

# Case-Finding Instruments for Depression in Primary Care Settings

Cynthia D. Mulrow, MD, MSc; John W. Williams Jr., MD, MHS; Meghan B. Gerety, MD; Gilbert Ramirez, DrPH; Oscar M. Montiel, MD; and Caroline Kerber, MD

■ **Objective:** To evaluate the usefulness of case-finding instruments for identifying patients with major depression in primary care settings.

■ **Data Sources:** A MEDLINE search of the English-language medical literature; bibliographies of selected papers; and experts.

■ **Study Selection:** Studies that were done in primary care settings with unselected patients and that compared case-finding instruments with accepted diagnostic criterion standards for major depression were selected.

■ **Data Synthesis:** 9 case-finding instruments were assessed in 18 studies. More than 15 000 patients received screening with a case-finding instrument; approximately 5300 of these received criterion standard assessment. Case-finding instruments ranged in length from 2 to 28 questions. Average administration times ranged from less than 2 minutes to 6 minutes. Sensitivities and specificities for detecting major depression ranged from 67% to 99% and from 40% to 95%, respectively. No significant differences between instruments were found. Overall sensitivity was 84% (95% CI, 79% to 89%); overall specificity was 72% (CI, 67% to 77%). If a case-finding instrument were administered to 100 primary care patients with a 5% prevalence of major depression, the clinician could expect that 31 patients would screen positive, that 4 of the 31 would have major depression, and that 1 patient with major depression would not be identified.

■ **Conclusions:** Several instruments with reasonable operating characteristics are available to help primary care clinicians identify patients with major depression. Because the operating characteristics of these instruments are similar, selection of a particular instrument should depend on issues such as feasibility, administration and scoring times, and the instruments' ability to serve additional purposes, such as monitoring severity or response to therapy.

*Ann Intern Med.* 1995;122:913-921.

From the Audie L. Murphy Memorial Veterans Affairs Hospital, San Antonio, Texas. For current author addresses, see end of text.

Depressive disorders are common, persistent, and recurring afflictions among primary care patients. They cause substantial suffering for patients and their families and are associated with a loss of personal productivity and a markedly increased risk for suicide. Further, the presence of depression puts persons with comorbid conditions, such as recent myocardial infarction, at increased risk for illness and death. Persons with depression spend more time with their physicians during office visits and use more health care than persons without depression (1). The annual health care cost associated with depression in the United States is estimated to be \$43.7 billion (2). Under-scoring the importance of identifying patients with depression is that the effectiveness of therapy, including antidepressants, psychotherapy, and counseling, has clearly been established (3).

Despite these issues, primary care providers fail to diagnose and treat as many as 35% to 50% of patients with depressive disorders (4, 5). Obstacles to the appropriate recognition of depression include inadequate provider knowledge of diagnostic criteria; competing comorbid conditions and priorities among primary care patients; time limitations in busy office settings; concern about the implications of labeling; poor reimbursement mechanisms; and uncertainty about the value, accuracy, and efficiency of screening mechanisms for identifying patients with depression. We address the last of these obstacles and assess the feasibility and operating characteristics of several case-finding instruments that have been used to detect depressive disorders in primary care settings. Our ultimate goal is to familiarize providers with the advantages and disadvantages of these instruments so that they can make informed decisions about incorporating them into practice.

## Methods

### Data Acquisition

We did a MEDLINE search of the English-language medical literature published from 1966 through February 1994. Search terms included "depressive disorder or depression," "diagnosis," and the specific names of each of 11 case-finding instruments cited in previous relevant reviews or bibliographies (6-9). Other sources were references identified from pertinent articles and national experts in the field of depression. Experts included authors of papers that were selected for review and two members of the Agency for Health Care Policy and Research Guideline Panel on Depression.

Of 906 articles identified through MEDLINE, 210 were deemed potentially relevant. These were reviewed to identify studies that met the following selection criteria.

