Recognizing Depression: A Comparison Of Family Physician Ratings, Self-Report, And Interview Measures

James C. Coyne, Ph.D., Thomas L. Schwenk, M.D., and Mark Smolinski, M.D.

Abstract: Major depressive disorder is the most common diagnosis encountered in family practice, yet family physicians are relatively unlikely to make the diagnosis. This study compared physician ratings of depression with scores from the Center for Epidemiological Studies-Depression (CES-D) questionnaire and with telephone interview diagnoses of depression using the 3rd revised edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R) criteria for major depressive disorder in a population of 266 patients in community-based family practices. Additional assessments were made of health status, stress, social support, prescribed psychotropic medication, and counseling. The prevalence of positive questionnaire scores in this population was 22.6 percent, and the prevalence of major depressive disorder (based on telephone interview) was 8 percent. Physician ratings of depression were relatively inaccurate when compared with either CES-D scores or telephone interview diagnoses. Optimum specificity (80 percent) and sensitivity (50 percent) with telephone interview diagnoses were achieved when physicians rated the patient as having any depression versus having no depression. Physician ratings of depression were correlated with their assessment of patient stress, social support, and physical health but not with more objective measures of these variables.

When compared with telephone interview diagnoses, the sensitivity and specificity of the CES-D scores were relatively poor, suggesting that the CES-D is not useful as a screening tool for selected populations. Finally, we found that family physicians base their assessments of depression more on distress than on depressive symptoms. Certain physician myths, barriers, and biases may exist that preclude the effective diagnosis of depression. (J Am Board Fam Pract 1991; 4:207-15.)

Epidemiological studies suggest that major depression surpasses hypertension as the most frequent illness encountered in family practice.1 It can be debilitating, and even in the absence of the full syndrome, depressive symptoms are associated more with functional impairment than arthritis, hypertension, or diabetes.2

Despite these findings, clinically significant depressive symptoms of most family practice patients go undetected and do not receive treatment.3,4 When compared with patient self-report of depressive symptoms on questionnaires, detection by family physicians is particularly low.5-7 If the discordance between physician ratings and screening questionnaires does indeed represent a pool of patients with undiagnosed depression, then the burden of suffering is large, particularly for those patients feeling helpless and hopeless, who experience impairment in work and social life, and for whom the associated somatic symptoms cause unnecessary medical evaluations and treatment.

Few family practice patients with significant depressive symptoms identify depression as the reason for their visit or complain directly of depression.5,8,9 Physicians must therefore either inquire routinely about depressive symptoms or rely on other information to signal that such an inquiry is warranted. Suggestions that physicians might improve their detection of depression by administering self-report screening questionnaires to their patients have met with considerable controversy.10,11 Whereas the sensitivity of self-report questionnaires is high, their specificity is low to moderate, and their positive predictive

From The Department of Family Practice, University of Michigan Medical Center, 1018 Fuller Street, Ann Arbor. Address reprint requests to James C. Coyne, Ph.D., University of Michigan Medical Center, Department of Family Practice, 1018 Fuller Street, Ann Arbor, MI 48109-0708.

This research was supported in part by a National Institute of Mental Health Grant 1-R01-MH43796-01A2.
value is low.\textsuperscript{11} Further, much of the depression that these instruments detect is so minor, situational, or transient that antidepressant medication would not be indicated.\textsuperscript{12} Thus, discordance between physician judgments and these instruments might in part reflect the physician appropriately ignoring symptoms for which major pharmacological or psychotherapeutic action would be inappropriate.

In an effort to shed light on this issue, we developed a study to determine the prevalence of depressive symptoms in a community-based family practice population using a standardized screening instrument. Patient scores were compared with physician ratings and, for selected patients, with the results of a diagnostic interview. Patients selected on the basis of elevated depression scores were given a structured telephone interview that contained questions derived from the \textit{Diagnostic and Statistical Manual of Mental Disorders} (DSM-III-R) criteria for major depression.\textsuperscript{13} Telephone administration of a diagnostic interview has produced results comparable to those obtained face to face.\textsuperscript{14,15} The telephone interview also included previously validated sets of questions about stress, social support, physical health problems, and mental health care utilization to allow us to explore sources of discrepancies among the findings of the self-report questionnaire, the physician ratings, and the telephone assessment of depression.

