A Self-Report Scale to Help Make Psychiatric Diagnoses

The Psychiatric Diagnostic Screening Questionnaire

Mark Zimmerman, MD; Jill I. Mattia, PhD

Background: The Psychiatric Diagnostic Screening Questionnaire (PDSQ) is a brief, psychometrically strong, self-report scale designed to screen for the most common DSM-IV Axis I disorders encountered in outpatient mental health settings. In the present report, we describe the diagnostic performance (sensitivity, specificity, and positive and negative predictive values) of the PDSQ in an outpatient setting.

Methods: Six hundred thirty psychiatric outpatients presenting for treatment were evaluated with the Structured Clinical Interview for DSM-IV after completing the PDSQ. Patients arrived approximately 20 minutes before the scheduled time of the appointment to complete the scale. Diagnostic raters were blind to responses on the scale.

Results: The PDSQ’s subscales’ diagnostic performance varied in a predictable manner according to the cutoff score—as the threshold for case identification increased, subscale sensitivity decreased and specificity increased. Mean subscale sensitivities of 80%, 85%, and 90% resulted in mean subscale specificities of 78%, 73%, and 66%, respectively, and negative predictive values of 95%, 96%, and 97%. Receiver operating curves were determined for each subscale and all areas under the curve were significant.

Conclusions: The PDSQ is a diagnostic aid designed to be used in clinical practice to facilitate the efficiency of conducting initial diagnostic evaluations. From a clinical perspective, it is most important that a diagnostic aid have good sensitivity, so that most cases are detected, and high negative predictive value, so that most noncases on the measure are indeed noncases. Our results indicate that most of the PDSQ subscales were able to achieve this goal.

Arch Gen Psychiatry. 2001;58:787-794
SUBJECTS AND METHODS

SUBJECTS

A subset of patients (n=630) evaluated in the Rhode Island Hospital Department of Psychiatry outpatient practice were interviewed by a trained diagnostic rater who administered the SCID and presented the results to the treating clinician. We previously reported that the patients who were and were not interviewed with the SCID were similar in their demographic characteristics and scores on self-report measures of symptom severity.16 The Rhode Island Hospital Institutional Review Committee approved the research protocol, and all patients provided informed, written consent. Details regarding interviewer experience, training, supervision, and reliability are presented in other reports from the MIDAS project.16,19,20

When scheduling their appointments, patients were told to arrive early to complete some standard forms. The PDSQ takes approximately 15 to 20 minutes to complete, and its administration did not disrupt usual clinical practice. Because we were planning to test the PDSQ’s validity by examining the relationship between subscale scores and psychiatric diagnoses, SCID interviewers were kept blind to subjects’ responses on the measure. The PDSQ was always completed before the SCID.

MEASURES

The PDSQ has undergone several rounds of study involving more than 3000 primary care and psychiatric outpatients. After each large validation study, the scale was revised based on a psychometric analysis of the subscales and items. The final version of the PDSQ consists of 126 questions assessing the symptoms of 13 DSM-IV disorders in 5 areas: eating disorders (bulimia/binge-eating disorder); mood disorders (major depressive disorder [MDD]); anxiety disorders (panic disorder, agoraphobia, PTSD, obsessive-compulsive disorder, generalized anxiety disorder [GAD], and social phobia); substance use disorders (alcohol abuse/dependence and drug abuse/dependence); and somatoform disorders (somatization disorder and hypochondriasis). In addition, there is a 6-item psychosis screen. The disorders chosen for coverage were selected because they are the most prevalent in epidemiological surveys of the community and the most frequently reported in large clinical samples.16,23,24 Three subscales (mania, dysthymic disorder, and anorexia) were dropped because of poor psychometric performance after extensive investigation.

