Clinical utility of the DSM-5 alternative model for borderline personality disorder: Differential diagnostic accuracy of the BFI, SCID-II-PQ, and PID-5

J. Christopher Fowler\textsuperscript{a,b,*}, Alok Madan\textsuperscript{a,b}, Jon G. Allen\textsuperscript{b}, Michelle Patriquin\textsuperscript{a,b}, Carla Sharp\textsuperscript{c}, John M. Oldham\textsuperscript{b}, B. Christopher Frueh\textsuperscript{b,d}

\textsuperscript{a}The Menninger Clinic, 12301 Main Street, Houston, TX 77035, United States
\textsuperscript{b}Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, United States
\textsuperscript{c}University of Houston, United States
\textsuperscript{d}University of Hawaii, 200 West Kawili St., Hilo, HI 96720, United States

Abstract

Background: With the publication of DSM 5 alternative model for personality disorders it is critical to assess the components of the model against evidence-based models such as the five factor model and the DSM-IV-TR categorical model. This study explored the relative clinical utility of these models in screening for borderline personality disorder (BPD).

Methods: Receiver operator characteristics and diagnostic efficiency statistics were calculated for three personality measures to ascertain the relative diagnostic efficiency of each measure. A total of 1653 adult inpatients at a specialist psychiatric hospital completed SCID-II interviews. Sample 1 (n = 653) completed the SCID-II interviews, SCID-II Questionnaire (SCID-II-PQ) and the Big Five Inventory (BFI), while Sample 2 (n = 1,000) completed the SCID-II interviews, Personality Inventory for DSM5 (PID-5) and the BFI.

Results: BFI measure evidenced moderate accuracy for two composites: High Neuroticism+ low agreeableness composite (AUC = 0.72, SE = 0.01, \( p < 0.001 \)) and High Neuroticism+ Low + Low Conscientiousness (AUC = 0.73, SE = 0.01, \( p < 0.0001 \)). The SCID-II-PQ evidenced moderate-to-excellent accuracy (AUC = 0.86, SE = 0.02, \( p < 0.0001 \)) with a good balance of specificity (SP = 0.80) and sensitivity (SN = 0.78). The PID-5 BPD algorithm (consisting of elevated emotional lability, anxiousness, separation insecurity, hostility, depressivity, impulsivity, and risk taking) evidenced moderate-to-excellent accuracy (AUC = 0.87, SE = 0.01, \( p < 0.0001 \)) with a good balance of specificity (SP = 0.76) and sensitivity (SN = 0.81).

Conclusions: Findings generally support the use of SCID-II-PQ and PID-5 BPD algorithm for screening purposes. Furthermore, findings support the accuracy of the DSM 5 alternative model Criteria B trait constellation for diagnosing BPD. Limitations of the study include the single inpatient setting and use of two discrete samples to assess PID-5 and SCID-II-PQ.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

During the run-up to the DSM 5 Alternative Model for personality [1], empirical reviews and new research focused on the best approach to conceptualizing and measuring personality pathology. Empirically-validated models including the five-factor model (FFM) demonstrated predictive validity in the relationship to clinical personality features [2–12].

In light of empirical evidence supporting dimensional models of personality and the superior performance of the FFM, the Diagnostic and Statistical Manual of Mental Disorders Personality and Personality Disorders Work Group proposed a hybrid, dimensional model for diagnosing personality disorders consisting of five broad, higher-order personality trait domains comprised of subordinate trait facets of pathological personality [13,14]. The trait facets were integrated into the Criteria B sets for specific diagnoses. In the case of borderline personality disorder (BPD) traits of emotional lability, anxiousness, separation insecurity, hostility, depressivity, impulsivity, and risk taking were proposed as core features of the disorder. In the wake of
the publication of DSM 5, research on the clinical utility of the Personality Inventory for DSM-5 (PID-5) the alternative model’s personality trait measure [15] has rapidly built a base of support for its construct validity [16] and clinical utility [17–21] as a broad-based measure of maladaptive personality. A recent study partially supported the proposed BPD trait facets in differentiating BPD from other personality disorders and healthy controls [22]. While informative, the regression models provide limited information about clinical decision making when diagnosing patients. If clinicians are to utilize DSM 5 emerging measures such as the PID-5 to aid in the determination of personality disorder diagnosis then data on trait composites and optimal cutoffs must be established. Of equal importance is establishing the clinical utility of the emerging measures relative to well-validated FFM measures and the traditional DSM-IV measures for PDs. The current study aimed to address this gap by leveraging admission data from the Menninger Outcomes Project [23] to assess the screening accuracy in detecting the presence of BPD among three measures representing three competing approaches to personality assessment: an FFM measure (Big Five Inventory: BFI), self-report DSM 5 polythetic categorical diagnostic criteria (Structured Clinical Interview for DSM-IV Axis II Personality Disorders Questionnaire: SCID-II-PQ) and the dimensional measure of Criteria B personality traits of the alternative model (PID-5). The large and voluntary inpatient sample (N = 1653) with a 19% prevalence of BPD provided an excellent test case for evaluating the screening properties of personality measures.

