

Diagnosis of Nonpsychotic Patients in Community Clinics

M. Katherine Shear, M.D., Catherine Greeno, Ph.D.,
Jasbir Kang, M.D., Diane Ludewig, Ph.D., Ellen Frank, Ph.D.,
Holly A. Swartz, M.D., and Miyako Hanekamp, M.S.

Objective: Diagnosis-specific, proven efficacious treatments are a major recent advance in psychiatry. Appropriate use of such treatments presupposes patients who meet the diagnostic criteria and clinicians who have accurately diagnosed the target disorder and comorbid conditions. Since little is known about whether these prerequisites are commonly met, the authors conducted a study at two community treatment sites to determine the frequency of various axis I diagnoses and the concordance between the diagnoses recorded in patient charts and those obtained by a structured interview. Given that a DSM diagnosis may not be sufficient to understand a patient's problems, the authors also obtained ratings of interpersonal functioning. **Method:** The subjects were 164 nonpsychotic patients who were seen at a rural (N=114) or urban (N=50) community treatment facility. Raters trained to reliably use the Structured Clinical Interview for DSM-IV (SCID) conducted diagnostic interviews. Clinical charts were reviewed to obtain clinical diagnoses. Patients completed questionnaires regarding interpersonal functioning. **Results:** Most (N=145, 88%) of the patients met the SCID criteria for a current axis I diagnosis, and 53% (N=87) met the criteria for two or more disorders. Clinical and SCID diagnoses had poor agreement. Evidence was found for interpersonal dysfunction. **Conclusions:** Most patients met the diagnostic criteria for conditions for which there are proven treatments; however, inaccurate diagnosis proved common. This barrier to optimal treatment could be ameliorated with the use of structured interviews for common diagnoses. Scores on social/interpersonal measures support the premise that DSM symptoms provide only part of the relevant information about patients' conditions.

(Am J Psychiatry 2000; 157:581–587)

Good medical treatment rests on a foundation of accurate diagnosis. Only two decades ago, however, it seemed that the reliability of psychiatric diagnosis was

Presented in part at the 126th annual convention of the American Public Health Association, Washington, D.C., November 15–19, 1998. Received Dec. 14, 1998; revisions received June 7, Jun 30, and Sept. 27, 1999; accepted Sept. 30, 1999. From the Clinical Services Research Program for Women and the Clinical Research Center for Mid-Life Mood Disorders, Western Psychiatric Institute and Clinic, Department of Psychiatry, University of Pittsburgh Medical Center. Address reprint requests to Dr. Shear, Western Psychiatric Institute and Clinic, Department of Psychiatry, University of Pittsburgh Medical Center, 3811 O'Hara St., Pittsburgh, PA 15213; shearmk@msx.umpc.edu (e-mail).

Supported by NIMH grant MH-53817, NIMH Clinical Research Center grant MH-52247, and a grant from the NIMH Mental Health Intervention Research Center for Mood and Anxiety Disorders (MH-30915).

The authors thank Barbara Kumer, Howard Stein, and Nancy Stack for technical assistance and Structured Clinical Interview for DSM-IV trainer Susan Wheeler and rater Beverly Sullivan for their help.

a “hopeless undertaking” (1). Studies demonstrated low interrater agreement with the use of DSM-II criteria (2, 3), thought to be because of both criteria and information variance (4, 5). Researchers addressed these problems by developing the Research Diagnostic Criteria (6) and the Schedule for Affective Disorders and Schizophrenia (7). In 1980, DSM-III brought clearly defined diagnostic criteria to the clinical arena. Improved reliability was demonstrated in early field trials (8), but semistructured interviews and procedures for training and certifying raters have since been used in studies to confirm the reliability of DSM diagnoses (9–12).

Proven efficacious treatments are now available for specific psychiatric disorders. Patients who participate in efficacy studies meet the DSM-IV criteria for targeted disorders, but it is not known whether most patients in community settings meet these criteria. In addition, the precision with which the DSM criteria are used in usual clinical practice is unknown. Although

studies have demonstrated that training in the use of DSM-III can produce good reliability (13), it is likely that little training has occurred in busy clinical settings. Moreover, most clinicians still evaluate patients using an unstructured, open-ended approach, which may not be aimed at establishing the presence of diagnostic criteria. Thus, the accuracy of clinical diagnosis may be low. One way to assess accuracy is to use research-tested methods to examine concordance with clinical diagnoses. We undertook such a project, comparing chart-recorded clinical diagnoses with those made by trained raters using the Structured Clinical Interview for DSM-IV (SCID) (14).

