Journal of Personality Disorders, 15(4), 351-357, 2001

DIAGNOSTIC EFFICIENCY OF THE IOWA PERSONALITY DISORDER SCREEN ITEMS IN A NONCLINICAL SAMPLE

Timothy J. Trull, PhD, and Marissa Amdur, BA

The diagnostic efficiency of the 11-item Iowa Personality Disorder Screen (IPDS; Langbehn et al., 1999) was evaluated in a nonclinical sample of young adults, 35% of whom met DSM-III-R criteria for a personality disorder, in a retrospective analysis of SIDP-R data. Results indicated that two IPDS item sets (i.e., combinations of items) produced hit rates of more than 80% along with good sensitivity, specificity, positive predictive power, and negative predictive power. Combined with the findings of Langbehn et al. (1999), these results suggest that the IPDS may be useful as a screening measure for personality disorder in both clinical and nonclinical populations.

The assessment and diagnosis of personality disorders has been of great interest to researchers and clinicians in the fields of psychiatry and clinical psychology. With hope of finding reliable and valid methods of diagnosing individuals with personality disorders, a number of diagnostic interviews have been developed. Currently, most consider structured and semistructured interviews the best means by which to establish personality disorder diagnosis (Langbehn et al., 1999). Clinicians use structured diagnostic interviews to ensure that every patient is asked the same questions in the same order and to allow for additional probes to clarify answers. Furthermore, structured diagnostic interviews tend to demonstrate higher reliability than other types of diagnostic instruments (e.g., unstructured clinical interviews). Structured interviews that are currently used to diagnose personality disorders include, for example, the Structured Interview for DSM-IV Personality Disorders (SIDP-IV; Pfohl, Blum, & Zimmerman, 1995), the Structured Clinical Interview for DSM Personality Disorders (SCID-II; First, Spitzer, Gibbon, & Williams, 1995), the Personality Disorder Examination (Loranger, 1988), the Diagnostic Interview for Personality Disorders (Zanarini, Frankenburg, Chauncey, & Gunderson, 1987), and the Personality Disorder Interview-IV (Widiger, Mangine, Corbitt, Ellis, & Thomas, 1995).

Although the use of Axis II structured interviews provides a useful and reliable method of diagnosis, there are drawbacks. One of the major problems

From the Department of Psychology, University of Missouri.

This research was supported by National Institute of Mental Health grants (R55 MH52695 and R01 MH52695) awarded to the first author.

Address correspondence to Tim Trull, 210 McAlester Hall, Department of Psychology, University of Missouri, Columbia, MO 65211; E-mail: TrullT@missouri.edu.

is the length and cost of these interviews. Structured diagnostic interviews are detailed in their analysis of personality disorders. This makes for not only a lengthy interviewing process but it also requires a specially trained individual to administer such a test (Langbehn et al., 1999). These two characteristics make the cost of administration high. Therefore, a brief screening measure that accurately identified those who are likely to have personality disorders would be desirable. Such a screening measure would help the researcher or clinician to quickly ascertain those who should be assessed further with a full structured Axis II interview. In addition, an Axis II screening measure would be useful to those who study personality disorder in community or nonclinical populations where the base rates for personality disorders are lower (i.e., compared with clinical populations). Instead of administering an entire structured interview to everyone in a community sample, it would be much less costly and much more efficient to identify those who may have personality disorders for more extensive assessment.

Toward this goal, several researchers have developed personality disorder screening measures, typically paper-and-pencil inventories (for a review see Langbehn et al., 1999). For example, Pilkonis, Kim, Proietti, and Barkham (1996) developed a screening measure using items from the Inventory of Interpersonal Problems (IIP) (Horowitz, Rosenberg, Baer, Ureno, & Villasenor, 1988). Because interpersonal problems characterize the personality disorders, Pilkonis et al. (1996) hypothesized that the IIP would be an ideal instrument for screening personality disorders. In a sample of 145 inpatients and outpatients, 79% of whom were diagnosed with a personality disorder, Pilkonis et al. (1996) developed five screening subscales derived from the IIP items (IIP-PD scales). Overall, Pilkonis et al. (1996) found that the IIP-PD performed well in identifying individuals with any personality disorder.