We selected BPD as the disorder to assess because in clinical populations it is the most common personality disorder with a relative risk rate of 28.5% and it has an international lifetime prevalence rate of 5.9% [24]. Longitudinally, BPD is associated with high rates of suicide and severe functional impairment [25–27], comorbid mental disorders [28], treatment utilization [29], and high costs to society [30]. Despite being the most extensively studied personality disorder [31–33] there remains significant gaps in the evaluation of clinically useful screening for BPD.

Measure selection was based on the competing models, and past evidence linking personality facets to BPD. The BFI was utilized due to the strong evidence of convergence between FFM model and personality disorders. In one of the first clinical studies linking the FFM to BPD diagnosis, Clarkin [34] found high neuroticism and low agreeableness were correlated with the categorical diagnosis, and that high neuroticism was correlated with 5 of 9 BPD criteria. Subsequent studies have demonstrated strong association with BPD diagnosis [35] with high neuroticism scores functioning as a distinguishing characteristic of BPD [36]. Reflecting its biological underpinnings, research indicated that the genetic factors that influence individual differences in neuroticism, agreeableness, conscientiousness, and extraversion account for all genetic liability to borderline personality disorder, predominantly through high neuroticism and low agreeableness [37]. Based on these findings FFM traits of high neuroticism and low agreeableness were hypothesized to be associated with BPD diagnosis. Individual FFM traits and an alternative FFM model (high neuroticism, low agreeableness, and low conscientiousness) were also explored.

Borderline items from DSM-IV SCID-II-PQ [38] demonstrated good sensitivity (0.78–0.87) and specificity (Range 0.75–0.78) in predicting BPD [39–42]. Borderline items from the SCID-PQ are expected to provide the best screening properties because the items reflect the SCID-II interview criteria. The PID-5 is a dimensional measure of personality pathology developed by members of the Diagnostic and Statistical Manual of Mental Disorders Personality and Personality Disorders Work Group to map on to personality disorders including BPD. According to the alternative model, elevations in traits of emotional lability, anxiousness, separation insecurity, hostility, depressivity, impulsivity, and risk taking are hypothesized to be predictive of BPD. The PID-5 constellation (summed average scores) of the above trait facets will be referred to as the PID-5 BPD algorithm and was hypothesized to produce adequate screening properties for BPD.

2. Methods

2.1. Participants

Sample 1 consisted of 653 adult patients consecutively admitted (November 2010–June 2012) to the Menninger Clinic who completed a SCID II research interview and completed the BFI and self-report SCID-II QP. The SCID-II QP was not used as a screen for the SCID-II interview. Rather, all criteria for 6 PDs (without skip-outs) were assessed using the research version of the Structured Clinical Interview for DSM-IV Axis II Personality Disorders. Based on SCID –II interview 127 (19.4%) patients were diagnosed with BPD. Gender distribution was relatively even with 50.6% female. Average age was 35.9 years (SD = 14.6). Participants were Caucasian (90.3%), multiracial (5.8%), African American (1.9%), Asian (1.2%), American Indian (0.4%) and Pacific Islander (0.4%), and 7.3% identified as being of Hispanic or Latino ethnicity. Education level was above the national average with 85.5% indicating some college experience. The majority (62.7%) of participants were not working in the 30 days prior to admission.

