The Predictive Validity of a Screening Measure for identifying individuals with Personality Disorders

Jennifer Mulcahy-Avery, McNair Scholar, The Pennsylvania State University

Faculty Research Advisor: Dr. Kenneth N. Levy
Assistant Professor of Psychology
Department of Psychology
College of Liberal Arts
The Pennsylvania State University

Abstract

We examined the predictive validity of a personality disorders screening instrument, the International Personality Disorder Screener (IPDE-S). One thousand and fourteen undergraduates at a large Northeastern urban university completed the IPDE-S as part of a larger study (Posner et al., 2002). A subset of 66 individuals were interviewed with the International Personality Disorder Examination (IPDE), a semi-structured interview for assessment for the DSM, designed by NIMH and WHO to assess DSM-IV and ICD (International Classification Diagnosis). The predictive validity of the IPDE-S for identifying individuals with personality disorders on the IPDE was determined using the conditional probabilities of positive predictive power, negative predictive power, sensitivity, and specificity. Findings indicate the screener had high sensitivity but low specificity.

Introduction

Recent research suggests that personality disorders (PD's) are highly prevalent, frequently comorbid with other psychiatric disorders and negatively affect the outcome of otherwise efficacious treatment PD's are even highly prevalent in non-clinical epidemiological samples (Grant, et al., 2004; Skodol, et al., 2002). In college samples, PD's have shown to negatively impact students psychosocial and academic functioning (Bagge et al., 2004; Lenzenweger; 1999). Given these facts, it has become increasingly important to be able to identify individuals suffering from personality disorders. In response, a number of clinical researchers have developed brief screening measures to identify individuals who may be suffering from PDs (Loranger, 1994; Zanarini et al. 2003). These screening measures assess a variety of PDs, showing different levels of sensitivity, specificity, positive predictive power and negative predictive power.

Zanarini et al. (2003) developed a screening measure for borderline personality disorder (BPD). The screener consisted of 10 true/false questions that assessed criteria based on the DIPD-IV (diagnostic interview for the DSM-IV personality disorders) module. The study looked for high sensitivity and specificity rates for their screener. At the optimal cutoff score of 7, the screener yielded a sensitivity of .81 and .85, showing that the screener was able to correctly identify all those who met diagnostic criteria for BPD on the interview, as well as correctly identifying all the individuals who did not meet criteria for BPD on the interview. The adequate levels of both sensitivity and

specificity make the screener a viable tool in indentifying which individuals should be evaluated more thoroughly for the presence BPD.

In another study, (Dalrymple and Zimmerman, 2008) screened individuals for the prevalence of social fears and a lifetime history of SAD (social anxiety disorder) using the screening question in the SAD module of the SCID I (SCID; First, Spitzer, Williams, & Gibbon, 1997). They measured the screener using the four conditional probabilities of sensitivity, specificity, positive predictive power and negative predictive power. The researchers modified the screener by adding a list of 13 social fears. Subjects were asked the first question "Was there ever anything that you have been afraid to do or felt uncomfortable doing in front of other people, like speaking, eating, or writing?" If the question was answered affirmatively, the additional 13 questions were assessed. The results showed that with the additional 13 questions, the sensitivity was extremely high at 100%, specificity moderate at 66.4%, positive predictive power at 58.2 % and negative predictive power at 100%. With the social fears questions, the screener produced a great ability to correctly identify all those who met criteria for SAD along with being able to moderately identify all those who do not meet criteria for SAD.

A more ambitious study (Morse & Pilkonis, 2007) examined three different screeners for PD's including the Inventory of personal problems personality disorder scale (IIP-PD; Horowitz, Rosenberg, Ureno & Villaseno, 1988), the Self-Directedness Scale: Temperament and Character Inventory (TCI-SD; Cloninger et al., 1994) and the Iowa Personality Disorder Screen (IPDS; Langbehn et al., 1999). The sample consisted of psychiatric, and non-psychiatric and community populations. After screeners were completed, the participants completed a semi-structured interview, the Interpersonal Relations Assessment (IRA; Heape, Pilkonis, Lambert & Proietti, 1989) and a structured interview, the Structured Interview for DSM-IV personality (SID IV; Pfohl, Blum, & Zimmerman, 1997). The results of the conditional probabilities for the three screening measures showed that in the psychiatric population, the screeners were moderately able to correctly identify all those who met criteria for a PD on semi-structured and structured interviews. In the non-psychiatric sample, the screeners had a more varied mix, showing lower levels of sensitivity and specificity. The goal of the study was to see if using more than one screener would increase validity. However, it was found that no one screener was superior to any other.