**Methods**

Figure 1 summarizes the design of the study. The study sample of patients was obtained through the cooperation of an informal research network of private family physicians practicing in southeastern Michigan. The 6 participating physicians were young, residency-trained, board-certified, and practicing in rural and suburban areas. A research assistant was stationed in their reception rooms for 4-hour periods during which consecutive patients were approached and asked to participate in the study. Patients were told that the investigators were interested in studying the discussions patients have with their physicians and that involvement in the study might possibly include a subsequent telephone interview. The written description of the study given to patients described some of the telephone interview without mentioning depression. The study design and recruitment procedure were reviewed and approved by the institutional human subjects committee. Consent for the screening was obtained from 266 patients, more than 85 percent of the patients approached. The demographic characteristics and stated reasons for visit are summarized in Table 1. Because most patients were being seen for acute care, the physicians’ task of detecting depression tended to occur in a brief encounter structured around the articulation and diagnoses of other problems.

In the reception rooms, the patients completed the Center for Epidemiological Studies–Depression questionnaire (CES-D),\textsuperscript{16} as well as a set of demographic questions and 5-point self-rating scales for depression, stress, and social support. Checklists were attached to the charts of consenting patients, and immediately after the office visit, the physicians rated the patients for depression on a 5-point Likert-like scale ranging from 1 (no depression) to 5 (severe depression). Physicians also rated the patients’ physical health, stress level, and adequacy of social support, and they indicated the patients’ reasons for the visit and whether they were receiving either psychopharmacologic treatment or counseling.

Patients scoring more than 15 on the CES-D were called for a telephone interview within 10 days. This interview was conducted by a 3rd-year medical student familiar with DSM-III-R criteria for depression. Of the 61 patients who scored more than 15 (23 percent), 49 (80 percent) agreed to participate in the interview. In addition to depression, the telephone interview also included previously validated measures of stressful life events,\textsuperscript{17} social support,\textsuperscript{9,18} physical health,\textsuperscript{19} and mental health care utilization.\textsuperscript{20} The assessment of depression was not as detailed as the semi-structured interview schedules used in psychiatric
research, which can take 2 to 3 hours to administer and score. Instead, questions were designed to replicate the DSM-III-R criteria of duration of mood disturbance and the presence of psychological and vegetative symptoms about which a physician might inquire when depression is suspected. There were also questions concerning prescriptions for antidepressants.

**Measures**

**Depressive Symptoms**
The 20-item instrument that was used for screening, the CES-D, is internally consistent and well validated. Using a cutoff of 16 or greater, Hough, et al. found sensitivity to be 0.769 and specificity to be 0.528 for major depressive disorder. With the same score cutoff, Schulberg, et al. found that sensitivity of the CES-D was 0.963 and specificity was 0.386 for depressive disorders in primary medical care patients. In that study, 26 percent of the primary medical care patients scoring more than 15 and only 1 percent of those scoring 15 or lower were considered depressed when given a semistructured interview. We decided therefore that for the present study only those patients with scores more than 15 on the CES-D would be interviewed.

**Depressive Disorder**
The assessment of major depression during the telephone interview began with a series of questions inquiring whether in the past 6 months patients had experienced a period of 2 weeks or more in which they were sad, blue, moody, or down every day. For those answering "yes," there were additional questions concerning insomnia, fatigue, appetite, weight loss or gain, pessimism, and suicidal thoughts. Patients were considered to be depressed if they gave an affirmative answer to the question about duration of depressed mood and indicated the presence of at least four other symptoms consistent with DSM-III-R criteria. Because our diagnostic procedure failed to exclude those patients for whom depression is precipitated by organic factors, such as hypothyroidism, there could be a slight overdiagnosis of depression.

