Validity of Prototype Diagnosis for Mood and Anxiety Disorders

Jared A. DeFife, PhD; Joanne Peart, PhD; Bekh Bradley, PhD;
Kerry Ressler, MD, PhD; Rebecca Drill, PhD; Drew Westen, PhD

Context: With growing recognition that most forms of psychopathology are best represented as dimensions or spectra, a central question becomes how to implement dimensional diagnosis in a way that is empirically sound and clinically useful. Prototype matching, which involves comparing a patient's clinical presentation with a prototypical description of the disorder, is an approach to diagnosis that has gained increasing attention with forthcoming revisions to both the *DSM* and the *International Classification of Diseases.*

Objective: To examine prototype diagnosis for mood and anxiety disorders.

Design, Setting, and Patients: In the first study, we examined clinicians' *DSM-IV* and prototype diagnoses with their ratings of the patients' adaptive functioning and patients' self-reported symptoms. In the second study, independent interviewers made prototype diagnoses following either a systematic clinical interview or a structured diagnostic interview. A third interviewer provided independent ratings of global adaptive functioning. Patients were recruited as outpatients (study 1; N=84) and from primary care clinics (study 2; N=143).

Main Outcome Measures: Patients' self-reported mood, anxiety, and externalizing symptoms along with independent clinical ratings of adaptive functioning.

Results: Clinicians' prototype diagnoses showed small to moderate correlations with patient-reported psychopathology and performed as well as or better than *DSM-IV* diagnoses. Prototype diagnoses from independent interviewers correlated on average $r=.50$ and showed substantial incremental validity over *DSM-IV* diagnoses in predicting adaptive functioning.

Conclusions: Prototype matching is a viable alternative for psychiatric diagnosis. As in research on personality disorders, mood and anxiety disorder prototypes outperformed *DSM-IV* decision rules in predicting psychopathology and global functioning. Prototype matching has multiple advantages, including ease of use in clinical practice, reduced artifactual comorbidity, compatibility with naturally occurring cognitive processes in diagnosticians, and ready translation into both categorical and dimensional diagnosis.

RESEARCHERS HAVE SPENT more than 3 decades trying to improve the classification of psychiatric disorders, leading to substantial empirical advances. Strikingly absent, however, has been a similar body of research on how to maximize diagnostic reliability, validity, and clinical utility in every-

Author Affiliations:
Departments of Psychology
(Drs DeFife and Westen) and
Psychiatry (Drs Bradley, Ressler,
and Westen), Emory University,
Atlanta, Georgia; New York
Presbyterian Hospital, New
York (Dr Peart); and
Department of Psychiatry,
Cambridge Health Alliance and
Harvard Medical School,
Cambridge, Massachusetts
(Dr Drill). Dr Peart is now with
Greystone Park Psychiatric
Hospital, Morris Plains,
New Jersey.

CME available online at www.jamanetworkcme.com and questions on page 135

day practice. Although the procedures designed for diagnosing patients since *DSM-III* have produced greatly enhanced reliability in research protocols, they have not done so in clinical practice, where cumbersome lists of symptoms with complex coding algorithms that vary by disorder have proven largely untenable.[1-4]

Perhaps not surprisingly, a manual that was not designed to take into account the cognitive processing parameters of the diagnostician[2,5-7] is frequently not used, or not used as intended, by clinicians. Clinicians instead tend to diagnose in everyday practice by pattern matching rather than counting criteria and applying cut-

offs for categorical diagnoses.[5,6,8] When clinicians do apply the complex procedures required for diagnosis in *DSM-IV*, their interrater reliability tends to be disappointing.[9,10] Even for research purposes, those decision rules tend to relegate much of psychopathology to "not otherwise specified" categories.[11]

Prototype matching is an alternative approach that more naturally fits the way humans (including clinicians) categorize.[3,12]

This approach has been tested extensively with personality disorders (PDs) and has outperformed *DSM-IV* diagnosis in interrater reliability, validity, and ratings of clinical utility in numerous studies by multiple research teams.[1,2,13-18] Prototype diagnosis entails assessing the extent to which the patient's clinical presentation matches paragraph-length descriptions of disorders using a simple 5-point scale (eAppendix, http://www.jamapsych.com). A diagnosis of 4 or 5 means the patient resembles the diagnosis enough to be described as having the disorder ("caseness"); a diagnosis of 3 means the patient has features of the disorder; and the default value of 1 means the patient does not resemble the prototype. This system combines the advantages of categorical diagnosis (eg, patients can be described as having major depressive disorder [MDD]) with the advantages of dimensional diagnosis (ie, patients can be rated for the extent to which they have the disorder), reflecting the wealth of data suggesting that most psychopathology is better represented as dimensional than categorical.[4]

This article reports a 2-part study of the validity of prototype diagnosis in 2 independent samples using distinct methods, extending prior work on prototype diagnosis with PDs to a broader range of psychiatric disorders (mood and anxiety disorders). Since the work of Robins and Guze,[19] researchers have identified a number of criteria that can be used to validate a classification system and compare alternative approaches to classification and diagnosis (similar to assessing construct validity[20,21]). External criteria address the question of whether proposed diagnostic groupings predict theoretically relevant criterion variables (eg, interview data, laboratory tests, adaptive functioning, etiology, molecular genetics and epigenetics, and prognosis and/or treatment response).[19,21] The aim of the first study was to assess the validity of a prototype-matching approach compared with patient self-reports (ie, whether prototype diagnoses made by clinicians would correlate with patient self-reported dimensions with which they should correlate but not with those with which they should not). In the second study, we assessed the validity of prototype diagnoses across 3 independent research interviews. Independent interviewers completed prototype diagnoses following either a systematic clinical or life history interview or structured diagnostic interview for *DSM-IV* disorders, and a third interviewer assessed adaptive functioning. By examining the validity of prototype diagnosis compared with clinician-rated adaptive functioning and patient self-reported symptoms (study 1) and across multiple independent diagnostic interviewers (study 2), we extend a growing body of evidence supporting the use of dimensional prototype diagnosis to a broader range of psychiatric syndromes.

