a patient who illustrates a given personality disorder in its "purest" or ideal form (7, 18). We aggregated the resulting personality disorder descriptions to obtain a single composite description for each personality disorder.

The *empirical prototype* method reflects a purely empirical approach to identifying personality disorders, without regard to existing DSM-IV diagnostic categories. A national sample of experienced clinicians used the SWAP-200 to describe a specific personality disorder patient in their care (19). We applied a statistical procedure (Q-analysis) (20) to the resulting descriptions to identify empirically distinct diagnostic groupings, reflecting "natural" cleavages in the patient sample. Q-analysis is computationally identical to factor analysis, except that the data matrix is transposed, so that patient cases are factored over items rather than the items being factored over cases. The resulting Q-factors describe kinds of people (i.e., the characteristics shared by people with similar personality profiles). An empirical prototype is a statistically generated composite description of patients identified empirically whose profiles are similar to one another.

METHOD

CLINICIAN PARTICIPANTS

As part of an NIMH-funded project on the classification of personality pathology, we contacted a random national sample of psychiatrists and

psychologists with at least 5 years of experience postresidency or postlicensure from the membership registers of the American Psychiatric Association and the American Psychological Association, including clinicians targeted in prior solicitations. Approximately 35% of the clinicians agreed to participate; those who submitted completed materials received a consulting fee of \$200.

We asked the clinicians to describe "an adult patient you are currently treating or evaluating who has enduring patterns of thought, feeling, motivation or behavior—that is, personality problems—that cause distress or dysfunction." To obtain a broad range of examples of personality pathology, we emphasized that the patients need not have a personality disorder diagnosis. Patients had to meet the following additional inclusion criteria: age ≥ 18 years, not currently psychotic, and known well by the clinician (using the guideline of ≥ 6 clinical contact hours but ≤ 2 years, to minimize confounds imposed by personality change with treatment). To minimize selection biases, we directed the clinicians to consult their calendars to select the last patient they saw during the previous week who met study criteria.

MEASURES

Clinical Data Form. The Clinical Data Form (Table 1) is a clinician-report form developed to assess a range of variables relevant to demographic characteristics, diagnosis, and etiology (7). Clinicians rate the patient's adaptive functioning and also rate developmental and family history variables with which clinicians who have met with a patient over a number of hours are likely to be familiar (e.g., history of foster care, family history of criminality). In prior studies, clinicians' judgments on these variables have predicted theoretically relevant criterion variables and reflected reasonable (and conservative) decision rules (21). To investigate an exploratory appendix to the Clinical Data Form, we also asked the clinicians to indicate whether they were treating the patient with psychotherapy and/or with any of several classes of medication. The clinicians rated the effectiveness of each treatment using a 5-point scale.

Axis II checklist. To maximize the accuracy of the clinicians' DSM-IV personality disorder diagnoses, we presented the clinicians with a randomly ordered checklist of the criteria for all axis II disorders. In prior studies, this method has produced results that mirror findings based on structured interviews (22, 23). To generate categorical diagnoses, we applied DSM-IV decision rules. To generate DSM-IV dimensional diagnoses, we summed the number of criteria met per disorder.

Table 1. Domains and Selected Items of the Clinician-Report Clinical Data Form

Domain	Selected Items
Clinician demographics	Discipline (psychiatry or psychology), theoretical orientation, employment sites (e.g., private practice, inpatient unit, school), sex
Patient demographics	Age, sex, ethnicity, marital status, education level, socioeconomic status
Adaptive functioning	Global Assessment of Functioning Scale score; rating of level of personality disturbance, based loosely on Kernberg's concept of personality organization; ratings of employment history (1 = unable to keep a job, 3 = unstable, 5 = stable, 7 = working to potential); quality of romantic relationships and quality of friendships (1 = very poor, 7 = close and loving); history of suicide attempts, hospitalizations, arrests, and recent job losses for interpersonal reasons (coded 0/1 for absent/present); and social support (number of close friends in whom the patient feels comfortable confiding)
Developmental history	Parental divorce, adoption, foster care, lengthy separations from primary attachment figure, residential placement (coded 0/1 for absent/present); 7-point ratings of family stability and warmth; physical and sexual abuse (coded 0/1/2 for absent, unsure, present)
Family history	Ratings of first- and second-degree biological relatives for psychosis, bipolar disorder, major depression, anxiety disorder, alcohol abuse, prescription drug abuse, illicit substance abuse, criminality, suicide attempts, and completed suicide (coded 0/1/2 for absent, unsure, present)

Personality disorder construct ratings. As part of the broader project, we designed a measure to allow clinicians to rate the extent to which the patient resembled each DSM-IV personality disorder *construct*, irrespective of specific diagnostic criteria. The clinicians rated each DSM-IV personality disorder using the same 5-point rating system that is depicted in Figure 1. However, for each diagnosis, we reproduced only the single-sentence summary that introduces the disorder in the text of DSM-IV (e.g., "The essential feature of Borderline Personality Disorder is a pervasive pattern of instability of interpersonal relationships, self-image, and affects, and marked impulsivity that begins by early adulthood and is present in a variety of contexts."). For the present study, we used these ratings to address the rival hypothesis that our two methods of prototype diagnosis might outperform DSM diagnosis because they contain more information.