Study samples had to have been composed of primary care patients attending clinic. Patients were excluded if they had been selected because they had specific conditions (such as chronic pain or cancer) or because they had specific demographic char-

**Table 1. Characteristics of Case-Finding Instruments That Have Been Used to Detect Depression in Primary Care Settings\***

Instrument	Scope	Items, n†	Response Format	Time Frame of Questions	Score Range	Usual Cut-point‡	Literacy Level§	Administration Time, min	Monitor Severity or Response
BDI	Depression-specific	21	4 statements of symptom severity per item	Today	0-63	10 mild 20 moderate 30 severe	Easy	2-5	Yes
CES-D	Depression-specific	20	4 frequency ratings: "less than one day" to "most or all (5-7) days"	Past week	0-60	16	Easy	2-5	Yes
GHQ	Psychiatric illness (several versions)	28	4 frequency ratings: "not at all" to "much more than usual"	Past few weeks	0-28	4	Easy	5-10	Yes
HSCL	Multiple versions and multiple components with depression category	25	4 frequency ratings: "not at all" to "much more than usual"	Past week	25-100	43	Average	2-5	Yes
MOS-D	Depression-specific	8	Frequency ratings; same format as CES-D	Past week	0-1 (logistic regression)	0.06	Average	< 2	Yes
ID	Depression-specific	15	2 items, yes or no; 3 statements of normal, overt, or covert symptomatology	Recently	0-15	10	Easy	2-5	Unknown
PRIME-MD	Multiple components with depression category	2	Yes or no	Past month	0-2	1	Average	< 2	Unknown
SDDS-PC	Multiple components with depression category	5	Yes or no	Past month	0-4	2	Easy	< 2	Unknown
SDS	Depression-specific	20	4 frequency ratings: "little of the time" to "most of the time"	Recently	25-100 (sum/80 × 100)	50 mild 60 moderate 70 severe	Easy	2-5	Yes

\* BDI = Beck Depression Inventory; CES-D = Center for Epidemiologic Studies Depression Screen; GHQ = General Health Questionnaire; HSCL = Hopkins Symptoms Checklist; MOS-D = Medical Outcomes Study Depression Screen; ID = Popoff Index of Depression; PRIME-MD = Primary Care Evaluation of Mental Disorders; SDDS-PC = Symptom Driven Diagnostic System-Primary Care; SDS = Zung Self-Assessment Depression Scale.

† Item numbers for the PRIME-MD and SDDS-PC refer to depression questions only; item numbers for the HSCL refer to depression plus anxiety questions.

‡ Cut-point is the number at or above which the test is considered positive.

§ Easy = 3rd- to 5th-grade reading level; average = 6th- to 9th-grade reading level according to Fog Formula (38).

acteristics (for example, they were immigrants in a particular ethnic group). Both a case-finding instrument and a diagnostic criterion standard had to have been administered. The criterion standard had to have had formal standardized diagnostic criteria for depression. Accepted criterion standards were the Diagnostic Schedule Manual-3 criteria (DSM-III or DSM-III-R) and the Research Diagnostic Criteria, or a close approximation of these. Standard interview procedures, such as the Diagnostic Interview Schedule or the Structured Clinical Interview for DSM-III, had to have been used to arrive at the diagnosis. Chart or physician diagnoses of depression that were made without specified formal interview procedures and diagnostic criteria were excluded.

Nineteen studies involving nine case-finding instruments met the selection criteria: Fourteen were found during the MEDLINE search; 1 came from a relevant bibliography; and 4 were unpublished at the time of the search and came from experts (10-28). Of the remaining articles screened, 92% were excluded because they did not involve primary care patients, 6% were excluded because they had used an inadequate criterion standard, and 2% were excluded because they involved selected populations (29, 30) or because they had tested modified and unvalidated versions of case-finding instruments (31).

#### Data Extraction

Articles were abstracted by two independent reviewers. Determination of study quality was made on the basis of 1) whether

the case-finding instrument was administered and interpreted independently of the criterion standard and 2) whether the proportion of persons receiving the criterion standard assessment was less than or more than 50% of those approached for criterion standard assessment. Quality assessment addressed methodologic issues relevant to the evaluation of diagnostic tests (such as independent assessment and selection bias) and did not necessarily reflect the ability of studies to address their original aims. There were no disagreements about quality assessments.

#### Data Synthesis

Established cut-points for case-finding instruments (Table 1) were used. Two-by-two tables were constructed that categorized numbers of screened-positive and screened-negative persons who did and did not meet criterion standard diagnosis for major depression and major depression or dysthymia. Kraemer's method (32) was used to adjust for verification bias for studies that used two-stage assessment techniques; whereby the criterion standard was administered only to a random sample of persons who screened negative on case-finding instruments (33). The authors of all but one study provided us with additional data and analyses when two-by-two tables could not be derived from abstraction of the published article. This one study (28) was dropped from further review because its authors could not be contacted and tables could not be derived from published information.