In a recent editorial, Tucker (15) noted that current DSM diagnoses provide only part of the information we need to treat patients. In particular, he drew attention to the context of the patient's social/interpersonal life. We agree that in addition to accurate diagnosis, the patient's interpersonal life is clinically important and is not addressed by a structured diagnostic interview. Thus, we included assessment of several dimensions of interpersonal functioning in our study. We used a research-oriented approach to this assessment, choosing areas with evidence for an impact on psychiatric symptoms and measures known to have good psychometric properties.

The purpose of this article is to present data from two community mental health facilities (one rural and one urban) in western Pennsylvania, including rates of various axis I diagnoses, as determined by trained raters using the SCID, and the concordance between SCID diagnoses and those recorded in patient records. In addition, we report scores on self-reported measures of interpersonal problems—social support, partner abuse, and overall life functioning.

METHOD

Subjects

We recruited consecutive consenting nonpsychotic adults, aged 18–65 years, who were seen for outpatient treatment at either a free-standing clinic in rural Pennsylvania (N=114) or the open treatment clinic at our urban academic medical center (N=50). According to 1990 census information (16), the rural clinic services a county with a population of roughly 186,000 and a median income of \$29,455. The urban site services a catchment area of approximately 250,000 within a county with a population of 1,336,449 and a median income of \$35,136. Women living alone with children, a group at risk for depression (17), have an estimated income of \$11,439 and \$14,464 in the rural and urban counties, respectively.

Data collection occurred between April 1, 1996, and August 6, 1997. Subjects meeting the study criteria were identified by a research associate after their initial clinical evaluation and were consecutively recruited by telephone. A minimum of three telephone calls were made over a 2-week period, including at least one evening call. During an 8-month recruitment period, 422 patients were screened at the rural clinic. Fifty-four were psychotic and were thus excluded, 28 had serious medical problems and were considered unable to participate, and 27 were excluded for a variety of other reasons. Of the 313 eligible subjects, we were unable to reach 119, and 35 refused to participate. SCID interviews were scheduled with the remaining 159 patients. Forty-five of these failed to keep their ap-

pointments and/or complete the interview, resulting in 114, including 78 women and 36 men, who completed the assessment.

At the urban academic medical center, 251 patients were screened during a 3-month enrollment period. Sixty were excluded because of participation in other research projects, one because of a severe medical illness, two for unclear reasons, and one refused to be contacted. Of the 187 eligible subjects, we were unable to reach 85, and 20 refused to participate. SCID interviews were scheduled with the remaining 82. Thirty-two of these failed to keep their appointments and/or complete the interview, resulting in 50 who completed the assessment.

Clinical charts, available for 309 of the 313 eligible rural subjects and all 187 of the eligible urban subjects, were reviewed to obtain clinical diagnoses, demographics, and insurance coverage. The first listed clinical diagnosis in the charts was "primary diagnosis," the likely focus of treatment. Clinicians sometimes indicated "rule out" as a modifier of the diagnosis, and this was noted in the chart review.

There were no differences between the participants and nonparticipants in mean ages, proportions of women, racial makeups, rates of primary clinical diagnoses, or types of insurance coverage, categorized as government, private, or self-pay. Patients in the rural clinic were significantly less well educated, tended to have lower incomes, and had significantly more children than those at the urban clinic but did not differ on other variables, such as age, sex, marital status, or insurance coverage.

Procedures

Intake assessment at the rural clinic was performed by a master's-level clinician hired by the clinic specifically for this purpose. Procedures called for a confirmatory diagnostic examination by a physician. However, because of physician shortages during the study period, only 61 of 114 (54%) of the rural subjects had a physician assessment recorded in their charts within the 3-month review period, and the mean time between clinician and physician evaluations was 35 days. All clinical assessments at the urban clinic were completed by nonphysician mental health professionals in consultation with a supervising psychiatrist, who co-signed the assessment. In-person assessment by the physician was not mandated, and when it occurred, the physician's opinion was reflected in the clinician's diagnosis.