Along with the various measures mentioned previously, there is a screener to accompany the International Personality Disorder Examination (IPDE), the IPDE-S. In an initial validation study, Lenzenweger and colleagues, (Lenzenweger et al., 1997) used a two-stage application process to determine the efficacy of the IPDE-S and the IPDE interview. The researchers used the IPDE-S to determine how well it could be used as a screener for PDs on the IPDE interview. Their results found that the screener had high sensitivity at .81, moderate specificity at .61, extremely high negative predictive power at .98 and low positive predictive power at .21. The Lenzenweger study represents an initial first step in validating the IPDE-S; however because the sample in the study consisted of a group of homogenous upper class elite students, little is known about the predictive validity of the measure in a more racially, ethnically and economically diverse samples.

In the present study, I examined the predictive validity of the IPDE screener in a

diverse, working-class sample of students using the four conditional probabilities, specificity, sensitivity and positive and negative predictive power. Specificity is the probability that, given an absence of a PD on the interview, the threshold for a PD was not met on the screener. Sensitivity is the probability that, given the presence of a positive diagnosis of a PD on the interview, the threshold was met for a probable PD on the screener. Positive predictive power is the probability of receiving a probable diagnosis of a PD on the interview, given that there was a positive diagnosis of a PD on the screener. Negative predictive power is the probability of not receiving a PD diagnosis on the interview given that there was a negative diagnosis of a PD on the screener. Examining the predictive validity of the IPDE-S on a racially, ethnically and economically diverse sample allows for the testing of the generalizability of the Lenzenweger finding and provides further validity for the IPDE-S.

Method

Participants

Eleven hundred and fourteen undergraduates at a large urban Northeastern University completed the International Personality Disorder Examination Screener (IPDE-S) as part of a larger study (Posner et al., 2002). A subset of 66 of these individuals completed the interview for the International Personality Disorder Examination (IPDE). Seven hundred and twenty two (71.1%) of participants were women, 349 (34.3%) were Caucasian, 206 (20.3%) were of African American descent, 211 (20.8%) were of Asian descent, and 250 (24.6%) were Latino/a. Five hundred and twenty-seven (51.9%) students were employed. Complete Demographic characteristics for the subgroups can be seen in table 1.

Measures

International Personality Disorder Examination Screener **(**IPDE-S) (Loranger, 1991).

The screener for the IPDE-S is a 77 item True/False paper and pencil measure designed to assess for the presence of pathological personality traits. The IPDE-S screens for the ten DSM-IV personality disorders; cluster A (paranoid, schizoid, schizotypal), cluster B (antisocial, borderline, histrionic and narcissism), and cluster C (avoidant, dependent, obsessive-compulsive). The screener focuses on six different areas of personality and behavior. These areas are work, self, interpersonal relations, affects, reality testing, and impulse control. The questions on the screener are scored based on the sum of endorsed items.

International Personality Disorder Examination (Loranger, 1994).

International Personality Disorder Examination (IPDE, Loranger, Sartorius, Andreoli, & Berger, 1994). The IPDE is a semi-structured diagnostic interview for diagnosing personality disorders. It consists of 99 items arranged in six categories (e.g., Self or Work), along with a detailed scoring manual (Loranger et al., 1994). Each item assesses part or all of a *DSM-IV* personality-disorder criterion and is rated on a three-

point scale: 0 = *absent or normal*, 1 = *exaggerated or accentuated*, 2 = *meets criteria or pathological*. Items consist of one or several primary questions and follow-up questions. All positive responses are followed by requests for examples. After the provided questions are exhausted, the clinical interviewer is free to ask additional questions until he or she is able to score the item. The IPDE generates probable (subthreshold number of *DSM-IV* criteria met) and definite diagnoses for each of the *DSM-IV* diagnoses. It also generates dimensional scores for each diagnosis by adding the ratings on all the criteria composing a diagnosis.

