# Results of the Search for Personality Disorder Screening Tools: Clinical Implications

## Sara Germans, MD, PhD; Guus L. Van Heck, PhD; and Paul P. G. Hodiamont, MD, PhD

**ABSTRACT**

**Objective:** To examine the characteristics, validity, posttest probabilities, and screening capabilities of 8 different instruments used to predict personality disorders.

**Method:** Screening instruments were examined in 3 prospective, observational, test-development studies in 3 random samples of Dutch psychiatric outpatients, using the Structured Clinical Interview for *DSM-IV* Axis II Disorders (SCID-II) as the "gold standard." The studies were performed from March 2004 to March 2005 (study 1: N = 195, mean age = 32.7 years), October 2006 to January 2007 (study 2: N = 79, mean age = 34.3 years), and January 2008 to October 2009 (study 3: N = 102, mean age = 33.7 years). The following 8 assessment instruments were examined: 3 short questionnaires (a self-report form of the Standardized Assessment of Personality-Abbreviated Scale [SAPAS-SR], the self-report Iowa Personality Disorder Screen [IPDS], and a short self-report version of the SCID-II [S-SCID-II]); 2 longer questionnaires (the self-report SCID-II Personality Questionnaire [SCID-II-PQ] and the NEO Five-Factor Inventory [NEO-FFI]); 1 short semistructured interview (the Quick Personality Assessment Schedule [PAS-Q]); and 2 informant-based interviews (the Standardized Assessment of Personality [SAP] and the Standardized Assessment of Personality-Abbreviated Scale for Informants [SAPAS-INF]).

**Results:** The SCID-II rate of identification of personality disorders in the 3 studies was between 48.1% and 64.1%. The SAPAS-SR, the IPDS, and the PAS-Q had the best sensitivity (83%, 77%, and 80%, respectively) and specificity (80%, 85%, and 82%, respectively). Moreover, these 3 instruments correctly classified the largest number of patients. Using the SAPAS-SR, the IPDS, or the PAS-Q raises the odds from 50% to between 80% and 84% that a patient in a psychiatric outpatient population will receive a personality disorder diagnosis.

**Conclusions:** The results provide evidence for the usefulness of the SAPAS-SR, IPDS, and PAS-Q instruments for personality disorder screening. Because the PAS-Q takes a longer time and requires qualified personnel to administer it, we recommend use of the SAPAS-SR or the self-report version of the IPDS.

In Western countries, the median prevalence for personality disorders is 13% for general populations, 50% for outpatient populations, and 70% for inpatient and forensic populations.[1,2] Early recognition of these frequently occurring personality disorders is extremely important as they cause serious psychosocial problems and can hinder the course and the treatment of psychiatric disorders.[3–7] Judging from these statistics, it would seem that personality disorders should be a frequent diagnosis in the daily praxis of psychiatric hospitals, in both inpatient and outpatient care. However, this appears not to be the case, and personality disorders are often underdiagnosed in the first consultation.[8] An important reason for this underdiagnosis might be the lack of 2 aspects of an adequate diagnostic procedure reflecting content and form. As to the content, doctors generally feel more at ease with the fluctuating state aspects of Axis I (*DSM-III* and *DSM-IV*) than with the more enduring aspects of Axis II. Regarding the form, because the diagnosis of personality disorder is based on the presence of long-existing characteristics, clinicians might be reluctant to diagnose personality disorder in the first encounter with a patient who is complaining about Axis I problems. Patients with personality disorder consume a lot of hospital staff time during office hours and beyond. If the personality disorder of a patient is not taken into account, then the overall treatment will probably stagnate or produce a reverse effect.[9,10] To attain the most efficient and adequate treatment, it is important that personality disorders are detected early.

Literature shows that the reliability of clinical assessment in determining psychiatric disorders, including personality disorders, has often been found to be rather dubious.[11–13] Attempts to identify this unreliability led to 4 sources of variance: (1) information variance, (2) observation variance, (3) interpretation variance, and (4) criterion variance.[14,15] Information variance can occur if different clinicians use different information sources about the patient or if the patient gives them varying information. Observation and interpretation variance implies that different clinicians who get the same information will remember or describe or weight the information differently and, therefore, interpret the information differently. Criterion variance occurs in those situations in which clinicians use different criteria for categories of psychopathological phenomena. The publication of *DSM-III* successfully cancelled out criterion variance. The introduction of standardized clinical-psychiatric interviews (along with training in these) led to substantially reduced variance in information, as well as in observation and interpretation. The disadvantage of standardized clinical-psychiatric interviews, however, is that they are often time-consuming and always have to be conducted by experienced, well-trained professionals.[16] A screening tool can be used to limit these disadvantages.

The screening principle means that people are subjected to a quick test in order to differentiate between likely cases and noncases. It should be kept in mind that screening tests have a global

diagnostic value. Specific diagnoses can be attained only by a much more far-reaching procedure, which of course takes more time and requires extensive expertise.

Therefore, a screening instrument can be useful in a 2-stage procedure for case identification. A highly sensitive screener in the first or case-finding phase will yield a maximum of potential cases to be confirmed or rejected by means of the diagnostic instrument in the second or case-identification phase. False positives in the first phase are not particularly a problem because they will be identified as noncases in the second phase. False negatives, however, are to be taken seriously, for no other reason than that they will have no diagnostic follow-up and will therefore be missed; a high negative predictive value is of paramount importance. Patients with a positive result on the screening scale should be interviewed subsequently with a detailed structured or semistructured interview aimed at the assessment of a specific personality disorder.