Diagnostic prototype questionnaire. Approximately one-half of the participating clinicians rated their patients using the cluster B *clinician prototypes*, and the remaining clinicians rated their patients using the *empirical prototypes*. To construct paragraph-long clinician prototypes for the present study, we selected the Q-sort items with the highest average ranking for each disorder from our prior study (7) and wove them into paragraph form. For the empirical prototypes, we similarly selected the highest-ranked items for the four diagnostic groupings identified empirically by means of Q-factor analysis that resembled the axis II cluster B disorders (which were replicated in both adolescent and adult samples) (7, 22): antisocial-psychopathic, emotionally dysregulated (borderline), histrionic, and narcissistic.

PROCEDURE

After completing the Clinical Data Form, the axis II checklist, the personality disorder construct ratings, and other measures, the clinicians read a brief (three-paragraph) overview of the prototype matching system and rated their patients using either the cluster B *clinician prototypes* (presented to 147 clinicians) or the cluster B *empirical prototypes* (presented to 144 clinicians). The clinicians diagnosed the patient on all four disorders using the rating scale depicted in Figure 1. The clinicians then compared the prototype matching system to the standard DSM-IV procedure (the count/cutoff approach) using 5-point ratings on four clinical criteria: ease of use, usefulness for communication with other clinicians, ability to capture important information about the patient, and general clinical utility. The ratings were anchored in relation to the current DSM-IV diagnostic procedure (1 = much worse, 3 = about the same, 5 = much better).

DATA ANALYSIS

We first examined *internal criteria*, focusing on whether the four diagnostic methods (DSM-IV categorical diagnosis, DSM-IV dimensional diagnosis, clinician prototypes, and empirical prototypes) differed in identifying comorbidity among the cluster B disorders. Next we compared these methods on *external criteria*, assessing the correlation between each diagnosis in each of the four systems and the following three sets of variables selected a priori: adaptive functioning, treatment response, and etiology. To see whether axis II

Table 2. Characteristics of a Random National Sample of Clinicians (N=291) Who Provided Descriptions of a Patient With Personality Problems and of the Patients Described by the Clinicians

Characteristic	Value	
	N	%
Clinician characteristics		
Discipline		
Psychiatry	88	30.2
Psychology	203	69.8
Sex		
Female	137	47.0
Male	154	53.0
Theoretical orientation		
Eclectic	151	51.8
Psychodynamic	82	28.1
Cognitive behavioral	49	17.0
Biological	9	3.1
	Mean	SD
Years of experience	20.0	9.5
Patient characteristics		
Age (years)	42.9	12.4
	N	%
Sex		
Female	184	63.2
Male	107	36.8
Ethnicity/race		
Caucasian	257	88.4
African American	15	5.1
Hispanic	9	3.1
Other (e.g., Asian)	9	3.1
Socioeconomic status		
Poor	21	7.2
Working class	71	24.3
Middle class	110	38.0
Upper/upper middle class	89	30.5
Marital status		
Married or cohabiting	120	41.4
Single or divorced	171	58.6
Primary Axis I diagnosis		
Major depression	101	34.7
Dysthymic disorder	136	46.7
Generalized anxiety disorder or anxiety disorder not otherwise specified	110	37.8
Adjustment disorder	56	19.2
Substance use disorder	47	16.2
	Mean	SD
Global Assessment of Functioning Scale score	58.4	9.6
Treatment characteristics Length (months) ^a	16.6	16.5
	N	%
Clinical setting		
Private practice	228	78.4
Clinic	48	16.4
Inpatient/residential setting	5	1.7
Forensic setting	3	1.0
Other	7	2.5

^a Median = 15.0 months.

diagnoses derived using each system showed incremental validity in predicting criterion variables over and above axis I diagnosis, we used hierarchical linear regression to predict a composite measure of adaptive functioning, entering axis I diagnosis in step 1 and each set of four axis II diagnoses in step 2. To see whether a personality health prototype would be a useful addition to axis II, we included a personality health prototype in step 3 of each regression analysis. The personality health prototype is a measure of personality strengths and adaptive resources, which we have proposed for inclusion in DSM-V (19). Finally, we compared prototype diagnosis with DSM-IV diagnosis on *clinical criteria*, using the clinicians' ratings of variables such as ease of use and general clinical utility.

RESULTS

Table 2 summarizes the clinician and patient characteristics. The clinicians were evenly split among men and women, reported a range of theoretical orientations, and were highly experienced (mean = 20 years of experience). The patients resembled those seen in the community; approximately two-thirds were female, 88.4% were Caucasian, and the majority had an axis I mood disorder.

INTERNAL CRITERIA

As Table 3 shows, rates of comorbidity assessed categorically by using DSM-IV criteria were high and were comparable to rates reported in studies using structured interviews. The median rate of comorbidity among the cluster B disorders was 44.7%. For patients who received a cluster B diagnosis using DSM-IV criteria (50.2% of the patients), the average number of cluster B diagnoses was 1.7 (SD = 0.98). The two prototype approaches assigned fewer cluster B diagnoses overall (35.9% and 32.7% of patients, respectively, using the cutoff for clinicians' ratings of ≥ 4), and the average patient who received a personality disorder diagnosis received fewer comorbid diagnoses (for the clinician prototypes: mean = 1.31, SD = 0.67; for the empirical prototypes: mean = 1.21, SD = 0.41). The number of cluster B diagnoses assigned by using DSM-IV criteria was significantly greater than the number assigned by using the two prototype approaches (clinician prototypes: $t = 5.23$, $df = 143$, $p < 0.001$; empirical prototypes: $t = 5.84$, $df = 146$, $p < 0.001$).