A scattergram (Figure 1) plotting true-positive against false-positive rates was constructed to visually evaluate variability among studies (34). To provide a visual reference for the consistency of study results, we modeled a summary receiver-operating curve based on the logit transformations of the true-positive and false-positive rates.

Average sensitivities and specificities, weighted by study size and corrected for two-stage assessment techniques when indicated, were computed both by case-finding instrument and by overall instruments (35). The point estimates and 95% CIs were calculated using a linear random-effects model (36, 37). Approximate 95% CIs were estimated using quadratic root formulae because most of the point estimates were near unity (37). Differences in weighted average sensitivities and specificities between case-finding instruments were evaluated using the z statistic with the Scheffe multiple-comparison adjustment (37). Stratified analyses were done to evaluate whether estimated sensitivities and specificities varied between high-quality studies and those with major selection bias or lack of independent assessment. Regression analysis was used to determine associations between reported study prevalences and sensitivity estimates (37).

## Results

### Descriptions of Case-Finding Instruments

Characteristics of the nine case-finding instruments that have been evaluated in primary care settings are pre-

sented in Table 1. All of the questionnaires are written either at the easy (3rd to 5th grade) or average (6th to 9th grade) reading level (38). Almost all can be self-administered in less than 5 minutes. Except for the General Health Questionnaire, all include specific questions aimed at assessing depressed mood or whether a patient feels sad or blue. All include questions assessing anhedonia. Most are available in languages other than English, such as Spanish.

The Beck Depression Inventory, the Center for Epidemiologic Studies Depression Screen, and the Zung Self-Assessment Depression Scale are three commonly used, traditional instruments that were developed specifically to identify depression. They include similar numbers of questions and use response formats that rely either on ranking symptom severity or on classifying frequency of symptoms. The time frames of questions are "today" for the Beck Depression Inventory, "over the past week" for the Center for Epidemiologic Studies Depression Screen, and "recently" for the Zung Self-Assessment Depression Scale. These three instruments have been used in numerous settings (including the community, the clinic, and the hospital) not only to identify depression but also to rate severity of depression and to monitor response to therapy.

**Table 2. Studies Evaluating Case-Finding Instruments for Depression in Primary Care Settings\***

Year	Study (Reference)	Instrument	Sample Size†	Population‡	Criterion Standard	Quality Score§
1990	Perez-Stable et al. (10)	BDI	265	Mixed	DSM-III	4
		CES-D	214			
1990	Zich et al. (11)	BDI	31	Mixed	DSM-III	4
		CES-D				
1993	Kirmayer et al. (12)	CES-D	685	Academic	DSM-III	1
1983	Schulberg et al. (13)	CES-D	294	Community	DSM-III	4
1994	Hendrie et al. (14)	CES-D	125	Academic (age > 59 years)	DSM-III-R	2
1990	Hough et al. (15)	CES-D	525	HMO	DSM-III	2
			525			
		GHQ, 28 items	525			
		HSCL				
1994	Fechner-Bates et al. (16)	CES-D	425	Community	DSM-III-R	2
1979	Hoeper et al. (17)	GHQ, 30 items	247	Academic	RDC	2
1987	Von Korff et al. (18)	GHQ, 28 items	809	Academic	DSM-III	1
1987	Goldberg and Blackwell (19)	GHQ, 28 items	283	Community	DSM-III	1
1977	Finlay-Jones and Murphy (20)	GHQ, 30 items	100	Community (women only)	DSM-III-R-like algorithm	2
1982	Okimoto et al. (21)	ID	55	Veterans Affairs	DSM-III	3
		SDS	55	(age > 59 years)		
1994	Williams et al. (22)	MOS-D	99	Mixed	DSM-III-R	2
1988	Burnam et al. (23)¶	MOS-D	501	HMO	DSM-III	4
1994	Broadhead et al. (24)¶	MOS-D	388	Community	DSM-III-R	
		SDDS-PC	388			
1994	Broadhead et al. (24)	SDDS-PC	257	Mixed	DSM-III-R	3
1994	Spitzer et al. (25)	PRIME-MD	431	Mixed	DSM-III-R	1
			337			
		SDS				
1990	Magruder-Habib et al. (26)	SDS	206	Veterans Affairs	DSM-III	4
1979	Raft et al. (27)	SDS	69	Academic medical Screening clinic	DSM-III-like Criteria	4