There are 2 kinds of screening instruments for personality disorders: short structured or semistructured interviews and questionnaires. Examples of structured interviews are the Standardized Assessment of Personality-Abbreviated Scale (SAPAS),[17] the Iowa Personality Disorder Screen (IPDS),[18] the Rapid Personality Assessment Schedule (PAS-R),[19] and the Quick Personality Assessment Schedule (PAS-Q).[20,21] These instruments employ the same source of information: the patient. Consequently, the quality of the data collected is very much dependent on the capability and willingness of the patient to provide a factual picture and a truthful report. Furthermore, it should be kept in mind that the reports might be colored by the psychiatric problems of the patients.[12] A solution could be found in employing a screening instrument that uses 1 or more informants as sources of information. Examples of such short informant-based interviews are the Standardized Assessment of Personality (SAP)[22,23] and the Standardized Assessment of Personality-Abbreviated Scale for Informants (SAPAS-INF).[22,23]

Questionnaires to be filled in by patients themselves obviously do not take much of the clinician's time. With respect to the reliability issue, the interviewer's observer and interpretation variance have been excluded; on the other hand, the respondent's interpretation variance plays a major role. To minimize the criterion variance, it is important that the questionnaires are based on a standardized diagnostic system, such as the *DSM-IV*. An example of a short questionnaire that can be filled in within 10 minutes is the self-report version of the IPDS.[24] A longer self-report questionnaire is the Structured Clinical Interview for *DSM-IV* Axis II Disorders (SCID-II) Personality Questionnaire (SCID-II-PQ),[25] which is based on a categorical system.[26–27] An example of a questionnaire based on a dimensional system is the NEO Five-Factor Inventory (NEO-FFI).[28,29]

In this article, we compare 8 different screening instruments, taking into account the different practical circumstances and the psychometric values. In addition, we discuss the clinical implications of the outcomes of these comparisons. Data were collected in 3 different studies.

## METHOD

### Participants

All 3 prospective, observational, test-development studies included psychiatric outpatients who were referred between 2004 and 2009 to GGZ Breburg, a community mental health center in Tilburg, The Netherlands. The studies were approved by the regional medical ethics committee.

The first study was performed between March 2004 and March 2005 and had 195 participants, 5.8% of whom dropped out. The second study took place from October 2006 to January 2007 and involved 79 participants, with a dropout percentage of 8.9%. The third study was carried out between January 2008 and October 2009 and had 102 participants, with a dropout percentage of 25.3%.

The distribution according to sex was 42.6% male and 57.4% female (study 1), 43.0% male and 57.0% female (study 2), and 40.2% male and 59.8% female (study 3). The mean age of the participants was 32.7 years (study 1), 34.3 years (study 2), and 33.7 years (study 3).

### Measures

Study 1 examined 3 short questionnaires (the self-report form of the SAPAS [SAPAS-SR],[30] the IPDS,[31] and a short self-report version of the SCID-II [S-SCID-II]),[32] as well as a longer questionnaire (the NEO-FFI) and a structured interview (the PAS-Q). Study 2 focused on a longer questionnaire, the SCID-II-PQ.[33] Study 3 employed 2 informant-based interviews: the SAP and the SAPAS-INF. Table 1 depicts the different characteristics of these screening instruments.

***SAPAS-SR.*** The SAPAS consists of 8 dichotomously rated items, which are taken from the opening section of an informant-based semistructured interview, the SAP.[22,23] Each item is scored as 0 (absent) or 1 (present), and the sum of these scores generates the overall score, ranging from 0 to 8. Moran et al[17] validated the SAPAS in a sample of

**Table 1. Practical Characteristics of the Different Personality Disorder Screening Instruments**

| Variable | SAPAS-SR | IPDS | S-SCID-II | SCID-II-PQ | PAS-Q | NEO-FFI | SAP | SAPAS-INF |
|---|---|---|---|---|---|---|---|---|
| Type of instrument | Self-rating | Self-rating | Self-rating | Self-rating | Semistructured interview | Self-rating | Semistructured interview | Structured interview |
| Ratee | Patient | Patient | Patient | Patient | Patient | Patient | Informant | Informant |
| Rater | Patient | Patient | Patient | Patient | Patient | Patient | CS (Q) | NCS (NQ) |
| Length, min | 5–10 | 5–10 | 5–10 | 30–45 | 15 | 30 | 30 | 5–10 |
| Number of items | 8 | 11 | 10 | 119 | 6[a] 6–32[b] | 60 | 14[a] 14–90[b] | 8 |
| Personnel | NQ | NQ | NQ | NQ | Q | Q | Q | NQ |
| Classification system | *ICD-10/DSM-IV* | *DSM-IV* | *DSM-IV* | *DSM-IV* | *ICD-10* | NA | *ICD-10/DSM-IV* | *ICD-10/DSM-IV* |
| Results/score | Total score | Total score | Total score | Total score and subscores for all *DSM-IV* personality disorders | Total score and subscores for all *ICD-10* personality disorders | 5 domain scores according to the Big Five factors | Total score and subscores for all the *ICD-10* and *DSM-IV* personality disorders | Total score |

[a]General questions. [b]Minimum-maximum questions in total.

Abbreviations: CS = clinical staff, IPDS = Iowa Personality Disorder Screen, NA = not applicable, NCS = nonclinical staff, NEO-FFI = NEO Five-Factor Inventory, NQ = nonqualified staff, PAS-Q = Quick Personality Assessment Schedule, Q = qualified staff, SAP = Standardized Assessment of Personality, SAPAS-INF = Standardized Assessment of Personality-Abbreviated Scale for Informants, SAPAS-SR = Standardized Assessment of Personality-Abbreviated Scale Self-Report, SCID-II-PQ = Structured Clinical Interview for *DSM-IV* Axis II Disorders Personality Questionnaire, S-SCID-II = short version of the Structured Clinical Interview for *DSM-IV* Axis II Disorders.

60 adult psychiatric patients recruited from outpatient and inpatient units, using the SCID-II[34] as the "gold standard." When validators used a cutoff score of 3, the sensitivity and specificity of the SAPAS were 0.94 and 0.85, respectively, and the positive and negative predictive values were 0.89 and 0.92, respectively.[17] Even short interviews, however, require specific clinical training. Therefore, we believed that the uptake of the SAPAS might improve if it were administered as a short self-report measure (SAPAS-SR). The original version of the SAPAS was translated into the Dutch language by the authors and translated back into English by the translation center of Tilburg University, Tilburg, The Netherlands (to assure accuracy of the Dutch version).