METHODS

STUDY 1

Participants

Participants were psychiatric outpatients receiving treatment in clinics in the Departments of Psychiatry and Psychology at Emory University and in the Department of Psychiatry at Harvard Medical School at the Cambridge Health Alliance. They included 50 women and 34 men, ranging in age from 18 to 60 years (mean [SD] age, 37.9 [12.3] years). Patients were primarily white (81 patients [97%]), represented a wide array of socioeconomic and ethnic groups (35 middle-class patients [42%]), and varied in levels of functioning as evidenced by Global Assessment of Functioning (GAF) scores ranging from 28 to 90 (mean [SD] GAF score, 62.8 [10.8]). Clinician participants were licensed or postgraduate clinicians with at least 2 years of clinical experience.

Procedure

Patients who agreed to participate and signed a consent form completed a packet of materials, after which the research coordinator contacted the treating clinician to complete a parallel packet. Clinicians were blind to any data obtained from the self-report instruments.

Measures

Clinician-Report Measures. *Clinical Data Form–Clinician Form.* The Clinical Data Form–Clinician Form is a questionnaire designed for expert informants that gathers information on a wide range of demographic, diagnostic, and etiological variables, including a rating of the *DSM-IV* GAF scale. The Clinical Data Form–Clinician Form has been used for a variety of studies.[22] Prior research finds ratings of adaptive functioning to be highly reliable and strongly correlated with ratings made by independent interviewers.[23,24] Clinician-reports on the Clinical Data Form–Clinician Form strongly correlate with patient self-reports, with high diagnostic efficiency.[25]

Prototype Diagnosis. Clinicians made prototype diagnoses on 6 disorders selected for their prevalence in outpatient samples: MDD, dysthymic disorder, bipolar disorder, panic disorder, generalized anxiety disorder, and posttraumatic stress disorder. Prototypes were constructed by weaving *DSM-IV* criteria together into paragraph-length descriptions (except for exclusion criteria that would prioritize one diagnosis over another or arbitrary time frames that could not be better captured with language signifying duration more generically).

DSM-IV *Mood Disorder Diagnosis.* Clinicians also made present vs absent mood disorder diagnoses for the 3 mood disorders (MDD, dysthymic disorder, and bipolar disorder) so we could compare them with their prototype diagnoses. To ensure that clinicians used *DSM-IV* criteria, we reproduced the relevant pages from the *DSM-IV* and asked the clinicians to review each criterion and the decision rules and then to indicate whether the disorder was present or absent.

Patient Self-Report Measures. Patients completed a range of well-validated measures. Based on prior research, these were predicted to correlate or not correlate with different disorders (convergent and discriminant validity).

Personality Assessment Inventory. The Personality Assessment Inventory[26] (PAI) is an omnibus measure designed to assess a wide range of psychopathology. For the present study, we examined the following scales and subscales: depression, suicidal ideation, anxiety, and traumatic stress for convergent validity; substance use, aggression, and antisocial personality disorder for discriminant validity.

Positive and Negative Affect Schedule–Trait Version. The patients completed the Positive and Negative Affect Schedule–Trait Version[27] (PANAS-T). This is a 20-item self-report measure consisting of 10 positive and 10 negative emotions.

Anxiety Sensitivity Index. The patients completed the Anxiety Sensitivity Index.[28] It is a 16-item scale that measures fear

Table 1. Hypothesized Correlations Between Prototype Diagnoses and Criterion Variables

Measure	MDD	Dysthymic Disorder	GAD	Panic Disorder	PTSD
			Prototype Diagnosis		
GAF	−	−	−	−	−
PAI subscale					
Depression	++	++	NS	NS	NS
Suicidal ideation	++	+	NS	NS	NS
Anxiety	+	+	++	++	+
Traumatic stress	NS	NS	NS	NS	++
PANAS-T subscale					
Positive affect	−−	−	NS	NS	NS
Negative affect	++	+	+	+	+
Anxiety Sensitivity Index	+	NS	NS	++	+
Discriminant indices on PAI					
Aggression	NS	NS	NS	NS	NS
Alcohol problems	NS	NS	NS	NS	NS
Drug problems	NS	NS	NS	NS	NS
Antisocial personality disorder	NS	NS	NS	NS	NS

Abbreviations: GAD, generalized anxiety disorder; GAF, Global Assessment of Functioning; MDD, major depressive disorder; NS, nonsignificant; PAI, Personality Assessment Inventory; PANAS-T, Positive and Negative Affect Schedule–Trait Version; PTSD, posttraumatic stress disorder; +, small positive correlation; ++, moderate positive correlation; −, small negative correlation; −−, moderate negative correlation.

of anxiety-related sensations, particularly those common in panic disorder.[29]

Data Analysis

The aim of the study was to assess the validity of a prototype-matching approach to prevalent Axis I disorders. Our first goal was to assess whether prototype diagnosis would demonstrate validity, first compared with clinician-rated adaptive functioning and second compared with self-reported symptoms (ie, whether clinicians' prototype diagnoses would correlate with patients' self-reports where they should but not where they should not). These hypotheses were tested conservatively given the high base-rate intercorrelations of mood and anxiety symptoms and disorders. (Because of low base-rate prevalence in the study 1 sample, we did not carry forward analyses using bipolar disorder.) Thus, in a first analysis, we correlated clinician-diagnosed prototypes with GAF scores and self-reported psychopathology.

Table 1 presents the hypothesized direction and magnitude of correlations between prototype diagnoses and criterion variables. For example, as the GAF is a nonspecific measure of functioning,[30,31] we hypothesized small to moderate inverse correlations across all diagnoses. With respect to criterion variables expected to distinguish particular disorders, a wealth of data has linked high negative affect to both mood and anxiety disorders but low positive affect uniquely to mood disorders.[32] We thus hypothesized PANAS-T negative affect to be significantly correlated with both anxiety and depression prototypes but positive affect to show significant inverse correlations with depression diagnoses alone. Similarly, based on prior research,[33] we expected the PAI traumatic stress subscale to uniquely discriminate a diagnosis of posttraumatic stress disorder. With respect to other measures, research has identified anxiety sensitivity as assessed by the Anxiety Sensitivity Index to be a prominent feature in panic disorder and secondarily in posttraumatic stress disorder[34,35]; recent research has shown moderately elevated Anxiety Sensitivity Index scores in some patients with MDD as well.[36-38] To test the discriminant valid-

ity of prototype diagnoses, we included psychopathology variables *not* expected to be uniquely associated with specific mood and anxiety diagnoses (ie, substance abuse, aggression, and antisocial personality disorder as assessed from the PAI).

Next, to examine the degree of agreement between clinicians' prototype diagnoses and their categorical *DSM-IV* diagnoses on the same disorders, we calculated diagnostic efficiency statistics for the 2 mood disorders on which we had both carefully obtained but naturalistic *DSM-IV* diagnoses and sufficient variance in the sample (MDD and dysthymic disorder).