**Life Events**
For a measure of life events, we used the 53 undesirable life event items of the PERI Life Events Scale. The original PERI instrument consists of 102 life events that were generated from inductive interviews in an urban New York sample. Forty-nine of the items refer to desirable life changes, however, and only undesirable changes have been consistently related to depression.

**Social Support**
The measure of social support was drawn from questions previously validated in community surveys conducted by the Institute for Social Research, as well as a recent study of depressive symptoms in family practice patients. Questions refer to whether spouse, kin, and friends express an interest in and care about the respondents.

**Health Status**
The Health Status Questionnaire is a self-reported measure of specific somatic complaints and chronic conditions that we adopted with minimal modification from the questionnaire validated in past epidemiological studies. The questionnaire has satisfactory test-retest reliability and correlation with medical records, as well as a high correlation between patient reports and physician ratings using the same items.

### Table 1. Characteristics of Sample Screened in Family Physicians’ Reception Rooms (n = 266).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>No. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>37.8</td>
</tr>
<tr>
<td>Range</td>
<td>18–76</td>
</tr>
<tr>
<td>Employment</td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>206 (77)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>60 (23)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>94 (35)</td>
</tr>
<tr>
<td>Women</td>
<td>172 (65)</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>52 (20)</td>
</tr>
<tr>
<td>Married</td>
<td>181 (68)</td>
</tr>
<tr>
<td>Other</td>
<td>33 (13)</td>
</tr>
<tr>
<td>Reason for visit</td>
<td></td>
</tr>
<tr>
<td>Acute medical care</td>
<td>130 (49)</td>
</tr>
<tr>
<td>Medical follow-up</td>
<td>56 (21)</td>
</tr>
<tr>
<td>General physical examination</td>
<td>45 (17)</td>
</tr>
<tr>
<td>Gynecologic care</td>
<td>11 (4)</td>
</tr>
<tr>
<td>Other</td>
<td>24 (9)</td>
</tr>
</tbody>
</table>
In the present study, each respondent received three scores based on information obtained from the questionnaire. The first score was the total number of somatic symptoms reported. Symptoms included chest pain, back trouble, headaches, and recurrent abdominal pain. The second score was based on the number of chronic health problems reported, and the third was based on the number of such health-related impairments as hearing loss and limited mobility.

Mental Health Care Utilization
The measure of mental health care utilization was adapted from the Epidemiologic Catchment Area study.20 It inquires about seeking help for “problems with stress, emotions, nerves, drugs, alcohol or your mental health” from both lay and professional sources.

Results
As noted above, 61 (23 percent) of the 266 patients who were screened scored in the depressed range on the CES-D. This figure falls in the middle range of what has been reported in other studies using screening instruments (12 to 56 percent).29 CES-D scores were not related to reason for visit, age, or sex. As found in a previous study,9 our total sample had more women than men, but approximately equal proportions of women and men had elevated depressive symptom scores. Thus, the additional women with depressive symptoms were captured at consultation rather than within the reception room sample. Within this sample of high scorers, the formerly married were more represented among patients with high depression scores than were the single and currently married (χ² = 8.05, P < 0.005). Similarly, the unemployed were more represented than the employed (χ² = 6.15, P < 0.05). With a sample of 266, even small associations among continuous variables proved statistically significant, and so we limited our consideration of Pearson correlations to those of at least 0.32 (i.e., explaining 10 percent of the variance). CES-D scores were related to patient self-ratings of depression (r = 0.61), stress (r = 0.39), and physical health (r = 0.35), but not support.

There was only a moderate relation between physician ratings of depression and CES-D scores (r = 0.39, P < 0.001); thus, physician ratings of patient depression accounted for only 16 percent of the variance in screening questionnaire scores. Physician ratings of patient depression were related to physician ratings of patient health (r = 0.51, P < 0.001) and stress (r = 0.59, P < 0.001). Physician ratings were also related to patient self-ratings of depression (r = 0.39, P < 0.001).