In determining the length of the PDSQ subscales, we tried to balance the desire to keep the scale brief so that it would be feasible to incorporate it into routine clinical practice with the desire to make the scale comprehensive so that most or all diagnostic criteria of the included disorders were assessed. The MDD subscale was the longest PDSQ subscale, at 22 items, because it assessed each of the 9 DSM-IV symptom criteria, and a separate question for each element of compound MDD criteria was included (eg, the sleep disturbance criterion includes questions of both increased and decreased sleep). The reason for including this level of detail was the potential treatment implications of the presence of vegetative and reverse vegetative symptoms of MDD. The number of items on the other PDSQ subscales were as follows: PTSD (n=15), bulimia/binge-eating disorder (n=10), obsessive-compulsive disorder (n=8), panic disorder (n=8), psychosis (n=6), agoraphobia (n=11), social phobia (n=15), alcohol abuse/dependence (n=6), drug abuse/dependence (n=6), GAD (n=10), somatization (n=3), and hypochondriasis (n=5).

The PDSQ inquires about current and recent symptoms. Because the DSM-IV symptom-duration requirement varies by disorder, we adopted multiple time frames. We thought it would be too confusing and awkward to follow all of the DSM-IV duration requirements; therefore we simplified this by using 2 different time frames. All but 2 questions on the first 6 subscales (MDD, PTSD, bulimia/binge-eating disorder, obsessive-compulsive disorder, panic disorder, and psychosis) refer to the past 2 weeks. The exceptions are the first 2 questions on the PTSD subscale, which ask about having ever experienced or witnessed a traumatic event. The time frame for the other 7 subscales (agoraphobia, social phobia, alcohol abuse/dependence, drug abuse/dependence, GAD, somatization, and hypochondriasis) is 6 months before the evaluation. We chose a longer time frame for these domains of psychopathology because the symptoms of some of these disorders are more intermittent and may not have been present during the 2 weeks before the evaluation. For example, problems associated with drugs and alcohol are often sporadic, in contrast to the daily symptoms of MDD. Similarly, phobic situations may be encountered on an irregular basis. For GAD and hypochondriasis, symptoms are assessed for the past 6 months because DSM-IV requires a 6-month duration to be diagnosed.

The PDSQ was intended to be administered and scored in the office before the initial diagnostic evaluation. Respondents easily understand the yes/no response format of the scale. To illustrate the content of the scale, Table 1 lists the items and format of the generalized anxiety disorder subscale. Items answered yes are scored 1, items answered no, 0. The items on each diagnostic subscale are grouped, and the subscales are clearly demarcated from each other. This organization of the PDSQ facilitates rapid hand-scoring that makes it feasible to be used in routine clinical practice. (Copies of the PDSQ are available from the authors.)

Fifth, the changing health care delivery system has placed increasing time constraints on conducting diagnostic evaluations. When clinicians’ time to conduct diagnostic evaluations is reduced, it is more likely that additional...
In the validity study of the final version of the PDSQ, 994 psychiatric outpatients completed the scale.17 The 13 PDSQ subscales demonstrated good to excellent levels of internal consistency. Cronbach’s α was greater than 0.80 for 12 of the 13 subscales, and the mean of the α coefficients was 0.86. Test-retest reliability was examined in the 185 subjects who completed the PDSQ 2 times less than a week apart. Test-retest reliability coefficients were greater than 0.80 for 9 subscales, and the mean of the test-retest correlation coefficients was 0.83. The convergent and discriminant validity of the PDSQ subscales27 was examined in 361 patients who completed a package of questionnaires at home less than a week after completing the PDSQ. The booklet included measures of symptoms related to each of the PDSQ symptoms domains.7,8,26-40

Every PDSQ subscale was more highly correlated with the concordant validity scale assessing the same symptom domain vs other symptoms domains. Across all subscales, the mean correlation between the PDSQ subscales and their respective validity scale was 0.66, while the mean correlation between PDSQ subscales and measures of other symptom domains was 0.25. Finally, for each of the disorders assessed by the PDSQ, the mean diagnosis-specific subscale score in patients with and without that disorder assessed by the PDSQ, the mean diagnosis-related to each of the PDSQ symptoms domains.7,26-40