Sample 2 consisted of 1000 adult patients consecutively admitted (July 2012–May 2016) to the Menninger Clinic who completed a SCID II research interview as well as the BFI and PID-5 self-report. Based on SCID-II interview 191 (19.1%) patients were diagnosed with BPD. Gender distribution was relatively even with 46% female. Average age was 34.0 years (SD = 14.6). Participants were Caucasian (90.2%), multiracial (5.8%), African American (1.9%), Asian (1.2%), American Indian (0.4%) and Pacific Islander (0.6%), and 7.3% identified as being of Hispanic or Latino
ethnicity. Education level was above the national average with 88.7% indicating some college experience. The majority (62.7%) of participants were not working in the 30 days prior to admission.

2.2. Procedures

Data were collected as part of the hospital’s ongoing Adult Outcomes Project to assess treatment response. All measures used in the current study were collected within 72 h of admission. Assessments were conducted via hospital-wide web survey on laptop computers. This project was a hybrid clinical quality and research outcomes project, conducted with all patients; accordingly, all assessments were designed and implemented as an element of routine clinical care and integrated into treatment planning and monitoring of progress such that less than 4% of patients declined participation. Patients and their treatment teams were provided with personality profile scores and feedback within 24 h with the expressed intention that individual patient profiles would be used to inform treatment decisions. Patients and teams were informed that the findings would be used to evaluate the overall effectiveness of treatment and for research purposes. Use of the project’s data was approved by Baylor College of Medicine’s Institutional Review Board.

2.3. Measures

Demographic variables and history of psychiatric hospitalization and psychiatric service usage were assessed using a standardized patient information survey [23]. Personality disorder diagnoses were assessed using the research version of the Structured Clinical Interview for DSM-IV Axis II Personality Disorders [38]. Individual-level criteria were coded as absent (0) or present (1) for Antisocial, Avoidant, Borderline, Narcissistic, Obsessive-Compulsive, and Schizotypal with no skip-outs (other PDs were not coded due to base-rates below 1% in the hospital between 2010 and 2012). Psychiatric disorders were assessed using the research version of the Structured Clinical Interview for DSM-IV Disorders [43]. Master’s level researchers conducted all interviews and coded diagnoses after reviewing past psychiatric history, collateral information from family, psychosocial assessment, and nursing assessment. This process combined the ecologically valid longitudinal evaluation of the “all available data” diagnostic approach [44] with the rigorous research diagnostic interviews. In light of the significant overlap between BPD, and mood/anxiety disorders [45–49] interviewers were trained to discriminate between BPD’s hallmark of transient/reactive affective lability [45] and the more sustained symptoms of mood disorders. The BFI [50] is a 44-item questionnaire that assesses the FFM personality domains of neuroticism, agreeableness, conscientiousness, extraversion and openness. Domain scales demonstrate high reliability, clear factor structure, and strong convergence with the NEO-FFI [50]. The BFI yielded adequate internal consistency in the current sample for neuroticism (Cronbach’s $\alpha = 0.83$), agreeableness (Cronbach’s $\alpha = 0.80$), conscientiousness (Cronbach’s $\alpha = 0.85$), extraversion (Cronbach’s $\alpha = 0.88$) and openness (Cronbach’s $\alpha = 0.80$). The BPD items from the SCID-II-PQ [38] consist of 15 Yes/No questions keyed to the 9 BPD criteria such as: “Have you often become frantic when you thought that someone you really cared about was going to leave you?” Psychometric properties of the SCID-II-PQ demonstrated adequate internal consistency (Cronbach’s $\alpha = 0.88$) and two week test-retest (ICC = 0.87) reliability [38]. In the current sample, the SCID-II-PQ yielded adequate internal consistency (Cronbach’s $\alpha = 0.83$). The PID-5 [15] is 220-item dimensional measure comprised of 25 non-overlapping trait scales that load onto 5 higher-order dimensions (negative affect, detachment, antagonism, disinhibition, and psychotism). Assessments of the clinical utility of the PID-5 indicated that trait domains accounted for a substantial amount of variance in DSM-IV personality disorder severity and are linked to DSM-IV personality disorders [51], and demonstrated incremental validity in predicting DSM-IV PDs [52]. Recent findings indicated that PID-5 traits are highly stable, prospectively predictive of psychosocial functioning, and associated with psychosocial functioning over time [53]. The PID-5 yielded adequate internal consistency in the current sample (Cronbach’s $\alpha = 0.98$).