**IPDS.** The IPDS consists of 11 items originally derived from the Structured Interview for *DSM-III* Personality Disorders (SIDP).[35,36] The IPDS was validated in a group of 52 nonpsychotic inpatients and outpatients, and the outcome was compared with diagnoses based on the complete Structured Interview for *DSM-IV* Personality Disorders (SIDP-IV).[37]

In their original publication, Langbehn et al[18] did not report the sensitivity, specificity, and predictive values of the IPDS as a whole. Instead, these values were reported for each individual item. Moreover, optimal cutoff scores for specific subsets of items were presented; for instance, a subset of 6 a priori items was proposed as an overall screen. In so doing, the authors showed that the sensitivity, specificity, and predictive values differed considerably for specific subsets of items. Excellent sensitivity (92%) and good specificity (79%) were reached with IPDS items 4–8, whereas a subset consisting of item 1 and items 3–8 (ie, all the items that individually showed evidence of discriminability) showed sensitivity and a specificity of 79% and 86%, respectively. Because of these promising results, Langbehn et al advised further experimentation with all 11 items of the IPDS.[18]

Morse and Pilkonis[24] and Trull and Amdur[38] examined the utility of such a self-report version of the IPDS using the SIDP-IV as reference. Morse and Pilkonis[24] concluded that their self-report version was quite satisfactory in both psychiatric and nonpsychiatric samples. For instance, for a subset of IPDS items (items 1–6), sensitivity and specificity were 97% and 46%, respectively, with a positive predictive value of 90% and a negative predictive value of 71%.

The original version of the IPDS was translated into the Dutch language by the authors and translated back into English by the translation center of Tilburg University, Tilburg, The Netherlands (to assure accuracy of the Dutch version).

**S-SCID-II.** For the development of the S-SCID-II, we used the set of data that was collected by Masthoff and Trompenaars.[39] Axis II diagnoses were determined using the SCID-II interview. Their study included 533 participants, of whom 495 completed the test booklet (92.9%). For the identification of those items that best predicted SCID-II diagnoses, as a first step, a series of logistic regression analyses were performed. For every single personality disorder, only those items were selected from the total sets of SCID-II items that were intended to measure a particular personality disorder with the best discriminating function for predicting caseness, ie, the absence or presence of any personality disorder, according to the full SCID-II interview. Thereafter, again using logistic regression analyses, this set of potential predictors was used to predict caseness. The set of 10 items consisted of the following: paranoid (item 1), narcissistic (item 1), borderline (items 4, 5, and 8), avoidant (item 2), dependent (item 2), and depressive (items 2, 4, and 6). There was a good model fit on the basis of these 10 predictors ($\chi^2_{10} = 228.23$; N = 495; $P < .001$). The overall predictive rate was 76.0%. Inspection of regression coefficients, Wald statistics, and significance levels of the individual items reveals

that all 10 items contributed significantly to the prediction of the presence of *any* personality disorder.[39] Therefore, it was decided to accept this set of 10 items as a useful screening instrument for personality disorders. The S-SCID-II was administered as a self-report measure.

***SCID-II-PQ.*** The SCID-II-PQ is a questionnaire filled in by patients themselves. It has 119 items that match the questions in the SCID-II interview, with the introductory questions and observation items removed. With affirmative or negative answers, the respondent determines whether the feature is present. Three international studies examined the use of the SCID-II-PQ as a screening instrument.[25,40,41] Ekselius et al[25] did a study with 69 psychiatric patients and compared the SCID-II interview and the SCID-II-PQ. They suggested an adaptation of the cutoff scores for the SCID-II-PQ because of a high overrating of 19%. With the adapted cutoff scores, there was an overrating of 4% and a sensitivity and specificity of 87% and 75%, respectively. They found an overall κ of agreement of 0.75, and the correlation between number of criteria fulfilled in the SCID interview and the SCID-II-PQ was 0.84.[25] Similar data were found in the study conducted by Jacobsberg et al,[40] in which the SCID-II-PQ was examined with the Personality Disorder Examination as the "gold standard."

***NEO-FFI.*** One of the best-known models for defining personality, using a dimensional approach, is the Big Five model. The Big Five model is a general comprehensive framework for structuring individual differences.[42,43] The model is seen as pervasive across cultures. The 5 dimensions reflect sociability (*extraversion*), interpersonal interaction (*agreeableness*), self-discipline and impulse control (*conscientiousness*, describing task- and goal-directed behavior), personal adjustment (*neuroticism*, contrasting emotional stability with anxiety, anger, and other negative feelings), and openness to new experiences (*openness*, reflecting the breadth, depth, and complexity of mental and experiential life).

Costa and McCrae[42] have suggested that the 5-factor model of personality is highly relevant to the conceptualization and assessment of personality disorders. They have proposed to let the Big Five model replace the categorical system for identifying personality disorders in *DSM-IV*. Several authors support these claims.[44,45] Costa and McCrae[46] have described how personality disorder can be understood in terms of the Big Five dimensions. The NEO-FFI[29] is grounded in the Big Five dimensions and is a self-report instrument with 60 items.

***PAS-Q.*** The PAS-Q[20] is a shortened version of the *ICD-10* version of the Personality Assessment Schedule and takes about 15 minutes to complete. We previously described[21] the association between the 8 PAS-Q sections and the corresponding *ICD-10* categories, as well as the "translation" into the *DSM-IV-TR* classification system. The PAS-Q interview starts with open questions about character, relationships, job performance, drug problems, and law-breaking behavior to complete possibly missing information about the patient. Then, there are 8 specific sections for personality

disorders, namely (1) suspiciousness and sensitivity, (2) aloofness and eccentricity, (3) aggression and callousness, (4) impulsiveness and borderline, (5) childishness and lability, (6) conscientiousness and rigidity, (7) anxiousness and shyness, and (8) resourcelessness and vulnerability. To identify a certain personality disorder in each section, there are 2 screening questions, a positive answer to which leads to probing questions and eventually to scoring the characteristics in question. The interviewer assesses the severity of the personality disorder in every section, taking into account the answers to the introductory questions, the specific questions, and the background information on the patient. The PAS-Q distinguishes 4 levels of severity: (0) no personality disorder, (1) personality difficulty, (2) simple personality disorder, and (3) diffuse or complex personality disorder. The original version of the PAS-Q was translated into the Dutch language by the authors and translated back into English by the translation center of Tilburg University, Tilburg, The Netherlands. The result of the latter translation was nearly identical to that of the original version.

***SAP.*** The SAP[22,23] is a brief semistructured interview with an informant. The informant is asked an opening sequence of 14 questions that might suggest particular keywords. These keywords in turn lead to different categories of personality disorder. This process will happen by asking questions to find out whether enough criteria are met and whether there is enough evidence for these criteria to indicate the presence of a distress or handicap. If no keywords appear in the 14-item introduction phase, then the interview is terminated and no personality disorder is assumed.