To test whether prototype diagnosis performs as well as categorical *DSM-IV* diagnosis for the disorders on which we had categorical data (the 2 mood disorders), we performed 2 additional sets of analyses. First, we examined the raw correlations between categorical diagnoses and the same criterion variables used to assess the prototypes. Second, to determine whether prototype diagnoses might show incremental validity over categorical *DSM-IV* diagnoses (ie, whether they might account for more variance in predicting relevant criterion variables over and above categorical *DSM-IV* diagnoses), we used hierarchical multiple regression, entering categorical *DSM-IV* diagnoses for each of the 2 mood disorders in step 1 (coded 0/1) and prototype diagnoses in step 2, to predict clinician-rated GAF scores and self-reported PAI depression.

STUDY 2

Participants

Successive adult patients in the waiting rooms of the primary care and obstetrical-gynecological clinics of a university-affiliated urban public hospital were approached while either waiting for their medical appointments or waiting with others scheduled for medical appointments (typically significant others for obstetrics appointments). Exclusion criteria were minimal (mental retardation and active psychosis). The 143 participants (47 male [32%], 96 female [67%]) ranged in age from 18 to 74 years (mean [SD] age, 43.3 [13.4] years) and were predominantly African American (129 patients [90%]). They varied in levels of functioning as evidenced by GAF scores, ranging from 30 to 95 (mean [SD] GAF score, 70.4 [12.2]).

Procedure

Patients meeting criteria for participation based on a brief battery of questionnaires were scheduled for an initial assessment for the larger project. This project included 2 to 3 hours of self-reports followed by 2 full days of biological and psychiatric assessments.[39,40]

Measures

Structured Clinical Interview for *DSM-IV* Axis I Disorders. The Structured Clinical Interview for *DSM-IV* Axis I Disorders[41] (SCID) is a structured clinical interview designed to assess Axis I diagnoses in psychiatric population studies. The reliability of the procedure is well established.[42,43] The SCID interviews were conducted by trained clinical raters blinded to other assessment data.

Clinical Diagnostic Interview. The Clinical Diagnostic Interview (CDI) is a 2- to 3-hour systematic clinical interview designed to standardize the kind of interviewing approach typically used by experienced clinicians.[44,45] Following initial questions about the nature and history of current symptoms, the interviewer asks patients about significant interpersonal relationships from the past and present, work history, moods and

Table 2. Correlations Between Prototype Diagnoses or *DSM-IV* Mood Disorder Categorical Diagnoses and Criterion Indices in 84 Patients

| | Pearson Correlation Coefficient, r^a | | | | | | |
| | Prototype Diagnosis | | | | | *DSM-IV* Mood Disorder Categorical Diagnosis | |
Measure	MDD	Dysthymic Disorder	GAD	Panic Disorder	PTSD	MDD	Dysthymic Disorder
GAF	**−0.42**[b]	**−0.33**[b]	**−0.15**	**−0.11**	**−0.22**[c]	−0.40[b]	−0.14
PAI subscale							
Depression	**0.47**[b]	**0.37**[b]	0.12	0.11	0.07	0.50[b]	0.18
Suicidal ideation	**0.42**[b]	**0.39**[b]	0.12	−0.04	−0.04	0.32[b]	0.21
Anxiety	**0.27**[c]	**0.21**	**0.33**[b]	**0.37**[b]	**0.17**	0.16	−0.07
Traumatic stress	0.17	0.12	0.19	0.15	**0.30**[b]	0.21	−0.03
PANAS-T subscale							
Positive affect	**−0.34**[b]	**−0.27**[c]	−0.18	0.05	0.01	−0.44[b]	−0.16
Negative affect	**0.33**[b]	**0.24**[c]	**0.26**[c]	**0.26**[c]	**0.11**	0.22	0.01
Anxiety Sensitivity Index	**0.22**[c]	0.03	0.15	**0.23**[c]	**0.11**	0.29[b]	−0.09
Discriminant indices on PAI							
Aggression	0.15	0.03	0.07	−0.04	0.03	0.11	−0.04
Alcohol problems	0.07	0.07	0.14	−0.06	−0.09	0.06	0.14
Drug problems	0.04	−0.01	0.09	−0.03	0.01	−0.08	0.01
Antisocial personality disorder	−0.07	−0.07	−0.11	−0.11	−0.11	−0.03	−0.07

Abbreviations: GAD, generalized anxiety disorder; GAF, Global Assessment of Functioning; MDD, major depressive disorder; PAI, Personality Assessment Inventory; PANAS-T, Positive and Negative Affect Schedule–Trait Version; PTSD, posttraumatic stress disorder.
[a] Hypothesized correlations are shown in bold.
[b] $P < .05$.
[c] $P < .01$.

emotions, and characteristic ways of thinking (to assess clinical and subclinical thinking disturbances).

Prototype Diagnosis. In contrast to study 1, where the treating clinician provided the clinical diagnostic data, study 2 used diagnostic data from independent clinical research interviewers. The SCID interviewers (blinded to all other interview and self-report data) assigned both categorical diagnoses according to *DSM-IV* algorithms and prototype diagnoses for the same 6 mood and anxiety disorders rated in study 1. The CDI interviewers (also blinded to all other interview and self-report data) completed prototype diagnoses for the 6 disorders.

Longitudinal Interval Follow-Up Evaluation–Baseline Version. The Longitudinal Interval Follow-up Evaluation–Baseline Version[46] is a structured interview for assessing adaptive functioning across multiple domains. The validity of the Longitudinal Interval Follow-up Evaluation–Baseline Version has been previously demonstrated in multiple studies.[46,47] The Longitudinal Interval Follow-up Evaluation–Baseline Version interview was conducted and scored by a third interviewer who provided a GAF rating and was blinded to all data from the SCID and CDI.

Data Analysis

We first assessed the cross-observer validity of prototype diagnoses by correlating prototype diagnoses made following standard administration of the SCID with prototype diagnoses made from the systematic clinical interview (the CDI) by a different interviewer. We also examined the diagnostic efficiency of prototype ratings from the SCID interviewer to the standard categorical diagnoses obtained from SCID scoring algorithms.