\* BDI = Beck Depression Inventory; CES-D = Center for Epidemiologic Studies Depression Screen; GHQ = General Health Questionnaire; HMO = health maintenance organization; HSCL = Hopkins Symptom Checklist; MOS-D = Medical Outcomes Study Depression Screen; ID = Popoff Index of Depression; PRIME-MD = Primary Care Evaluation of Mental Disorders; SDDS-PC = Symptom Driven Diagnostic System-Primary Care; SDS = Zung Self-Assessment Depression Scale.

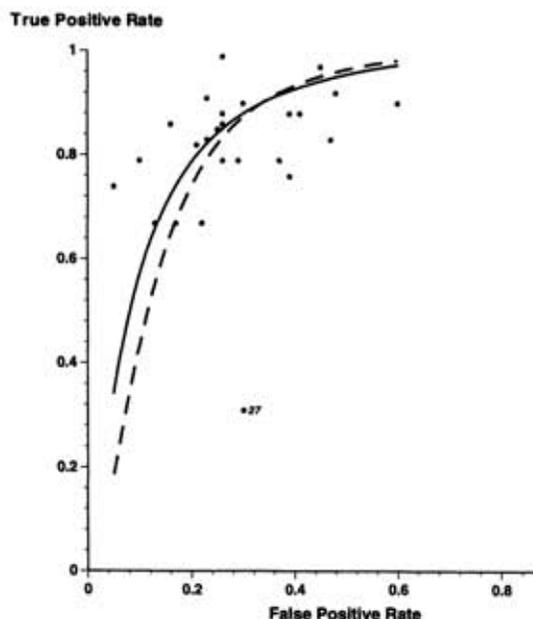
† The sample size refers to the actual number who received the criterion standard.

‡ Mixed-community and university-affiliated clinics; academic-university-affiliated clinics; community-private practice clinics.

§ Lower scores indicate higher quality.

¶ This study used the same sample as that used in the study by Hough and coworkers (15).

¶ The study by Broadhead and colleagues is listed twice because it included both an initial test set of patients and a validation set of patients.



**Figure 1.** Plot of true-positive rate against false-positive rate for case-finding instruments to detect major depression. The numbered data point represents the outlier study (27). The dashed curve represents the summary receiver-operating curve through all data points with the outlier study included; the solid curve represents the summary receiver-operating curve through all data points with the outlier study excluded.

The General Health Questionnaire and the Hopkins Symptom Checklist are questionnaires that screen for general psychiatric illness; the Hopkins Checklist has a specific category for depression. Both of these instruments have several versions with different numbers of questions. The Medical Outcomes Study Depression Screen is a depression-specific screening instrument that was developed by combining two questions from the Diagnostic Interview Schedule (39) with six questions from the Center for Epidemiologic Studies Depression Screen. A logistic regression scoring method is used; this requires a calculator.

The Primary Care Evaluation of Mental Disorders (PRIME-MD) and the Symptom Driven Diagnostic System-Primary Care instruments are recently developed, multidimensional questionnaires. Each has screening questions arranged in several categories (for example, mood or depression, anxiety, alcohol abuse, and somatization) that are used to trigger more extensive diagnostic interviewing sections for specific DSM-III-R diagnoses. The depression components of these two instruments include the fewest questions of all case-finding instruments that have been studied in primary care settings.

#### Descriptions of Studies and Findings

The 18 studies that we reviewed are outlined in Table 2. These studies involved more than 15 000 primary care patients who were screened with a case-finding instrument; approximately 5300 of these patients were administered an acceptable criterion standard for diagnosing major depression. For several reasons, fewer persons received a criterion standard than were screened. First, to maximize efficiency, investigators often administered the criterion standard to random samples of persons who

screened negative on the case-finding instrument. Second, persons sometimes failed to keep criterion standard appointments or refused the criterion standard. Last, the criterion standard was occasionally administered only to self-selected volunteers. Six of the 18 studies had potential major selection biases because more than 50% of persons did not receive criterion standard assessment because they refused the assessment or did not volunteer for it, or because they failed to keep an appointment (10, 11, 13, 21, 24, 26).