The core of the diagnostic evaluation was the January 1995 DSM-IV patient version of the SCID. The Axis I version of the SCID covers 7 DSM-IV sections: mood, psychotic, substance use, anxiety, somatof orm, adjustment, and eating disorders. For 3 symptom domains—psychosis, bulimia, and somatization—we combined patients with related diagnoses. The psychosis group included patients diagnosed as having schizophrenia, schizoaffective disorder, psychotic disorder not otherwise specified (NOS), MDD with psychotic features, and bipolar depression with psychotic features. The bulimia group included subjects diagnosed as having binge-eating disorder (a DSM-IV appendix diagnosis that would otherwise be captured as part of the eating disorder NOS category), as well as bulimia nervosa. The somatization group included patients diagnosed as having somatization disorder and undifferentiated somatoform disorder. Finally, when examining the PDSQ MDD subscale, we combined patients with MDD, bipolar I depression, bipolar II depression, and 4 patients who met full criteria for MDD but also had nonbizarre delusions outside of the mood episode. According to DSM-IV, these 4 patients were diagnosed as having psychotic disorder NOS and depressive disorder NOS.

DATA ANALYSIS

There are several excellent articles describing the descriptive statistics of test performance.5,45 Despite these, in several studies of the performance of self-administered screening tests, incorrect definitions and miscalculations of these statistics were found46; therefore, we present a brief overview of this area.

Sensitivity refers to a test’s ability to correctly identify individuals with the disorder, whereas specificity refers to a test’s ability to identify persons who are not ill. Sensitivity and specificity provide useful psychometric information about a test; however, the clinically more meaningful conditional probabilities are positive and negative predictive values. These values indicate the probability that an individual is ill or not ill given that the test identifies them as ill or not ill. Accordingly, positive predictive value is the percentage of individuals classified as ill by the test who truly are ill, whereas negative predictive value is the percentage of individuals classified not ill by the test who truly are not ill.

Sensitivity, specificity, and positive and negative predictive values are not invariant properties of a test—they are a function of the cutoff point used to distinguish cases from noncases, they are influenced by disease prevalence, and they are related to each other. Four axioms characterize these relationships: (1) Lowering a test’s cutoff score to identify cases increases the test’s sensitivity and decreases its specificity. (2) Conversely, raising the test threshold to identify cases decreases the test’s sensitivity and increases its specificity. (3) At constant sensitivity and specificity, a test’s positive predictive value is higher in samples where disease prevalence is greater. (4) At constant sensitivity and specificity, a test’s negative predictive value is higher in samples where disease prevalence is lower.

Depending on the instrument’s purpose, cutoff scores might be selected to optimize the sensitivity or specificity of the scale.57,48 In the present report, we describe the diagnostic performance of the PDSQ subscales across the range of cutoff scores. We determined the average specificity and positive and negative predictive value across the PDSQ subscales when sensitivity was 80%, 85%, and 90%. When the values of sensitivity were not exactly 80%, 85%, or 90%, we extrapolated the values of the other diagnostic performance statistics. For example, the agoraphobia subscale had no corresponding cutoff for a sensitivity of 85%. We extrapolated from the sensitivity of 82% and 88% (specificity of 83% and 75%), and estimated that at a sensitivity of 85% the agoraphobia subscale would have a specificity of 79%. We conducted receiver operating curve analyses to completely determine the subscales’ diagnostic performance across the range of cutoff points and to allow us to evaluate the diagnostic performance of the different subscales by examining their areas under the curve (AUCs).45,47
The data in Table 2 show the demographic and diagnostic characteristics of the sample. The majority of the subjects were white, female, married or single, and had some college education. The mean (SD) age of the sample was 37.8 (11.9) years. The most frequent DSM-IV diagnoses were MDD (47.9%), social phobia (26.5%), GAD (17.5%), and panic disorder (17.0%).