Next, we assessed the incremental validity of prototype diagnosis over SCID *DSM-IV* diagnosis in a highly conservative way, by predicting GAF scores assessed from the Longitudinal Interval Follow-up Evaluation–Baseline Version interview from

prototype diagnosis made by the SCID interviewer, entering the SCID-based prototype diagnosis in step 2 of a hierarchical regression for each disorder after entering the SCID interviewer's categorical *DSM-IV* diagnosis in step 1. Given that the prototypes were constructed using *DSM-IV* criteria, the only difference between the 2 diagnoses for each disorder was diagnostic method (whether the interviewer was evaluating each criterion individually and applying cutoffs for categorical caseness or was judging the match between the pattern of the patient's symptoms and the same criteria woven together into a diagnostic prototype).

RESULTS

STUDY 1

Table 2 shows correlations between clinician-rated prototype and categorical *DSM-IV* diagnoses with clinician-rated adaptive functioning and self-reported psychopathology. Although we expected substantial cross-correlations between mood and anxiety disorders because of the common factor of negative affect they share,[48-50] prototype diagnoses provided compelling evidence for convergent and discriminant validity, with moderate correlations on the same dimensions assessed by self-report, small correlations on indices with which they have proven comorbid in prior research, and correlations near or below 0 with variables included for discriminant validity.

For example, the MDD prototype showed significant positive correlations with measures assessing depression and negative affectivity, notably the PAI depression scale, the PAI suicidal ideation scale, and the PANAS-T negative affect scale. As predicted based on prior

research on comorbidity, MDD correlated secondarily with PAI anxiety and anxiety sensitivity scores; however, unlike the anxiety disorder prototypes, it negatively correlated with the PANAS-T positive affect scale as predicted. As expected, MDD was also significantly associated with lower GAF scores.

Much like the MDD prototype, the dysthymic disorder prototype showed significant positive associations with measures related to negative affect, including the PAI depression scale, the PANAS-T negative affect scale, and the PAI suicidal ideation scale. Also as predicted, the dysthymic disorder prototype negatively correlated with the GAF score. Across measures, dysthymic disorder correlations were uniformly lower than those between the MDD prototype and the same criterion variables, as would be predicted by the severity of the depressive state.

The anxiety disorder prototypes functioned largely as expected, with GAD score associated positively with negative affect and the PAI anxiety scale. The PTSD prototype was uniquely associated with the PAI traumatic stress subscale, and the panic disorder prototype was most strongly associated with anxiety sensitivity. All prototypes were negatively related to the GAF score, although to somewhat different degrees, as expected by the severity of illness.

Table 2 summarizes the associations between *DSM-IV* categorical mood diagnoses and the same criterion variables. The magnitudes of the correlations for *DSM-IV* MDD were similar to its corresponding prototype. However, the categorical *DSM-IV* dysthymic disorder diagnosis showed substantially weaker correlations with criterion variables than the dysthymic disorder prototype. Thus, the criterion validity results suggest that prototype diagnoses of MDD and dysthymic disorder performed as well as or better than *DSM-IV* diagnoses for the same disorders.

To assess the degree of agreement between clinicians' prototype diagnoses and their categorical *DSM-IV* diagnoses of the same disorders, we used diagnostic efficiency statistics (using receiver operating curve characteristics). Areas under the curve for MDD and dysthymic disorder prototype diagnosis in predicting *DSM-IV* categorical diagnosis ratings were 0.90 and 0.79, respectively (both significant at $P < .001$). Using the prototype's caseness cutoff of 4 or higher indicating a positive diagnosis, the MDD prototype had an overall correct classification rate of 0.85, with a sensitivity of 0.77 and specificity of 0.89. The dysthymic disorder prototype had an overall correct classification rate of 0.87, with a sensitivity of 0.60 and specificity of 0.97. These data suggest that clinicians are diagnosing essentially the same constructs whether using prototype diagnosis or applying the more familiar *DSM-IV* diagnostic criteria and diagnostic algorithms for categorical diagnosis specified in the manual.

Finally, to determine whether prototype diagnosis might actually show incremental validity compared with *DSM-IV* categorical diagnosis (ie, whether prototype diagnosis captures information that categorical *DSM-IV* diagnosis does not), we conducted a series of hierarchical multiple regressions to predict clinician-rated GAF scores and PAI depression for the 2 mood disorders, entering categorical diagnosis in step 1 and prototype diagnosis

in step 2. The combination of prototype diagnosis and categorical diagnosis accounted for significant variance in PAI depression (MDD, $R^2 = 0.29$; dysthymic disorder, $R^2 = 0.13$) and GAF scores (MDD, $R^2 = 0.19$; dysthymic disorder, $R^2 = 0.09$). Adding prototype diagnosis (step 2) to categorical *DSM-IV* diagnosis alone (step 1) resulted in significant increments in predicted variance for clinician-reported adaptive functioning (MDD prototype, F change = 3.91, $P < .05$, $\Delta R^2 = 0.05$, $\beta = -0.29$, $P < .05$; dysthymic disorder prototype, F change = 6.02, $P < .05$, $\Delta R^2 = 0.07$, $\beta = -0.33$, $P < .05$). Prototype diagnosis similarly accounted for additional variance in self-reported depression (MDD prototype, F change = 4.70, $P < .05$, $\Delta R^2 = 0.04$, $\beta = 0.29$, $P < .05$; dysthymic disorder prototype, F change = 7.91, $P < .05$, $\Delta R^2 = 0.09$, $\beta = 0.37$, $P < .01$). In most cases, the standardized β values for categorical diagnosis accounted for nonsignificant variance when prototype diagnosis was included in the model (the exception being PAI depression and MDD, where prototype diagnosis was the stronger predictor but categorical diagnosis was predictive as well, $\beta = 0.31$, $P < .01$).

Given the degree of overlap between our categorical and dimensional diagnostic predictors (the prototypes were constructed from *DSM-IV* diagnostic criteria), to eliminate the possibility of statistical artifacts, we ran collinearity diagnostics with each regression. For each analysis, the magnitude of the variance inflation factor was well under 2. Variance inflation factors of 5 or greater raise concerns about multicollinearity[51]; thus, the validity results do not reflect overlap in the predictor variables.