Research diagnostic assessments at both sites were performed by one of two experienced nonphysician clinicians (registered nurse or licensed social worker) trained in the use of the SCID and certified according to the standards of the Biometrics Division of the New York State Psychiatric Institute. Such certification requires 100% concordance with a certified SCID rater for the primary diagnosis and for the presence or absence of comorbidity on four consecutive interviews. On this instrument, a given axis I diagnosis is coded as present or absent for the lifetime and, if present in the past month, as current. The primary diagnosis is the one that, in the judgment of the interviewer, should be the focus of treatment. All study SCID interviews were audiotaped and reviewed by the SCID trainer. In addition, when questions arose, an interview was reviewed during weekly visits by a supervisor. Any unresolved questions or disagreements were discussed and resolved with the first author.

All subjects completed self-report questionnaires assessing functional impairment on the Sheehan Disability Scale (18, 19) and the Medical Outcomes Study 36-item Short-Form Health Survey (20). To evaluate interpersonal life contexts, we used three well-validated measures. The Inventory of Interpersonal Problems is a lengthy questionnaire designed to rate interpersonal problems for which individuals seek psychotherapy. We used a shortened version of this instrument (21–24), which profiles interpersonal sensitivity, ambivalence, and aggression and can be used to estimate the presence of personality disorder. Because low social support has been associated with mental and physical illness (25), we included the Interpersonal Support Evaluation List (26), which measures four dimensions of support: tangible (instrumental or material support), appraisal (availability of someone to talk to about problems), self-esteem (positive comparison of oneself with others), and belonging (people one can do things with). Norms for the four subscales in general population studies range from 32.9 to 34.4 (SD=5.0–6.0) (26). We also ad-

ministered the Partner Abuse Scale (27, 28) to detect domestic violence, identified at the time by the surgeon general as "the number one public health problem in America" (29); it is associated with a risk for mental illness (30). Reported mean scores for a comparison population are 0.90 (SD=5.13) for physical and 6.94 (SD=12.29) for verbal abuse.

Statistical Methods

Comparisons between the two clinics were conducted by using *t* tests for continuous and chi squares for categorical variables. Agreement between SCID and clinical diagnoses was evaluated by using the kappa, a test for the independence of frequencies that corrects for the base rate (31). Comparisons of subsets of subjects (e.g., participants versus nonparticipants, patients with versus patients without physician assessment, or diagnostic match versus no diagnostic match) were conducted by using *t* tests when two groups were compared or with factorial analysis of variance when relationships among multiple factors were sought. All tests were two-tailed.

Permission to conduct the study was granted by the University of Pittsburgh's institutional review board. After a complete description of the study to the subjects, written informed consent was obtained. The subjects were reimbursed \$20 for completion of the structured interview and the self-report questionnaire. We also obtained permission to review the charts of subjects who were not participants to determine whether they were significantly different from those of participants in demographic or diagnostic characteristics. Identifying information for these subjects has been deleted from the data set.

RESULTS

DSM-IV Axis I Diagnosis on SCID Interview

It is somewhat surprising that there was little difference in the diagnostic profiles of patients from the urban academic clinic and patients from the rural clinic; data from the two groups are combined for most analyses. For the structured interview, 145 of 164 (88%) of the patients met the full SCID criteria for at least one current axis I disorder. Among the 19 who did not, 12 had a past DSM-IV diagnosis that did not meet all of the criteria for a current diagnosis at the time of the interview, but they were considered by the interviewer to have symptoms sufficiently severe to warrant treatment. The frequencies of primary diagnoses—grouped by mood disorder, anxiety disorder, adjustment disorder, and all other disorders—are presented in table 1.

The majority (N=96 of 164, 59%) of the patients met the SCID criteria for a primary depressive disorder. Depression was four times as common as any anxiety disorder (N=23 of 164, 14%) as a primary diagnosis; all other primary diagnoses were far less common. Somewhat more women than men (N=69 of 111, 62%; N=26 of 53, 49%, respectively) were diagnosed with a primary depressive disorder ($\chi^2=2.90$, *df*=1, *p*=0.09).