The correlations between dimensional DSM-IV diagnoses made by using the number of symptoms met for each disorder were also high (median $r = 0.47$). The two prototype-matching systems

Table 3. Comorbidity of DSM-IV Categorical Personality Disorder Diagnoses (N=290) and Correlation of DSM-IV Dimensional Personality Disorder Diagnoses (N=290) With Clinician Prototypes and Empirical Prototypes (N=143–147)

Variable and Diagnosis	Antisocial Personality Disorder ^a	Borderline Personality Disorder ^a	Histrionic Personality Disorder ^a	Narcissistic Personality Disorder ^a
	% Comorbid	% Comorbid	% Comorbid	% Comorbid
Comorbidity with DSM-IV categorical diagnoses				
Antisocial personality disorder	—	66.0	44.7	51.1
Borderline personality disorder	30.7	—	28.7	29.7
Histrionic personality disorder	50.0	59.0	—	57.1
Narcissistic personality disorder	40.7	50.8	40.7	—
	r	r	r	r
Correlation				
DSM-IV dimensional diagnoses				
Antisocial personality disorder	—	0.46***	0.53***	0.46***
Borderline personality disorder	—	—	0.52***	0.32***
Histrionic personality disorder	—	—	—	0.53***
Narcissistic personality disorder	—	—	—	—
Clinician prototypes				
Antisocial personality disorder	—	0.18*	0.31***	0.55***
Borderline personality disorder	—	—	0.52***	0.05
Histrionic personality disorder	—	—	—	0.24**
Narcissistic personality disorder	—	—	—	—
Empirical prototypes				
Antisocial personality disorder	—	0.18*	0.09	0.38***
Borderline personality disorder	—	—	0.34***	0.11
Histrionic personality disorder	—	—	—	0.15
Narcissistic personality disorder	—	—	—	—

^aFor analysis of comorbidity, DSM-IV categorical diagnoses were used. For correlations, separate analyses were conducted for DSM-IV dimensional diagnoses, clinician prototypes, and empirical prototypes. All correlations were conducted with Pearson's product-moment correlation.

* $p \leq 0.05$ (two-tailed). ** $p \leq 0.01$ (two-tailed). *** $p \leq 0.001$ (two-tailed).

fared better. As Table 3 shows, for the clinician prototypes, the median correlation between disorders was 0.28; for the empirical prototypes, the median correlation was 0.17. To make a rough estimate of the significance of these differences, we compared the median intercorrelations for DSM-IV dimensional diagnoses with the median intercorrelations for each prototype approach using Fisher's z . The differences were significant or near-significant even in two-tailed analyses (clinician prototypes: $z = 1.87$, $p = 0.06$; empirical prototypes: $z = 2.85$, $p = 0.004$).

In light of the reduced comorbidity with the prototype approaches, we correlated prototype diagnoses with DSM-IV dimensional diagnoses (number of symptoms met) to determine if the prototype approaches were in fact diagnosing constructs similar to the constructs assessed using DSM-IV criteria. The coefficients in boldface type in Table 4 reflect convergence across dimensional diagnostic methods. Both prototype methods clearly converged with DSM-IV

dimensional diagnosis, although the empirical prototypes showed slightly greater convergence (median: $r = 0.76$) and discriminant validity (median coefficient off the diagonal: $r = 0.29$, lower than the median correlation of the DSM dimensional diagnoses with each other). Thus, the prototypes provided a reasonable proxy for DSM-IV dimensional diagnoses as widely operationalized (number of symptoms met) but did so with less diagnosis of comorbidity.

EXTERNAL CRITERIA

Although prototype diagnosis appears advantageous in minimizing comorbidity, an important question is whether using prototype diagnosis leads to offsetting losses in validity (predicting *external criteria*). Thus, we examined the correlations of the disorders as diagnosed by using the four methods with ratings of adaptive functioning, treatment response, and developmental and family history variables.

Table 4. Correlation of Clinician Prototypes and Empirical Prototypes With DSM-IV Dimensional Personality Disorder Diagnoses (N=143–147)^a

Diagnostic Approach and Diagnosis	DSM-IV Dimensional Diagnosis			
	Antisocial Personality Disorder	Borderline Personality Disorder	Histrionic Personality Disorder	Narcissistic Personality Disorder
Clinician prototypes				
Antisocial personality disorder	0.54***	0.17*	0.32***	0.44***
Borderline personality disorder	0.38***	0.76***	0.44***	0.18*
Histrionic personality disorder	0.25**	0.40***	0.49***	0.17*
Narcissistic personality disorder	0.25**	0.05	0.27***	0.72***
Empirical prototypes				
Antisocial personality disorder	0.79***	0.29***	0.27***	0.48***
Borderline personality disorder	0.30***	0.77***	0.33***	0.20*
Histrionic personality disorder	0.24**	0.31***	0.53***	0.19*
Narcissistic personality disorder	0.35***	0.16	0.34***	0.74***

^a Coefficients in boldface type reflect convergence across diagnostic methods.