The average overall interrater reliability (Cohen κ) for the SAP is 0.76, with a range from 0.60 to 0.82.[47] The interinformant reliability varies from 0.96 to 0.93.[48] The positive and negative predictive values of the SAP were 47% and 97%, respectively.[23] It was concluded that the SAP is a potentially adequate screening instrument in a 2-phase approach in epidemiologic assessment of personality disorder. The original version of the SAP was translated into the Dutch language by the authors and translated back into English by the translation center of Tilburg University, Tilburg, The Netherlands (to assure accuracy of the Dutch version).

***SAPAS-INF.*** The authors translated the items of the original SAPAS, a structured interview, and created a self-report questionnaire, the SAPAS-SR.[17,30] The authors transformed the Dutch SAPAS-SR into a structured interview for informants (SAPAS-INF).

***SCID-II.*** In all 3 studies, the SCID-II[34,49] was the "gold standard." The SCID-II interview is a semistructured interview to determine regular personality disorders, according to *DSM-IV-TR* criteria, as well as passive-aggressive and depressive personality disorders, as stated in the *DSM-IV* appendix.[50] The interview starts with a series of open questions intended to provide the interviewer with insight into the behavior, the interpersonal relationships, and the reflective abilities of the patient. Then, there are 134 items with more structured questions, grouped around the specific personality disorders. In scoring these, the interviewer has

to take into account the level of deviation, continuity, and pervasiveness. In the case of schizotypal, schizoid, theatrical, and narcissistic personality disorders, the interviewer is also required to take the patient's observed behavior into account.

A personality feature can be scored as (1) not present, (2) present to a limited extent, or (3) present. In scoring, not only is the patient's answer to the question important, but the interviewer also has to take all available sources of information into account. The interrater reliability and internal consistency of the SCID-II interview have proven to be satisfactory—and also for the Dutch population.[49,51,52]

To adequately conduct the SCID-II interview, the researchers, all of whom were psychiatrists, were trained in the technical aspects of conducting an interview. The staff of the Regional Institute for Continuing Education and Training, Eindhoven, The Netherlands, offered this certified training.

## Procedure

In all 3 studies, the procedure was roughly the same. The process of randomization consisted of 1 daily blind draw from the full set of referrals, performed by the secretary of the intake desk. After the drawing, the inclusion and exclusion criteria were checked by the secretary, and, in cases of eligibility, the invitation letter was sent. In cases of noneligibility, no second drawing was done that day. Exclusion criteria included the inability to undergo the protocol due to severe mental illness, illiteracy, dyslexia, mental retardation, severe visual or auditory handicap, cerebral damage, or refusal to participate. In addition to the invitation letter, there was a meeting between the patients and researchers in which eligible patients received verbal information along with the opportunity to ask questions. After this introductory procedure, all patients were asked to sign an informed consent form. The SAPAS-SR, IPDS, S-SCID-II, SCID-II-PQ, NEO-FFI, and PAS-Q were completed at the initial clinical appointment. The researcher who conducted the SCID-II interview was blinded to the results of the SAPAS-SR, IPDS, S-SCID-II, and SCID-II-PQ. The SCID-II interview was conducted 1 week later than the initial screening tests. The 4 self-rated screening tests were repeated 2 to 3 weeks after the initial assessment.

The SAP and the SAPAS-INF were conducted as face-to-face interviews with an informant in a routine standardized diagnostic process. The researcher was blinded to earlier obtained information concerning the patient or the SCID-II interview results.

## Analysis

All data were analyzed using the Statistical Package for the Social Sciences, version 12 (SPSS Inc, Chicago, Illinois). Test-retest reliability at the level of total SAPAS-SR, IPDS, and S-SCID-II scores was determined with Pearson correlation coefficients. Test-retest reliability of the separate items was determined using phi coefficients for binary data. Internal consistency was examined using Cronbach α coefficients.[53]

Cronbach α will generally increase when the correlations between the items of a scale increase.[54]

Receiver operating characteristic (ROC) analysis was used to study the effect on the predictive values for the presence of a personality disorder, as diagnosed with the SCID-II at the cutoff levels of scores on the SAPAS-SR, the IPDS, the S-SCID-II, and the PAS-Q. The ROC analysis relies heavily on sensitivity and specificity values and is a widespread method for examining the overall performance of a test.[55] Each point on the curve corresponds to a specific pairing of sensitivity and specificity. Inspection of the curve will be useful for finding an optimal cutoff value for use in decision-making. The total area under the ROC curve is a measure of the performance of the diagnostic instrument since it reflects the test performance at all possible cutoff levels.[56]

To compare the different screening instruments, likelihood ratios were calculated. The likelihood ratio incorporates the sensitivity and specificity of the test and provides a direct estimate of how much a test result will change the odds of having a personality disorder. The likelihood ratio for a positive result (LR+) specifies how much the odds of having a personality disorder increase when a test is positive. The likelihood ratio for a negative result (LR−) indicates how much the odds decrease when the test is negative.

The combination of the likelihood ratio and information about the prevalence of personality disorder and characteristics of the patient pool determines the posttest odds of personality disorder. The posttest probability (PTP) describes the proportions of patients with that particular test result who either have or do not have a personality disorder [posttest odds/(1 + posttest odds)].

## RESULTS

The prevalence of personality disorder in the different studies was as follows: 50% in study 1, with a mean number of personality disorders of 1.8 in patients diagnosed with any personality disorder; 64.1% in study 2, with a mean number of personality disorders of 2.2 in patients diagnosed with any personality disorder; and 48.1% in study 3, with the mean number of personality disorders of 1.6 in patients diagnosed with any personality disorder.

Table 2 shows the psychometric values of the different screening instruments. With the prescribed cutoff scores, the SCID-II-PQ overrated dramatically. When we increased the cutoff score by 3, the percentage of patients that were correctly classified increased from 62% to 75%. The SAPAS-SR, the IPDS, and the PAS-Q performed the best in terms of sensitivity and specificity for having *any* personality disorder. Moreover, these 3 reached the highest number of patients that were correctly classified. The test-retest coefficient turned out to be high for the following 4 screening instruments: the SAPAS-SR, the IPDS, the S-SCID-II, and the PAS-Q.