Table 2 shows the frequency of all current axis I diagnoses. Examined in this way, depression is still the most common condition, but anxiety disorders are almost as frequent. Thus, comorbid anxiety is common. Considering all current diagnoses, 53% (N=87) of the patients met the criteria for two or more current axis I diagnoses, and 29% (N=48) met the criteria for three or more. One hundred twenty-eight (78%) of the pa-

TABLE 1. Agreement of Categorical SCID and Clinical Primary Diagnoses for 164 Nonpsychotic Patients in Two Community Clinics

Category of SCID Primary Diagnosis	Category of Clinical Primary Diagnosis							
	Mood Disorder		Anxiety Disorder		Adjustment Disorder		Other Disorder	
	N	%	N	%	N	%	N	%
Mood disorder (N=96)	49	51	4	4	32	33	11	12
Anxiety disorder (N=23)	5	22	6	26	9	39	3	13
Adjustment disorder (N=19)	6	32	0	0	10	53	3	16
Other disorder (N=26)	6	23	0	0	7	27	13	50
Total (N=164)	66	40	10	6	58	36	30	18

tients were diagnosed with depression and/or anxiety, of whom 32% (N=41) met the SCID criteria for depression without anxiety, 25% (N=32) for anxiety without depression, and 43% (N=55) for both anxiety and depression. Women were almost twice as likely as men (N=82, 50%; N=44, 27%, respectively) to meet the criteria for comorbid anxiety and depression ($\chi^2=5.91$, *df*=2, *p*=0.05).

Although we did not formally evaluate axis II disorders, patients completed an abbreviated form of the Inventory of Interpersonal Problems (21). A mean score of 1.1 or higher on this instrument has been reported to indicate probable personality disorder (22–24). Seventy-seven percent (N=126) of all subjects and 91% (N=79 of 87) of those with two or more axis I disorders scored in this range. Subjects with scores on the Inventory of Interpersonal Problems of 1.1 or higher were more likely to have two or more concurrent axis I diagnoses than were those with lower scores (N=105, 64%; N=43, 26%, respectively) ($\chi^2=17.07$, *df*=3, *p*=0.001).

Concordance Between Clinical and SCID Diagnoses

Chart diagnosis often did not concur with results on the SCID. The kappa was 0.24 for interrater reliability of primary diagnosis (categorized as depressive disorder, anxiety disorder, adjustment disorder, or other disorder).

Among those with a primary diagnosis of an anxiety disorder per the SCID, 26% (N=6 of 23) of chart diagnoses also identified an anxiety disorder. Concordance was better for a primary SCID diagnosis of depression, for which 51% (N=49 of 96) of the clinical records also identified depression; however, there were also more false positive diagnoses of depression. Among the 49% (N=47 of 96) of the patients for which depression was not recorded on the chart, 68% (N=32) were assigned "adjustment disorder," "V code," or "no clinical diagnosis." The remainder received diagnoses of anxiety disorder, substance abuse, or bipolar disorder. The SCID diagnoses for the 16 cases of unconfirmed clinical depression included adjustment dis-

TABLE 2. Frequencies of All Current SCID and Clinical Diagnoses, and Concordance of Diagnostic Categories, for 164 Non-psychotic Patients in Two Community Clinics

Diagnosis	Patients With Diagnosis			Kappa	95% Confidence Interval
	SCID Only	Clinical Only	Both SCID and Clinical		
Bipolar					
I	4	4	2		
II	1	4	0		
Other	0	0	1		
Major depressive disorder	45	15	39		
Dysthymia	10	12	2		
Depression not otherwise specified	2	11	0		
Mood disorder due to medical condition	1	0	0		
Substance-induced mood disorder	2	1	0		
Schizophrenia	0	1	0		
Schizoaffective disorder	0	1	0		
Substance-induced psychosis	2	0	0		
Psychosis not otherwise specified	3	1	0		
Substance abuse or dependence					
Alcohol	3	14	3		
Sedative	1	1	0		
Cannabis	5	3	1		
Opioid	1	1	1		
Cocaine	3	4	1		
Hallucinogen	2	0	0		
Polysubstance	2	0	0		
Panic disorder	25	6	2		
Agoraphobia	2	0	0		
Social phobia	22	1	1		
Specific phobia	22	2	1		
Obsessive-compulsive disorder	14	1	1		
Posttraumatic stress disorder	22	1	1		
Generalized anxiety disorder	18	3	1		
Anxiety disorder not otherwise specified	11	4	0		
Somatization	0	1	0		
Undifferentiated	2	0	0		
Hypochondriasis	1	1	0		
Body dysmorphism	0	0	1		
Anorexia nervosa	1	0	0		
Bulimia nervosa	2	2	0		
Binge eating disorder	6	2	0		
Adjustment disorder	5	58	6		
Other disorder	5	7	0		
Summary					
Any bipolar disorder	6	8	2	0.18	−0.09 to 0.44
Any mood disorder	40	16	55	0.33	0.19 to 0.47
Any anxiety disorder	70	5	17	0.12	0.03 to 0.22
Any substance abuse or dependence	11	17	9	0.29	0.10 to 0.49
Any eating disorder	6	3	2	0.28	−0.05 to 0.61
Any adjustment disorder	5	58	6	0.07	−0.04 to 0.18

order (N=5), anxiety disorder (N=5), substance abuse (N=3), and other disorder (N=3).