Data Analysis

Chi-square analyses were used to compare those who were above threshold for a PD on the IPDE-S and those meeting criteria for a PD on the IPDE. Based on the chi-square analyses the conditional probabilities of positive predictive power, negative predictive power, sensitivity and specificity were calculated.

Results

Diagnosis

As shown in table 2, nine (13.6%) of the 66 interviewed participants met criteria for either a probable or definite diagnosis of a personality disorder on the IPDE interview. These nine individuals had a total of 18 PD's. Six participants met criteria for a definite diagnosis, with those PD's being- borderline, histrionic, dependent and PDNOS. Additionally, there were 7 participants that met criteria for a probable diagnosis; with those PD's being paranoid, borderline, histrionic, narcissistic, and PDNOS. Due to comorbidity of PD's, participants often had two or three diagnoses, resulting in various combinations of definite and probable scores.

Conditional Probabilities

As shown in table 3, four sets of conditional probabilities-positive predictive power, negative predictive power, sensitivity and specificity were calculated for the diagnosis of a PD.

Sensitivity was extremely high (100%) with all nine participants who met criteria for either a probable or definite PD also having met threshold for a PD on the IPDE screener. However, specificity was extremely low (5.9%). Similarly, negative predictive power was extremely high (100%) while positive predictive power was low (15.8%). These probabilities were assessed based on the IPDE-S manual suggested cutoff score of 3 (Loranger, 1991). Cutoff scores were later raised, to see if the probabilities would improve. A cutoff score of 4 also yielded extremely high sensitivity (100%), but improved the specificity slightly (17.6%). Negative predictive power (100%) and positive predictive power (17.6%) were similar to a cutoff score of 3. A cutoff score of 5 also yielded extremely high sensitivity (100%), and a moderate level of specificity (37.3%). Negative predictive power was extremely high (100%) and positive predictive power was low (22.0%), but slightly improved. A cutoff score of 6 yielded moderate sensitivity (55.6%), but much improved specificity (64.7%). Negative predictive power remained high (89.2%) but positive predictive power was still low (21.7%). The Kappa statistics for the cutoff scores were .018 for a cutoff score of 3, .060 for a cutoff score of 4, .151 for a cutoff score of 5 and .124 for a cutoff score of 6.

Discussion

The IPDE-S has shown adequate predictive validity in a homogenous, primarily Caucasian economically advantaged and elite student sample. However, replications of that finding are needed and questions remain regarding the generalizability of previous studies to more diverse groups. Therefore, the IPDE-S was examined to test its predictive validity in a diverse sample.

The IPDE-S showed excellent sensitivity with cutoff scored of 3, 4 and 5 correctly identifying all those who were diagnosed with a PD based on the IPDE interview. However, its specificity was very low, (especially for cutoff scores of 3, 4 and 5) and thus misidentified large numbers of individuals who did not meet criteria for a PD based on the IPDE interview. Although specificity was better with a 6 criteria cutoff, it was not adequate and additionally the sensitivity decreased to an inadequate level. A cutoff score of 5 however, yielded the best overall results. Likewise, the strongest kappa value is for a cutoff score of 5 at .151, indicating the screener and the interview matched best at this score.

Results from similar studies (Lenzenweger et al., 1997) show that efforts to reduce false- positive rates by increasing the threshold resulted in higher false-negative rates. The findings of the current study suggest that a higher cutoff score than that indicated in the IPDE-S manual might be preferable. A cutoff score of 5 might be useful in a two-stage process to identify those likely to have a PD, but it is unlikely that any cutoff score is sufficient for clinical use. Other studies have found weak conditional probabilities with other PD screening measures (Morse et al., 2007).