STUDY 2

The data from study 1, particularly the incremental validity data, are compelling in that clinicians were highly familiar with *DSM-IV* diagnosis but not with prototype diagnosis, yet prototypes showed substantial incremental validity in predicting not only GAF scores as rated by the clinician but also psychopathology as independently rated by a second informant, the patient. Nevertheless, the data were limited in several respects, most importantly the small number of disorders on which we could assess the relative validity of prototype diagnosis and the lack of reliable interview data for making categorical diagnoses. Further, although prior research has shown that prototype ratings of Axis II disorders show high interrater reliability and cross-observer convergence more generally,[17] no data are available for Axis I syndromes, and it is unclear to what extent prototype diagnosis following the kind of less structured interviewing routinely conducted in clinical practice would correlate with data derived from a structured interview. We addressed these limitations in a second study, with patients sampled from a very different patient population and data obtained from 3 independent clinical research interviewers.

In the first set of analyses, we correlated prototype diagnoses from the SCID and the CDI. As **Table 3** shows, prototype diagnosis made by 2 independent interviewers using 2 entirely different interviewing methods on different days correlated on average $r = 0.50$, whereas off-

Table 3. Correlations Between Prototype Ratings From the Structured Clinical Interview for *DSM-IV* Axis I Disorders and the Clinical Diagnostic Interview in 67 to 70 Patients[a]

	Pearson Correlation Coefficient for CDI-Rated Prototype, *r*					
SCID-Rated Prototype	**MDD**	**Dysthymic Disorder**	**Bipolar Disorder**	**GAD**	**Panic Disorder**	**PTSD**
MDD	**0.56[b]**	0.52[b]	0.18	0.27[c]	0.41[b]	0.22
Dysthymic disorder	0.48[b]	**0.51[b]**	0.13	0.41[b]	0.14	0.32[b]
Bipolar disorder	0.12	0.13	**0.73[b]**	−0.05	−0.01	−0.02
GAD	0.23	0.42[b]	0.10	**0.38[b]**	0.24[c]	0.24
Panic disorder	0.15	0.16	0.05	0.31[b]	**0.35[b]**	0.10
PTSD	0.31[b]	0.21	0.14	0.12	0.12	**0.43[b]**

Abbreviations: CDI, Clinical Diagnostic Interview; GAD, generalized anxiety disorder; MDD, major depressive disorder; PTSD, posttraumatic stress disorder; SCID, Structured Clinical Interview for *DSM-IV* Axis I Disorders.

[a] Mean convergent validity coefficients (correlations in bold, on the diagonal, *r* = .50) and mean discriminant validity coefficients (*r* = .15) are shown.

[b] *P* < .01.

[c] *P* < .05.

diagonal (discriminant validity) correlations, even for highly comorbid disorders, averaged only *r* = 0.15. With the exception of generalized anxiety disorder, which showed relatively nonspecific correlates in comparison with the other prototypes, the prototype diagnoses converged and diverged as expected, much as in study 1.

Also as in study 1, diagnostic efficiency analyses of SCID ratings demonstrated that dimensional prototype diagnoses for each disorder strongly converged with categorical diagnoses obtained using *DSM-IV* criterion scoring algorithms from the same clinical material (areas under the curve ranged from 0.87-0.99). A caseness prototype rating of 4 or higher resulted in strong diagnostic efficiency statistics across all disorders, with overall correct classifications from 0.91 to 0.98, sensitivities ranging from 0.53 to 1.00, and specificities between 0.95 and 1.00.

The next set of analyses used hierarchical multiple regression to assess the incremental validity of *DSM-IV* vs prototype diagnoses based on the same interview data (the SCID), essentially holding constant the method of interview, the interviewer, and the diagnostic criteria and varying only the way of assigning diagnoses (*DSM-IV* algorithms vs prototype matching). Table 3 and **Table 4** report the incremental validity for prototype ratings based on the SCID interview after holding constant categorical *DSM-IV* diagnoses of the same disorders based on the same SCID interview. (Again, collinearity diagnostics were conducted and variance inflation factors were acceptably low, <2 in all cases.)

What is striking from both Table 3 and Table 4 and particularly from **Table 5** (anxiety disorders) is the extent to which categorical diagnosis drops out as a predictor when prototype diagnosis is entered into the equation (or in the case of dysthymic disorder yields a β value in the wrong direction). (For bipolar disorder, variance was once again limited in this primary care sample, leading neither prototype nor *DSM-IV* diagnosis to be particularly predictive of GAF scores.) These analyses suggest that categorical diagnosis using *DSM-IV* decision rules loses a substantial amount of information obtained from an interview designed for *DSM-IV* categorical diagnosis (the SCID) that can be captured by prototype matching.

COMMENT

The data lead to the following conclusions. First, in comparison with the categorical *DSM-IV* diagnostic procedures with which clinicians are familiar, a novel prototype diagnosis approach correlates with patient self-reported psychopathology in ways that generate validity coefficients that are as strong or stronger. Prototype diagnosis is associated with coherent patterns among criterion variables, including in highly conservative analysis, such as the predicted positive associations between both MDD and dysthymic disorder with negative affect, the predicted positive associations between anxiety disorders and negative affect, and the unique negative associations of only MDD and dysthymic disorder with positive affect.

Second, prototype diagnosis using an interview designed to systematize the kind of clinical interviewing widely used in everyday practice (the CDI) correlates in predictable ways with prototype diagnosis made by the most widely used structured interview in psychiatric research (the SCID). Prototype diagnosis also converges with clinical caseness treated categorically.

Third, prototype diagnosis demonstrates substantial incremental validity over and above categorical *DSM-IV* diagnosis in predicting patient self-reported psychopathology as well as both clinicians' and independent interviewers' ratings of global functioning on the GAF. The increments in validity are not modest. With the exception of 1 analysis for MDD in the first study, the predictive value of *DSM-IV* diagnosis dropped to 0 when prototype diagnoses were included in the equation, and prototype diagnosis for each of the disorders examined in the second study eliminated all variance accounted for by *DSM-IV* diagnosis using the algorithms prescribed in the diagnostic manual, with precisely the same information available from the same structured interview. Perhaps the most striking example was dysthymic disorder, in which prototype diagnosis provided incremental validity above and beyond categorical diagnosis for GAF score and PAI depression in the first study and for GAF score (the only criterion variable examined) in the second. In both studies, categorical diagnoses, whether by

Table 4. Incremental Validity of Structured Clinical Interview for *DSM-IV* Axis I Disorders–Based Prototype Ratings Over Structured Clinical Interview for *DSM-IV* Axis I Disorders Categorical Diagnoses for Mood Disorders in 136 to 143 Patients

