



Published in final edited form as:

J Pers Disord. 2007 April ; 21(2): 179–198. doi:10.1521/pedi.2007.21.2.179.

SCREENING FOR PERSONALITY DISORDERS

Jennifer Q. Morse, PhD and Paul A. Pilkonis, PhD

From the Department of Psychiatry, Western Psychiatric Institute and Clinic, University of Pittsburgh Medical Center.

Abstract

A brief but valid self-report measure to screen for personality disorders (PDs) would be a valuable tool in making decisions about further assessment and in planning optimal treatments. In psychiatric and nonpsychiatric samples, we compared the validity of three screening measures: the PD scales from the Inventory of Interpersonal Problems, a self-report version of the Iowa Personality Disorder Screen, and the self-directedness scale of the Temperament and Character Inventory. Despite their different theoretical origins, the screeners were highly correlated in a range from .71 to .77. As a result, the use of multiple screeners was not a significant improvement over any individual screener, and no single screener stood out as clearly superior to the others. Each performed modestly in predicting the presence of any PD diagnosis in both the psychiatric and nonpsychiatric groups. Performance was best when predicting a more severe PD diagnosis in the psychiatric sample. The results also highlight the potential value of multiple assessments when relying on self-reports.

Personality disorders (PDs) are often diagnosed in patients with Axis I psychiatric disorders (Dunayevich et al., 2000; Leibbrand, Hiller, & Fichter, 1999; O'Brien & Vincent, 2003; Starcevic, Bogojevic, Marinkovic, & Kelin, 1999; Verheul, 2001). Comorbid PDs complicate first-line treatments for these Axis I disorders (Bank & Silk, 2001; Bell, 2002; Corruble, Ginestet, & Guelfi, 1996; Dunayevich et al., 2000; Grant et al., 2004; Leibbrand et al., 1999; Mulder, 2002; Reich, 2003; Rosenvinge, Martinussen, & Ostensen, 2000; Thase, 1996; Verheul, 2001) as well as the treatment of medical illnesses (Adler, 1990; Gross et al., 2002; Hueston, Werth, & Mainous, 1999; Moran, Rendu, Jenkins, Tylee, & Mann, 2001). In addition, providing care for patients with PDs imposes an interpersonal burden upon clinicians that requires an enhanced set of therapeutic skills (cf. Clarkin, Yeomans, & Kernberg, 1999; Linehan, 1993).

This recognition of the prognostic and clinical importance of PDs creates a need for improved methods of case identification. There is considerable debate, however, about the choice and feasibility of methods for diagnosing PDs, especially since *best practices*, usually involving interviews and case formulation, are often too labor intensive for routine clinical purposes and even some research agendas. It would be valuable to have a first-stage screening measure for PDs to inform decisions about investing further resources for assessment, treatment planning, and therapy. Therefore, the goal of the work reported here was to examine the utility of three brief, self-report screening instruments for PDs in both psychiatric and nonpsychiatric samples.

PREVIOUS EFFORTS TO SCREEN FOR PERSONALITY DISORDERS

INTERVIEW METHODS

Although we used self-reports in the present investigation to maximize efficiency, it should be noted that short interviews have been developed for screening purposes, largely by extracting a small subset of items from longer, structured diagnostic interviews. These interviews include the Iowa Personality Disorder Screen (IPDS, with 11 diagnostic criteria), the Standardised Assessment of Personality-Abbreviated Scale (SAPAS, with 8 criteria), and a group of 15 DSM criteria described by Nurnberg et al. (2000). In psychiatric samples across 4 studies (Langbehn et al., 1999; Moran et al., 2003; Nurnberg et al., 2000; Walters, Moran, Choudhury, Lee, & Mann, 2004), the brief interviews demonstrated sensitivities ranging from .69 to .96 and specificities ranging from .50 to .91.

SELF-REPORT METHODS

A number of self-report measures have been developed to identify patients with PDs. Such measures, however, have not usually been intended as screening tools. Therefore, they range from 140 to 395 items and ask in an inclusive way about Axis II diagnostic features. Such instruments include the Schedule for Nonadaptive and Adaptive Personality (SNAP; Clark, 1993), the Dimensional Assessment of Personality Pathology (DAPP; Livesley, Jackson, & Schroeder, 1991; Pukrop, Gentil, Steinbring, & Steinmeyer, 2001), the Temperament and Character Inventory (TCI; Cloninger, Przybeck, Svrakic, & Wetzell, 1994), the Personality Disorder Questionnaire (PDQ-4; Hyler, 1994), the Wisconsin Personality Disorders Inventory (WISPI; Klein et al., 1993), the International Personality Disorder Examination Screen (IPDES; Loranger et al., 1994), which has 250 true-false items despite its name, and the DSM-IV and ICD-10 Personality Questionnaire (DIP-Q; Ottosson et al., 1995). The length of these measures, however, is inconsistent with the purpose of finding a short but also valid instrument to use for initial screening.

In previous attempts to address the problem of screening for PDs, we developed a briefer self-report measure (with 47 items and 5 subscales) derived from the Inventory of Interpersonal Problems (Pilkonis, Kim, Proietti, & Barkham, 1996; Scarpa et al., 1999; Stern, Kim, Trull, Scarpa, & Pilkonis, 2000). Our conceptual framework was an interpersonal one, and the general hypothesis guiding the work was that one of the best markers of PDs is chronic difficulty in interpersonal relationships. In addition, we regard persistent interpersonal problems as a critical focus for clinical purposes because interpersonal difficulties play an important role in maintaining and, at times, exacerbating other symptomatology on both Axis I (e.g., depression, anxiety) and Axis II (e.g., anger, suicidality, impulsive behavior).

For the current study, our goal was not only to cross-validate our previous work with the IIP-PD scales in a psychiatric sample (where the scales had been developed) but also to examine their operating characteristics in samples where the prevalence of PDs is likely to be lower (e.g., samples of medical patients, samples from community sources). Research in lower base rate samples provides a stringent test of the generalizability of any screener: "The statistical problems that emerge when one moves from psychiatric to primary-care [i.e., lower base rate] populations almost all arise from the fact that the prevalence of many specific disorders is low or very low. . . . The essential point is that diagnostic methods that work well in settings with moderate to high base rates of disorder may not work well when imported to settings with low rates of psychiatric disorder" (Shrout, 1992, p. 294).

In addition, we wanted to examine the utility of the IIP-PD scales compared to other screening measures reflecting different theoretical backgrounds and to investigate whether the use of multiple measures increased validity. For this purpose, we chose one measure linked to the

DSM (since PDs are diagnosed categorically with the criteria of the DSM) and one measure linked to general models of personality (given the widespread interest in examining relationships between general models of personality and specific expressions of psychopathology). In the former case, we selected the Iowa Personality Disorder Screen (which we translated into a self-report format of 19 items), and in the latter case, the 44-item self-directedness scale from the Temperament and Character Inventory (TCI-SD). Cloninger and colleagues (1994; Cloninger, Svrakic, & Przybeck, 1993) have developed a *psychobiological* model of temperament and personality and an instrument, the TCI, that includes four scales for temperament (novelty seeking, harm avoidance, reward dependence, and persistence) and three scales for character (cooperativeness, self-directedness, and self-transcendence). Svrakic, Whitehead, Przybeck, & Cloninger (1993) reported that scores from the character scales differentiated patients with any versus no PD, whereas scores from the temperament scales differentiated among types of PD (e.g., clusters A, B, and C). In addition, they provided raw scores on the self-directedness scale for screening thresholds in both community and clinical samples.