* $p \leq 0.05$. ** $p \leq 0.01$. *** $p \leq 0.001$.

Adaptive functioning. We first examined adaptive functioning, including an aggregated measure of global functioning (obtained by standardizing and summing the following five ratings selected a priori: Global Assessment of Functioning [GAF], severity of personality dysfunction, quality of romantic relationships, quality of friendships, and occupational functioning) and three relatively non-inferential measures (history of suicide attempts, psychiatric hospitalizations, and arrests). Table 5 reports the partial correlations between each personality disorder diagnosis (with adjustment for the other three diagnoses within each set) and measures of adaptive functioning, with coefficients in boldface type indicating primary hypothesized relationships. (We covaried for other cluster B diagnoses to provide a more accurate portrait of associations with particular personality disorders, although the raw correlations produced generally similar patterns.) The correlations were similar across the four approaches, although they were somewhat larger where predicted for the empirical prototypes.

If a personality axis is to be useful, it must predict variance in adaptive functioning beyond axis I diagnosis. Thus, in a second set of analyses, we used hierarchical linear regression to determine whether 1) any of the four diagnostic methods predicted variance in global adaptive functioning after adjustment for the most prevalent axis I diagnoses in the sample (prevalence $\geq 10\%$); 2) the four systems differed in the amount of variance accounted for, and 3) addition of a personality health prototype (i.e., a measure of personality strengths and adaptive resources) to axis II accounted for additional variance after holding constant both axis I and axis II diagnoses.

We performed four regression analyses (one for each diagnostic method) using our aggregated measure of global functioning as the criterion variable. (Data using GAF scores alone produced similar findings.) In step 1 we entered axis I diagnoses; in step 2, axis II (cluster B) diagnoses; and in step 3, personality health prototype ratings. As Table 6 shows, axis I diagnoses routinely accounted for about 10% of the variance in adaptive functioning, which is substantial. However, in all four analyses, adding the four cluster B diagnoses in step 2 yielded a significant improvement in the model, with multiple Rs increasing incrementally from categorical DSM diagnosis to dimensional DSM diagnosis to clinician prototypes to empirical prototypes. Adding the personality health prototype in step 3 led to large and statistically significant increments in prediction in all four analyses.

Treatment response. Because 95.5% of patients in the study received psychotherapy and 67.7% were treated with antidepressant medication (87.2% of those who received antidepressants received selective serotonin reuptake inhibitors), we were able to conduct analyses using treatment response as a criterion variable. We consider these analyses to be preliminary, both because of the preliminary nature of the measures and because of the dearth of prior research to inform our hypotheses (that borderline and antisocial features would negatively predict outcome for both psychotherapy and medication). Nevertheless, treatment response is a key variable in validating a diagnostic system (2), and diagnoses, to be clinically useful, should inform treatment decisions. Once again we report partial correlations, with adjustment for other three diagnoses in each set.

Table 5. Correlation Between Personality Disorder Diagnoses Made With Four Diagnostic Methods and Measures of Adaptive Functioning and Treatment Response^a

Diagnostic Approach and Diagnosis	Adaptive Functioning			Treatment Response		
	Global Functioning	Suicide	Psychiatric Hospitalization	Arrest	Psychotherapy	Antidepressants
DSM-IV categorical diagnosis						
Antisocial personality disorder	-0.25***	-0.03	0.14*	0.49***	-0.13*	-0.08
Borderline personality disorder	-0.23***	0.37***	0.27***	-0.02	-0.07	-0.13
Histrionic personality disorder	0.09	-0.04	-0.04	-0.07	-0.03	0.13
Narcissistic personality disorder	0.01	-0.00	-0.08	-0.09	-0.12	-0.05
DSM-IV dimensional diagnosis						
Antisocial personality disorder	-0.20**	-0.03	0.01	0.28***	-0.12*	-0.10
Borderline personality disorder	-0.23**	0.43***	0.35***	-0.02	-0.06	-0.16*
Histrionic personality disorder	0.08	-0.11	-0.00	0.04	0.07	0.19**
Narcissistic personality disorder	0.04	-0.00	-0.10	-0.09	-0.12	-0.05
Clinician prototypes						
Antisocial personality disorder	-0.22**	0.06	0.21*	0.34***	-0.14†	-0.18†
Borderline personality disorder	-0.41***	0.35***	0.27***	0.09	-0.12	-0.11
Histrionic personality disorder	0.13	-0.15†	-0.05	0.02	0.10	0.04
Narcissistic personality disorder	0.02	-0.13	-0.27***	-0.06	-0.14†	-0.02
Empirical prototypes						
Antisocial personality disorder	-0.29***	0.01	0.21*	0.42***	-0.03	0.02
Borderline personality disorder	-0.44***	0.62***	0.47***	0.02	-0.22**	-0.29**
Histrionic personality disorder	0.10	-0.06	-0.10	0.06	0.09	0.09
Narcissistic personality disorder	0.12	-0.10	-0.25**	-0.10	-0.05	0.15

^a For DSM-IV categorical and dimensional diagnoses, N=290 except for psychotherapy effectiveness (N=270) and antidepressant effectiveness (N=197). For clinician prototypes and for empirical prototypes, N=143–144, except for treatment response for psychotherapy (N=133–137) and for antidepressants (N=90). Coefficients in bold represent primary hypothesized relationships.