To assess the screening potential of the various instruments in a consistent way, 5 characteristics and the balance between them are important: sensitivity, specificity, positive

**Table 2. Sensitivity, Specificity, and Power to Predict *Any* Personality Disorder for the Different Screening Instruments**

| Variable | SAPAS-SR | IPDS | S-SCID-II | SCID-II-PQ | PAS-Q | NEO-FFI | SAP | SAPAS-INF |
|---|---|---|---|---|---|---|---|---|
| Sensitivity, % | 83 | 77 | 78 | 100/78[a] | 80 | 63 | 69 | 76 |
| Specificity, % | 80 | 85 | 78 | 27/78[a] | 82 | 35 | 76 | 58 |
| Positive predictive value, % | 80 | 83 | 78 | 100/75[a] | 81 | 48 | 84 | 77 |
| Negative predictive value, % | 82 | 79 | 78 | 27/74[a] | 81 | 50 | 58 | 57 |
| Correctly classified, % | 81 | 81 | 78 | 62/75[a] | 80 | 49 | 72 | 70 |
| Internal consistency | 0.45 | 0.64 | 0.67 | NA | 0.35 | NA | NA | NA |
| Test-retest reliability | 0.89 | 0.87 | 0.94 | NA | 0.94 | NA | NA | NA |

[a]Adjusted cutoff score (+3).
Abbreviations: IPDS = Iowa Personality Disorder Screen, NA = not applicable, NEO-FFI = NEO Five-Factor Inventory, PAS-Q = Quick Personality Assessment Schedule, SAP = Standardized Assessment of Personality, SAPAS-INF = Standardized Assessment of Personality-Abbreviated Scale for Informants, SAPAS-SR = Standardized Assessment of Personality-Abbreviated Scale Self-Report, SCID-II-PQ = Structured Clinical Interview for *DSM-IV* Axis II Disorders Personality Questionnaire, S-SCID-II = short version of the Structured Clinical Interview for *DSM-IV* Axis II Disorders.

**Table 3. Algorithm for Assessment of the Screening Instruments[a]**

| Category[b] | Criterion 1 | | Criterion 2 |
|---|---|---|---|
| ++ | 5 of the 5 ≥ 0.80 | | |
| + | 4 of the 5 ≥ 0.80 | And | 0 of the 5 ≤ 0.50 |
| | Or | | |
| | 5 of the 5 ≥ 0.70 | And | 0 of the 5 ≤ 0.50 |
| ± | 3 of the 5 ≥ 0.80 | And | 0 of the 5 ≤ 0.50 |
| | Or | | |
| | 4 of the 5 ≥ 0.70 | And | 0 of the 5 ≤ 0.50 |
| | Or | | |
| | 5 of the 5 ≥ 0.60 | And | 0 of the 5 ≤ 0.50 |
| - | 2 of the 5 ≥ 0.80 | And | Max 1 of the 5 ≤ 0.50 |
| | Or | | |
| | 3 of the 5 ≥ 0.70 | And | Max 1 of the 5 ≤ 0.50 |
| | Or | | |
| | 4 of the 5 ≥ 0.60 | And | Max 1 of the 5 ≤ 0.50 |
| – | 0 or 1 of the 5 ≥ 0.80 | And | Max 2 of the 5 ≤ 0.50 |
| | Or | | |
| | 2 of the 5 ≥ 0.70 | And | Max 2 of the 5 ≤ 0.50 |
| | Or | | |
| | 3 of the 5 ≥ 0.60 | And | Max 2 of the 5 ≤ 0.50 |
| — | 3 or more ≤ 0.60 | | |

[a]The 5 characteristics are sensitivity, specificity, positive predictive value, negative predictive value, and the percentage correctly classified; for example, 5 of the 5 characteristics ≥ 0.80 means that all the 5 characteristics have a value of 0.80 or more.
[b]The category symbols represent a progression from the best screening capacity (++) down to the worst screening capacity (—) in terms of number of characteristics and extent of fulfillment.
Abbreviation: Max = maximum.

predictive value (PPV), negative predictive value (NPV), and percentage of correctly classified patients. Table 3 shows the algorithm using the 5 characteristics to evaluate the screening capacity of the examined instruments in terms of the number of characteristics and the extent to which the characteristics were fulfilled. For example, the first category (++) requires that all 5 characteristics (sensitivity, specificity, PPV, NPV, and percentage correctly classified) are equal to or exceed 0.80.

Table 4 shows the screening capacity results for all the screening instruments for (1) any personality disorder, (2) a specific cluster of personality disorders, and (3) a specific personality disorder. The SAPAS-SR, the IPDS, the S-SCID-II, and the PAS-Q were the best screeners for *any* personality disorder. When normal cutoff scores were used, the SCID-II-PQ overrated dramatically and thus was classified in the fourth category (-). Only after the cutoff

scores were adjusted by +3 could the SCID-II-PQ rise to the second category (+). The NEO-FFI was classified as the poorest screener for *any* personality disorder.

If there is a need to screen for a specific personality disorder, eg, borderline personality disorder or antisocial personality disorder, one might use the SAP and the SCID-II-PQ, respectively.

Table 5 shows the LRs and PTPs for a positive and negative test outcome. The SAPAS-SR, the IPDS, and the PAS-Q appeared to have the best PTPs and raised the 50% odds in the outpatient population[1,2] to 80%–84% after a positive test outcome. The SAPAS-SR and the PAS-Q reduced the odds from 50% in the outpatient population[1,2] to 9% after a negative test outcome.

## DISCUSSION

The goal of this project was to provide busy clinicians a powerful screening tool for personality disorders that is time-efficient and easy to administer, while also accurate and, therefore, useable in clinical practice. The SAPAS-SR and the IPDS perform the best and are easy to administer. They do not require qualified personnel and take only 5 minutes to complete.

The findings should be interpreted in light of a number of limitations. First, not all the personality disorders were present in all the studies; notably, the schizoid and schizotypal personality disorders were absent. Participants with a single cluster A personality disorder could easily have become false negatives. This problem, however, is a minor one because only a small number of participants had a single cluster A personality disorder, not only in our samples, but also in other studies (eg, see Bernstein et al[57]). The fact that some cluster B personality disorders (eg, histrionic) are not represented probably is also a minor limitation due to comorbidity with other personality disorders.