Strengths of the current study include a diverse urban sample. Previous studies, (Lenzenweger, et al., 1997; Lenzenweger 1999; Bragge et al., 2004) were conducted using homogenous samples. One important limitation is the relatively low number of individuals meeting criteria for a PD, thus these results need to be replicated in a larger sample.

References

Bagge, C., Nickell, A., Stepp, S., Durrett, C., Jackson, K., & Trull, T. (2004). Borderline Personality Disorder Features Predict Negative Outcomes 2 Years Later. *Journal of Abnormal Psychology, 113*, 279-288.

Cloninger,C.R., Przybeck, T., Svrakic, D.M., &Wetzel, R. (1994). The Temperament and Character inventory (TCI): A guide to its development and use. Unpublished manuscript, St. Louis, MO.

First, M.B., Spitzer, R.L., Williams, J.B.W., & Gibbon, M. (1997). Structured Clinical Interview for DSM-IV (SCID). Washington (DC). American Psychiatric Assoication.

Dalrymple, K.L., & Zimmerman, M. (2008). Screening for social fears and social anxiety disorder in psychiatric outpatients. *Comprehensive Psychiatry, 49*, 399-406.
Grant, B. F., Hasin, D.S., Stinson, F.S., Dawson, D.A., Chou, S.P., Ruan, W.J., et al. (2004). Prevalence, Correlates, and Disability of Personality Disorders in the United

States: Results from the National Epidemiological Survey on Alcohol and Related Conditions. *Journal of Clinical Psychiatry, 65,* 948-958.

Heape, C.L., Pilkonis, P.A., Lambert, J., & Prioetti, J.M. (1989). Interpersonal relations Assessment. Unpublished manuscript, Department of Psychiatry, University of Pittsburg, PA.

Horowitz, L.M., Rosenberg, S.E., Ureno, G., & Villaseno, V.S. (1988). Inventory of interpersonal problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology, 56,*885-892.

Langbehn, D.R., Pfohl, B.M., Reynolds, S., Clark, L.A., Battaglia, M., Bellodi, L., et al. (1999). The Iowa personality disorder screen: Development and preliminary validation of a brief screening interview. *Journal of Personality Disorders, 13*, 75-89

Lenzenweger, M.F. (1999). Stability and Change in Personality Disorder Features. *Archives of General Psychiatry, 56,* 1009-1015.

Lenzenweger, M.F., Loranger, A.W., Korfine, L., & Neff, C. (1997). Detecting Personality Disorders in a Nonclinical Population. *Archives of General Psychiatry, 54*, 345-351.

Loranger, A.W., Sartorius, N., Andreoli, A., Berger, P., Buchheim, P., Channabasavanna, S.M., et al. (1994). The International Personality Disorder Examination. *Archives of General Psychiatry, 51*, 215-224.

Morse, J.Q., Pilkonis, P.A. (2007). Screening for Personality Disorders. *Journal of Personality Disorders, 21,* 179-198.

Pfohl, B., Blum, N., & Zimmerman, M. (1997). Structured Interview for DSM-IV personality (SIDP-IV). Washington, DC: American Psychiatric Press, Inc.

Skodol, A.E., Siever, L.J., Livesley, W.J., Gunderson, J.G., Pfohl, B., & Widiger, T.A. (2002). The Borderline Diagnosis II: Biology, Genetics, and Clinical Course. *Society of Biological Psychiatry, 51,* 951-963.

Zanarini, M.C., Vujanovic, A.A., Parachini, E.A., Boulanger, J.L., Frankenburg, F.R., Hennen, J. (2003). A Screening Measure for BPD: The McClean Screening Instrument for Borderline Personality Disorder (MSI-BPD). Journal of Personality Disorders, 17, 568-573.

Table 1.