Table 2 also summarizes the similarities and differences between current diagnoses when made by means of the SCID rater or a clinical rater. The first column of table 2 shows the 36 diagnoses given to any patient by the SCID rater or by clinical diagnosis. The second column shows the number of times each diagnosis was given by the SCID rater only, the third column shows the number of times each diagnosis was given by a clinical rater only, and the fourth column shows the number of times the raters agreed. Several features of table 2 are of note. Raters agreed in a pronounced minority of cases. Overall, use of the SCID resulted in more diagnoses than did standard clinical procedures. Anxiety disorders, in particular, were much more likely

to be diagnosed by the SCID rater than by a clinical rater. The one notable exception was "adjustment disorder," which was diagnosed much more frequently by a clinical rater than by the SCID rater.

When the 36 diagnoses were considered separately, kappas could not be computed consistently because of the many empty cells. The diagnoses were collapsed into groups for the calculation of kappas. When the diagnoses were collapsed, the SCID rater and clinical raters were considered to agree if, for example, the clinical rater diagnosed major depression and the SCID rater diagnosed dysthymia. Correspondingly, the number of agreements is larger in the summary section of table 2. Nonetheless, kappas are still very low, and in half of the cases, the confidence intervals for the kappas include zero.

Neither demographic variables (sex, age, race, marital status, income, and education) nor severity of illness were associated with diagnostic agreement. Indices of environmental stress (low social support, partner abuse, and exposure to violence) did not predict assignment to the adjustment disorder category. However, agreement on a primary diagnosis (mean axis I diagnoses for a match=1.6, SD=0.8; mean axis I diagnoses for no match=2.0, SD=1.5) was associated with significantly fewer axis I diagnoses ($t=2.04$, $df=148.4$, $p=0.04$), suggesting that a more complex pattern of symptoms was associated with more disagreement on a DSM diagnosis.

In the rural clinic, separate diagnoses were provided by physicians for a subset of the patients. The 61 patients who had a physician diagnosis recorded within the review period had significantly more treatment visits than did the 53 with no physician diagnosis (mean=4.0, SD=2.8; mean=2.4, SD=3.4, respectively) ($t=-2.77$, $df=110.0$, $p=0.007$). In addition, the patients who received physician assessments were more likely than those without physician assessments to meet the SCID criteria for depression ($N=43$ of 61 participants, 70%; $N=23$ of 53 nonparticipants, 43%) ($\chi^2=9.14$, $df=2$, $p=0.01$), to meet the criteria for more current SCID diagnoses (mean=2.1, SD=1.2; mean=1.6, SD=1.7) ($t=-2.07$, $df=112$, $p=0.04$), and to achieve higher scores on the SCID self-report measures of impairment. Agreement between the SCID and physician diagnoses was no better than that between the clinician and SCID diagnoses. The kappa for interrater reliability, calculated as agreement between the physician and SCID rater on the primary diagnosis, as previously, was 0.15. The kappa for agreement between physician and clinician diagnoses was similarly low.

Clinicians frequently used a designation of "rule out." Specifically, 49% of the clinicians' depression diagnoses from patient charts were designated "rule out," as well as more than 50% of the recorded anxiety disorders diagnoses. The kappa was not improved by allowing for a match of any SCID diagnosis with any clinical diagnosis, including rule outs.

Interpersonal Functioning and Indices of Illness Severity and Impairment

Questionnaire results indicated that social support was low (all subscale means ≤ 20 , SDs ≤ 7), and there was evidence of partner abuse in a significant subgroup (mean score on Partner Abuse Scale, verbal: mean=16.4, SD=22.5; physical: mean=3.4, SD=10.2).