Second, the validation studies were performed with an interviewer who was blinded to the outcome of the different instruments, except for the PAS-Q. For practical reasons, the interviews were performed by the same person (S.G.). To minimize possible bias, this interviewer refrained from reviewing the results of the interview and from filing the information in the patients' dossiers. We are aware that this procedure, forced by practical considerations reflecting the

**Table 4. Performance of the Different Screening Instruments in Predicting Different Categories of Personality Disorder From Any Personality Disorder to a Specific Personality Disorder (algorithm for assessment of screening instruments is in Table 3)**

| Disorder | SAPAS-SR | IPDS | S-SCID-II | SCID-II-PQ[a] | PAS-Q | NEO-FFI[b] | SAP | SAPAS-INF |
|---|---|---|---|---|---|---|---|---|
| Any personality disorder | ++ | + | + | + | ++ | — | ± | – |
| Cluster A personality disorder | – | – | - | - | – | | – | — |
| Cluster B personality disorder | - | ± | - | - | + | | – | – |
| Cluster C personality disorder | – | - | – | - | ± | | — | – |
| Specific personality disorder | | | | | | | | |
|   Paranoid | - | – | - | ± | – | | - | — |
|   Schizoid | — | – | — | ...c | – | | ...c | ...c |
|   Schizotypal | — | – | — | ...c | ...d | | ...c | ...c |
|   Borderline | - | - | - | + | ± | | + | – |
|   Histrionic | — | - | — | ...c | - | | – | — |
|   Narcissistic | - | - | - | – | ...d | | – | — |
|   Antisocial | - | – | - | ++ | - | | – | — |
|   Avoidant | - | - | - | – | - | | – | — |
|   Dependent | – | – | - | – | - | | – | — |
|   Obsessive-compulsive | – | – | — | – | – | | - | — |
|   Not otherwise specified | – | - | – | – | ...d | | – | — |

[a]With the adjusted cutoff scores.  [b]The NEO-FFI is not based on a classification system for personality disorders; therefore, it is not possible to give details about clusters or specific personality disorders.  [c]This specific personality disorder was not present in the study sample.  [d]This personality disorder is not present in the *ICD-10* and, therefore, is not measured.

Abbreviations: IPDS = Iowa Personality Disorder Screen, NEO-FFI = NEO Five-Factor Inventory, PAS-Q = Quick Personality Assessment Schedule, SAP = Standardized Assessment of Personality, SAPAS-INF = Standardized Assessment of Personality-Abbreviated Scale for Informants, SAPAS-SR = Standardized Assessment of Personality-Abbreviated Scale Self-Report, SCID-II-PQ = Structured Clinical Interview for *DSM-IV* Axis II Disorders Personality Questionnaire, S-SCID-II = short version of the Structured Clinical Interview for *DSM-IV* Axis II Disorders.

**Table 5. Likelihood Ratios and Posttest Probabilities of the Different Screening Instruments in Different Populations**

| Variable | SAPAS-SR | IPDS | S-SCID-II | SCID-II-PQ[a] | PAS-Q | NEO-FFI | SAP | SAPAS-INF |
|---|---|---|---|---|---|---|---|---|
| LR+ | 4.1 | 5.1 | 3.5 | 3.2 | 4.4 | 1.0 | 2.9 | 1.8 |
| LR− | 0.2 | 0.3 | 0.3 | 0.4 | 0.2 | 1.1 | 0.4 | 0.4 |
| PTP+ in general populations, % | 38 | 43 | 34 | 32 | 40 | 13 | 30 | 21 |
| PTP+ in outpatient populations, % | 80 | 84 | 78 | 76 | 81 | 49 | 74 | 64 |
| PTP+ in inpatient populations, % | 91 | 92 | 89 | 88 | 91 | 69 | 87 | 81 |
| PTP− in general populations, % | 15 | 21 | 21 | 26 | 15 | 49 | 26 | 26 |
| PTP− in outpatient populations, % | 9 | 13 | 13 | 17 | 9 | 35 | 17 | 17 |
| PTP− in inpatient populations, % | 6 | 8 | 8 | 11 | 6 | 25 | 11 | 11 |

[a]Adapted cutoff scores.

Abbreviations: IPDS = Iowa Personality Disorder Screen, LR− = likelihood ratio for a negative result, LR+ = likelihood ratio for a positive result, NEO-FFI = NEO Five-Factor Inventory, PAS-Q = Quick Personality Assessment Schedule, PTP− = posttest probability showing proportion of patients who do not have a personality disorder, PTP+ = posttest probability showing proportion of patients who have a personality disorder, SAP = Standardized Assessment of Personality, SAPAS-INF = Standardized Assessment of Personality-Abbreviated Scale for Informants, SAPAS-SR = Standardized Assessment of Personality-Abbreviated Scale Self-Report, SCID-II-PQ = Structured Clinical Interview for *DSM-IV* Axis II Disorders Personality Questionnaire, S-SCID-II = short version of the Structured Clinical Interview for *DSM-IV* Axis II Disorders.

institute's daily clinical practice, does not represent the best possible design. However, we feel that the risk of bias is presumably low due to the fact that the number of interviewees was rather high, the time interval between the interviews was rather lengthy, and inspection of patients' records in preparation of the interviews did not take place. Moreover, the fact that there was high correspondence between the PAS-Q and SCID-II interviews also provides a convincing argument for the relative absence of bias.

Third, the algorithm that we used to assess the screening instruments does not have a theoretical background. There is, as far as we know, no model known in the international literature. We are aware that, with the use of such a model, we simplify the reality—not in all situations is it important to

have a good balance between the 5 characteristics. In specific situations, one might prefer a particular highly rated characteristic at the cost of other characteristics. But, for a more global evaluation of the available screening instruments, we chose to compare them categorically by this model.

Finally, a potential pitfall for these kinds of instrument-validating studies is spectrum bias, meaning that the test is evaluated in a population composed of a mix of very ill patients and healthy controls. In such a population, the test obviously performs better in distinguishing the ill from the healthy than it does in actual practice.[58] The most appropriate design in these cases is a cross-sectional study that includes a spectrum of patients similar to those to whom the test will be administered in clinical practice. Although

the instruments were validated in 3 separate cross-sectional studies, all subjects in these studies were randomly selected from the entire group of outpatients that had been referred to the psychiatric hospital—ie, the target population for these instruments.