We compared the performance of these measures with a design involving two stages. First, we examined their characteristics in a large stage-1 screening sample ($N = 624$) of psychiatric patients and nonpsychiatric participants, who included medical patients with diabetes and people in the university community receiving no treatment of any kind. The correlations between the measures and the estimates of the prevalence of PDs using previously established thresholds were of particular interest in the initial sample. Second, we examined the screening properties of the measures (sensitivity, specificity, efficiency, positive and negative predictive value, and kappa coefficients) using receiver operating characteristic (ROC) analyses in a stage-2 sample ($N = 151$) where we established the presence of PDs with a structured interview and a best-estimate diagnosis. These analyses allowed us to determine if the screening properties reported previously for the measures were supported by the results from this new sample.

METHOD

SAMPLE

The sample was collected in two stages (see Figure 1). The stage-1 screening sample consisted of 624 participants (253 psychiatric patients, 110 diabetic patients, and 261 persons from the university community) who completed the three self-report screening measures. The use of a diverse initial screening sample allowed us to examine the performance of the screening measures in subsamples where the base rate of PDs was likely to vary. The stage-2 interview sample consisted of 70 psychiatric patients and 81 nonpsychiatric subjects (23 diabetic patients and 58 subjects from the university community) who had been stratified on the basis of their screening self-reports and were then randomly selected to participate. The stratified random sampling was intended to produce a prevalence of 50% for diagnosable PDs in both the psychiatric and nonpsychiatric groups, which would provide an optimal test of the operating characteristics of the screeners. It was also intended to promote objectivity among the clinical evaluators conducting stage-2 assessments since the probability of any participant having a PD was assumed to be equal across the subsamples. Clinical evaluators were not blind to subsample, but they were blind to participants' scores on the screeners. Figure 1 demonstrates that approximately half of those interviewed in both the psychiatric and nonpsychiatric samples had positive screens.

Stratification was done using all three screening measures. For every seven participants, five were stratified on the basis of their IIP-PD scale scores, one based on the self-directedness scale of the TCI, and one based on the Iowa PD Screen. The goal of stratifying with all the different screening measures was to avoid relying exclusively on one theoretical perspective

on PDs for this purpose. Following stratification, participants were randomly selected to undergo an intensive face-to-face assessment at intake and at follow-up (six months later). Nonpsychiatric participants were more likely to refuse the invitation to complete the Stage 2 interview (43% versus 28%, $\chi^2_{(1)} = 5.37$, $p < .05$). This reflects differences in the expectations held by the two groups with respect to participation in psychiatric research. While psychiatric outpatients are accustomed to the demands of research assessments (lengthy interviews, self-report questionnaires, etc.), medical patients and university controls are less acclimated to these requests and were less willing to comply. The two-stage procedure also allowed us to examine the test-retest reliability of the IIP-PD and TCI-SD scales. On the IIP-PD scales, 63 psychiatric patients (90%) and 77 nonpsychiatric participants (95%) provided complete data at intake as well as screening; for the TCI-SD scale, the comparable numbers were 58 psychiatric patients (83%) and 70 nonpsychiatric participants (86%). The test-retest reliability of the IPDS was not examined since it was given as a self-report measure at screening and the full Structured Interview for DSM-IV Personality (SIDP-IV) (in which the items from the IPDS are embedded) was used as the structured Axis II interview at the intake assessment.

The final best-estimate PD diagnosis available for each participant was used as the “gold standard” in all analyses, with the large majority of these coming from a diagnostic conference after the 6-month follow-up evaluation. The best-estimate PD diagnoses were made blind to the results of the screening measures. Of the original 151 stage-2 subjects, 132 participated in the follow-up evaluations (with 6 people having partial evaluations, e.g., abbreviated interviews by telephone). Thus, the overall rate for completing the entire protocol was 87% (132/151), with no differences between the 2 subsamples (86% for psychiatric patients and 89% for the nonpsychiatric subjects).

Psychiatric Sample—Psychiatric outpatients from 21 to 60 years old were solicited from the general adult psychiatry clinic at Western Psychiatric Institute and Clinic (WPIC), Pittsburgh, PA. Patients with psychotic disorders, organic mental disorders, and mental retardation were excluded, as were patients with major medical illnesses that influence the central nervous system and might be associated with organic personality disturbance (e.g., Parkinson’s disease, cerebrovascular disease, and seizure disorders).

Nonpsychiatric Sample—The nonpsychiatric sample was comprised of diabetic patients and participants from the university community. Diabetic outpatients from 21 to 60 years old were solicited from the Center for Diabetes and Endocrinology at the University of Pittsburgh Medical Center, Pittsburgh, PA. Staff from the research team were introduced to potential participants while these patients waited for appointments with their physician. Diabetic patients currently receiving psychiatric treatment were excluded.

Comparison subjects from the university community were recruited from mailing lists of over 2,000 staff and faculty maintained by the University of Pittsburgh and the University of Pittsburgh Medical Center, Pittsburgh, PA. A letter informed recipients about the existence and purposes of the study and provided a prepaid postcard to return if they were willing to complete the three screening measures for a \$10 reimbursement. (Members of all three subsamples received the same \$10 payment for completing the screening questionnaires.) Fourteen percent of the faculty and staff receiving a letter returned the postcard. Of those, 88% returned the screening questionnaires; of those returning the questionnaires, 86% completed them satisfactorily and were not receiving ongoing medical or psychiatric care. Note that these respondents, who were a small fraction (11%) of the original mailings, were not intended to be a representative community sample but were intended only to serve as a comparison group not involved currently in either psychiatric or medical treatment.

Demographic characteristics of the participants in the two subsamples are summarized in Table 1. In the stage-1 sample ($N = 624$), the modal participant was a 38-year-old, white, married woman. However, differences between the subsamples did appear for minority and marital status. Psychiatric subjects were more likely to be members of minority groups and less likely to be married than participants in the nonpsychiatric sample. In the smaller stage-2 sample ($N = 151$), the subsamples differed only on marital status, with patients again less likely to be married. Among the 151 stage-2 participants, 98 (65%) nominated a significant other (SO) to serve as an informant; 64 of these SOs (65% of those eligible) completed an evaluation. Thus, 42% of the assessments (64/151) included collateral information from a spouse, relative, or good friend.

STAGE-1 SCREENING MEASURES

IIP Personality Disorder Scales (IIP-PD)—The original Inventory of Interpersonal Problems (IIP; Horowitz, Rosenberg, Ureño, & Villaseño, 1988) contains 127 items derived from content analysis of intake interviews with psychiatric outpatients. Items are rated from 0 *not at all* to 4 *extremely distressing*. Conventional scoring includes six subscales (problems with intimacy, assertiveness, sociability, submissiveness, interpersonal control, and excessive responsibility for others). The five personality disorder scales (Pilkonis et al., 1996) are scored from 47 of the original 127 items. These scales are *interpersonal sensitivity*, *interpersonal ambivalence*, *aggression*, *need for social approval*, and *lack of sociability*. Previous analyses indicated that the mean of the first three scales distinguished best between patients with any versus no PD. The final two scales were developed to distinguish cluster C PDs from other PDs but are not the focus here.

Self-Directedness Scale: Temperament and Character Inventory (TCI-SD)—The Temperament and Character Inventory (Cloninger et al., 1994) is a 226-item, true–false questionnaire that measures seven dimensions of personality. Novelty seeking, harm avoidance, reward dependence, and persistence are considered dimensions of temperament, whereas self-directedness, cooperativeness, and self-transcendence are considered dimensions of character. The current study included only the 44-item self-directedness scale. Unlike the IIP and IPDS, lower scores on the TCI-SD are associated with PD traits.