Although several self-report Axis II screening measures are available, some clinicians and researchers may prefer to use interview-based screening questions for personality disorder. A relatively new screening measure, based on interview items, is the Iowa Personality Disorder Screen (IPDS) (Langbehn et al., 1999). Briefly, this 11-item instrument was initially developed using existing Axis II interview data (primarily from the SIDP-R; Pfohl, Blum, Zimmerman, & Stangl, 1989) on clinical patients from a variety of research centers (n = 1,203). Selection of items and subsets of items for inclusion in the IPDS was based on several criteria, including (a) ability to empirically discriminate between those with and those without a personality disorder diagnosis (overall, and for Cluster B and Clusters A and C, respectively); and (b) ease of administration and scoring with reference to the target's long-standing pattern of beliefs and behavior (Langbehn et al., 1999). To cross-validate the screening measure, the IPDS was administered to a second sample of psychiatric inpatients and outpatients (validation sample; n = 52; M age = 39.6 years, 73% women). Diagnostic judgments based on responses to the 11-item IPDS were compared with those derived from the full SIDP-IV (Pfohl et al., 1995). The sensitivity and specificity were determined for each individual item, in addition to various item sets (i.e., subsets of the items). Langbehn et al. (1999) presented preliminary evidence suggesting that several IPDS item sets may serve as a good general screen for personality disorder in clinical samples.

Although results of this initial IPDS study show promise for this particular screening measure, further research on the diagnostic efficiency and utility of the IPDS was suggested (Langbehn et al., 1999). For example, it is important to evaluate the diagnostic efficiency of the IPDS in samples that differ from the original validation sample in terms of base rate of personality disorder and of demographic features (e.g., age, clinical status, gender composition). With this in mind, the main objective of the current study was to evaluate further the ability of the IPDS to identify those who meet diagnostic criteria for a personality disorder. Toward this end, we conducted a retrospective analysis of the diagnostic efficiency of IPDS items using data from the full SIDP-R. Instead of focusing on the identification of individuals with personality disorders in a clinical sample (e.g., psychiatric inpatients and outpatients) as did Langbehn et al. (1999), however, the focus of this study was to determine the ability of the IPDS to successfully screen individuals in a nonclinical sample. Such a study would provide preliminary data on the utility of the IPDS for Axis II research in nonclinical populations and in populations with a lower base rate of personality disorder. Similar to Langbehn et al. (1999), we evaluated the sensitivity and specificity for each of the 11 items and for selected item sets (i.e., those evaluated by Langbehn et al., 1999). In addition, we also calculated the hit rate (percentage of correct classifications), the positive predictive power (PPP), and the negative predictive power(NPP) for IPDS items and item sets.

METHOD

PARTICIPANTS

The data used in evaluating the Iowa Personality Disorder Screen (IPDS) were previously collected in a study described in Trull (1995). Briefly, Trull (1995) conducted a study of features of borderline personality disorder (BPD) in a nonclinical, young adult sample. Approximately 1,700 college students completed a self-report measure of borderline personality disorder features, the Personality Assessment Inventory-Borderline features scale (PAI-BOR; Morey, 1991). Based on their scores on this inventory, potential study participants were divided into those scoring above threshold (≥ 70 *T*, raw score ≥ 38) and below threshold, respectively. These cut offs (based on community norms) are used to identify individuals with some degree of borderline features but not necessarily a BPD diagnosis (Morey, 1991). This ensured some degree of personality pathology in the final sample, consistent with the goals of the original study. Trull (1995) randomly selected individuals from the above- and the below-threshold lists to contact regarding participation in a three-hour laboratory study that involved the completion of several structured diagnostic interviews and self-report inventories. Those who agreed to participate first completed the PAI-BOR a second time to identify those who retained the same threshold classification (above vs. below) at retest and to eliminate those whose scores appeared state-like (vs. trait-like). A total of 103 participants received the same classification at the mass testing and at the laboratory session and subsequently completed the