*p≤0.05. **p≤0.01. ***p≤0.001. †p≤0.05 (one-tailed, where predicted for analyses with smaller Ns).

As Table 5 shows, antisocial personality disorder, borderline personality disorder, or both were negatively correlated with response to psychotherapy across all four diagnostic approaches. (We did not address differences among therapeutic orientations, given the limited numbers of patients treated by therapists with each orientation.) Once again, the four approaches produced similar coefficients, although DSM categorical diagnoses tended to be least predictive of response to both psychotherapy and pharmacotherapy (which is not surprising, given the psychometric disadvantages of dichotomous variables), whereas the borderline personality disorder empirical prototype had the largest (negative) correlations with both psychotherapy and pharmacotherapy response.

Etiology. We next compared the four diagnostic methods on associations with the following variables shown in prior research to be relevant to the etiology of cluster B disorders, particularly antisocial personality disorder and borderline personality

disorder: physical abuse, sexual abuse, and family history of internalizing disorders (mood and anxiety), externalizing disorders (criminality, alcohol abuse, and illicit drug abuse), and suicide. (Little information in this area is available for narcissistic personality disorder and histrionic personality disorder.) Table 7 reports partial correlations between each diagnosis and these etiological variables, with predicted correlations presented in boldface type. Once again the results were similar across diagnostic approaches.

Ruling out a rival hypothesis. The data suggest that prototype diagnosis in everyday practice minimizes findings of comorbidity with no offsetting cost in validity. One might argue, however, that the prototype approaches tested here have the advantage of richer item sets (i.e., more information than the eight or nine criteria per disorder in DSM-IV). The ability to include 18 to 20 criteria per disorder is in fact an advantage of prototype diagnosis, because inclusion of that many criteria per disorder would render the count/cutoff approach unusable, as

Table 6. Hierarchical Linear Regression Analyses of Predictors of Adaptive Functioning From Four Methods for Diagnosis of Personality Disorders (N=142–146)^a

Variable	DSM-IV Categorical Diagnosis				DSM-IV Dimensional Diagnosis				Clinician Prototypes				Empirical Prototypes			
	R	R ²	F	p	R	R ²	F	p	R	R ²	F	p	R	R ²	F	p
Regression models																
Step 1: axis I	0.30	0.09	2.21	<0.05	0.30	0.09	2.21	<0.05	0.30	0.09	2.23	<0.05	0.36	0.13	3.50	0.003
Step 2: axis II	0.46	0.21	4.99	0.001	0.52	0.27	8.11	<0.0001	0.54	0.29	9.49	<0.0001	0.57	0.32	9.74	0.000
Step 3: personality health prototype	0.59	0.35	27.75	<0.0001	0.61	0.37	22.01	<0.0001	0.65	0.43	26.94	<0.0001	0.71	0.50	46.78	<0.0001
			β	p			β	p			β	p			β	p
Final step predictors																
Major depressive disorder			-0.19	0.01			-0.17	<0.03			-0.19	0.008			0.05	0.67
Dysthymic disorder			-0.03	0.72			-0.03	0.71			-0.07	0.31			0.13	0.06
Generalized anxiety disorder			-0.03	0.70			-0.05	0.54			-0.05	0.51			-0.12	0.06
Anxiety disorder not otherwise specified			0.03	0.72			0.02	0.77			0.00	0.98			0.06	0.41
Substance use disorder			-0.02	0.81			0.00	0.98			0.01	0.93			-0.01	0.82
Adjustment disorder			0.06	0.40			0.04	0.58			0.00	0.99			-0.01	0.93
Antisocial personality disorder			-0.19	0.02			-0.20	<0.04			-0.23	0.007			-0.12	0.10
Borderline personality disorder			-0.08	0.35			-0.23	<0.02			-0.32	<0.0001			-0.34	<0.0001
Histrionic personality disorder			0.08	0.35			0.19	0.08			0.10	0.25			0.09	0.20
Narcissistic personality disorder			-0.14	<0.09			-0.14	0.12			-0.02	0.83			0.15	<0.04
Personality health prototype			0.41	<0.0001			0.36	<0.0001			0.37	<0.0001			0.47	<0.0001

^aSimultaneous multiple regression, in which standardized betas reflect the residual contribution of each predictor with the other predictors held constant, was used.

determination of presence/absence would be required for each of 200 criteria across disorders. Nevertheless, we tested this rival hypothesis by examining clinicians' *personality disorder construct ratings*—5-point prototype ratings of *single-sentence* summaries of each cluster B disorder from DSM-IV that convey less information than the diagnostic criteria for each disorder. The data were strikingly similar to those we obtained with the clinical and empirical prototypes: the rate of comorbidity was substantially lower than with the DSM-IV dimensional diagnoses (median $r=0.24$), and the pattern of external correlates was equivalent.