Iowa Personality Disorder Screen (IPDS)—The Iowa Personality Disorder Screen was developed as a brief screening interview (Langbehn et al., 1999). Its 11 DSM criteria were selected following analyses of a large sample of 1,203 interviews conducted at six research sites. Different subsets of 4 to 7 items were selected to distinguish patients with PDs from those without PDs and to attempt to distinguish patients with Cluster B PDs from those with other PDs. For this study, the 11 criteria from the screening interview were translated into 19 self-report items, from which each criterion could be scored.

STAGE-2 BEST-ESTIMATE DIAGNOSTIC PROCEDURES

Complete details of the assessment methodology are provided in previous reports (Pilkonis et al., 1995; Pilkonis, Heape, Ruddy, & Serrao, 1991; Pilkonis et al., 1996). A briefer summary is included here. Assessments at intake required at least three sessions with every participant, and each session lasted approximately two hours, for a total of six or more hours of face-to-face contact. Session 1 included the administration of the Structured Clinical Interview for DSM-IV (First, Gibbon, Spitzer, & Williams, 1997). No constraint other than the exclusion for psychosis was placed on Axis I diagnosis.

Table 2 summarizes the Axis I diagnoses present at intake in the stage-2 sample. As Table 2 illustrates, the most prevalent current diagnoses among psychiatric patients were affective disorders alone (major depressive, bipolar I and II, and dysthymic disorders, $N = 27$; 39%) and

comorbid affective and anxiety disorders ($N = 13$; 19%). The prevalences of any current diagnosis among diabetic patients and untreated community subjects were 39% and 36%, respectively (and 37% overall in the nonpsychiatric group).

In session 2, a detailed social and developmental history was taken, using a semi-structured interview, the Interpersonal Relations Assessment (IRA) developed for this purpose (Heape, Pilkonis, Lambert, & Proietti, 1989). A brief assessment of life events and strains during the past year was also done to elicit information about the social context in which the acute psychiatric problems had occurred. During session 3, the Structured Interview for DSM-IV Personality (SID IV; Pfohl, Blum, & Zimmerman, 1997) was administered. Following the subject assessment, the SIDP-IV was completed (in the third person) with an informant (when available) who knew the participant well and could report about his or her characteristic personality features.

Following the SO evaluation, the primary interviewer presented the case at a two-hour diagnostic conference with colleagues from the research team of interviewers and other judges. All available data (historical and concurrent) were reviewed and discussed at the conference, and each clinician voted independently about the presence or absence of a PD. Judges were given access to all data that had been collected (other than initial screening information): current and lifetime Axis I information, symptomatic status, social and developmental history, life events, personality features acknowledged on the Axis II interviews, and SO reports about personality and functioning. They were asked to integrate this information, with a particular emphasis on functional impairment (e.g., difficulties in major social roles, problems in interpersonal relationships) and longitudinal consistency (e.g., chronic difficulties not restricted to episodes of Axis I problems, markers of persistent social disability).

The most critical measure that emerged from the best-estimate consensus was the overall decision about the presence versus absence of a PD on clinical grounds, and it is this measure that served as the *gold standard* for PD status. Most decisions were unanimous (79% at intake, and 92% at 6 months), but in cases of disagreement, a majority rule was used to assign the clinical diagnosis. Because of the large prevalence of multiple PD diagnoses, many of which did not segregate neatly within the DSM clusters, a mutually exclusive decision was made on clinical grounds about the cluster to which each PD patient belonged: 1% were assigned to cluster A (odd, eccentric, schizotypal personalities), 50% to cluster B (dramatic, expressive, externalizing personalities), and 49% to cluster C (anxious, withdrawn, internalizing personalities). An overall rating of severity of Axis II pathology was also done by consensus on a 100-point scale (an analog to the Global Assessment Scale, Endicott, Spitzer, Fleiss, & Cohen, 1976) developed for this purpose. The entire procedure was repeated at the 6-month follow-up evaluation, except for the IRA (developmental history).

RESULTS

STAGE-1 SCREENING

Means and standard deviations in the stage-1 sample for the three individual screeners (the mean of the first three IIP-PD scales, the score from the TCI-SD scale, and the total number of criteria endorsed on the IPDS) are presented in Table 3. The two subsamples differed significantly on each of the three individual screeners (IIP-PD: $t_{(421.3)} = -13.87$, $p < 0.0001$; TCI-SD: $t_{(428.6)} = 16.56$, $p < 0.0001$; IPDS: $t_{(376.7)} = -14.21$, $p < 0.0001$). As expected, post hoc Tukey comparisons showed that the psychiatric sample always differed from the other two groups, which never differed between themselves.

Correlations Between the Screening Measures—The three screeners were highly correlated in a range (of absolute values) from .71 to .77 (IIP-PD with TCI-SD, $-.71$; IIP-PD

with IPDS, .77; and TCI-SD with IPDS, $-.74$). This finding documented excellent (even surprising) convergent validity among the measures. The large correlations between the individual screeners suggest a higher-order latent construct of psychological and social impairment that is tapped by each measure despite theoretical differences between them. The IIP-PD scales focus on distress associated with interpersonal problems (things that one “does too much” or finds “hard to do” with others); the TCI-SD scale focuses on themes of personal responsibility and purpose, internal locus of control, and behavioral regulation; and the DSM criteria included in the IPDS focus on themes of affect regulation and interpersonal anxiety. Despite these differences, the strong association among the screeners is probably best explained by their relationship to a hierarchical latent construct.

Prevalence of PDs Estimated by the Screeners Using Previously Established Thresholds—Each of the screening measures has thresholds reported previously to indicate a greater likelihood of PD diagnosis: a mean of the first three PD scales on the IIP of 1.1 or greater (Pilkonis et al., 1996); a self-directedness score of less than 20 on the TCI for definite PDs or a score less than 28 for both probable and definite PDs (Svrakic et al., 1993); or two positive criteria from the first six criteria on the IPDS (Langbehn et al., 1999). The prevalence of PDs estimated by each of the three screeners using these thresholds is presented in Table 4; note that the Langbehn et al. (1999) paper also reported on the operating characteristics of other subsets of criteria included in Table 4. The IIP-PD scales and the IPDS produced similar prevalence estimates: in the ranges of 65–74% for psychiatric patients and 17–27% for nonpsychiatric subjects. By contrast, the TCI-SD scale was more conservative.

STAGE-2 ASSESSMENT

Prevalence of Personality Disorders—The prevalence of PDs in the subsamples varied despite our attempts to stratify so that the rate of PDs would be 50% in both groups. Stratification was more successful in the nonpsychiatric groups than with psychiatric patients. According to the final best-estimate consensus, 59 of 70 psychiatric patients (84%) were diagnosed with PDs and 36 of 81 of the nonpsychiatric participants (44%). Table 5 summarizes the distribution of Axis II diagnoses according to the consensus judgments for the two groups.

Receiver Operating Characteristics (ROC) Analyses—For the purpose of screening, sensitivity is more critical than specificity. Therefore, presentation of the results from the ROC analyses focuses on the thresholds from the three screeners needed to identify participants with any PD in the stage-2 sample at sensitivities of .70, .80, and .90. Table 6 and Table 7 summarize the results for the psychiatric subsample, and Table 8 summarizes the results for the nonpsychiatric subsample. In addition to the usual operating characteristics from ROC analyses, we include two kappa coefficients reflecting the level of agreement between the screener and the best-estimate diagnostic consensus: an unweighted kappa that accounts for all errors (both false positives and false negatives) and a weighted kappa that focuses only on false negatives (i.e., poor sensitivity).