interview/laboratory phase of the study. Of the 103 participants, 54 participants scored above threshold and 49 participants scored below threshold on the PAI-BOR. As for the demographic features, the total sample included 53 females (51.5%) and 50 males, with a mean age of 19 years (SD = 1.36). Most of the participants were Caucasian (88.3%) and single (98.1%); several participants reported previous psychiatric hospitalization (4.9%) or previous outpatient treatment (28.2%).

PROCEDURE

Each participant completed the SIDP-R (Pfohl, Blum, Zimmerman, & Stangl, 1989), a structured interview used to diagnose DSM-III-R personality disorders. Four researchers (two women and two men) served as SIDP-R interviewers for this study. Before the study, each interviewer received extensive training in the administration of the SIDP-R and each interviewer reliably scored 10 previously conducted SIDP-R interviews by independently reviewing audiotapes. Trull (1995) presents more detail regarding interviewer training and the reliability of the SIDP-R ratings for this study. In this particular sample, 35% of the participants met diagnostic criteria for at least one DSM-III-R personality disorder. The most prevalent DSM-III-R personality disorder diagnoses in this sample were (a) passive-aggressive (11.7%); (b) histrionic (9.7%); (c) borderline (6.8%); (d) obsessive-compulsive (6.8%); (e) narcissistic (5.8%), and (f) paranoid (5.8%). To evaluate the diagnostic efficiency of the IPDS items and item sets, SIDP-R items and DSM-III-R criteria corresponding to the IPDS items were identified (and aggregated in the case of the IPDS item sets). Diagnostic efficiency was calculated using the presence of any SIDP-R personality disorder diagnosis as the criterion. Specifically, we calculated the hit rate, sensitivity, specificity, PPP, and NPP for each IPDS item and each IPDS item set.

RESULTS

The results of this study are presented in Tables 1 and 2. Table 1 lists each individual item of the IPDS and the corresponding DSM-III-R criteria along with its hit rate, sensitivity, specificity, PPP, and NPP in this sample. For comparison purposes, the sensitivity and specificity rates reported for Langbehn et al.'s (1999) validation sample are also presented. For each IPDS item, the sensitivity was relatively low (range = 16.7 to 50.0), indicating that the majority of those with a DSM-III-R personality disorder did not exhibit each individual symptom. Conversely, the specificity was much higher and eight of the 11 items had specificity rates of 90% or greater. Except for IPDS item 5 (excessive social anxiety), this general pattern of lower sensitivity and higher specificity for IPDS items was also reported in Langbehn et al.'s (1999) validation sample. In Table 1, PPP refers to the probability that an individual had a personality disorder diagnosis given the endorsement of a particular IPDS item. NPP is the probability of no personality disorder diagnosis given that the IPDS item was not present. In general, the values for these two diagnostic efficiency statistics were acceptably high, suggesting that many individual IPDS items can reasonably identify