Predicting GAF Score	β	F Change	R	R²
MDD				
Step 1: MDD categorical diagnosis		39.74[a]	0.48	0.23
Step 2: MDD categorical and prototype diagnosis		39.50[a]	0.64	0.40
MDD categorical diagnosis	−0.08			
MDD prototype diagnosis	−0.58[a]			
Dysthymic disorder				
Step 1: dysthymic disorder categorical diagnosis		3.33	0.15	0.02
Step 2: dysthymic disorder categorical and prototype diagnosis		43.07[a]	0.50	0.25
Dysthymic disorder categorical diagnosis	0.19[b]			
Dysthymic disorder prototype diagnosis	−0.59[a]			
Bipolar disorder				
Step 1: bipolar disorder categorical diagnosis		1.33	0.10	0.01
Step 2: bipolar disorder categorical and prototype diagnosis		1.40	0.14	0.02
Bipolar disorder categorical diagnosis	−0.01			
Bipolar disorder prototype diagnosis	−0.13			

Abbreviations: GAF, Global Assessment of Functioning; MDD, major depressive disorder.
[a] *P* < .001.
[b] *P* < .05.

Table 5. Incremental Validity of Structured Clinical Interview for *DSM-IV* Axis I Disorders–Based Prototype Ratings Over Structured Clinical Interview for *DSM-IV* Axis I Disorders Categorical Diagnoses for Anxiety Disorders in 136 to 143 Patients

Predicting GAF Score	β	F Change	R	R²
GAD				
Step 1: GAD categorical diagnosis		16.22[a]	0.32	0.10
Step 2: GAD categorical and prototype diagnosis		24.63[a]	0.48	0.23
GAD categorical diagnosis	0.00			
GAD prototype diagnosis	−0.49[a]			
Panic disorder				
Step 1: panic disorder categorical diagnosis		1.93	0.12	0.01
Step 2: panic disorder categorical and prototype diagnosis		6.34[b]	0.24	0.06
Panic disorder categorical diagnosis	0.05			
Panic disorder prototype diagnosis	−0.27[b]			
PTSD				
Step 1: PTSD categorical diagnosis		5.96[c]	0.27	0.07
Step 2: PTSD categorical and prototype diagnosis		3.99[c]	0.34	0.12
PTSD categorical diagnosis	−0.01			
PTSD prototype diagnosis	−0.34[c]			

Abbreviations: GAD, generalized anxiety disorder; GAF, Global Assessment of Functioning; PTSD, posttraumatic stress disorder.
[a] *P* < .001.
[b] *P* < .01.
[c] *P* < .05.

the treating clinician or structured interview, failed to predict most criterion variables with prototype diagnosis held constant, whereas the reverse was not the case. (With respect to bipolar disorder, future research using inpatient samples is clearly needed.)

Fourth, prototype diagnosis appears to be relatively robust across experience level of diagnosticians, type of patients (psychiatric vs primary care/obstetrics), and populations (primarily white vs primarily African American). It is notable that clinicians in the first study were relatively inexperienced (eg, postgraduate years III and IV) and had not only been relatively recently trained in the use of the *DSM* but lacked the rich knowledge structures to "fill out" mental prototypes that come with years of seeing patients with the forms of psychopathology they

were diagnosing, which should have attenuated correlations between prototype diagnoses and criterion variables. In fact, one of the advantages of providing clinicians with standardized diagnostic prototypes in a diagnostic manual is that they can then hang experience developed over time on shared prototypes rather than develop their own idiosyncratic mental prototypes that lead to unreliable diagnoses in clinical practice.[18] The data presented here do not support the frequently asserted argument that the problem with the *DSM* is that clinicians are not following it. Rather, the problem is that the *DSM* is wedded to complex algorithms yielding categorical diagnoses that are less predictive in both clinical practice and research based on structured interviews of the most basic criterion variables described years ago by Robins

and Guze[19] as essential to establishing the validity of an approach to classification and diagnosis.

The studies reported here have a number of limitations, potential criticisms, and implications. The primary 2 limitations pertain to the samples and disorders examined. First, because all patients were outpatients, there was clear restriction of range for bipolar disorder. Second, we tested only a subset of disorders, focusing on highly prevalent mood and anxiety disorders. This study cannot address the extent to which the results generalize to other disorders, although it is worth noting the increasing evidence of the validity, reliability, and clinical utility of prototype diagnosis in research on personality disorders[1,13,52,53] as well as mood disorders, including bipolar spectrum disorders.[54]

Third, it may well be that, as in the *International Statistical Classification of Diseases, 10th Revision* and proposals for the *International Statistical Classification of Diseases, 11th Revision*, we could produce even more reliable and valid findings if researchers and clinicians were to use versions of the same diagnoses that differ only in their method of application depending on the setting and goal. In multiple studies of several classes of disorder,[1,2,14,18] clinicians have now rated prototype diagnosis far more clinically useful and less cumbersome than the complex algorithms prescribed in the *DSM-IV* (eg, for mood, anxiety, eating, and personality disorders). However, single-item prototype ratings are likely to be less reliable than multi-item scales, preferably scales (criterion sets) derived empirically using procedures such as factor analysis.[52,54] Rather than relying on kinds of complex diagnostic procedures used since the *DSM-III*, however, researchers could rate the same criteria used to construct empirical prototypes for clinical use but do so using multi-item diagnostic scales that can be aggregated across items (much like the Hamilton Rating Scales, which do not require interviewers to tally 3 items from criterion A, 4 from criterion B, and so forth).

A final potential criticism is that these studies confounded a comparison between prototype diagnosis and *DSM-IV* diagnosis, with a comparison of dimensional and categorical diagnosis. A reasonable question is whether we can be certain that the prototypes did not outperform categorical diagnosis the same way other dimensional approaches might have. Two responses mitigate this argument. First, for better or worse, the current diagnostic system is categorical, and our goal was to test it against a practical dimensional alternative that also yields categorical diagnoses useful for clinical communication. The burden of proof is on advocates of other dimensional systems to show that they not only can match the incremental validity shown here but can also be implemented in ways that are clinically feasible. Second, research on personality disorders has clearly demonstrated that prototype approaches substantially outperform other dimensional approaches on measures of clinical utility without sacrificing validity.[1,13,14] How other dimensional models might look for Axis I disorders and whether they could be implemented in reliable and clinically useful ways with disorders such as the mood and anxiety disorders examined here are unknown. Counting symptoms of Axis I disorders (the only dimensional alternative for Axis I of which we are aware that preserves the syndromal approach familiar to clinicians and provides continuity with 30 years of research) would be just as cumbersome as, if not more so, than the current categorical system, requiring clinicians to inquire about and tally hundreds of symptoms.