The data in Tables 3 and 4 show that the PDSQ subscales' diagnostic properties varied in a predictable manner according to the cutoff score—as the threshold increased sensitivity decreased and specificity increased. At respective cutoff scores resulting in a sensitivity of 80%, subscale specificities ranged from a high of 91% for the bulimia and drug abuse/depression subscales to a low of 58% for the somatization subscale. The mean specificity across all subscales when subscale sensitivity was 80%, 85%, and 90% was 78%, 73%, and 66%, respectively. (It should be noted that the psychosis subscale did not achieve a sensitivity of 80%. For this subscale, our analysis included the corresponding diagnostic statistics for the subscale's maximum sensitivity of 75%.)

In our other analyses, if the subscale did not achieve a sensitivity of 85% or 90%, the subscale was not included in the calculation of average specificity and predictive values across subscales. When subscale sensitivity was 80%, 85%, and 90%, the mean positive predictive value across the subscales was 32%, 31%, and 30%, and the mean negative predictive value was 95%, 96%, and 97%, respectively.

Receiver operating characteristic curves were determined for each subscale and all AUCs were significant (Figure). All areas under the curve were above 0.75, ranging from 0.76 to 0.92 with a mean of 0.85.

In the present article, we have described the diagnostic properties of the PDSQ, a self-report scale designed to assess the most common DSM-IV Axis I disorders presenting in outpatient settings. Longer, multidimensional questionnaires, such as the Minnesota Multiphasic Personality Inventory 25 and the Millon Clinical Multiaxial Inventory,23 have been used as diagnostic aids; however, they were not designed to be congruent with the current diagnostic nomenclature. Moreover, these inventories are too long, and their scoring too time-consuming, to be routinely completed in the early 1980s and scored in an office waiting area before the initial evaluation. Other scales...
have been developed to assess specific DSM-IV Axis I disorders such as MDD, PTSD, and bulimia, but they are limited to only one type of pathology. The self-report version of the Primary Care Evaluation of Mental Disorders assesses multiple disorders, but it was developed for use in primary care settings.
The PDSQ was intended as a diagnostic aid to be used in clinical practice to facilitate the efficiency of conducting the initial diagnostic evaluation. Consequently, we recommend that a cutoff resulting in diagnostic sensitivity of 90% be chosen when using the scale in clinical practice. Table 4 highlights the cutoff scores on each subscale corresponding to a sensitivity of 90%. From a clinical perspective, it is most important that the diagnostic aid have good sensitivity and corresponding high negative predictive value. With high negative predictive value, the clinician can be confident that when the test indicates that the disorder is not present there is little need to inquire about that disorder’s symptoms. False positives are less of a problem for a screening questionnaire because their major cost is the time a clinician takes to determine that the disorder is not present. Presumably, this is time the clinician would have nonetheless spent for the same purpose. Based on the cutoffs resulting in a sensitivity of 90%, the mean negative predictive value of the PDSQ subscales was 97%, and the false positive rate was 34%.

It is important to understand what might contribute to false-positive results. In our analyses, patients diagnosed as having “subthreshold” DSM-IV disorders (i.e., partial remission, or NOS) were not counted as cases. Cases below the threshold of full diagnostic criteria were not rare. Elsewhere, we examined the frequency of anxiety disorders in depressed outpatients and found that 10.7% of the patients had an anxiety disorder in partial remission at the time of the evaluation and 15.3% had a current anxiety disorder NOS.53 Not surprisingly, patients with subthreshold disorders scored significantly higher on the PDSQ subscale than patients without the disorder. This indicates that the test is sensitive to the symptoms of the disorder even when it is not fully present. However, it is important to be cautious in interpreting these results, as the inclusion of subthreshold cases may lead to an overestimation of the disorder’s prevalence in clinical practice. Based on the cutoffs, the mean negative predictive value of the PDSQ subscales was 97%, and the false positive rate was 34%.

Receiver operating curves for each of the Psychiatric Diagnostic Screening Questionnaire subscales in 630 psychiatric outpatients. All curves were significant at P<.001. AUC indicates area under the curve; please see the first footnote to Table 3 for additional subscale abbreviations.
disorder, although lower than patients who met full diagnostic criteria. Thus, some of the patients who were false positives had clinically significant symptoms of the disorder being assessed, although they did not meet full diagnostic criteria for a current disorder.