TABLE 1. Individual Items of the Iowa Personality Disorder Screen

Item	Hit Rate	Sensitivity	Specificity	PPP	NPP
1. Marked shifts in moods (BRD3)	66.0	50.0 [54.2]	74.6 [75.0]	51.4	73.5
2. Uncomfortable when not the center of attention (HST5)	67.0	25.0 [12.5]	89.6 [92.9]	56.3	69.0
3. Actions directed toward immediate satisfaction (HST7)	72.8	30.6 [45.8]	95.5 [78.6]	78.6	71.9
4. Reluctant to confide in others (PAR5)	69.9	16.7 [54.2]	98.5 [82.1]	85.7	68.8
5. Excessive social anxiety (AVD2 AVD4)	71.8	22.2 [79.2]	98.5 [67.9]	88.9	70.2
6. Unwilling to get involved unless certain of being liked (AVD3)	73.8	27.8 [66.7]	98.5 [89.3]	90.9	71.7
7. Lack of stable image (BRD6)	70.9	36.1 [45.8]	89.6 [92.9]	65.0	72.3
8. Prone to overemphasize importance (NAR3 NAR4)	69.9	36.1 [54.2]	88.1 [89.3]	61.9	72.0
9. Expects to be exploited or harmed by others (PAR1)	69.9	16.7 [8.3]	98.5 [92.9]	85.7	68.8
10. Bears grudges or is unforgiving of insults (PAR4)	66.0	38.9 [29.2]	80.6 [85.7]	51.9	71.1
11. Insensitive to concerns of others (NAR8)	69.9	16.7 [25.0]	98.5 [82.1]	85.7	68.8

Note. Hit rate = total number of correct classifications/*n*; PPP = positive predictive power; NPP = negative predictive power. Corresponding personality disorder criteria are indicated in parentheses after each item (e.g. , BRD3 = third criterion of borderline personality disorder). Numbers in brackets indicate Sensitivity and Specificity rates , respectively , reported in the validation sample in Langbehn et aL. (1999).

those with a personality disorder diagnosis. However, even in these cases, the overall correct classification rate (hit rate) did not greatly exceed the value of the hit rate that would be expected by chance alone (i.e., 65%; the hit rate if one predicted that no one in the sample had a personality disorder diagnosis).

Table 2 presents the diagnostic efficiency of selected subsets of the IPDS items as proposed by Langbehn et al. (1999). Specifically, Langbehn et al. (1999) evaluated the diagnostic efficiency of three sets of IPDS items using different cut off points (e.g., ≥ 2 items rated as present, ≥ 3 items rated as present). Similar to the results for individual IPDS items, Table 2 indicates that sensitivity of IPDS item sets was somewhat low, regardless of the cut off. Again, this indicates that there are a number of individuals who received a personality disorder diagnosis based on a full Axis II interview but who were not identified as such by IPDS item sets. However, the PPP results are more promising. Here, the IPDS item sets were highly accurate in their predictions of who meets diagnostic criteria for a personality disorder. Finally, NPP results suggest that the IPDS item sets do an adequate job of eliminating a personality disorder diagnosis.

DISCUSSION

The results of this study indicate that the IPDS works reasonably well in predicting which individuals in a nonclinical sample were judged to have a personality disorder based on a full Axis II interview. Although individual

TABLE 2. Diagnostic Efficiency of IPDS Subscales.

Item Set	Hit Rate	Sensitivity	Specificity	PPP	NPP
Cutoff = 2					
1-6	81.6	52.8	97.0	90.5	79.3
4-8	76.7	38.9	97.0	87.5	74.7
1, 3-8	83.5	69.4	91.0	80.6	84.7
Cutoff = 3					
1-6	72.8	22.2	1.00	1.00	70.5
4-8	70.9	16.7	1.00	1.00	69.1
1, 3-8	76.7	36.1	98.5	92.9	74.2

Note. Hit Rate = Total number of correct classifications/*n*; PPP= positive predictive power; NPP= negative predictive power.

IPDS items were not particularly sensitive to a personality disorder diagnosis and produced a range of PPP and NPP rates, the diagnostic efficiency of selected sets of IPDS items was, in general, better. First, regardless of the item set or the cut off, the hit rates exceeded those predicted by base rate in this sample alone (i.e., 65%). Second, using a cut off score of 2, IPDS item sets 1–6, and 1, 3–8 exhibited adequate sensitivity, and good specificity, PPP, and NPP. Taken together, these results suggest that administering either of these sets of IPDS items with a cut off of 2 may work well for many purposes.