Physicians reported that 8 (3 percent) of the 266 patients were prescribed antidepressants, 5 (2 percent) were prescribed anxiolytics, and 7 (3 percent) received counseling. The relation between physician ratings of depression and these therapeutic actions was modest (r = 0.29, P < 0.001 for prescribing antidepressants; r = 0.24, P < 0.001 for prescribing anxiolytics; and r = 0.32, P < 0.001 for counseling). The mean CES-D score for patients taking antidepressants was 22 versus 10.7 for patients who were not taking antidepressants (t = 2.78, P < 0.01). The three interventions tended to be mutually exclusive; none of the patients taking antidepressants received anxiolytics, and only 1 received counseling.

To calculate the physicians’ sensitivity and specificity of diagnosis, we dichotomized their ratings of patient depression by choosing a cutoff for a rating of nondepressed versus depressed. We

Table 2. Sensitivity and Specificity of Physician Ratings of Depression Compared with CES-D Score and Telephone Interview Diagnosis, 5-Point Scale Using Various Cut Points.

<table>
<thead>
<tr>
<th>Physician Rating of Depression on 5-point Likert-Type Scale</th>
<th>Elevated CES-D Score (≥ 16) (n = 266)</th>
<th>Telephone Interview Diagnoses of Depression (CES-D ≥ 16) Subsample (n = 49)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity</td>
<td>Specificity</td>
</tr>
<tr>
<td>2</td>
<td>0.72</td>
<td>0.31</td>
</tr>
<tr>
<td>3</td>
<td>0.36</td>
<td>0.37</td>
</tr>
<tr>
<td>4</td>
<td>0.20</td>
<td>0.57</td>
</tr>
<tr>
<td>5</td>
<td>0.08</td>
<td>0.83</td>
</tr>
</tbody>
</table>
were much more likely to have talked to their family physicians about stress or mental health problems, than by their physicians who were distressed but not depressed. Of the 20 patients who had discussed such problems with their physician, 15 were depressed. Ten patients reported discussing problems with a mental health professional or social worker; 8 had also discussed such problems with their physicians, and 7 were depressed.

With regard to patient reports of psychotropic medication prescriptions, 11 said that they had received a prescription in the past 6 months because of problems with stress, emotions, nerves, drugs, alcohol, or mental health. Ten of these patients indicated that they were taking the medication as directed, and 9 reported that they had received the prescription from the family physician in whose reception room they had been first screened for the study. Of the 2 patients taking antidepressants not prescribed by their family physician, 1 had received the prescription during a psychiatric hospitalization, and the other had gone for evaluation to another physician recommended by a psychotherapist. Thus, family physicians play a key role in prescribing psychotropic medication, particularly antidepressants. Four of the 8 patients who reported taking antidepressants were not depressed in the telephone interview.

Overall, of the 21 patients who were depressed, 15 were not taking any psychotropic medication or receiving counseling from their family physicians. Taking into account intervention by the family physician, medication prescribed by other physicians, and counseling or psychotherapy from any source, 11 of the depressed patients did not receive intervention for their condition.

A statistically significant correlation occurred between the 5-point physician rating and telephone interview diagnosis of depression in this subsample of 49 distressed patients, but it was clinically unimpressive (r = 0.29, P < 0.05). Likewise, there was a modest relation between physician rating of depression and whether the patient had experienced a 2-week period of depressed mood (r = 0.29, P < 0.05), as well as one between physician rating and the total number of criterion symptoms (r = 0.36, P < 0.01). Physician ratings of depression were correlated 0.63 with patients’ reports of receiving a prescription for psychotropic medication. This high correlation is not surprising because these family physicians were the primary source of such pre-
Physician ratings of patient depression were not related to the patients' evaluations of the quality of care they received, but they were positively related to patients' evaluations of the quality of the physician-patient relationship ($r = 0.26, P < 0.05$).