Participant Demographic Characteristics

	Screened Participants (N=1014)	Interviewed (N=66)	Diagnosed with PD (N=9)
Mean Age (SD)	20.03 (4.00)	20.8 (4.89)	19.22 (1.20)
Female (%)	722 (71.1)	54 (81.8)	6 (66.6)
Ethnicity (%)			
Caucasian	349 (34.3)	31 (46.9)	4 (44.4)
African Descent	206 (20.3)	8 (12.1)	1 (11.1)
Asian	211 (20.8)	9 (13.6)	1 (11.1)
Latino/a	250 (24.6)	11 (16.6)	2 (22.2)
Other		1 (1.5)	1 (11.1)
Employed (%)			
Yes	527 (51.9)	29 (43.9)	5 (55.5)
Education Level of Father (%)			
Less than Jr. High School	47 (4.6)	4 (6.3)	1 (11.1)
Junior High School	55 (5.4)	1 (1.6)	
Partial High School	96 (9.4)	4 (6.3)	1 (11.1)
High School Graduate	236 (23.2)	20 (31.7)	4 (44.4)
GED	27 (2.7)	1 (1.6)	
Partial 2 yr. college	49 (4.8)		
Partial 4 yr. college	56 (5.5)		
Technical School	57 (5.6)	4 (6.3)	
Associate Degree	35 (3.4)	1 (1.6)	
Standard College (BA, BS, AB)	173 (17.0)	8 (12.7)	3 (33.3)
Masters Level (MA, MS, MSW, MPH)	78 (7.7)	9 (14.3)	

Doctoral Degree (PhD, MD, JD)	40 (3.9)	2 (3.2)	
Education Level of Mother (%)			
Less than Jr. High School	68 (6.7)	5 (7.9)	
Junior High School	50 (4.9)	1 (1.6)	
Partial High School	91 (9.0)	6 (9.5)	
High School Graduate	261 (25.7)	16 (25.4)	3 (33.3)
GED	25 (2.5)	2 (3.2)	2 (22.2)
Partial 2-yr college	72 (7.1)	1 (1.6)	
Partial 4 yr. college	59 (5.8)	2 (3.2)	1 (11.1)
Technical School	32 (3.1)	1 (1.6)	
Associate Degree	48 (4.7)	4 (6.3)	
Standard College (BA, BS, AB)	165 (16.2)	8 (12.7)	2 (22.2)
Masters Level (MA, MS, MBA MPH)	95 (9.4)	8 (12.7)	1 (11.1)
Doctoral Level (PhD, MD, JD)	16 (1.6)	1 (1.6)	

Note: Data on gender were missing for 117. Data on employment were missing for 198. Data on education were missing for 132 participants. Of the 66 interview participants, data were missing for age on 6 participants, for ethnicity on 6 participants, for employment on 12 participants and for education of parents on 12 participants.

Table 2

Prevalence of a Personality Disorder on the IPDE.

Interviewed Subgroup N=66	Definite N (%)	Probable N (%)	Total N (%)
Paranoid	0 (0.0)	2 (28.5)	2 (33.3)
Schizoid	0 (0.0)	0 (0.0)	0 (0.0)
Schizotypal	0 (0.0)	0 (0.0)	0 (0.0)
Antisocial	0 (0.0)	0 (0.0)	0 (0.0)
Borderline	1 (16.6)	3 (42.8)	4 (44.4)
Histrionic	2 (33.3)	3 (42.8)	5 (55.5)
Narcissistic	0 (0.0)	1 (14.2)	1 (11.1)
Dependent	2 (33.3)	0 (0.0)	2 (22.2)
Ob-Compulsive	0 (0.0)	0 (0.0)	0 (0.0)
PDNOS	3 (2.0)	1 (14.2)	4 (44.4)
AnyPD	6 (9.1)	7 (10.6)	9 (13.6)

Note: Due to comorbidity, the number of PD's is greater than the number of subjects.

Table 3

Conditional Probabilities for the IPDE-S Cutoff Scores.

	3 Criterion	4 Criterion	5 Criterion	6 Criterion
PPP	15.8%	17.6%	22.0%	21.7%
NPP	100%	100%	100%	89.2%
Sensitivity	100%	100%	100%	55.6%
Specificity	5.0%	17.6%	37.3%	64.7%
True Pos.	9	7.6	9	5
False Pos.	48	43.4	32	18
True Neg.	3	7.6	19	33
False Neg.	0	1.4	0	4
Kappa	.018	.060	.151	.124