In contrast, prototype diagnosis has the advantage of other dimensional approaches in that it can minimize artifactual comorbidity and the use of "not otherwise specified" diagnoses while permitting clinicians to diagnose clinically important subthreshold pathology. At the same time, it preserves continuity with the current diagnostic system without burdening clinicians with procedures that cannot be implemented in everyday practice, an advantage that, empirically, clinicians tend to prefer.[1,13,14,55] The current studies are by no means a definitive statement on dimensional diagnosis or prototype diagnosis, but they do highlight the cross-observer, concurrent, and incremental validity of prototype diagnosis as well as its viability as an alternative to a diagnosis approach that has been used for 30 years but has never been tested against a systematic empirical alternative and that has proven increasingly problematic.

REFERENCES

1. Westen D, Shedler J, Bradley R. A prototype approach to personality disorder diagnosis. *Am J Psychiatry*. 2006;163(5):846-856.
2. Westen D, Heim A, Morrison K, Patterson M, Campbell L. Simplifying diagnosis using a prototype-matching approach: implications for the next edition of the *DSM*. In: Beutler LE, Malik ML, eds. *Rethinking the DSM: A Psychological Perspective*. Washington, DC: American Psychological Association; 2002:221-250.
3. Ahn W, Kim NS. The causal status effect in categorization: an overview. In: Medin DL, ed. *The Psychology of Learning and Motivation: Advances in Research and Theory*. San Diego, CA: Academic Press; 2001:23-65.
4. Widiger TA, Clark LA. Toward *DSM-V* and the classification of psychopathology. *Psychol Bull*. 2000;126(6):946-963.
5. Jampala VC, Sierles FS, Taylor MA. The use of *DSM-III* in the United States: a case of not going by the book. *Compr Psychiatry*. 1988;29(1):39-47.
6. Morey LC, Ochoa ES. An investigation of adherence to diagnostic criteria: clinical diagnosis of the *DSM-III* personality disorders. *J Pers Disord*. 1989;3(3):180-192. doi:10.1521/pedi.1989.3.3.180.
7. Lewczyk CM, Garland AF, Hurlburt MS, Gearity J, Hough RL. Comparing DISC-IV and clinician diagnoses among youths receiving public mental health services. *J Am Acad Child Adolesc Psychiatry*. 2003;42(3):349-356.
8. Lipkowitz MH, Idupuganti S. Diagnosing schizophrenia in 1982: the effect of *DSM-III*. *Am J Psychiatry*. 1985;142(5):634-637.

9. Lahey BB, Applegate B, McBurnett K, Biederman J, Greenhill L, Hynd GW, Barkley RA, Newcorn J, Jensen P, Richters J. *DSM-IV* field trials for attention deficit hyperactivity disorder in children and adolescents. *Am J Psychiatry*. 1994;151(11):1673-1685.

10. Widiger TA, Cadoret R, Hare R, Robins L, Rutherford M, Zanarini M, Alterman A, Apple M, Corbitt E, Forth A, Hart S, Kultermann J, Woody G, Frances A. *DSM-IV* antisocial personality disorder field trial. *J Abnorm Psychol*. 1996;105(1):3-16.

11. Thompson-Brenner H, Westen D. A naturalistic study of psychotherapy for bulimia nervosa, part 1: comorbidity and therapeutic outcome. *J Nerv Ment Dis*. 2005;193(9):573-584.

12. Ahn W, Kim NS, Lassaline ME, Dennis MJ. Causal status as a determinant of feature centrality. *Cogn Psychol*. 2000;41(4):361-416.

13. Rottman BM, Ahn WK, Sanislow CA, Kim NS. Can clinicians recognize *DSM-IV* personality disorders from five-factor model descriptions of patient cases? *Am J Psychiatry*. 2009;166(4):427-433.

14. Spitzer RL, First MB, Shedler J, Westen D, Skodol AE. Clinical utility of five dimensional systems for personality diagnosis: a "consumer preference" study. *J Nerv Ment Dis*. 2008;196(5):356-374.

15. Westen D, Weinberger J. When clinical description becomes statistical prediction. *Am Psychol*. 2004;59(7):595-613.

16. Ortigo KM, Bradley B, Westen D. An empirically based prototype diagnostic system for *DSM-V* and *ICD-11*. In: Millon T, Krueger R, Simonsen E, eds. *Contemporary Directions in Psychopathology: Scientific Foundations of the* DSM-V *and* ICD-11. New York, NY: Guilford; 2010:374-390.

17. Westen D, DeFife JA, Bradley B, Hilsenroth MJ. Prototype personality diagnosis in clinical practice: a viable alternative for *DSM-V* and *ICD-11*. *Prof Psychol Res Pr*. 2010;41(6):482-487.

18. Westen D. Prototype diagnosis of psychiatric syndromes. *World Psychiatry*. 2012;11(1):16-21.

19. Robins E, Guze SB. Establishment of diagnostic validity in psychiatric illness: its application to schizophrenia. *Am J Psychiatry*. 1970;126(7):983-987.

20. Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull*. 1955;52(4):281-302.

21. Livesley WJ, Jackson DN. Guidelines for developing, evaluating, and revising the classification of personality disorders. *J Nerv Ment Dis*. 1992;180(10):609-618.

22. Westen D, Shedler J. Revising and assessing Axis II, part I: developing a clinically and empirically valid assessment method. *Am J Psychiatry*. 1999;156(2):258-272.

23. Westen D, Muderrisoglu S, Fowler C, Shedler J, Koren D. Affect regulation and affective experience: individual differences, group differences, and measurement using a Q-sort procedure. *J Consult Clin Psychol*. 1997;65(3):429-439.

24. Dutra L, Campbell L, Westen D. Quantifying clinical judgment in the assessment of adolescent psychopathology: reliability, validity, and factor structure of the Child Behavior Checklist for clinician report. *J Clin Psychol*. 2004;60(1):65-85.

25. DeFife JA, Drill R, Nakash O, Westen D. Agreement between clinician and patient ratings of adaptive functioning and developmental history. *Am J Psychiatry*. 2010;167(12):1472-1478.

26. Morey LC. *The Personality Assessment Inventory: Professional Manual*. Odessa, FL: Psychological Assessment Resources; 1991.

27. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. *J Pers Soc Psychol*. 1988;54(6):1063-1070.