Studies were conducted in various settings, including managed care clinics of health maintenance organizations, general practices in the community, and clinics affiliated with medical universities. Most studies used DSM-III or DSM-III-R criteria to diagnose depression and had categorized diagnoses for current major depression within the previous 2 to 4 weeks. Prevalences of major depression ranged from 4% to 17%, except for two small studies with major selection biases, which had prevalences of 31% and 42%, respectively (21, 27).

The General Health Questionnaire and the Center for Epidemiologic Studies Depression Screen were the most widely evaluated, that is, they were evaluated in the most studies, in the most patients, and in both academic and community primary care clinics. Although slightly more than half the studies were of reasonably high quality for diagnostic test evaluations, eight either did not clearly administer the criterion standard independent of case-finding results or had major selection bias.

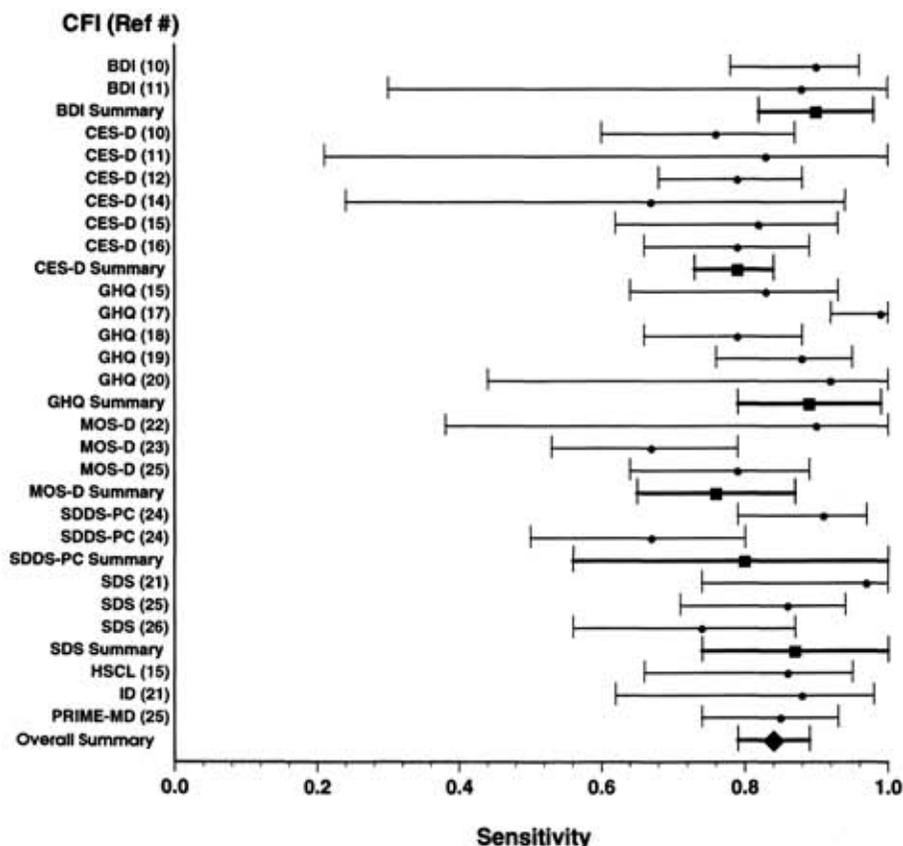
Figure 1 is a scattergram that plots study results about true-positive and false-positive rates for case-finding instruments to detect major depression. The usual cut-points for instruments were used for these calculations (Table 1). The cut-point for mild depression was used for the two scales with three listed cut-points (the Beck Depression Inventory [10] and the Zung Self-Assessment Depression Scale [50]); the study by Raft and coworkers (27) only had information on a cut-point of 60 for the Zung Scale. The scattergram suggests that one study varies systematically from the others (27); this was the small study by Raft and coworkers, which used the higher cut-point for the Zung Scale. This study had an unusually low sensitivity (31%; CI, 16% to 51%) for detecting major depression.