CLINICAL CRITERIA

Next we compared the two prototype systems with the count/cutoff approach on clinical criteria, using ratings of ease of use, usefulness for clinical

communication, ability to capture important information about the patient's personality, and clinical utility. The results were virtually identical for the two prototype systems (Figure 2 and Figure 3). The clinicians strongly preferred prototype diagnosis to the count/cutoff method on every dimension assessed, with roughly 70% of the clinicians rating the prototype systems as better or much better than the DSM-IV approach, 10% preferring the more familiar DSM system, and the remaining 20% rating the two diagnostic approaches as equivalent.

DISCUSSION

As in research using structured interviews, categorical axis II diagnosis in clinical practice produces substantial diagnostic overlap and generally shows similar or lower correlations with relevant criterion

Table 7. Correlation Between Personality Disorder Diagnoses Made With Four Diagnostic Methods and Measures of Developmental History and Family History in First- and Second-Degree Relatives^a

Diagnostic Approach and Diagnosis	Developmental History		Family History		
	Physical Abuse	Sexual Abuse	Externalizing Disorders	Internalizing Disorders	Suicide
DSM-IV categorical diagnosis					
Antisocial personality disorder	0.02	0.05	0.12*	−0.06	0.10
Borderline personality disorder	0.12*	0.21***	0.18**	0.15**	0.11
Histrionic personality disorder	0.08	0.08	0.01	−0.00	0.02
Narcissistic personality disorder	−0.08	−0.08	−0.06	0.04	−0.07
DSM-IV dimensional diagnosis					
Antisocial personality disorder	0.08	0.05	0.20***	−0.04	0.13*
Borderline personality disorder	0.17**	0.32***	0.15*	0.12*	0.07
Histrionic personality disorder	0.02	0.01	0.00	0.07	0.02
Narcissistic personality disorder	−0.12*	−0.17**	−0.11	−0.03	−0.05
Clinician prototypes					
Antisocial personality disorder	0.01	−0.01	0.20*	−0.14	0.02
Borderline personality disorder	0.15†	0.36***	0.15†	0.24**	0.03
Histrionic personality disorder	0.02	−0.12	−0.03	−0.00	0.09
Narcissistic personality disorder	−0.03	−0.12	−0.17*	0.01	−0.09
Empirical prototypes					
Antisocial personality disorder	0.24**	0.10	0.16†	0.02	0.22**
Borderline personality disorder	0.10	0.40***	0.24**	0.05	0.17*
Histrionic personality disorder	−0.06	−0.04	−0.05	0.07	−0.02
Narcissistic personality disorder	−0.13	−0.02	0.01	0.03	0.09

^aFor DSM-IV categorical and dimensional diagnoses, N=290. For clinician prototypes and empirical prototypes, N=143 or 144. Coefficients in bold represent primary hypothesized relationships.

*p≤0.05. **p≤0.01. ***p≤0.001. †p≤0.05 (one-tailed, where predicted for analyses with smaller Ns).

variables, compared with dimensional diagnosis, operationalized in multiple different ways. Given this consistent finding in the literature and widespread evidence of subthreshold personality pathology that is not diagnosable by using axis II (24), it is difficult to argue that DSM-V should retain a primarily categorical approach to diagnosing personality pathology.

Prototype diagnosis reduced findings of comorbidity without decrements in validity. All four diagnostic approaches yielded similar estimates of validity. However, where the empirically derived prototype diagnoses differed in their external correlates from the DSM-IV diagnoses (both categorical and dimensional), they tended to be slightly superior in predicting clinically meaningful variables such as adaptive functioning.

Clinicians rated a prototype matching diagnostic method, even one using unfamiliar (empirically derived) personality descriptions, as easier to implement and more clinically meaningful than the count/cutoff approach. Spitzer, First, and Skodol

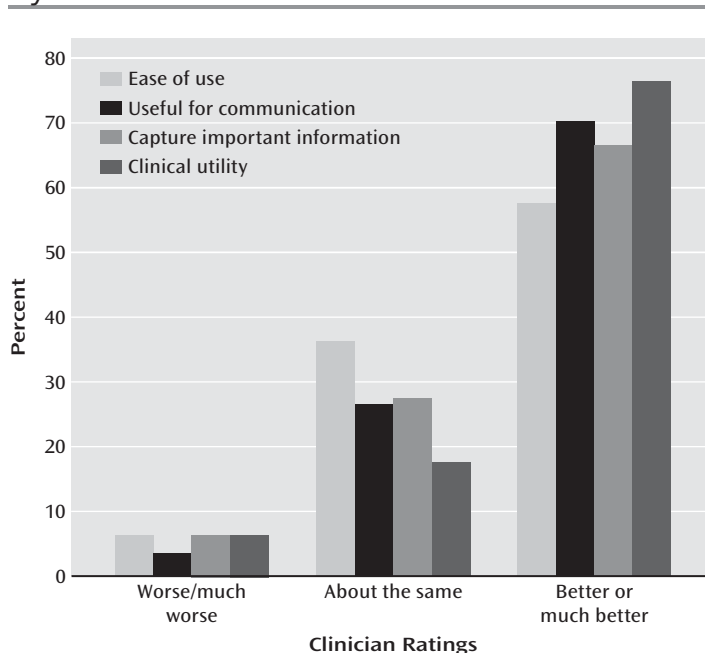
(unpublished data) have similarly found that experienced psychiatrists and psychologists rate prototype approaches as more clinically useful than both the current DSM approach and alternative dimensional (trait) models.