28. Reiss S, Peterson RA, Gursky DM, McNally RJ. Anxiety sensitivity, anxiety frequency and the prediction of fearfulness. *Behav Res Ther*. 1986;24(1):1-8.

29. McNally RJ. Anxiety sensitivity and panic disorder. *Biol Psychiatry*. 2002;52(10):938-946.

30. Tungström S, Söderberg P, Armelius BA. Relationship between the Global Assessment of Functioning and other *DSM* axes in routine clinical work. *Psychiatr Serv*. 2005;56(4):439-443.

31. Hilsenroth MJ, Ackerman SJ, Blagys MD, Baumann BD, Baity MR, Smith SR, Price JL, Smith CL, Heindselman TL, Mount MK, Holdwick DJ Jr. Reliability and validity of *DSM-IV* Axis V. *Am J Psychiatry*. 2000;157(11):1858-1863.

32. Watson D, Clark LA, Carey G. Positive and negative affectivity and their relation to anxiety and depressive disorders. *J Abnorm Psychol*. 1988;97(3):346-353.

33. McDevitt-Murphy ME, Weathers FW, Flood AM, Eakin DE, Benson TA. The utility of the PAI and the MMPI-2 for discriminating PTSD, depression, and social phobia in trauma-exposed college students. *Assessment*. 2007;14(2):181-195.

34. Peterson RA, Reiss S. *Anxiety Sensitivity Index Manual*. Washington, OH: IDS Publishing; 1992.

35. Taylor S, Koch WJ, McNally RJ. How does anxiety sensitivity vary across the anxiety disorders? *J Anxiety Disord*. 1992;6(3):249-259. doi:10.1016/0887-6185(92)90037-8.

36. Otto MW, Pollack MH, Fava M, Uccello R, Rosenbaum JF. Elevated Anxiety Sensitivity Index scores in patients with major depression: correlates and changes with antidepressant treatment. *J Anxiety Disord*. 1995;9(2):117-123. doi:10.1016/0887-6185(94)00035-2.

37. Rodriguez BF, Bruce SE, Pagano ME, Spencer MA, Keller MB. Factor structure and stability of the Anxiety Sensitivity Index in a longitudinal study of anxiety disorder patients. *Behav Res Ther*. 2004;42(1):79-91.

38. Taylor S, Koch WJ, Woody S, McLean P. Anxiety sensitivity and depression: how are they related? *J Abnorm Psychol*. 1996;105(3):474-479.

39. Gillespie CF, Bradley B, Mercer K, Smith AK, Conneely K, Gapen M, Weiss T, Schwartz AC, Cubells JF, Ressler KJ. Trauma exposure and stress-related disorders in inner city primary care patients. *Gen Hosp Psychiatry*. 2009;31(6):505-514.

40. Bradley B, DeFife JA, Guarnaccia C, Phifer J, Fani N, Ressler KJ, Westen D. Emotion dysregulation and negative affect: association with psychiatric symptoms. *J Clin Psychiatry*. 2011;72(5):685-691.

41. First M, Spitzer R, Gibbon M, Williams J. *Structured Clinical Interview for* DSM-IV *Axis I Disorders*. New York: New York State Psychiatric Institute; 1997.

42. Zanarini MC, Skodol AE, Bender D, Dolan R, Sanislow C, Schaefer E, Morey LC, Grilo CM, Shea MT, McGlashan TH, Gunderson JG. The Collaborative Longitudinal Personality Disorders Study: reliability of Axis I and II diagnoses. *J Pers Disord*. 2000;14(4):291-299.

43. Lobbestael J, Leurgans M, Arntz A. Inter-rater reliability of the Structured Clinical Interview for *DSM-IV* Axis I Disorders (SCID I) and Axis II Disorders (SCID II). *Clin Psychol Psychother*. 2011;18(1):75-79.

44. Westen D, Muderrisoglu S. Assessing personality disorders using a systematic clinical interview: evaluation of an alternative to structured interviews. *J Pers Disord*. 2003;17(4):351-369.

45. DeFife JA, Westen D. Empirically informed clinical interviewing for personality disorders. In: Levy R, Ablon J, Kächele H, eds. *Psychodynamic Psychotherapy Research: Evidence-Based Practice and Practice-Based Evidence*. New York, NY: Springer; 2012:553-569.

46. Keller MB, Lavori PW, Friedman B, Nielsen E, Endicott J, McDonald-Scott P, Andreasen NC. The Longitudinal Interval Follow-up Evaluation: a comprehensive method for assessing outcome in prospective longitudinal studies. *Arch Gen Psychiatry*. 1987;44(6):540-548.

47. Warshaw MG, Keller MB, Stout RL. Reliability and validity of the Longitudinal Interval Follow-up Evaluation for assessing outcome of anxiety disorders. *J Psychiatr Res*. 1994;28(6):531-545.

48. Krueger RF, Hicks BM, Patrick CJ, Carlson SR, Iacono WG, McGue M. Etiologic connections among substance dependence, antisocial behavior, and personality: modeling the externalizing spectrum. *J Abnorm Psychol*. 2002;111(3):411-424.

49. Kendler KS, Prescott CA, Myers J, Neale MC. The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Arch Gen Psychiatry*. 2003;60(9):929-937.

50. Watson D, Clark LA. Negative affectivity: the disposition to experience aversive emotional states. *Psychol Bull*. 1984;96(3):465-490.

51. O'Brien RM. A caution regarding rules of thumb for variance inflation factors. *Qual Quant*. 2007;41(5):673-690. doi:10.1007/s11135-006-9018-6.

52. Westen D, Shedler J, Bradley B, DeFife JA. An empirically derived taxonomy for personality diagnosis: bridging science and practice in conceptualizing personality. *Am J Psychiatry*. 2012;169(3):273-284.

53. Skodol AE, Bender DS, Morey LC, Clark LA, Oldham JM, Alarcon RD, Krueger RF, Verheul R, Bell CC, Siever LJ. Personality disorder types proposed for *DSM-5*. *J Pers Disord*. 2011;25(2):136-169.

54. Westen D, Malone JC, DeFife JA. An empirically derived approach to the classification and diagnosis of mood disorders. *World Psychiatry*. 2012;11(3):172-180.

55. Rottman BM, Kim NS, Ahn WK, Sanislow CA. Can personality disorder experts recognize *DSM-IV* personality disorders from five-factor model descriptions of patient cases? *J Clin Psychiatry*. 2011;72(5):630-639.