It should be noted that the prevalence of personality disorders is a powerful determinant of how useful a particular diagnostic instrument will be. The prevalence of personality disorder in study 3 was higher than in the other 2 studies. The prevalences in study 1 and study 2 were more or less similar to the results of other international studies (eg, see Zimmerman et al[2] and Masthoff and Trompenaars[39]). Furthermore, the mean number of personality disorders in patients who had any personality disorder was higher in study 3 in comparison with studies 1 and 2. It seems that the sample in study 3 was slightly different—they seemed sicker, which can be due to the fact that in study 3 there was a higher percentage of dropouts. For future research, it is therefore important that all screening instruments are examined in the same sample.

We concluded that for dichotomous case findings (a 2-step procedure of personality disorder identification) the SAPAS-SR and the IPDS are best suited. For case findings for a particular personality disorder or in situations dictated by practicality, eg, when the patient cannot be examined, another screening instrument might be preferred.

## REFERENCES

1. Adel A, Grimm G, Mogge NL, et al. Prevalence of personality disorders at a rural state psychiatric hospital. *J Rural Comm Psychol*. 2006;E9(1). http://www.marshall.edu/jrcp/9_1_Adel_Grimm_Mogge.htm. Verified November 1, 2001.
2. Zimmerman M, Rothschild L, Chelminski I. The prevalence of *DSM-IV* personality disorders in psychiatric outpatients. *Am J Psychiatry*. 2005; 162(10):1911–1918.
3. Alnaes R, Torgersen S. Personality and personality disorders predict development and relapses of major depression. *Acta Psychiatr Scand*. 1997;95(4):336–342.
4. Farmer RF, Nelson-Gray RO. Personality disorders and depression: hypothetical relations, empirical findings, and methodological considerations. *Clin Psychol Rev*. 1990;10(4):453–476.
5. Shea MT, Widiger TA, Klein MH. Comorbidity of personality disorders and depression: implications for treatment. *J Consult Clin Psychol*. 1992; 60(6):857–868.
6. Gunderson JG, Morey LC, Stout RL, et al. Major depressive disorder and borderline personality disorder revisited: longitudinal interactions. *J Clin Psychiatry*. 2004;65(8):1049–1056.
7. Gunderson JG, Stout RL, McGlashan TH, et al. Ten-year course of borderline personality disorder: psychopathology and function from the Collaborative Longitudinal Personality Disorders Study.

8. *Arch Gen Psychiatry*. 2011;68(8):827–837.
8. Meyerson D. Is borderline personality disorder underdiagnosed? Presented at the 162nd Annual Meeting of the American Psychiatric Association; May 16–21, 2009; San Francisco, CA. Abstract SCR17-R051.
9. Moran P, Walsh E, Tyrer P, et al. Impact of comorbid personality disorder on violence in psychosis: report from the UK700 trial. *Br J Psychiatry*. 2003;182(2):129–134.
10. Newton-Howes G, Tyrer P, Johnson T. Personality disorder and the outcome of depression: meta-analysis of published studies. *Br J Psychiatry*. 2006;188(1):13–20.
11. Spitzer RL, Fleiss JL. A re-analysis of the reliability of psychiatric diagnosis. *Br J Psychiatry*. 1974;125:341–347.
12. Zimmerman M. Diagnosing personality disorders: a review of issues and research methods. *Arch Gen Psychiatry*. 1994;51(3):225–245.
13. Hart SD, Forth AE, Hare RD. Performance of criminal psychopaths on selected neuropsychological tests. *J Abnorm Psychol*. 1990;99(4):374–379.
14. Hodiamont PPG. *Het Zoeken van Zieke Zielen* [dissertation]. Nijmegen, The Netherlands: Katholieke Universiteit Nijmegen; 1986.
15. Rijnders CATh. *Case Counting Considered: Psychiatric Epidemiology and Clinical Judgement* [dissertation]. Tilburg, The Netherlands: Tilburg University; 2008.
16. Dingemans PMAJ, Sno HN. Meetinstrumenten bij persoonlijkheidsstoornissen. *Tijdschr voor Psychiatr*. 2004;46(10):705-709.
17. Moran P, Leese M, Lee T, et al. Standardised Assessment of Personality-Abbreviated Scale (SAPAS): preliminary validation of a brief screen for personality disorder. *Br J Psychiatry*. 2003;183(3):228–232.
18. Langbehn DR, Pfohl BM, Reynolds S, et al. The Iowa Personality Disorder Screen: development and preliminary validation of a brief screening interview. *J Pers Disord*. 1999;13(1):75–89.
19. Van Horn E, Manley C, Leddy D, et al. Problems in developing an instrument for the rapid assessment of personality status. *Eur Psychiatry*. 2000;15(suppl 1):29–33.
20. Tyrer P. Quick Personality Assessment Schedule: PAS-Q. In: Tyrer P, ed. *Personality Disorders: Diagnosis, Management and Course*. 2nd ed. London, England: Arnold; 2000:181–190.
21. Germans S, Van Heck GL, Hodiamont PPG. Quick Personality Assessment Schedule (PAS-Q): validation of a brief screening test for personality disorders in a population of psychiatric outpatients. *Aust N Z J Psychiatry*. 2011;45(9):756–762.
22. Mann AH, Jenkins R, Cutting JC, et al. The development and use of standardized assessment of abnormal personality. *Psychol Med*. 1981; 11(4):839–847.
23. Mann AH, Raven P, Pilgrim J, et al. An assessment of the Standardized Assessment of Personality as a screening instrument for the International Personality Disorder Examination: a comparison of informant and patient assessment for personality disorder. *Psychol Med*. 1999;29(4):985–989.
24. Morse JQ, Pilkonis PA. Screening for personality disorders. *J Pers Disord*. 2007;21(2):179–198.
25. Ekselius L, Lindström E, von Knorring L, et al. SCID II interviews and the SCID screen questionnaire as diagnostic tools for personality disorders in *DSM-III-R*. *Acta Psychiatr Scand*. 1994;90(2):120–123.
26. Patrick J, Links P, Van Reekum R, et al. Using the PDQ-R BPD scale as a brief screening measure in the differential diagnosis of personality disorder. *J Pers Disord*. 1995;9(3):266–274.
27. Fossati A, Maffei C, Bagnato M, et al. Brief communication: criterion validity of the Personality Diagnostic Questionnaire-4+ (PDQ-4+) in a mixed psychiatric sample. *J Pers Disord*. 1998;12(2):172–178.
28. Costa PJ, McCrae R. *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) Professional Manual*. Odessa, FL: Psychological Assessment Resources; 1992.
29. Hoekstra HA, Ormel J, De Fruyt F. *NEO-FFI Big Five Persoonlijkheidsvragenlijst*. Lisse, The Nederlands: Swets & Zeitlinger; 2003.
30. Germans S, Van Heck GL, Moran P, et al. The Self-Report Standardized Assessment of Personality-Abbreviated Scale: preliminary results of a brief screening test for personality disorders. *Pers Ment Health*. 2008;2(2): 70–76.
31. Germans S, Van Heck GL, Langbehn DR, et al. The Iowa Personality Disorder Screen (IPDS): preliminary results of the validation of a self-administered version in a Dutch population. *Eur J Psychol Assess*. 2010; 26(1):11–18.
32. Germans S, Van Heck GL, Masthoff ED, et al. Diagnostic efficiency among psychiatric outpatients of a self-report version of a subset of screen items of the Structured Clinical Interview for *DSM-IV-TR* Personality Disorders (SCID-II). *Psychol Assess*. 2010;22(4):945–952.
33. Hilderson KMI, Germans S, Rijnders CATh, et al. Validatie van de