In the two right-hand columns of Table 2, the sensitivity and specificity of physician ratings of depression are compared with depression diagnosis from the telephone interview. Again, the best physician performance occurred with a cutoff of 2 (any depression versus no depression), yielding a sensitivity of 0.80 and a specificity of 0.50. The association between physician ratings and diagnosis was not statistically significant. We attempted to identify any systematic differences between physician-identified false-positive patients—patients who were identified as depressed by their family physicians but who were not found to be depressed in the telephone interview—and those for whom there was concordance, but these analyses failed to yield any significant results.

In the analyses presented thus far, the physicians were evaluated on the basis of their ability to distinguish patients who were depressed from those who had considerable distress but were not depressed. It is reasonable to suppose that it is more difficult to make such a distinction than it is to detect depression in an unselected sample of patients, so we tested this possibility. Recalling that Schulberg, et al.\textsuperscript{24} reported only 1 percent of the patients who had CES-D scores lower than 16 were found to be depressed in interviews, we recalculated the relations between physician ratings and diagnosis by assuming that the 205 patients with low CES-D scores would not have been found to be depressed if they had been interviewed. Unfortunately, these results did not present in a better light the physicians' ability to detect depression. The correlation between physician ratings and diagnosis remained modest, although statistically significant ($r = 0.27, P < 0.001$). Using a physician rating cutoff of 2 on the 5-point scale, the physician sensitivity was 0.80, but specificity was 0.12; for a cutoff of 3 on the 5-point scale, the sensitivity was 0.47, but the specificity was 0.17. Thus, when we make the empirical assumption that patients with low CES-D scores are not depressed, we find that the physicians perceived much more depression in their patients than was justified.

### Discussion

Our results are relevant to evaluations of both the CES-D as a routine screening instrument in family practice and the detection of depression by family physicians. It appears that our sample was similar to other study samples in terms of the prevalence of elevated CES-D scores and of diagnosable depression as identified by our two-step assessment by reception room screening and telephone interview. Overall, the CES-D identified more than one-fifth of the reception room sample as distressed on the basis of the usual cutoff score of 15, but only 43 percent of these patients were considered to be depressed in the telephone interview. Our prevalence figure for diagnosis of depression based on an interview fell within the usual range of 8 to 10 percent reported by other investigators. Given this prevalence rate and a false-positive rate for the CES-D of 0.57, the screening instrument does not appear to be an efficient way of identifying depressed patients.

Other authors have suggested using a higher cutoff for the CES-D. Yet following the suggestion of Husaini, et al.\textsuperscript{25} that a cutoff of 17 be used, we would lose 3 (14 percent) of the depressed patients. Using Husaini and colleagues' alternative suggestion of a cutoff of 23, we would lose 7 (33 percent) of the patients found to be depressed in the interview. Using the Schulberg, et al.\textsuperscript{24} suggestion of a cutoff of 27, we would lose 11 (52 percent). Thus, efforts to improve the specificity of the CES-D cause a considerable loss of sensitivity. We cannot recommend the CES-D for use in routine screening of patients for whom there is no other reason for suspecting depression. With the suggestion of such a low positive predictive value for the CES-D, we instead recommend simple inquiry based on the formal criteria for major depression, i.e., duration of mood disturbance and accompanying psychological and somatic symptoms.

The loss of sensitivity that occurred when we raised the cut point on the CES-D reflects the prevalence a relatively mild depression among family practice patients, as well as conceptual differences between the CES-D and the diagnosis of depression.\textsuperscript{29} Examining the actual items on the CES-D, one can see that a person who was feeling unhappy and lonely because of a recent rejection by a close friend could have a high score while
failing to have any of the criterion symptoms for major depressive disorder.

Our results contribute to the still limited literature concerning the differences between persons who have a diagnosable depression and those who have elevated scores on a self-report depression inventory but who are not depressed. Aside from having a 2-week duration of mood disturbance and criterion symptoms, depressed patients in this study had psychosocial disadvantages, notably twice as many stressful life events and low social support, as well as more physical health complaints. Depressed patients also tended to have higher CES-D scores, but if we had raised the cutoff score, we would have substantially diminished the sensitivity.