In the psychiatric subsample (see Table 6), the general results when screening for any PD were modest. The area under the ROC curve (AUC), which can be understood as the probability of correctly identifying a PD patient from a pair of randomly selected patients by using the screening test, is a useful summary measure. AUCs for the three screeners were .66 for the TCI-SD scale, .68 for the IIP-PD scales, and .73 for the IPDS. Only the last value was statistically significant (i.e., different from the chance level of .50), with $p < .02$. To achieve the sensitivities reported in Table 6, the thresholds for the IIP-PD scales and the TCI-SD scale were considerably less stringent than those used previously. This result occurred largely as a function of the high base rate (84%) of PDs in the psychiatric subsample. To capture such a large prevalence of positive cases, the threshold on any measure must be quite lenient. (Note

that no established thresholds exist for the sum of positive indicators from all 11 criteria from the IPDS; previous work has focused on subsets of the 11 criteria, but we preferred to focus on the entire pool of items to capitalize on all the information that could be gained from them.)

To mitigate this base-rate effect (and to provide a fairer test of the screeners in the psychiatric group), we did a median split of the psychiatric patients on our 100-point rating of Axis II severity. On this measure, a score of 61 was the threshold for a clinical diagnosis of any PD, and the range of 61–80 was anchored with the following descriptor: “Significant to marked personality pathology associated with chronic impairment and definite personality disorder.” Forty-seven percent of patients scored 70 or higher on the scale. (In the total stage-2 sample, an Axis II severity score of 70 identified the upper quartile of all 151 participants.) Table 7 summarizes the results when using the screeners to predict membership in the subgroup of psychiatric patients with more severe PDs. In this case, the screeners performed better, with AUCs of .70 for the IPDS, .73 for the IIP-PD scales, and .75 for the TCI-SD scale, with all these results significant at $p < .005$. The thresholds for the IIP-PD scales and the TCI-SD scale needed to achieve the sensitivities reported in Table 7 were also more consistent with previous reports, i.e., a range on the IIP-PD scales from 0.7 to 1.1 (Pilkonis et al., 1996) and a TCI-SD score of less than 28 (Svrakic et al., 1993).

In the nonpsychiatric subsample (see Table 8), a skewed base rate posed less of a problem since close to half (44%) of this subsample was diagnosed with a PD, the original intent of stratifying the stage-2 sample. Despite this advantage, the AUCs were .59 for the IPDS, .60 for the TCI-SD scale, and .62 for the IIP-PD scales, with none of these results statistically significant. To try to understand further the lack of effectiveness of the screeners in this context, we examined the test-retest reliability of the IIP-PD scales and the TCI-SD scale since these measures were administered a second time at the intake assessment (usually 3–4 weeks following the initial screening).

In psychiatric patients, trait (versus state) effects appeared to be more influential, that is, test-retest reliability was highest. The test-retest reliability of the three individual IIP-PD scales ranged from .67 to .88 in psychiatric subjects, with the mean of these measures (the screening index itself) having a test-retest reliability of .85 ($M = 1.28$ at screening, $SD = .73$, and 1.18 at interview, $SD = .72$). For the TCI-SD scale, the test-retest reliability was also .85 ($M = 25.7$ at screening, $SD = 9.2$, and 28.1 at interview, $SD = 9.2$). In nonpsychiatric participants, the reliability coefficients for the individual IIP-PD scales ranged from .63 to .72; for the mean, reliability was only .69 ($M = 1.13$ at screening, $SD = .51$, and $.79$ at interview, $SD = .47$). For the TCI-SD scale, the test-retest reliability was more acceptable at .76 ($M = 32.3$ at screening, $SD = 8.2$, and 34.0 at interview, $SD = 7.5$). The general point is that the test-retest reliability of the screeners was lower in the nonpsychiatric group, constraining their validity.

Given this information, we performed analyses to examine the impact of using data from multiple administrations of the screeners. In the first set of analyses, mean scores from the two administrations of the IIP-PD scales and the TCI-SD scale were used as the screening index, and these means did improve the AUCs in 4 of 6 ROC analyses (see Table 9). We also examined the usefulness of the second administration of the screeners in predicting the diagnosis of a PD. The AUCs were similar to those for the mean of two administrations. Since a mean of two assessments is more reliable than a single assessment and the pattern of results was similar, we have focused on the mean of the two assessments. In a second set of analyses, we examined results from repeated administration of the screeners in categorical fashion. That is, we identified participants who had 0, 1, or 2 positive screens, and we did this separately for the IIP-PD scales (using a threshold of 1.1) and the TCI-SD scale (using a score of 27 or less), which also improved the AUCs in the same 4 ROC analyses. In these analyses, two negative

screens performed well in predicting the absence of a PD diagnosis. These results were stronger in the psychiatric sample and when using the IIP-PD scales rather than the TCI-SD scale.

DISCUSSION

The first point to emphasize is the strong associations among the screeners, given their theoretical differences. We anticipated using the screeners in an incremental way to capitalize on the different kinds of information they appear to tap, but surprisingly, this turned out not to be relevant. We did perform several composite or multivariate analyses (e.g., by using the sum of positive screeners at initial screening as a total index, by performing logistic regression analyses to predict PD status with all the screeners as potential predictors, by using recursive partitioning techniques to identify potential interactions between the screeners to predict PD status). None of these analyses, however, showed that the use of multiple screeners was a significant improvement over any individual screener, and no single screener stood out as clearly superior to the others. These outcomes support the interpretation that all three screeners reflect the same latent construct of disordered personality functioning and that they do so to about the same degree.

The screeners were more effective in the psychiatric subsample, especially when predicting more severe PDs. The screeners have been applied primarily to clinical samples, and this is where their continued use is best justified. The results reflected some of the difficulties of exporting screening measures, developed in high base rate samples, to contexts with lower base rates. In such applications, not only may base rates differ, but the phenomenon under investigation may also differ qualitatively, cf. the controversy about whether the “nature” of depression—not only its severity—is different in primary care settings than in psychiatric settings (Barrett, Barrett, Oxman, & Gerber, 1988; Coulehan, Schulberg, Block, Janosky, & Arena, 1990; Coyne, Fechner-Bates, & Schwenk, 1994; Gerber, Barrett, Manheimer, Whiting, & Smith, 1989; Schwenk, Coyne, & Fechner-Bates, 1996). In the present case, for example, we found that the test-retest reliability of the screeners was lower in the nonpsychiatric subsample. It may be that diabetic and community subjects had “something on their minds” when they agreed initially to complete screening questionnaires but that their distress diminished—scores were more benign at stage-2 clinical assessment, especially with the IIP-PD scales, and diagnoses of actual PDs were less frequent and judged to be less severe when they did appear. As a practical consequence, it may be necessary to use multiple self-report assessments to screen adequately for personality disorders. Our analyses using both: (a) the mean of two administrations and (b) multiple outcomes scored in categorical fashion supported the value of repeated administrations, and this finding appeared to be true for both psychiatric and nonpsychiatric samples.

In addition to issues related to the performance of screening measures in psychiatric versus nonpsychiatric samples, another consideration is the difference between screening for principal versus comorbid diagnoses. Zimmerman and colleagues (Sheeran & Zimmerman, 2002; Zimmerman & Mattia, 2001; Zimmerman & Sheeran, 2003) suggested that a distinction be made between identifying a principal diagnosis (the chief complaint for which the patient seeks treatment) and additional diagnoses (comorbid disorders that are not the primary reason for treatment). They reported that identifying comorbid psychiatric disorders was more difficult for clinicians using unstructured interviews compared with structured diagnostic interviews (Zimmerman & Mattia, 1999a, 199b). They also reported patterns of lower AUC, sensitivity, and PPV for additional rather than primary psychiatric diagnoses when screening for comorbid PTSD or other psychiatric disorders using self-report screening measures (Sheeran & Zimmerman, 2002; Zimmerman & Sheeran, 2003). However, identifying comorbid psychiatric diagnoses is particularly important when a comorbid diagnosis jeopardizes the treatment of the primary disorder. This is certainly the case when screening for PDs: PDs are highly comorbid

with Axis I disorders, the latter are more likely to be the reason that patients seek treatment, and PDs typically interfere with the treatment of Axis I disorders.