SCID-II persoonlijkheidsvragenlijst binnen een poliklinische psychiatrische populatie. *Psychopraktijk*. 2011;2:32–36.

34. First MB, Spitzer RL, Gibbon M, et al. The Structured Clinical Interview for *DSM-III-R* Personality Disorders (SCID-II), pt 2: multi-site test-retest reliability study. *J Pers Disord*. 1995;9:92–104.

35. Pfohl BM, Blum N, Zimmerman M, et al. *Structured Interview for* DSM-III-R *Personality: SIDP-R*. Iowa City, IA: Author-published; 1989.

36. Stangl D, Pfohl B, Zimmerman M, et al. A structured interview for the *DSM-III* personality disorders: a preliminary report. *Arch Gen Psychiatry*. 1985;42(6):591–596.

37. Pfohl BM, Blum N, Zimmerman M. *Structured Interview for* DSM-IV *Personality*. Washington, DC: American Psychiatric Press; 1997.

38. Trull TJ, Amdur M. Diagnostic efficiency of the Iowa Personality Disorder Screen items in a nonclinical sample. *J Pers Disord*. 2001; 15(4):351–357.

39. Masthoff EDM, Trompenaars AJWM. *Quality of Life in Psychiatric Outpatients* [dissertation]. Tilburg, The Netherlands: Tilburg University; 2006.

40. Jacobsberg L, Perry S, Frances A. Diagnostic agreement between the SCID-II screening questionnaire and the personality disorder examination. *J Pers Assess*. 1995;65(3):428–433.

41. Nussbaum D, Rogers R. Screening psychiatric patients for Axis II disorders. *Can J Psychiatry*. 1992;37(9):658–660.

42. Costa PTJ, McCrae RR. The five-factor model of personality and its relevance to personality disorders. *J Pers Disord*. 1992;6(4):343–359.

43. Goldberg LR. An alternative "description of personality": the big-five factor structure. *J Pers Soc Psychol*. 1990;59(6):1216–1229.

44. Wiggins JS, Pincus AL. Conceptions of personality disorders and dimensions of personality. *Psychol Assess*. 1989;1(4):305–316.

45. Costa PTJ, McCrae RR. Personality disorders and the five-factor model of personality. *J Pers Disord*. 1990;4(4):362–371.

46. Costa PTJ, McCrae RR. Looking backwards: changes in the mean levels of personality traits from 80 to 12. In: Cervone D, Mischel W, eds. *Advances in Personality Science*. New York, NY: Guilford Press; 2002: 219–237.

47. Pilgrim JA, Mellers JD, Boothby HA, et al. Inter-rater and temporal reliability of the Standardized Assessment of Personality and the influence of informant characteristics. *Psychol Med*. 1993;23(3): 779–786.

48. McKeon J, Roa B, Mann A. Life events and personality traits in obsessive-compulsive neurosis. *Br J Psychiatry*. 1984;144(2):185–189.

49. Weertman A, Arntz A, Kerkhofs MLM. *Gestructureerd Klinisch Interview voor de* DSM-III *Persoonlijkheidsstoornissen*. Lisse, The Netherlands: Swets Test Publisher; 1997.

50. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition. Washington, DC: American Psychiatric Association; 1994.

51. Maffei C, Fossati A, Agostoni I, et al. Interrater reliability and internal consistency of the Structured Clinical Interview for *DSM-IV* Axis II Personality Disorders (SCID-II), Version 2.0. *J Pers Disord*. 1997; 11(3):279–284.

52. Weertman A, Arntz A, Dreessen L, et al. Short-interval test-retest interrater reliability of the Dutch version of the Structured Clinical Interview for *DSM-IV* Personality Disorders (SCID-II). *J Pers Disord*. 2003;17(6):562–567.

53. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16(3):297–334.

54. Schmitt N. Uses and abuses of coefficient alpha. *Psychol Assess*. 1996;8(4):350–353.

55. Hanley JA. Receiver operating characteristic (ROC) methodology: the state of the art. *Crit Rev Diagn Imaging*. 1989;29(3):307–335.

56. Westin LK. *Receiver Operating Characteristic (ROC) Analysis: Evaluating Discriminance Effects Among Decision Support Systems*. Umeå, Sweden: Umeå University, Department of Computing Science; 2001.

57. Bernstein DP, Useda D, Siever LJ. Paranoid personality disorder. In: Lively WJ, ed. *The* DSM-IV *Personality Disorders*. London, England: Guilford Press; 1995:45–57.

58. Gray, GE. *Concise Guide to Evidence-Based Psychiatry*. Washington, DC: American Psychiatric Association; 2004.