Unprotected t tests were used to compare patients who met the SCID criteria for a primary diagnosis of a depressive disorder with those who did not. The alpha levels of the t tests were significant for scores on the Medical Outcomes Study Short-Form Health Survey (general health and mental health), the Inventory of Interpersonal Problems, the Sheehan Disability Scale, and the Interpersonal Support Evaluation List. Patients with depression reported more impairment (lower

scores) than patients without depression on the general health (mean=54, SD=23; mean=65, SD=25, respectively) ($t=2.86$, $df=159$, $p=0.009$) and mental health (mean=31, SD=19; mean=54, SD=21) ($t=8.81$, $df=159$, $p=0.001$) summary scores of the Medical Outcomes Study health survey and more impairment (higher scores) on the Sheehan Disability Scale (mean=22.9, SD=8.9; mean=17.5, SD=11.3) ($t=-2.85$, $df=115$, $p=0.006$). Similarly, those with depressive disorders had lower social support than those without depressive disorders, as indexed by the Interpersonal Support Evaluation List (e.g., total mean=66.2, SD=23.3; total mean=81.0, SD=20.2, respectively) ($t=3.96$, $df=121$, $p=0.001$). Interpersonal problems did not differentiate depressed from nondepressed individuals.

In women ($N=89$), but not in men ($N=47$), scores for verbal and physical abuse were correlated with the ratings for illness burden and impairment. Physical abuse correlates included the number of current diagnoses ($r=0.34$, $p<0.01$; $r=0.18$, n.s., respectively), the Medical Outcomes Study Short-Form Health Survey's pain index ($r=-0.27$, $p<0.02$; $r=0.16$, n.s.), and the Sheehan Disability Scale ($r=0.30$, $p<0.01$; $r=0.06$, n.s.); verbal abuse correlates included the Medical Outcomes Study health survey's pain index ($r=-0.30$, $p<0.01$; $r=0.22$, n.s.), social functioning (Medical Outcomes Study's Short-Form Health Survey) ($r=-0.37$, $p<0.01$; $r=-0.17$, n.s.), the Sheehan Disability Scale ($r=0.42$, $p<0.01$; $r=0.08$, n.s.), and social support (Interpersonal Support Evaluation List) ($r=0.32$, $p<0.01$; $r=0.16$, n.s.).

DISCUSSION

Diagnosis-specific treatments are important innovations in psychiatry. Clinicians increasingly are expected to follow treatment guidelines designating the use of such treatments (32, 33). However, such guidelines can be properly used only for the treatment of patients who meet the relevant DSM criteria, rather than mixed symptoms or syndromes, mild, subsyndromal disorders, or adjustment disorders. We need to know the frequency of conditions such as major depression and panic disorder, for which guidelines have been published. The proper implementation of treatment guidelines also presupposes an accurate clinical diagnosis. Little is currently known about such accuracy.

This report is one of the first diagnostic studies of patients who were seen at community mental health treatment facilities. We found that most patients did meet the criteria for a DSM-IV axis I diagnosis, as determined by a research-quality structured diagnostic interview. Primary diagnoses and concurrent disorders were strikingly similar for patients treated in a free-standing, rural community mental health clinic and an urban university-affiliated open treatment clinic. In both settings, most patients met the criteria for more than one current diagnosis, and a majority were depressed. Given the high frequency of potentially treatable axis I diagnoses, we conclude that it is important

to find ways to ensure that guidelines recommending proven efficacious treatments are implemented in the community.

We found that clinicians in the community are unlikely to record diagnoses in patient charts that are concordant with those obtained by a trained rater using a semistructured interview. Agreement between clinical and SCID diagnoses was remarkably low, with kappas in the range of 0.1–0.3. Even considering a broad definition of agreement for a prevalent disorder—depression—nearly one-half still failed to agree. Why might this be?

It is possible that chart diagnoses did not reflect the clinicians' true assessment (e.g., chart diagnosis made for insurance purposes). This seems unlikely since agreement was low on depression, a diagnosis made frequently by both clinicians and the SCID rater. Diagnostic assessments in these community clinics were performed by nonphysician clinicians, and these individuals may place less emphasis on diagnosis as a critical part of treatment than do physicians. However, in the subset of patients in the rural clinic for whom we could evaluate physician diagnoses, agreement with the SCID was also low. We believe a likely explanation is a lack of specific diagnostic training for community practitioners and the fact that clinicians do not use a standardized method of eliciting criteria. There is good reason to believe that such procedures substantially improve diagnostic accuracy. Given the frequency of depression and comorbid anxiety disorders, it would make sense to implement structured interviews for these common disorders. We are aware that many clinicians believe that an unstructured interview is needed for the development of therapeutic rapport and that the use of a structured interview might even interfere with the therapeutic relationship. Several of the clinicians whom we trained as SCID raters for this project approached their work with this expectation. These individuals were surprised to report that patients seemed to appreciate the SCID and often thanked the interviewer for asking good questions that helped the patients feel understood.