Figures 2 and 3 show sensitivities and specificities, respectively, for detecting major depression using study and case-finding instruments. The average weighted sensitivity of all studies, except for that by Raft and coworkers (27), was 84% (CI, 79% to 89%); the average weighted specificity was 72% (CI, 67% to 77%). There were no statistically significant differences for either sensitivities or specificities among the different case-finding instruments ( $P > 0.05$ ).

Only seven studies had data on the ability of instruments to identify persons with either dysthymia or major depression (10, 13, 15, 18, 23–25). Operating characteristics for detecting this combined category were not statistically significantly different from those for detecting major depression alone, although slightly lower sensitivity and higher specificity were seen. Average weighted sensitivity was 78% (CI, 73% to 83%); average weighted specificity was 77% (CI, 71% to 82%).

Stratified analyses were done to evaluate whether sen-

**Figure 2. Sensitivities of case-finding instruments (CFI) for identifying major depression.** Point estimates with 95% CIs are presented for single studies (\*), for summary data on case-finding instruments (■), and overall (◆). BDI = Beck Depression Inventory; CES-D = Center for Epidemiologic Studies Depression Screen; GHQ = General Health Questionnaire; HSCL = Hopkins Symptoms Checklist; MOS-D = Medical Outcomes Study Depression Screen; ID = Popoff Index of Depression; PRIME-MD = Primary Care Evaluation of Mental Disorders; SDDS-PC = Symptom Driven Diagnostic System-Primary Care; and SDS = Zung Self-Assessment Depression Scale.



sensitivities and specificities derived from studies with selection bias and from studies that did not assess criterion standard diagnosis independently of case-finding score differed from sensitivities and specificities derived from higher-quality studies. Sensitivities ranged from 79% to 83%; specificities ranged from 66% to 74%. Analyses stratified according to whether the DSM-III (11 studies) or DSM-III-R (6 studies) criterion standard was used showed sensitivities of 84% and 83%, respectively, and specificities of 69% and 76%, respectively. Studies that examined current depression (depression within the previous 2 to 4 weeks) had a sensitivity of 84% and a specificity of 75%, and those that either examined depression over the previous year or were unclear about their diagnostic time frame had a sensitivity of 86% and a specificity of 62%. None of the stratified analyses discussed above showed statistically significant differences between groups. Finally, regression analyses showed no significant associations between reported study prevalences of major depression and sensitivity estimates.

#### Expected Yield of Case-Finding Instrument Strategy

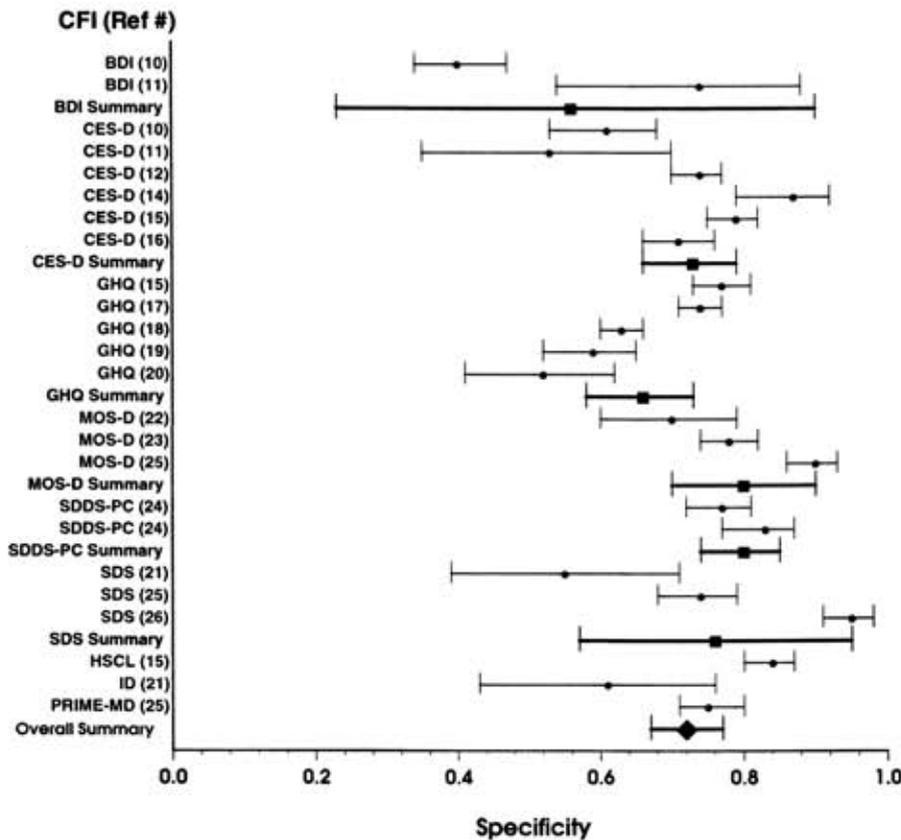
Figure 4 shows what a clinician could expect if he or she were to use a case-finding instrument in a clinic setting. This scenario assumes that the clinician sees 100 adult patients (a reasonably booming outpatient practice for a week) and that the prevalence of major depression in the clinic is 5% (a prevalence commonly found in primary care settings). If the sensitivity of the instrument were 80% and the specificity were 72% (Figures 2 and 3), 31 patients would score at or above the usual cut-point for the case-finding instrument and would be considered