2.4. Data analysis

All analyses were conducted in IBM SPSS version 22.0. Descriptive statistics and independent sample t-tests and chi-square analyses were performed to identify potential differences between Sample 1 and 2. Receiver Operating Characteristics (ROC) analyses were carried out to obtain the area under the curve (AUC) and standard error (SE) using the non-parametric method to assess the accuracy of the screening measures in identifying BPD diagnosis. Diagnostic efficiency statistics [54] were calculated for five metrics of screening properties: 1. Sensitivity (SN: the ability of a “positive” test result to correctly identify treatment non-response); 2. Specificity (SP: the ability of a “negative” test result to correctly identify those individuals without treatment non-response); 3. Positive predictive power (PPP: the probability that an individual has BPD when the test result is “positive”); 4. Negative predictive power (NPP: the probability that an individual does not have BPD when the test result is “negative”); and 5. Odds Ratio (OR: the odds that BPD is predicted when the test result is “positive”, compared to the odds of diagnosing BPD when the test result is negative).

3. Results

3.1. Descriptive statistics

There were no significant differences in the samples for age, gender, prevalence of borderline personality disorder.
diagnosis, or number for BPD criteria. Sample 1 was diagnosed with more clinical diagnoses (t = 2.40, p = 0.02; Cohen’s d = 0.12); however, the effect size was small. From the total sample, 82% of patients were diagnosed with at least two co-occurring Axis I disorders with average of 3.5 (SD = 2.1). The most prevalent disorders included major mood disorders (MDD Spectrum = 64.3%; Bipolar Spectrum = 18.0%), anxiety spectrum disorders (61.7%), substance use disorders (55.4%), and personality disorders (34.7%). Borderline PD was diagnosed in 19% of the sample.

3.2. Receiver Operator Characteristic (ROC) analyses

Separate ROC analyses were conducted for three BFI composites (t-scores for individual traits, neuroticism + agreeableness [reverse scored], neuroticism + agreeableness [reverse scored] + conscientiousness [reverse scored]), SCIP-PQ, and PID-5 BPD algorithm. Area under the ROC curve indicated poor accuracy of the FFM neuroticism (AUC = 0.69, SE = 0.01, p < 0.0001), agreeableness (AUC = 0.67, SE = 0.02, p < 0.0001), conscientiousness (AUC = 0.65, SE = 0.02, p < 0.0001), extraversion (AUC = 0.45, SE = 0.02, p < 0.003) and openness (AUC = 0.50, SE = 0.02, p < 0.0001). High neuroticism + low agreeableness composite evidenced moderate accuracy (AUC = 0.72, SE = 0.01, p < 0.001). The composite model of high neuroticism + low agreeableness + low conscientiousness evidenced moderate accuracy (AUC = 0.73, SE = 0.01, p < 0.0001). The 15-item SCID-II-PQ evidenced moderate-excellent accuracy (AUC = 0.86, SE = 0.02, p < 0.0001). The PID-5 BPD algorithm evidenced moderate-excellent accuracy (AUC = 0.87, SE = 0.01, p < 0.0001).

3.3. Diagnostic efficiency statistics

The clinical utility of any screening or diagnostic test is its ability to detect or predict any given individual’s likelihood of having or acquiring the malady in question [55]. Performance of the FFM algorithms (Table 1) indicated modest sensitivity, specificity and odds ratios. Results indicate scores ≥ 8 on the borderline subscale from the SQID-IIQP was associated with 78% sensitivity, 80% specificity, and odds ratio (OR = 13.67). Similarly, the PID-5 BPD algorithm indicate scores ≥ 11 was associated with 81% sensitivity, 76% specificity, and odds ratio (OR = 13.26).

4. Discussion

In light of the estimated 34.6 million adults that received mental health care in 2014 [56] accurate self-report BPD screening measures could be a cost-effective solution. To date, no BPD screener has emerged as the gold standard because each self-report screener evaluated thus far has failed to produce an adequate balance of sensitivity and specificity [57]. In medical practice, screening tests are often used to make provisional diagnosis until more definitive diagnostic tests are completed. In the case of psychiatric emergency departments and outpatient clinics, relatively accurate screening tests could be used to screen out most individuals with questionable borderline personality disorders and identify the smaller percentage of adults who should undergo more costly diagnostic interviews.