We are also cognizant of the fact that many clinicians have been trained to conduct more open-ended, broad-ranging interviews, leading to a greater awareness of interpersonal dysfunction. This sensibility may lead clinicians to believe there is a need to assess and treat problems other than DSM-IV axis I disorders, and we believe such a viewpoint may be correct (34). We found clear evidence of serious disturbances in the social/interpersonal lives of study subjects, which was more pronounced in depressed patients. For some, this included clinically significant partner abuse. Although empathic, open-ended interviewing for the evaluation of a patient's life context is often used, we advocate the inclusion of standard, reliable measures of social functioning as part of the outcome assessment. Such a strategy is more likely to convince outsiders of the reliability of the assessment and will likely result in better intersite reliability.

A limitation of this study is that in spite of repeated efforts and the provision of monetary reimbursement, we were able to reach only about one-third of the eligible patients to recruit them for participation. Without underestimating the importance of recruiting a high percentage of eligible subjects in a study such as this, we were somewhat reassured when our chart review of eligible individuals suggested that there were few systematic differences between participants and nonparticipants. Participation in our study did appear to mirror treatment attendance, because nonparticipants had significantly fewer clinical visits than participants. Thus, in addition to the impact on the generalization of research results, low accrual in this study draws attention to the serious problem of attrition in treatment settings, a well-known, but poorly understood, phenomenon. Also of note, in many studies the pool of eligible subjects is not identified, and the rate of enrollment is not reported. We suspect that our rates are similar to usual research accrual rates and suggest that researchers include this information in reports of study results.

In summary, a high percentage of nonpsychotic patients seen for treatment in community mental health settings met the DSM-IV criteria for major depression, often with co-occurring anxiety disorders. Thus, patients in these clinics are sufficiently similar to individuals who have participated in clinical research studies in academic medical centers to warrant attempts to disseminate proven treatments for anxiety and depression. However, inaccurate diagnosis was frequent and may be an important barrier to the implementation of such treatments. We suggest that this problem could be ameliorated with focused training and the use of structured interviews. In addition, there was clear evidence for disturbances in the social/interpersonal lives of the patients we assessed. We thus urge that standardized information about interpersonal lives, including social support and questioning about domestic violence, be considered an essential component of mental health evaluations and outcome assessment.

REFERENCES

1. Grove WM, Andreasen NC, McDonald-Scott P, Keller MB, Shapiro RW: Reliability studies of psychiatric diagnosis: theory and practice. *Arch Gen Psychiatry* 1981; 38:408–413
2. Beck AT, Ward CH, Mendelson M, Mock JE, Erbaugh JK: Reliability of psychiatric diagnosis, 2: a study of consistency of clinical judgments and ratings. *Am J Psychiatry* 1962; 119: 351–357
3. Kendell RE, Cooper JE, Gourlay AJ, Copeland JR, Sharpe L, Gurland BJ: Diagnostic criteria of American and British psychiatrists. *Arch Gen Psychiatry* 1971; 25:123–130
4. Segal DL, Falk SB: Structured interviews and rating scales, in *Behavioral Assessment: A Practical Handbook*, 4th ed. Edited by Bellack AS, Hersen M. Needham Heights, Mass, Allyn & Bacon, 1998, pp 158–178
5. Spitzer RL, Endicott J, Robins E: Clinical criteria for psychiatric diagnosis and DSM-III. *Am J Psychiatry* 1975; 132:1187–1192
6. Spitzer RL, Endicott J, Robins E: Research Diagnostic Criteria: rationale and reliability. *Arch Gen Psychiatry* 1978; 35: 773–782