However, conceptualizing trajectories of interpersonal and symptomatic complaints in terms of “presenting problems” suggests that one might ask participants about interpersonal problems more consistently and listen for persistent problems in the context of fluctuating symptomatic complaints. As a screening guideline, this pattern would require assessing both markers of PDs (using any of these three screeners) and a general symptom measure and repeating both. We would expect that persistent interpersonal and symptomatic complaints would mark a PD case, whereas persistent symptomatic complaints in the absence of chronic interpersonal problems would mark an Axis I psychiatric case without a PD diagnosis, and occasional interpersonal or symptomatic distress mark a nonpsychiatric, normal case. Unfortunately the current study provides information about Axis I symptoms (interview and self-report) only at the Stage 2 interview evaluation and hence cannot answer this question, which must be left for future research.

Though the performance of the screeners reported here was more modest than previous reports, the mean levels on the screeners were comparable. The current psychiatric sample had a mean IIP-PD score similar to a previous clinical sample, $M = 1.38$ (Pilkonis et al., 1996), and higher than the score for university students who met SIDP-R criteria for any PD, $M = 0.79$ (Stern et al., 2000). The current nonpsychiatric sample had scores similar to students not meeting SIDP-R criteria for any PD, $M = 0.67$ (Stern et al., 2000), and lower than another sample of university students who did not receive any other PD assessment, $M = 1.19$ (Scarpa et al., 1999). In addition, the psychiatric sample had a mean TCI self-directedness score similar to a previous sample of patients, $M = 26.4$ (Svarkic et al., 1993). The current nonpsychiatric sample had TCI self-directedness scores higher than scores for community adult respondents recruited at a shopping mall, $M = 30.7$ (Cloninger et al., 1993).

In this context, one can speculate about the high base rate of PDs in our psychiatric sample (84%) despite our attempt at stratification. Our previous work with psychiatric patients in the same setting found PD prevalence rates of 70 to 79% with no attempts to stratify (Pilkonis et al., 1995; Pilkonis et al., 1991; Pilkonis et al., 1996). Although high compared with some studies of psychiatric samples that had PD prevalence rates ranging from 43% to 55% (Keown, Holloway, & Kuipers, 2002; Moran et al., 2003; Nurnverg et al., 2000; Svarkic et al., 1993), this rate is understandable at our open-treatment tertiary care clinic, in part, because research treatment protocols focused on *cleaner* Axis I disorders divert cases with fewer Axis II features.

The current results should also be put into the broader context of assessment and treatment decisions in both primary care and psychiatric settings. Ideally a screening tool will efficiently identify cases from noncases in one administration. Unfortunately, this ideal may not be achieved for PDs, at least not for brief self-report measures. Aside from the comorbidity between PDs and Axis I disorders, the number of PD diagnoses represented in the DSM complicates the task of screening for the presence or absence of a PD. While the general description of PDs includes prominent interpersonal dysfunction, there are 12 different specific patterns for any single PD diagnosis (including the provisional PDs listed in DSM-III-R) as well as the possibility of a *mixed* presentation. Thus the challenge of screening for the presence or absence of any PD is rather more daunting than screening for the presence of a specific diagnosis. Nonetheless, the current results provide some confidence that someone who scores below the thresholds on any of these general screeners does *not* have a PD diagnosis; further assessment is not likely warranted unless first line interventions for any Axis I disorder fail. A positive screen on one of the three measures examined here could be followed by a self-report assessing PD traits more generally (e.g., the SNAP, WISPI, DAPP, PDQ4, IPDE-S, DIP-Q), a clinical interview, or one of the instruments that screens for a particular PD (McLean

Screening Instrument for BPD, MSI-BPD; Zanarini et al., 2003). In short, a more complicated screening process may be needed and a simple cutoff on a single screener may not be sufficient.

Acknowledgments

Data collection and manuscript preparation by Dr. Pilkonis was supported by NIMH grant R01 MH56888, Screening for Personality Disorders (PI: P. A. Pilkonis). Manuscript preparation by Dr. Morse was supported in part by NIMH grants T32 MH18269, Clinical Research Training for Psychologists (PI: P. A. Pilkonis) and T32 MH19986, Clinical Research Training in Late-Life Mood Disorders (PI: C. F. Reynolds, III). We thank the staff of the Personality Studies Program, especially Joseph Proietti and Summer Neal, for the execution of the project, and we thank David Klonsky for his comments on an earlier draft of this manuscript.

References

- Adler DA. Personality disorders: Treatment of the nonpsychotic chronic patient. *New Directions for Mental Health Services* 1990;47:3–15. [PubMed: 2277610]
- Bank PA, Silk KR. Axis I and Axis II interactions. *Current Opinion in Psychiatry* 2001;14:137–142.
- Barrett JE, Barrett JA, Oxman TE, Gerber PD. The prevalence of psychiatric disorders in a primary care practice. *Archives of General Psychiatry* 1988;45:1100–1106. [PubMed: 3264145]
- Bell L. Does concurrent psychopathology at presentation influence response to treatment for bulimia nervosa? *Eating & Weight Disorders* 2002;7:168–181. [PubMed: 12452248]
- Clark, LA. *Manual for the schedule for nonadaptive and adaptive personality (snap)*. Minneapolis: University of Minnesota Press; 1993.
- Clarkin, JF.; Yeomans, FE.; Kernberg, F. *Psychotherapy for borderline personality*. Hoboken, NJ: John Wiley & Sons, Inc; 1999.
- Cloninger, CR.; Przybeck, T.; Svrakic, DM.; Wetzel, R. *The temperament and character inventory (TCI): A guide to its development and use*. St. Louis, MO: 1994. Unpublished manuscript
- Cloninger CR, Svrakic DM, Przybeck TR. A psychobiological model of temperament and character. *Archives of General Psychiatry* 1993;50:975–990. [PubMed: 8250684]
- Corruble E, Ginestet D, Guelfi JD. Comorbidity of personality disorders and unipolar major depression: A review. *Journal of Affective Disorders* 1996;37:157–170. [PubMed: 8731079]
- Couleshan JL, Schulberg HC, Block MR, Janosky JE, Arena VC. Depressive symptomatology and medical co-morbidity in a primary care clinic. *International Journal of Psychiatry in Medicine* 1990;20:335–347. [PubMed: 2086521]
- Coyne JC, Fechner-Bates S, Schwenk TL. Prevalence, nature, and comorbidity of depressive disorders in primary care. *General Hospital Psychiatry* 1994;16:267–276. [PubMed: 7926703]
- Dunayevich E, Sax KW, Keck PE Jr, McElroy SL, Sorter MT, McConville BJ, et al. Twelve-month outcome in bipolar patients with and without personality disorders. *Journal of Clinical Psychiatry* 2000;61:134–139. [PubMed: 10732661]
- Endicott J, Spitzer RL, Fleiss JL, Cohen J. The global assessment scale. *Archives of General Psychiatry* 1976;33:766–771. [PubMed: 938196]
- First, MB.; Gibbon, M.; Spitzer, RL.; Williams, JBW. *Structured clinical interview for DSM-IV Axis I disorders (SCID-I)*. Washington, DC: American Psychiatric Press; 1997.
- Gerber PD, Barrett J, Manheimer E, Whiting R, Smith R. Recognition of depression by internists in primary care: A comparison of internist and “gold standard” psychiatric assessments. *Journal of General Internal Medicine* 1989;4:7–13. [PubMed: 2915277]
- Grant BF, Hasin DS, Stinson FS, Dawson DA, Chou SP, Ruan WJ, et al. Prevalence, correlates, and disability of personality disorders in the united states: Results from the national epidemiologic survey on alcohol and related conditions. *Journal of Clinical Psychiatry* 2004;65:948–958. [PubMed: 15291684]
- Gross R, Olfson M, Gameraff M, Shea S, Feder A, Fuentes M, et al. Borderline personality disorder in primary care. *Archives of Internal Medicine* 2002;162:53–60. [PubMed: 11784220]
- Heape, CL.; Pilkonis, PA.; Lambert, J.; Proietti, JM. *Interpersonal relations assessment*. Pittsburgh, PA: Department of Psychiatry, University of Pittsburgh; 1989. Unpublished manuscript