The data also supported inclusion of a personality health prototype in DSM-V. Such an index is useful in calling attention to patients' strengths and in gauging progress over time in treatment. In this study, the personality health prototype accounted for substantial variance in adaptive functioning even after accounting for axis I and axis II diagnoses.

LIMITATIONS

This study had some limitations that should be considered in interpreting the data. First, for the analyses assessing external criteria, the clinicians provided both the diagnostic data and the data on adaptive functioning, etiology, and treatment response. Thus, we cannot be certain that their

Figure 2. Clinicians Ratings on Clinical Criteria of the Clinician Prototype Matching System, Compared With the DSM-IV Diagnostic System



diagnostic judgments were independent of these external criteria. However, if the clinicians' biases influenced their ratings of criterion variables, this factor would favor the most familiar diagnostic methods, namely those prescribed in DSM-IV. The fact that the least familiar diagnoses (i.e., the empirical prototypes) tended to yield the strongest results is inconsistent with the bias hypothesis. Further, a growing amount of research has suggested that clinicians can in fact make highly reliable and valid judgments if their observations are quantified and standardized (25). In data recently collected by our group, the average correlation between prototype ratings made by two clinicians (advanced graduate students) listening to the same data (initial psychotherapy hours) was 0.70. These data suggest that even relatively inexperienced clinicians can make prototype diagnoses reliably. Clearly, however, the data suggest two next steps. The first is to replicate the findings on external criteria by using a design in which diagnosticians are unaware of all other data. The second is to see whether prototype ratings are more or less useful clinically in guiding clinicians' thinking and interventions (e.g., whether patients of clinicians instructed to make prototype ratings at the beginning of treatment and at various milestones throughout the treatment fare better or worse than those instructed to make repeated DSM-IV diagnoses).

A second limitation is that we examined only the cluster B disorders and hence do not know to what extent similar findings would generalize to the other personality disorders. A third limitation is that we tested only two variations of the prototype matching method. We did not, for example, compare DSM-IV diagnosis with prototype descriptions comprising only the seven to nine criteria used in DSM-IV (because of the difficulty of weaving such a small number of criteria into coherent prototype descriptions). Future research should vary the number of criteria embedded in prototypes to optimize reliability, validity, and parsimony.

IMPLICATIONS

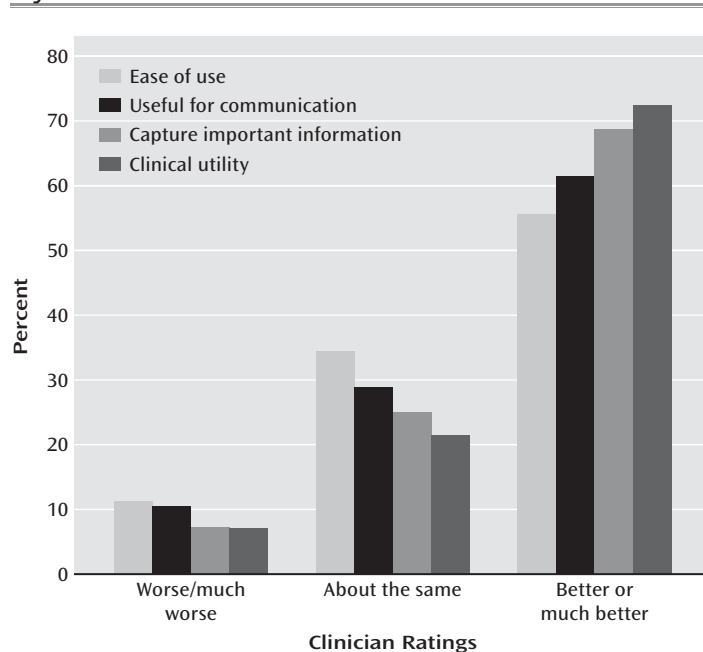
The count/cutoff approach was a tremendous improvement over DSM-II diagnosis. However, it had never previously been subjected to systematic testing against any other way of operationalizing diagnosis, particularly in clinical practice. Prototype diagnosis could be implemented with relatively minor taxonomic changes (by simply refining the prototypes tested here to match more closely the characteristics of patients with DSM-IV-defined personality disorders [7, 8]) or with more substantial changes (by deriving nonredundant diagnostic prototypes empirically). In either case, clinicians could make a complete axis II diagnosis in 1 or 2 minutes, generating a diagnostic profile (similar to an MMPI profile) that indicates, for each disorder, both the extent to which the patient resembles the prototype and whether the patient matches the prototype strongly enough to receive a categorical diagnosis useful for communication with other professionals. Prototype diagnosis has the parsimony of DSM-II diagnosis but lacks its disadvantages. For example, prototypes can be derived empirically, and, as noted earlier, they can be rated reliably.

