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

The Psychiatric Diagnostic Screening Questionnaire (PDSQ) is a self-report scale designed to screen for the most common DSM-IV Axis I disorders encountered in outpatient mental health settings. Five research and clinical factors occurring during the past 2 decades contributed to the development of the PDSQ. First, the publication of specific inclusion criteria to make psychiatric diagnoses, complemented by the development of standardized interviews to reliably assess the criteria, set the stage for the construction of self-administered questionnaires, such as the PDSQ, which screen for or make provisional psychiatric diagnoses. While research diagnostic interviews, such as the Schedule for Affective Disorders and Schizophrenia, Diagnostic Interview Schedule, and the Structured Clinical Interview for DSM-IV (SCID), are not infallible, they have been accepted as diagnostic standards to which the diagnostic performance of other tests (be they biological or self-report) are compared.

Second, several research groups demonstrated that it was possible to construct a self-report questionnaire that “diagnosed” individual DSM-IV disorders. One of the first such measures was constructed in the early 1980s by Zimmerman and colleagues, who developed the Inventory to Diagnose Depression to evaluate the DSM-III criteria for major depressive disorder (MDD). Their initial work on the Inventory to Diagnose Depression was replicated by other research groups, and subsequently other questionnaires have been designed to screen for specific DSM-IV Axis I disorders, such as posttraumatic stress disorder (PTSD) and bulimia nervosa.

The third influence on our decision to develop a questionnaire assessing several...
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. 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.

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 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. 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 the 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 past 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

*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. 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.

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 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.

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. Studies of comorbidity rates in patients whose conditions were diagnosed by clinicians in the routine clinical setting are one half to one third the comorbidity rates reported in studies using standardized research diagnostic interviews.

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
psychiatric disorders beyond the presenting complaint will not be detected.

These 5 factors were the impetus for the development of the PDSQ. Our goal was to develop a psychometrically sound, clinically useful instrument that was brief enough to be completed by patients before their initial diagnostic evaluation, yet comprehensive enough to cover the most common disorders for which individuals seek treatment. Finally, the scoring and organization of the measure should be clear and straightforward enough so that a clinician, or office worker can rapidly review and score the scale and obtain clinically meaningful information. In previous reports from the Rhode Island Methods to Improve Diagnostic Assessment and Services (MIDAS) project, we presented the psychometric properties of the successive versions of the PDSQ. In the present report from the MIDAS project, we describe the diagnostic performance of the PDSQ in an outpatient setting.

DATA ANALYSIS

There are several excellent articles describing the descriptive statistics of test performance. Despite these, in several studies of the performance of self-administered screening tests, incorrect definitions and miscalculations of these statistics were found; 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. 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 specificity 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).
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/dependence 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 2nd and the Millon Clinical Multiaxial Inventory, 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.

Table 3. Sensitivity and Specificity of Psychiatric Diagnostic Screening Questionnaire Subscales at Different Cutoff Scores in 630 Psychiatric Outpatients*

<table>
<thead>
<tr>
<th>Subscale Score</th>
<th>Sensitivity/Specificity, %†</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDD</td>
<td>Bulimia</td>
</tr>
<tr>
<td>1</td>
<td>100/5</td>
</tr>
<tr>
<td>2</td>
<td>100/10</td>
</tr>
<tr>
<td>3</td>
<td>100/18</td>
</tr>
<tr>
<td>4</td>
<td>100/25</td>
</tr>
<tr>
<td>5</td>
<td>99/34</td>
</tr>
<tr>
<td>6</td>
<td>99/42</td>
</tr>
<tr>
<td>7</td>
<td>98/52</td>
</tr>
<tr>
<td>8</td>
<td>95/63</td>
</tr>
<tr>
<td>9</td>
<td>90/67</td>
</tr>
<tr>
<td>10</td>
<td>81/73</td>
</tr>
<tr>
<td>11</td>
<td>70/78</td>
</tr>
<tr>
<td>12</td>
<td>63/84</td>
</tr>
<tr>
<td>13</td>
<td>53/89</td>
</tr>
<tr>
<td>14</td>
<td>43/92</td>
</tr>
<tr>
<td>15</td>
<td>31/94</td>
</tr>
<tr>
<td>16</td>
<td>22/97</td>
</tr>
<tr>
<td>17</td>
<td>13/99</td>
</tr>
<tr>
<td>18</td>
<td>5/99</td>
</tr>
<tr>
<td>19</td>
<td>2/100</td>
</tr>
<tr>
<td>20</td>
<td>0/100</td>
</tr>
<tr>
<td>21</td>
<td>0/100</td>
</tr>
</tbody>
</table>

*MDD indicates major depressive disorder; OCD, obsessive-compulsive disorder; PTSD, posttraumatic stress disorder; Agor, agoraphobia; Alcohol, alcohol abuse or dependence; Drug, drug abuse or dependence; GAD, generalized anxiety disorder; Somat, somatoform disorder; and Hypoch, hypochondriasis.

†Unshaded pairs of sensitivity/specificity values indicate the subscale cutoff score closest to 90% sensitivity, which is the recommended case identification cutoff score for use in clinical practice.