7. Endicott J, Spitzer RL: A diagnostic interview: the Schedule for Affective Disorders and Schizophrenia. *Arch Gen Psychiatry* 1978; 35:837–844
8. Spitzer RL, Forman JB, Nee J: DSM-III field trials, I: initial interrater diagnostic reliability. *Am J Psychiatry* 1979; 136:815–817
9. Williams JBW, Gibbon M, First MB, Spitzer RL, Davies M, Borus J, Howes MJ, Kane J, Pope HG, Rounsaville B, Wittchen H-U: The Structured Clinical Interview for DSM-III-R (SCID), II: multisite test-retest reliability. *Arch Gen Psychiatry* 1992; 49:630–636
10. Helzer JE, Robins LN, McEvoy LT, Spitznagel EL, Stoltzman RK, Farmer A, Brockington IF: A comparison of clinical and diagnostic interview schedule diagnoses: physician reexamination of lay-interviewed cases in the general population. *Arch Gen Psychiatry* 1985; 42:657–666
11. Di Nardo PA, Barlow DH: Anxiety Disorders Interview Schedule, Revised (ADIS-R). Albany, State University of New York at Albany, Phobia and Anxiety Disorders Clinic, 1988
12. Segal DL, Hersen M, Van Hasselt VB: Reliability of the Structured Clinical Interview for DSM-III-R: an evaluative review. *Compr Psychiatry* 1994; 35:316–327
13. Webb LJ, Gold RS, Johnstone EE, Diclemente CC: Accuracy of DSM-III diagnoses following a training program. *Am J Psychiatry* 1981; 138:376–378
14. First MB, Spitzer RL, Gibbon M, Williams JBW: Structured Clinical Interview for DSM-IV Axis I Disorders (SCID). New York, New York State Psychiatric Institute, Biometrics Research, 1995
15. Tucker GJ: Putting DSM-IV in perspective. *Am J Psychiatry* 1998; 155:159–161
16. Pennsylvania State Data Center: Beaver County, PA—1990 Census Profile <http://psdc.hbg.psu.edu/psdc/cou_profiles/c007.html>
17. Brown GW, Harris TO: Aetiology of anxiety and depressive disorders in an inner-city population, 1: early adversity. *Psychol Med* 1993; 23:143–154
18. Leon AC, Shear MK, Portera L, Klerman GL: Assessing impairment in patients with panic disorder: the Sheehan Disability Scale. *Soc Psychiatry Psychiatr Epidemiol* 1992; 27:78–82
19. Sheehan DV: The Anxiety Disease. New York, Scribner, 1983
20. Ware JE, Snow KK, Kosinski M, Gandek B: SF-36 Health Survey Manual and Interpretation Guide. Boston, Health Institute, New England Medical Center, 1993
21. Horowitz LM, Rosenberg SE, Baer BA, Ureno G, Villaseñor VS: Inventory of Interpersonal Problems: psychometric properties and clinical applications. *J Consult Clin Psychol* 1988; 56:885–892
22. Pilkonis PA, Kim Y, Proietti JM, Barkham M: Scales for personality disorders developed from the Inventory of Interpersonal Problems. *J Personal Disord* 1996; 10:355–369
23. Kim Y, Pilkonis PA: Selecting the most informative items in the IIP scales for personality disorders: an application of item response theory. *J Personal Disord* 1999; 13:157–174
24. Kim Y, Pilkonis PA, Barkham M: Confirmatory factor analysis of the personality disorder subscales from the Inventory of Interpersonal Problems. *J Pers Assess* 1997; 69:284–296
25. Wilson IB, Cleary PD: Linking clinical variables with health-related quality of life: a conceptual model of patient outcomes. *JAMA* 1995; 273:59–65
26. Cohen S, Mermelstein R, Kamarck T, Hoberman H: Measuring the functional components of social support, in *Social Support: Theory, Research and Application*. Edited by Sarason IG, Sarason BG. The Hague, the Netherlands, Martinus Nijhoff, 1985, pp 73–94
27. Attala JM, Hudson WW, McSweeney M: A partial validation of two short-form Partner Abuse Scales. *Women Health* 1994; 21:125–139
28. Hudson WW: The WALMYR Assessment Scales Scoring Manual. Tempe, Ariz, WALMYR, 1992
29. US Senate Judiciary Committee: Violence Against Women: A Week in the Life of America. Washington, DC, US Senate Judiciary Committee, Hart Office Bldg, Rm B04, 1992
30. Saunders DG: Posttraumatic stress symptom profiles of battered women: a comparison of survivors in two settings. *Violence Vict* 1994; 9:31–44
31. Bloch DA, Kraemer HC: 2×2 kappa coefficients: measures of agreement or association. *Biometrics* 1989; 45:269–287
32. American Psychiatric Association: Practice Guideline for Major Depressive Disorder in Adults. *Am J Psychiatry* 1993; 150(April suppl)
33. American Psychiatric Association: Practice Guideline for the Treatment of Patients With Panic Disorder. *Am J Psychiatry* 1998; 155(May suppl)
34. Cohen S, Hoberman HM: Positive events and social supports as buffers of life change stress. *J Applied Soc Psychol* 1983; 13:99–125