The utility of a screening measure ultimately depends on the stated purpose for the screening process. If one's purpose is to identify all of those who have a personality disorder diagnosis, and one accepts the possibility that there may be a number of false positives, then screening measures with high sensitivity may be preferred. Conversely, one might be concerned with the cost of administering a full diagnostic interview (especially to those who will not meet criteria for a personality disorder diagnosis) and therefore prefer to emphasize the accuracy of positive diagnostic predictions based on screening scores. In this case, screening measures with high PPP may be preferred.

In the present study, which sampled nonclinical participants, the IPDS item sets were modestly sensitive to a personality disorder diagnosis but were quite accurate in predicting which participants received a personality disorder diagnosis (i.e., produced high PPP values). These results suggest that the IPDS item sets might be useful to investigators who wish to minimize administration costs while maximizing predictive accuracy. Overall, based on the results from this particular nonclinical sample, it appears that the IPDS item set that includes items 1 and items 3 to 8 maximizes the hit rate of classification and provides the best balance of sensitivity, specificity, PPP, and NPP rates when using a cut off score of 2.

There are a several limitations that deserve comment. One limitation concerns the sample. Although it was a nonclinical sample, 35% of our sample met criteria for at least one personality disorder. This base rate is higher than that typically encountered in a nonclinical sample. Because base rates will affect diagnostic efficiency statistics, investigators should not assume

that our results will replicate in other nonclinical samples. Furthermore, our sample consisted of college students who were relatively young. This type of nonclinical sample may be less pathological than other types of samples (e.g., adults ages 25 to 30, clinical samples). Therefore, we encourage other researchers to assess the diagnostic efficiency of IPDS items in other populations and in nonclinical samples with lower base rates of personality disorder to evaluate fully the utility of the IPDS. Finally, it is important to note that the IPDS items were not administered separately from the full Axis II interview to the participants. Additional studies that adopt this research design (i.e., using the IPDS as a stand-alone measure) are needed because this is ultimately how an Axis II screening measure like the IPDS would be used.

REFERENCES

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1995). The structured clinical interview for DSM-III-R personality disorders (SCID-II). Part I: Description. *Journal of Personality Disorders*, *92*, 83-91.

Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureno, G., & Villasenor, 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., Cadoret, R., Grove, W., Pilkonis, P., & Links, P. (1999). The Iowa Personality Disorder Screen: Development and preliminary evaluation of a brief screening interview. *Journal of Personality Disorders*, *13*, 75-89.

Loranger, A. E. (1988). *Personality Disorder Examination (PDE) manual*. White Plains, NY: Author.

Morey, L. C. (1991). *Personality Assessment Inventory: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Pfohl, B. M., Blum, N., & Zimmerman, M. (1995). *Structured Interview for DSM-IV Personality: SIDP-IV*. Iowa City, IA: University of Iowa.

Pfohl, B. M., Blum, N., Zimmerman, M., & Stangl, D. (1989). *Structured interview for DSM-III-R Personality: SIDP-R*. Iowa City, IA: Author.

Pilkonis, P. A., Kim, Y., Proietta, J. M., & Barkham, M. (1996). Scales for personality disorders developed from the Inventory of Interpersonal Problems. *Journal of Personality Disorders*, *10*, 355-369.

Trull, T. J. (1995). Borderline personality disorder features in nonclinical young adults: 1. Identification and validation. *Psychological Assessment*, *7*, 33-41.

Widiger, T. A., Mangine, S., Corbitt, E. M., Ellis, C. G., & Thomas, G. V. (1995). *Personality Disorder Interview-IV: A semistructured interview for the assessment of personality disorders*. Odessa, FL: Psychological Assessment Resources.

Zanarini, M. C., Frankenburg, F.R., Chauncey, D. L., & Gunderson, J. G. (1987). The Diagnostic Interview for Personality Disorders: Interrater and test-retest reliability. *Comprehensive Psychiatry*, *28*, 467-480.