A question we did not address here is whether prototype diagnosis is suitable only for clinical practice (similar to ICD-10, which has different diagnostic procedures for research and practice). Although single-item diagnostic ratings may not provide data that are reliable enough for research purposes (although see reference 26), one way of augmenting the method tested here is to obtain secondary ratings for patients who receive a score >1 for a given disorder. For example, for borderline personality disorder, this augmentation might entail 5-point ratings of subdimensions or endophenotypes (e.g. emotional dysregulation, impulsivity, and attachment dysregulation) that are generated by factor analysis. Such ratings could be aggregated along with the prototype ratings to maximize

reliability or could be used as indicators of the latent construct in structural models. Alternatively, for research as well as clinical purposes, the prototypes could be rated along with a set of functional domains, such as motives and conflicts, cognition, emotional experience, emotion regulation, impulse regulation, relational functioning and representations, identity and self-experience, and adaptive strengths (27, 28). As we have shown elsewhere (29, 30), researchers can obtain high interrater reliability and validity for prototype diagnosis of personality disorder by applying the SWAP-200 Q-sort to data from a systematic clinical interview and correlating patients' profiles on the instrument with empirical prototypes; this procedure yields both a dimensional diagnosis and a functional assessment. We suspect that researchers could also derive prototype diagnoses from current axis II instruments, much as anxiety disorders researchers have derived dimensional ratings along with categorical diagnoses from structured interviews (31). An additional question is whether prototype diagnosis might be equally useful for axis I (13). We recently obtained similar findings with mood, anxiety, and eating disorders, but further research is clearly necessary.

Finally, these data raise the question of whether researchers may have too hastily invoked clinician error in explaining why clinicians tend not to use DSM-IV decision rules in assessing personality. Like all information processors, clinicians tend to elicit and organize the information they need to solve problems. Research in cognitive science has suggested that people tend to *satisfice* (a cross between *satisfy* and *suffice*), that is, to make a "good-enough" assessment for their purposes, and to make more precise determinations based on explicit decision rules if the need arises (32). Other research on categorization has suggested that the way people classify objects in a given domain reflects their goals (33), and such goals are overlapping but not identical in research and practice. In light of the dearth of research showing any treatment implications of clinical versus subthreshold symptoms (e.g., whether the patient meets four versus five criteria of a given personality disorder), making a "good-enough" assessment—particularly one that also captures subthreshold pathology not normally diagnosable using DSM-IV—may actually be a reasonable strategy. We suspect that clinicians already rely heavily on prototype matching in everyday diagnosis (16, 34, 35). Formalizing the prototypes clinicians use and selecting the attributes embedded in these prototypes empirically represent a way of minimizing idiosyncratic elements of diagnosis in clinical practice.

Figure 3. Clinicians Ratings on Clinical Criteria of the Empirical Prototype Matching System, Compared With the DSM-IV Diagnostic System



REFERENCES

1. Sokal RR: Classification: purposes, principles, progress, prospects. *Science* 1974; 185:1115–1123
2. Robins E, Guze S: The establishment of diagnostic validity in psychiatric illness: its application to schizophrenia. *Am J Psychiatry* 1970; 126:983–987
3. First MB, Pincus HA, Levine JB, Williams JBW, Ustun B, Peele R: Clinical utility as a criterion for revising psychiatric diagnoses. *Am J Psychiatry* 2004; 161:946–954
4. Feighner JP, Robins E, Guze SB, Woodruff RA, Winokur G, Munoz R: Diagnostic criteria for use in psychiatric research. *Arch Gen Psychiatry* 1972; 26:57–63
5. Spitzer RL, Endicott J, Robins E: Clinical criteria for psychiatric diagnosis and DSM-III. *Am J Psychiatry* 1975; 132:1187–1192
6. Widiger TA, Clark LA: Toward DSM-V and the classification of psychopathology. *Psychol Bull* 2000; 126:946–963
7. Westen D, Shedler J: Revising and assessing axis II, part 1: developing a clinically and empirically valid assessment method. *Am J Psychiatry* 1999; 156:258–272
8. Shedler J, Westen D: Refining personality disorder diagnoses: integrating science and practice. *Am J Psychiatry* 2004; 161:1350–1365
9. Ahn W, Kim NS: The causal status effect in categorization: an overview, in *The Psychology of Learning and Motivation: Advances in Research and Theory*, vol. 40. Edited by Medin DL. San Diego, Academic Press, 2001, pp 23–65
10. Jampala V, Sierles F, Taylor M: The use of DSM-III-R in the United States: a case of not going by the book. *Compr Psychiatry* 1988; 29:39–47
11. Westen D: Divergences between clinical and research methods for assessing personality disorders: implications for research and the evolution of axis II. *Am J Psychiatry* 1997; 154:895–903
12. Westen D, Shedler J: A prototype matching approach to diagnosing personality disorders: toward DSM-V. *J Personal Disord* 2000; 14:109–126
13. Westen D, Heim AK, Morrison K, Patterson M, Campbell L: Classifying and diagnosing psychopathology: a prototype matching approach, in *Rethinking the DSM: Psychological Perspectives*. Edited by Beutler L, Malik M. Washington, DC, American Psychological Association, 2002, pp 221–250
14. Westen D, Bradley R: Prototype diagnosis of personality, in *Handbook of Personality and Psychopathology*. Edited by Strack S. New York, John Wiley & Sons, 2005, pp 238–256
15. Rosch E, Mervis CB: Family resemblances: studies in the internal structure of categories. *Cognit Psychol* 1975; 7:573–605

- ## NOTES

This image shows a blank sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.