- Horowitz LM, Rosenberg SE, Ureño G, Villaseño VS. Inventory of interpersonal problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology* 1988;56:885–892. [PubMed: 3204198]
- Hueston WJ, Werth J, Mainous AGI. Personality disorder traits: Prevalence and effects on health status in primary care patients. *International Journal of Psychiatry in Medicine* 1999;29:63–74. [PubMed: 10376233]
- Hyler, SE. Personality diagnostic questionnaire for DSM-IV (PDQ-4). New York: New York State Psychiatric Institute;
- Keown P, Holloway F, Kuipers E. The prevalence of personality disorders, psychotic disorders and affective disorders amongst the patients seen by a community mental health team in London. *Social Psychiatry & Psychiatric Epidemiology* 2002;37:225–229. [PubMed: 12107714]
- Klein MH, Benjamin LS, Rosenfeld R, Treece C, Husted J, Greist JH. The Wisconsin personality disorders inventory: Development, reliability, and validity. *Journal of Personality Disorders* 1993;7:285–303.
- Langbehn DR, Pfohl BM, Reynolds S, Clark LA, Battaglia M, Bellodi L, et al. The Iowa personality disorder screen: Development and preliminary validation of a brief screening interview. *Journal of Personality Disorders* 1999;13:75–89. [PubMed: 10228929]
- Leibbrand R, Hiller W, Fichter M. Influence of personality disorders on therapy outcome in somatoform disorders at 2-year follow-up. *Journal of Nervous & Mental Disease* 1999;187:509–512. [PubMed: 10463069]
- Linehan, MM. Cognitive-behavioral treatment of borderline personality disorder. New York, NY: Guilford Press; 1993.
- Livesley WJ, Jackson DN, Schroeder ML. Dimensions of personality pathology. *Canadian Journal of Psychiatry—Revue Canadienne de Psychiatrie* 1991;36:557–562. [PubMed: 1742707]
- Loranger AW, Sartorius N, Andreoli A, Berger P, Buchheim P, Channaba-savanna SM, et al. The international personality disorder examination. *Archives of General Psychiatry* 1994;51:215–224. [PubMed: 8122958]
- Moran P, Leese M, Lee T, Walters P, Thornicroft G, Mann A. Standardised assessment of personality—abbreviated scale (SAPAS): Preliminary validation of a brief screen for personality disorder. *British Journal of Psychiatry* 2003;183:228–232. [PubMed: 12948996]
- Moran P, Rendu A, Jenkins R, Tylee A, Mann A. The impact of personality disorder in UK primary care: A 1-year follow-up of attenders. *Psychological Medicine* 2001;31:1447–1454. [PubMed: 11722159]
- Mulder RT. Personality pathology and treatment outcome in major depression: A review. *American Journal of Psychiatry* 2002;159:359–371. [PubMed: 11869996]
- Nurnberg H, Martin GA, Somoza E, Coccaro EF, Skodol AE, Oldham JM, et al. Identifying personality disorders: Towards the development of a clinical screening instrument. *Comprehensive Psychiatry* 2000;41:137–146. [PubMed: 10741893]
- O'Brien KM, Vincent NK. Psychiatric comorbidity in anorexia and bulimia nervosa: Nature, prevalence and causal relationships. *Clinical Psychology Review* 2003;23:57–74. [PubMed: 12559994]
- Ottosson H, Rodlund O, Ekselius L, von Knorring L, Kullgren G, Soderberg S. The DSM-IV and ICD-10 personality questionnaire (DIP-Q): Construction and preliminary validation. *Nordic Journal of Psychiatry* 1995;49:285–291.
- Pfohl, B.; Blum, N.; Zimmerman, M. Structured interview for DSM-IV personality (SIDP-IV). Washington, DC: American Psychiatric Press, Inc; 1997.
- Pilkonis PA, Heape CL, Proietti JM, Clark SW, McDavid JD, Pitts TE. The reliability and validity of two structured diagnostic interviews for personality disorders. *Archives of General Psychiatry* 1995;52:1025–1033. [PubMed: 7492254]
- Pilkonis PA, Heape CL, Ruddy J, Serrao P. Validity in the diagnosis of personality disorders: The use of the lead standard. *Psychological Assessment* 1991;3:46–54.
- Pilkonis PA, Kim Y, Proietti JM, Barkham M. Scales for personality disorders developed from the inventory of interpersonal problems. *Journal of Personality Disorders* 1996;10:355–369.
- Pukrop R, Gentil I, Steinbring I, Steinmeyer E. Factorial structure of the German version of the dimensional assessment of personality pathology-basic questionnaire in clinical and nonclinical samples. *Journal of Personality Disorders* 2001;15:450–456. [PubMed: 11723879]

- Reich JH. The effects of Axis II disorders on the outcome of treatment of anxiety and unipolar depressive disorders: A review. *Journal of Personality Disorders* 2003;17:387–405. [PubMed: 14632374]
- Rosenvinge JH, Martinussen M, Ostensen E. The comorbidity of eating disorders and personality disorders: A meta-analytic review of studies published between 1983 and 1998. *Eating And Weight Disorders: EWD* 2000;5:52–61. [PubMed: 10941603]
- Scarpa A, Luscher KA, Smalley KJ, Pilkonis PA, Kim Y, Williams WC. Screening for personality disorders in a nonclinical population. *Journal of Personality Disorders* 1999;13:345–360. [PubMed: 10633315]
- Schwenk TL, Coyne JC, Fechner-Bates S. Differences between detected and undetected patients in primary care and depressed psychiatric patients. *General Hospital Psychiatry* 1996;18:407–415. [PubMed: 8937906]
- Sheeran T, Zimmerman M. Screening for posttraumatic stress disorder in a general psychiatric outpatient setting. *Journal of Consulting & Clinical Psychology* 2002;70:961–966. [PubMed: 12182279]
- Shrout, PE. Identification of psychiatric cases in primary health care settings: The utility of two phase screening designs. In: Cooper, B.; Eastwood, MR., editors. *Primary health care and psychiatric epidemiology*. New York: Routledge; 1992. p. 293-306.
- Starcevic V, Bogojevic G, Marinkovic J, Kelin K. Axis I and Axis II comorbidity in panic/agoraphobic patients with and without suicidal ideation. *Psychiatry Research* 1999;88:153–161. [PubMed: 10622351]
- Stern BL, Kim Y, Trull TJ, Scarpa A, Pilkonis PA. Inventory of interpersonal problems personality disorder scales: Operating characteristics and confirmatory factor analysis in nonclinical samples. *Journal of Personality Assessment* 2000;74:459–471. [PubMed: 10900572]
- Svrakic DM, Whitehead C, Przybeck TR, Cloninger CR. Differential diagnosis of personality disorders by the seven-factor model of temperament and character. *Archives of General Psychiatry* 1993;50:991–999. [PubMed: 8250685]
- Thase ME. The role of Axis II comorbidity in the management of patients with treatment-resistant depression. *Psychiatric Clinics of North America* 1996;19:287–309. [PubMed: 8827191]
- Verheul R. Co-morbidity of personality disorders in individuals with substance use disorders. *European Psychiatry* 2001;16:274–282. [PubMed: 11514129]
- Walters P, Moran P, Choudhury P, Lee T, Mann A. Screening for personality disorder: A comparison of personality disorder assessment by patients and informants. *International Journal of Methods in Psychiatric Research* 2004;13:34–39. [PubMed: 15181485]
- Zanarini MC, Vujanovic AA, Parachini EA, Boulanger JL, Frankenburg FR, Hennen J. A screening measure for BPD: The McLean screening instrument for borderline personality disorder (MSI-BPD). *Journal of Personality Disorders* 2003;17:568–573. [PubMed: 14744082]
- Zimmerman M, Mattia JI. Is posttraumatic stress disorder under-diagnosed in routine clinical settings? *Journal of Nervous & Mental Disease* 1999a;187:420–428. [PubMed: 10426462]
- Zimmerman M, Mattia JI. Psychiatric diagnosis in clinical practice: Is comorbidity being missed? *Comprehensive Psychiatry* 1999b;40:182–191. [PubMed: 10360612]
- Zimmerman M, Mattia JI. A self-report scale to help make psychiatric diagnoses: The psychiatric diagnostic screening questionnaire. *Archives of General Psychiatry* 2001;58:787–794. [PubMed: 11483146]
- Zimmerman M, Sheeran T. Screening for principal versus comorbid conditions in psychiatric outpatients with the psychiatric diagnostic screening questionnaire. *Psychological Assessment* 2003;15:110–114. [PubMed: 12674730]