Table 4. Positive and Negative Predictive Value of Psychiatric Diagnostic Screening Questionnaire Subscales at Different Cutoff Scores in 630 Psychiatric Outpatients*

<table>
<thead>
<tr>
<th>Subscale Score</th>
<th>Positive/Negative Predictive Value, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDD</td>
<td>Bulimia</td>
</tr>
<tr>
<td>3</td>
<td>56/100</td>
</tr>
<tr>
<td>5</td>
<td>61/98</td>
</tr>
<tr>
<td>6</td>
<td>64/97</td>
</tr>
<tr>
<td>7</td>
<td>68/96</td>
</tr>
<tr>
<td>9</td>
<td>74/86</td>
</tr>
<tr>
<td>10</td>
<td>76/78</td>
</tr>
<tr>
<td>11</td>
<td>77/72</td>
</tr>
<tr>
<td>12</td>
<td>81/69</td>
</tr>
<tr>
<td>13</td>
<td>83/64</td>
</tr>
<tr>
<td>14</td>
<td>85/61</td>
</tr>
<tr>
<td>15</td>
<td>85/57</td>
</tr>
<tr>
<td>16</td>
<td>90/55</td>
</tr>
<tr>
<td>17</td>
<td>91/52</td>
</tr>
<tr>
<td>18</td>
<td>93/51</td>
</tr>
<tr>
<td>19</td>
<td>96/49</td>
</tr>
<tr>
<td>20</td>
<td>100/49</td>
</tr>
<tr>
<td>21</td>
<td>1/0</td>
</tr>
</tbody>
</table>

*MDD indicates major depressive disorder; OCD, obsessive-compulsive disorder; PTSD, posttraumatic stress disorder; Agor, agoraphobia; Alcohol, alcohol abuse or dependence; Drug, drug abuse or dependence; GAD, generalized anxiety disorder; Somat, somatoform disorder; and Hypoch, hypochondriasis.

†Could not be calculated because no patients scored 21 on the MDD subscale.
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 (ie, 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, although lower than patients who met full di-
agnostic 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 de-
marcating the boundary between different syn-
dromes,54–56 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 differ-
ent 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 de-
fining 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 evalua-
tion, but less than 6 months ago, might be false posi-
tive 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 pa-
tients without a history of alcohol or drug problems. These
post hoc analyses indicate that many of the false posi-
tives on the PDSQ are the result of the detection of clini-
cally 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 dis-
orders. Rhode Island has a strong community mental
health center network that treats most of the chroni-
cally mentally ill patients, and this accounts, in part, for
the low prevalence rates of psychotic disorders. The prac-
tice does not have a specialists in the treatment of sub-
stance use disorders; thus patients with a primary sub-
stance use problem are infrequently encountered. It will
be important to replicate and extend the present find-
ings to samples with different demographic and clinical
characteristics. Another direction for future research is
the performance of the PDSQ in primary care settings.
Studies of an earlier version of the scale found that it was
well received by primary care patients, and it possessed
favorable psychometric properties.57,58

We developed the PDSQ to aid clinicians in making
psychiatric diagnoses. It is common in medical practices
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 en-
counter, and the information should make it less likely
that areas of psychopathology are overlooked. Of course,
a thorough diagnostic interview is the diagnostic stan-
dard 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. How-
ever, clinicians often do not have the time to be as com-
prehensive 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 re-
quire more vs less assessment. Elsewhere, we described
how the PDSQ validly detected PTSD in patients whose
conditions were not diagnosed with PTSD by their treating
clinicians.59

During the past few years, a cottage industry has
arisen promoting products that assist in diagnostic and
outcome assessment. The American Psychological As-
sociation60 has clearly written guidelines for test devel-
opment; however, most of the tools marketed at profes-
sional meetings as DSM-IV diagnostic aids have not been
subjected to these rigorous test development proce-
dures. Research demonstrating the reliability and valid-
ity of these tools rarely has been published in peer-
reviewed journals. Nevertheless, these products are
advertised and sold, and some insurance companies re-
quire their use. Diagnostic (and outcomes) evaluations
are an important component of case formulation and treat-
ment 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 Na-
tional 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

1. Endicott J, Spitzer RL. A diagnostic interview: the Schedule for Affective Disor-
2. Robins LN, Helzer JE, Croughan J, Ratcliff KS. National Institute of Mental Health
Diagnostic Interview Schedule: its history, characteristics, and validity. Arch Gen
3. First MB, Spitzer RL, Williams JBW, Gibbon M. Structured Clinical Interview for
4. Zimmerman M, Coryell W, Corenthal C, Wilson S. A self-report scale to diag-
5. Goldston DB, O’Hara MW, Schartz HA. Reliability, validity, and preliminary nor-
mative data for the Inventory to Diagnose Depression in a college population.
6. Kuhner C, Veiel HO. Psychometrische und diagnostische Eigenschaften einer du-
etschsprachigen Version des Inventoary to Diagnose Depression (IDD). Diagnos-
7. Foa EB, Riggs DS, Dancu CV, Rothbaum BO. Reliability and validity of a brief
6:459-473.
8. Thelen MH, Farmer J, Wonderlich S, Smith M. A revision of the Bulimia Test: the
9. Maser JD, Cloninger CR. Comorbidity of Mood and Anxiety Disorders. Wash-