Criteria overlap among the DSM-IV disorders will result in false positives in any scale that follows the DSM-IV diagnostic criteria. As has been discussed by others, the high rates of comorbidity among the DSM-IV disorders may be caused, in part, by the difficulty in clearly demarcating the boundary between different syndromes, and some symptoms are inclusion criteria of more than 1 disorder.

Another assessment issue that could result in false-positive results on some of the PDSQ subscales is different time frames covered by the SCID and PDSQ. Our rule for distinguishing between current and past episodes on the SCID was the same for all disorders—after 2 months of symptom resolution, the disorder was considered a past diagnosis. This followed the DSM-IV suggestion for defining remission of a depressive episode. However, on the PDSQ, the questions assessing some disorders, such as alcohol and substance abuse/dependence, referred to the past 6 months. Thus, some subjects who quit drinking or abusing drugs more than 2 months before the evaluation, but less than 6 months ago, might be false positive on the PDSQ because we would have diagnosed them as having a past, not a current, substance use problem. In fact, it was our experience that even some individuals who quit abusing substances before 6 months ago would respond positively to the PDSQ questions. Consistent with this, patients with a past diagnosis of alcohol or drug abuse/dependence scored significantly higher than patients without a history of alcohol or drug problems. These post hoc analyses indicate that many of the false positives on the PDSQ are the result of the detection of clinically important symptoms.

A possible limitation of the study was our failure to randomize the order of presentation of the PDSQ and SCID assessments; thus, we were unable to examine the influence of order effects. Instead, we studied the PDSQ as we expect a screening measure to be used—preceding the more detailed diagnostic evaluation.

The present sample was drawn from a large general adult outpatient private practice setting in which the most common presenting problems were mood and anxiety disorders. Rhode Island has a strong community mental health center network that treats most of the chronically mentally ill patients, and this accounts, in part, for the low prevalence rates of psychotic disorders. The practice does not have a specialist in the treatment of substance use disorders; thus patients with a primary substance use problem are infrequently encountered. It will be important to replicate and extend the present findings to samples with different demographic and clinical characteristics. Another direction for future research is to have patients complete some initial paperwork that is reviewed before the initial visit. We recommend that the PDSQ be used in a similar way. That is, the responses to the scale should be reviewed before the face-to-face encounter, and the information should make it less likely that areas of psychopathology are overlooked. Of course, a thorough diagnostic interview is the diagnostic standard of care. There are no special questions on the PDSQ that allows it to detect psychopathology that otherwise would go undetected during a clinical evaluation. However, clinicians often do not have the time to be as comprehensive as they would like. It is our hope that the PDSQ can improve the efficiency of the diagnostic evaluation by guiding clinicians toward symptom areas that require more vs less assessment. Elsewhere, we described how the PDSQ validated PTSD in patients whose conditions were not diagnosed with PTSD by their treating clinicians.

During the past few years, a cottage industry has arisen promoting products that assist in diagnostic and outcome assessment. The American Psychological Association has clearly written guidelines for test development; however, most of the tools marketed at professional meetings as DSM-IV diagnostic aids have not been subjected to these rigorous test development procedures. Research demonstrating the reliability and validity of these tools rarely has been published in peer-reviewed journals. Nevertheless, these products are advertised and sold, and some insurance companies require their use. Diagnostic (and outcomes) evaluations are an important component of case formulation and treatment planning, and professional organizations might want to take a more active role in monitoring instruments that are intended to influence the diagnostic process.

Accepted for publication October 3, 2000.

Supported in part by grant MH56404 from the National Institute of Mental Health, Bethesda, Md (Dr Mattia).

Reprints: Mark Zimmerman, MD, Department of Psychiatry, Rhode Island Hospital, 235 Plain St, Suite 501, Providence, RI 02905 (e-mail: mzimmerman@lifespan.org).

REFERENCES