The current results indicate that the BPD items from the SCIDII PQ and PID-5 BPD algorithm provided an adequate balance of specificity and sensitivity; therefore, these measures could be used to screen for BPD. From a clinical perspective, it is highly advantageous to identify patients who display borderline traits using relatively low cost screenings tests especially in psychiatric and general medical settings. The equivalent diagnostic accuracy of the PID-5 and SCID-II QP also add to a growing body of research that supports the clinical validity and utility of the PID-5. Furthermore, the results add to an emerging literature supporting the continuity between traditional diagnosis and the AMPD model of BPD. Continuity between models is particularly advantageous in that new knowledge emerging from AMPD research can be synthesized with accrued knowledge from clinical, neurobiological, and epidemiological findings based on DSM-IV derived BPD diagnoses. Potential clinical utility of the PID-5 algorithm includes a more nuanced profile of the elevated personality traits that may inform intervention selection, but this speculation must await future study.

Relative to a sample of medical screening tests (Table 2), the SCIP-II-QP and PID-5 demonstrate superior screening characteristics compared to urine dipstick test for urinary tract infection [58] and comparable performance to the Multi-Test II in screening for timothy grass allergy [59], and the rapid tests for human influenza [60]. BPD screening instruments were inferior to mammography for detecting breast cancer [61] and blood tests for detecting HIV infection [62]. Based upon this cursory review, it is clear that, from a psychometric perspective, BPD screening can be substantially improved and currently fails to meet the gold standard of the blood tests for HIV.

A factor that may significantly impede further progress in advancing BPD screening is the fact that the reliability of the

Table 1

<table>
<thead>
<tr>
<th>BPD screener</th>
<th>SN</th>
<th>SP</th>
<th>PPP</th>
<th>NPP</th>
<th>OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFM composites (N = 1000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N + A ≥ 112</td>
<td>0.70</td>
<td>0.62</td>
<td>0.31</td>
<td>0.89</td>
<td>3.82</td>
</tr>
<tr>
<td>N + A + C ≥ 185</td>
<td>0.71</td>
<td>0.62</td>
<td>0.31</td>
<td>0.90</td>
<td>3.95</td>
</tr>
<tr>
<td>SCIDII PQ (n = 653)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCIDII PQ ≥ 8</td>
<td>0.78</td>
<td>0.80</td>
<td>0.49</td>
<td>0.94</td>
<td>13.67</td>
</tr>
<tr>
<td>PID-5 BPD algorithm (n = 1000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PID-5 ≥ 11</td>
<td>0.81</td>
<td>0.76</td>
<td>0.45</td>
<td>0.94</td>
<td>13.27</td>
</tr>
</tbody>
</table>

SP = specificity, the ability of a negative PHQ-9 to correctly identify non-responders; PPP = positive predictive power; NPP = negative predictive power; OR = odds ratio.
traditional polythetic approach to PD diagnosis is far from optimal. For example, the DSM 5 field trial yielded “questionable” inter-rater reliability (kappa = 0.34) across 11 academic centers in the US and Canada [63]. Furthermore, the DSM 5 polythetic system requiring five out of nine criteria to make a categorical BPD diagnosis results in massive heterogeneity within the disorder, thus increasing the complexity and error in determining a BPD diagnosis. It is important to note that the AMPD was designed to reduce overlap among personality disorder diagnoses, reduce heterogeneity within specific PD diagnoses, eliminate arbitrary diagnostic thresholds with little or no research basis, and provide diagnostic thresholds that are related to level of impairment in a meaningful way [64]. The current study provides the first data on research-derived cut-points for diagnostic thresholds for BPD based on PID-5 data; however, these findings require replication in general population and outpatient samples in order to justify clinical application. Future research applying the alternative model’s diagnostic criteria utilizing PID-5 data to anchor Criteria B decision making may help to address the problem of overlap (co-occurrence) of personality disorders.

While the current findings hold considerable promise, there are notable limitations that bear mention. First, results may not generalize to community outpatient samples because the current inpatient sample was predominantly Caucasian, better educated, and less psychotic. Second, only three screening instruments were evaluated whereas there are numerous personality scales that may yield comparable or better diagnostic properties. Third, despite efforts to ensure diagnostic accuracy of BPD during SCID-II interviewing, the high rates of co-occurring mood and anxiety disorders, so prevalent in the entire field [45–49], may have attenuated the strength of prediction of the screening measures. Finally, the study design was limited to the measurement of Criteria B (personality trait criteria). The addition of measurement of Criteria A (impairment in self/interpersonal functioning) with the Level of Personality Functioning Scale [65] would provide a more robust assessment of the alternative model.

References