This article has been cited by:

1. S.G. Selvik, B. Hummelen, U. Romild, D.R. Langbehn, G. Pedersen. 2018. The Iowa Personality Disorder Screen: A validation study in a psychiatric population that receives long-term group psychotherapy for personality related problems. *Personality and Mental Health* **12**:3, 229-240. [Crossref]
2. Andrea Fossati, Antonella Somma, Serena Borroni, Cesare Maffei, Kristian E. Markon, Robert F. Krueger. 2016. A Head-to-Head Comparison of the Personality Inventory for DSM-5 (PID-5) With the Personality Diagnostic Questionnaire-4 (PDQ-4) in Predicting the General Level of Personality Pathology Among Community Dwelling Subjects. *Journal of Personality Disorders* **30**:1, 82-94. [Abstract] [PDF] [PDF with links]
3. Mark Beitel, Skye Peters, Jonathan D. Savant, Christopher J. Cutter, John J. Cecero, Declan T. Barry. 2015. The Psychometric Properties of the Iowa Personality Disorder Screen in Methadone-Maintained Patients: An Initial Investigation. *Journal of Personality Disorders* **29**:1, 131-144. [Abstract] [PDF] [PDF with links]
4. Adrian Furnham, Rebecca Milner, Reece Akhtar, Filip De Fruyt. 2014. A Review of the Measures Designed to Assess DSM-5 Personality Disorders. *Psychology* **05**:14, 1646-1686. [Crossref]
5. Jo Ingul, Hans M Nordahl. 2013. Anxiety as a risk factor for school absenteeism: what differentiates anxious school attenders from non-attenders?. *Annals of General Psychiatry* **12**:1, 25. [Crossref]
6. Ingrid Olssøn, Alv A. Dahl. 2012. Avoidant personality problems—their association with somatic and mental health, lifestyle, and social network. A community-based study. *Comprehensive Psychiatry* **53**:6, 813-821. [Crossref]
7. Jo Magne Ingul, Christian A. Klöckner, Wendy K. Silverman, Hans M. Nordahl. 2012. Adolescent school absenteeism: modelling social and individual risk factors. *Child and Adolescent Mental Health* **17**:2, 93-100. [Crossref]
8. Ingrid Olssøn, Øystein Sørebø, Alv A Dahl. 2011. A cross-sectional testing of The Iowa Personality Disorder Screen in a psychiatric outpatient setting. *BMC Psychiatry* **11**:1. . [Crossref]
9. Christopher J. Miller, Sheri L. Johnson, Thomas R. Kwapil, Charles S. Carver. 2011. Three studies on self-report scales to detect bipolar disorder. *Journal of Affective Disorders* **128**:3, 199-210. [Crossref]
10. Alv A. Dahl, Christian Falk Dahl. 2010. Are there gender differences in impairment associated with high social anxiety? A community-based study. *Journal of Anxiety Disorders* **24**:5, 487-493. [Crossref]
11. Christian A. Falk Dahl, Alv A. Dahl. 2010. Lifestyle and social network in individuals with high level of social phobia/anxiety symptoms: a community-based study. *Social Psychiatry and Psychiatric Epidemiology* **45**:3, 309-317. [Crossref]
12. Sara Germans, Guus L. Van Heck, Douglas R. Langbehn, Paul P.G. Hodiamont. 2010. The Iowa Personality Disorder Screen. *European Journal of Psychological Assessment* **26**:1, 11-18. [Crossref]
13. I. Olssøn, A.A. Dahl. 2009. Personality problems are considerably associated with somatic morbidity and health care utilisation. *European Psychiatry* **24**:7, 442-449. [Crossref]
14. Caleb J. Siefert. Screening for Personality Disorders in Psychiatric Settings: Four Recently Developed Screening Measures 125-144. [Crossref]
15. Terry J. Lewin, Tim Slade, Gavin Andrews, Vaughan J. Carr, Charles W. Hornabrook. 2005. Assessing personality disorders in a national mental health survey. *Social Psychiatry and Psychiatric Epidemiology* **40**:2, 87-98. [Crossref]