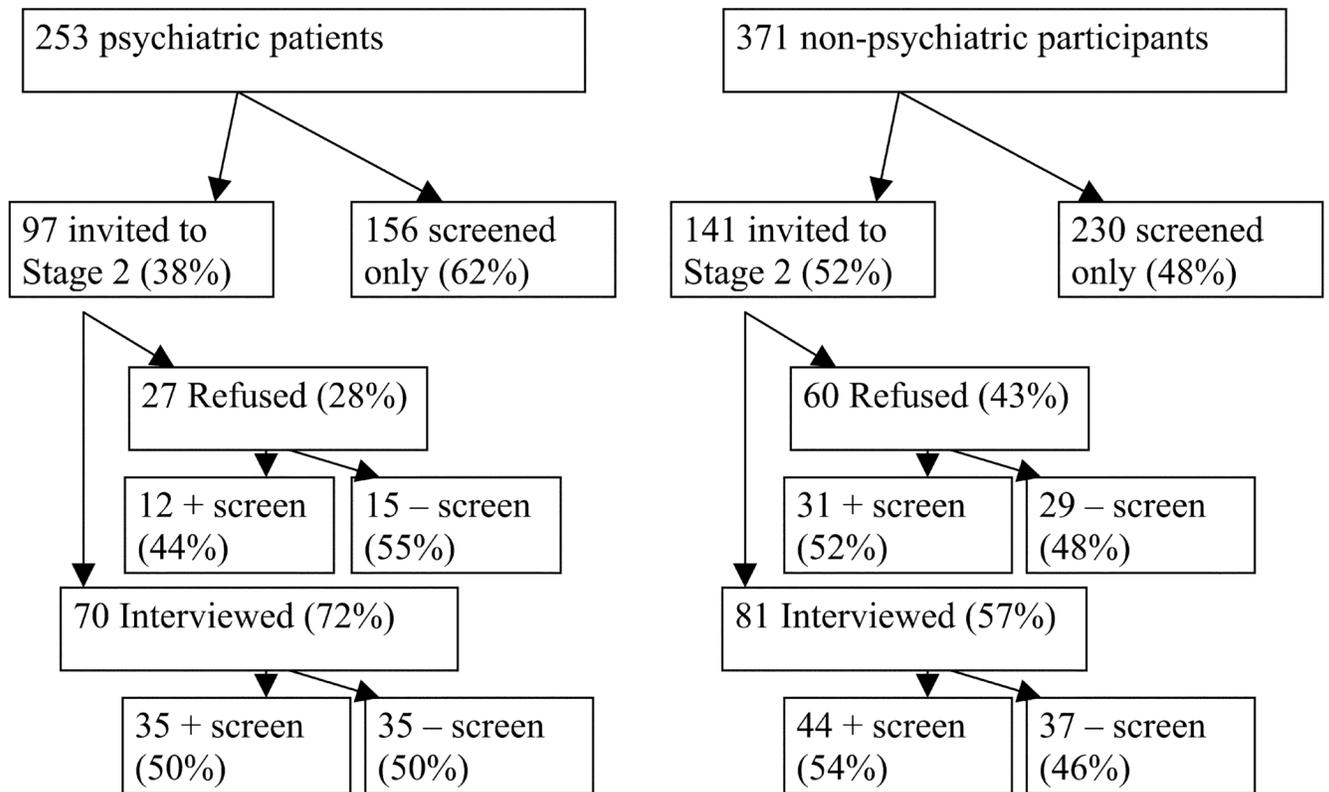


FIGURE 1.
Sampling from Stage 1 to Stage 2.

TABLE 1Demographic Characteristics of the Stage-1 Sample ($N = 624$)

	Psychiatric Patients ($n = 253$)	Nonpsychiatric Participants ($n = 371$)
Age (<i>sd</i>)	37.76 (10.86)	38.9 (10.67)
% Female (n)	69.2 (175)	67.7 (251)
Ethnicity (n) ¹		
% Caucasian	81.0 (204)	87.3 (324)
% African American	16.7 (42)	9.4 (35)
% Hispanic	0.4 (1)	0.5 (2)
% Asian	0.4 (1)	1.6 (6)
% Other	1.6 (4)	1.1 (4)
Marital Status (n)		
% Never married	43.1 (109)	30.2 (112)
% Married/long-term partner	31.6 (80)	58.8 (217)
% Separated/Divorced	23.3 (59)	10.2 (38)
% Widowed	2.0 (5)	1.1 (4)

¹ Ethnicity data was missing for one participant in the psychiatric patient sample.

TABLE 2Intake Axis I Diagnoses in the Stage-2 Sample ($N = 151$)

	Psychiatric Patients ($n = 70$)		Nonpsychiatric Participants ($n = 81$)	
	Frequency	Percent	Frequency	Percent
No current psychiatric disorders	4	5.7	51	63.0
Affective disorders only	27	38.6	4	4.9
Anxiety disorders only	5	7.1	7	8.6
Substance use disorders only	0	0.0	7	8.6
Affective and anxiety disorders	13	18.6	6	7.4
Affective and substance use disorders	7	10.0	0	0.0
Anxiety and substance use disorders	1	1.4	0	0.0
Affective, anxiety, and substance use disorders	2	2.9	0	0.0
Other disorders	11	15.7	6	7.4
Total	70	100.0	81	100.0

TABLE 3Means and Standard Deviations of Screening Scales in the Stage-1 Sample ($N = 624$)

Measure	Psychiatric Patients ($n = 253$)		Nonpsychiatric Participants ($n = 371$)	
	Mean	<i>SD</i>	Mean	<i>SD</i>
Mean of IIP PD1–3: Presence of Any PD	1.53	.72	0.80	.51
TCI: Self-Directedness	24.4	9.4	35.9	6.9
Iowa PD screen: Total of 11 Criteria	4.6	2.9	1.7	1.8

Note: The *ns* for each scale differ slightly due to differing numbers of valid cases for each measure.