No matter how we dichotomized the 5-point physician rating scale for patient depression, we were unable to find a cutoff point at which physicians displayed a good balance of sensitivity and specificity. Physicians performed somewhat better against the results of our telephone interview of patients with high CES-D scores than against the CES-D itself, but there was only a modest relation between physician ratings and CES-D scores, diagnosis based on the telephone interview, whether the patient reported the 2 weeks of mood disturbance required for a diagnosis of depression, or the total number of criterion symptoms reported by the patients.

Physician ratings were influenced as much by their sense of the patients' subjective stress and distress as they were by criterion symptoms of depression. Patients who discussed stress and mental health problems tended to be depressed, but physicians apparently missed syndromic depression syndrome even in a sizable proportion of these patients. We can suggest only limited reasons for the physicians' relatively poor performance. Despite a reliance on well-validated measures, our assessment of psychosocial and health variables in our telephone interview sheds little light on the problem. We recommend that future researchers concentrate on the actual patient-physician encounter, on the concept of depression held by physicians, and on myths and barriers preventing physicians from making accurate mental disorder diagnoses. We suspect that the physicians' concept of depression diverges from the DSM-III-R diagnostic criteria and that the physicians weigh subjective distress too heavily and pay too little attention to formal symptoms. Whether this divergence reflects a lack of knowledge or a rejection of the DSM-III-R criteria should be studied.

There are frequent claims in the literature that family physicians have low sensitivity to depression in their patients and therefore miss much of the depression that is presented to them. The findings of our study support this claim, but we are also concerned that family physicians demonstrate low specificity in their assessment of depression. If so, psychiatrists could also share this problem. A recent study found that psychiatric clinicians made 1.64 times the number of diagnoses of depression than that obtained with a structured interview.

Conclusion
In our study of 266 patients in community-based family practices, the patients showed a substantial prevalence of distress and depressive disorder, with physicians basing relatively inaccurate diagnoses of depression more on the former than the latter. A majority of patients likely to be suffering from major depression were receiving no treatment whatsoever, but those who did received it from their family physician. Significant physician myths, barriers, and covert biases could exist that preclude otherwise humanistic physicians from dealing effectively with an important primary care psychiatric problem. Such biases can be elucidated only with research that uses the physician and the physician-patient relationship as the units of study.

Our findings have a number of implications for practicing family physicians. Major changes are needed in the way family physicians are trained to diagnose depression. More emphasis should be placed on formal, criterion-based techniques and on the clarification of the difference between distress and depression as a diagnosis. Depressed patients are unlikely to specify depression as their presenting problem, and the routine use of screening questionnaires is a relatively inefficient means of detection. Such questionnaires could have some utility when patients provide an ambiguous presentation and there is some reason to suspect depression, yet the high prevalence of depression suggests that family physicians should look for signs of distress and depressive disorders in all patients. Family physicians should become
more familiar with the formal criteria for depression to improve their specificity, as well as their sensitivity, to depressive disorder in their patients. Perhaps the most straightforward approach for a family physician to improve performance in detecting depression is a readiness to inquire about depressive symptoms in a semistructured fashion, starting with whether a patient is experiencing the mood and energy disturbances that are requisite for a diagnosis.

This study suggests that family physicians are attentive to distress in their patients, but perhaps as a distraction from making a criterion-based diagnosis. The physician and patient may actually enter into a conspiracy in which psychosocial issues are addressed, but not as a legitimate basis for diagnosis, and a biomedical bias determines the health-care-seeking behavior of patients and the decision making of physicians.

References
NOTICE

Certificate of Added Qualifications in Geriatric Medicine
Examination April 10, 1992

The next examination for the American Board of Family Practice Certificate of Added Qualifications in Geriatric Medicine will be administered on April 10, 1992. Application materials for this exam will be available July 1, 1991. ABFP Diplomates interested in participating in the exam should request application materials by writing to:

Geriatric Medicine Examination
American Board of Family Practice
2228 Young Drive
Lexington, Kentucky 40505-4294