TABLE 4Estimated Prevalence of Personality Disorders in the Stage-1 Sample ($N = 624$) Using Previous Thresholds

Measure	Psychiatric Patients ($n = 253$)	Nonpsychiatric Participants ($n = 371$)
	%	%
Mean of IIP PD1–3:		
GE 1.1	71%	26%
TCI Self-Directedness:		
LE 19 (Definite)	32%	4%
LE 27 (Definite/Probable)	61%	11%
Iowa PD screen:		
Sum of criteria 1–6 GE 2	75%	27%
Sum of criteria 4–8 GE 2	71%	23%
Sum of criteria 1, 3–8 GE 3	65%	17%

TABLE 5Consensus Axis II Diagnoses in the Stage-2 Sample ($N = 151$)

	Psychiatric Patients ($n = 70$)		Nonpsychiatric Participants ($n = 81$)	
	<i>N</i>	%	<i>N</i>	%
Any PD	59	84.3%	36	44.4%
Paranoid	10	14.3%	6	7.4%
Schizoid	1	1.4%	6	7.4%
Schizotypal	1	1.4%	1	1.2%
Histrionic	2	2.9%	1	1.2%
Narcissistic	16	22.9%	10	12.3%
Antisocial	6	8.6%	1	1.2%
Borderline	18	25.7%	5	6.2%
Avoidant	12	17.1%	8	9.9%
Dependent	5	7.1%	2	2.5%
Obsessive Compulsive	9	12.9%	8	9.9%
Passive Aggressive	24	34.3%	11	13.6%
Depressive	7	10.0%	7	8.6%
Mixed	10	14.3%	6	7.4%
None	11	15.7%	45	55.6%

Note: The percentages sum to more than 100% because of multiple diagnoses. Among participants with an Axis II diagnosis, the mean number of Axis II diagnoses was similar in the two subsamples (psychiatric sample: $mean = 2.05$, $sd = 1.31$; nonpsychiatric sample: $mean = 2.00$, $sd = 1.06$).

TABLE 6
 Operating Characteristics of the Screening Scales Predicting Any PD in the Stage-2 Psychiatric Subsample ($n = 70$)

Measure	P	Q	EFF	SENS	SPEC	PPV	NPV	Unweighted κ	Weighted κ
Mean of IIP PDI-3:									
GE 0.82	.84	.66	.70	.71	.64	.91	.29	.24	.16
GE 0.56	.84	.81	.73	.81	.27	.86	.21	.08	.07
GE 0.41	.84	.91	.81	.93	.18	.86	.33	.14	.21
TCI Self-Directedness:									
LE 31	.84	.71	.67	.73	.36	.86	.20	.07	.05
LE 32	.84	.80	.76	.83	.36	.88	.29	.18	.15
LE 38	.84	.90	.83	.93	.27	.87	.43	.24	.32
Iowa PD Screen:									
Sum of 11 criteria GE 3	.84	.61	.66	.66	.64	.91	.26	.19	.12
Sum of 11 criteria GE 2	.84	.74	.76	.80	.55	.90	.33	.27	.21
Sum of 11 criteria GE 1	.84	.90	.89	.97	.46	.90	.71	.49	.66

Note: P = prevalence, Q = level of the test, EFF = diagnostic efficiency, SENS = sensitivity, SPEC = specificity, PPV = positive predictive value, NPV = negative predictive value. Unweighted κ gives equal importance to false positives and false negatives; weighted κ focuses exclusively on false negatives (poor sensitivity). AUC: Mean of IIP PD 1-3: .68^f, TCI self-directedness: .66^f, Iowa PD screen .73*
^f $p \leq .10$, * $p \leq .05$

Values for EFF, SENSE, SPEC, PPV and NPV at previously established thresholds of the IIP and TCI can be obtained from the authors.

TABLE 7

Operating Characteristics of the Screening Scales Predicting Any Severe PD in the Stage-2 Psychiatric Subsample ($n = 70$)

Measure	P	Q	EFF	SENS	SPEC	PPV	NPV	Unweighted κ	Weighted κ
Mean of IIP PD1-3:									
GE 1.25	.47	.50	.71	.73	.70	.69	.74	.43	.46
GE 0.82	.47	.66	.64	.82	.49	.59	.75	.30	.47
GE 0.41	.47	.90	.49	.91	.11	.48	.57	.02	.09
TCI Self-Directedness:									
LE 27	.47	.53	.71	.76	.68	.68	.76	.43	.49
LE 28	.47	.60	.70	.82	.60	.64	.79	.41	.55
LE 33	.47	.83	.56	.91	.24	.52	.75	.15	.47
Iowa PD screen:									
Sum of 11 criteria GE 3	.47	.61	.63	.76	.51	.58	.70	.27	.37
Sum of 11 criteria GE 2	.47	.74	.59	.85	.35	.54	.72	.19	.41
Sum of 11 criteria GE 1	.47	.90	.57	1.0	.19	.52	1.0	.18	1.0

Note: P = prevalence, Q = level of the test, EFF = diagnostic efficiency, SENS = sensitivity, SPEC = specificity, PPV = positive predictive value, NPV = negative predictive value. Unweighted κ gives equal importance to false positives and false negatives; weighted κ focuses exclusively on false negatives (poor sensitivity). AUC: Mean of IIP PD 1-3: .73***, TCI self-directedness: .75****, Iowa PD screen .70**

** $p \leq .01$, *** $p \leq .001$, **** $p \leq .0001$

TABLE 8

Operating Characteristics of the Screening Scales Predicting Any PD in the Stage-2 Nonpsychiatric Subsample ($n = 81$)

Measure	P	Q	EFF	SENS	SPEC	PPV	NPV	Unweighted κ	Weighted κ
Mean of IIP PDI-3:									
GE 1.01	.44	.61	.59	.72	.49	.53	.69	.20	.30
GE 0.89	.44	.67	.61	.81	.44	.54	.74	.24	.42
GE 0.72	.44	.79	.58	.92	.31	.52	.82	.21	.60
TCI Self-Directedness:									
LE 35	.44	.61	.59	.72	.49	.53	.69	.20	.30
LE 38	.44	.77	.51	.81	.27	.47	.63	.07	.17
LE 42	.44	.95	.44	.94	.04	.44	.50	-.13	-.01
Iowa PD screen:									
Sum of 11 criteria GE 3	.44	.46	.59	.56	.62	.54	.64	.18	.18
Sum of 11 criteria GE 2	.44	.65	.59	.78	.44	.53	.71	.21	.36
Sum of 11 criteria GE 1	.44	.78	.52	.83	.27	.48	.67	.09	.25

Note: P = prevalence, Q = level of the test, EFF = diagnostic efficiency, SENS = sensitivity, SPEC = specificity, PPV = positive predictive value, NPV = negative predictive value. Unweighted κ gives equal importance to false positives and false negatives; weighted κ focuses exclusively on false negatives (poor sensitivity). AUC: Mean of IIP PD 1-3: .62^f, TCI self-directedness: .60, Iowa PD screen .59

^f $p \leq .10$

TABLE 9

Areas Under the Curve for Repeated Administration of Screeners

Measure	One screen	Mean of two screens	Total number of positive screens (0, 1, or 2)
Psychiatric Subsample			
Predicting Any PD			
Mean of IIP PD1-3	.68 ^t	.73 [*]	.72 [*]
TCI-SD	.66 ^t	.70 [*]	.71 [*]
Predicting Severe PD			
Mean of IIP PD1-3	.73 ^{***}	.70 ^{**}	.72 ^{**}
TCI-SD	.75 ^{****}	.77 ^{***}	.76 ^{***}
Nonpsychiatric Subsample			
Predicting Any PD			
Mean of IIP PD1-3	.62 ^t	.66 [*]	.66 [*]
TCI SD	.60	.56	.59

^t
 $p \leq .10,$ ^{*}
 $p \leq .05,$ ^{**}
 $p \leq .01,$ ^{***}
 $p \leq .001,$ ^{****}
 $p \leq .0001.$