possibly depressed. Of these, 4 would meet the DSM-III-R criteria for major depression. Among the 69 patients who screened negative, 1 would actually have major depression. If the prevalence of major depression in the clinic setting was as high as 9%, 7 of the patients who screened positive would be correctly identified and 2 patients with major depression would not be diagnosed.

The above-mentioned sensitivity and specificity values convert to a likelihood ratio of 2.86 (sensitivity ÷ 1 - specificity). This likelihood ratio can be used to determine the post-test probability of a patient having major depression if he or she scores at or above the cut-point of the case-finding instrument. If a clinician practices in a setting where the prevalence or pretest probability of depression is 5%, a positive screening would increase the probability that a patient is suffering from major depression to 13%. If the prevalence in the clinician's setting is 9%, then a positive screening would increase the probability of major depression to 20%.

#### Discussion

Depression is one of the most common psychiatric diagnoses in primary care populations; major depressive disorders can be diagnosed in 6% to 9% of such patients (40). Depression is associated with significant burden and costs, and effective treatment is clearly available. Given that 50% to 60% of persons seeking help for depression are treated exclusively in the primary care setting, accurate detection in this setting is important (41). One strategy that may improve detection is the use of psychiatric screening or case-finding instruments in the physician's



**Figure 3. Specificities of case-finding instruments (CFI) for identifying major depression.** Point estimates with 95% CIs are presented for single studies (\*), for summary data on case-finding instruments (■), and overall ( ). BDI = Beck Depression Inventory; CES-D = Center for Epidemiologic Studies Depression Screen; GHQ = General Health Questionnaire; HSCL = Hopkins Symptoms Checklist; MOS-D = Medical Outcomes Study Depression Screen; ID = Popoff Index of Depression; PRIME-MD = Primary Care Evaluation of Mental Disorders; SDDS-PC = Symptom Driven Diagnostic System-Primary Care; and SDS = Zung Self-Assessment Depression Scale.

office as part of the routine health evaluation. In contrast to screening applied to unselected populations, case-finding occurs when clinicians search for disease by screening their own patients. Several things should be considered before this strategy is chosen.

First, are there case-finding instruments that are feasible to use in the primary care setting? The answer is yes. Several case-finding instruments are relatively short and can be easily self-administered by patients with at least grade-school to middle-school reading ability. For patients without such reading skills, instruments can be administered by trained office staff. Several of the instruments are available in different languages, including Spanish. All are easily scored, except for the Medical Outcomes Study Depression Screen, which requires a calculator.

Second, do the operating characteristics of the case-finding instruments enable the clinician to decrease the number of persons who need more intensive diagnostic interviewing without missing persons who might potentially benefit from a full diagnostic interview? The answer is probably yes. We have shown that in a group of 100 patients in a clinic with an average prevalence of 5% for major depression, the clinician would have to do a diagnostic interview in only 31 patients to identify the 4 with major depression, and only 1 person with major depression would be missed. It is unlikely that many busy clinicians could do or would want to do diagnostic interviews in all 100 patients so that they would not miss 1 case of depression. Further, when using a case-finding instrument, the clinician can expect to miss about 20% of patients with depression; current practice often misses 50% of such patients (4, 5).

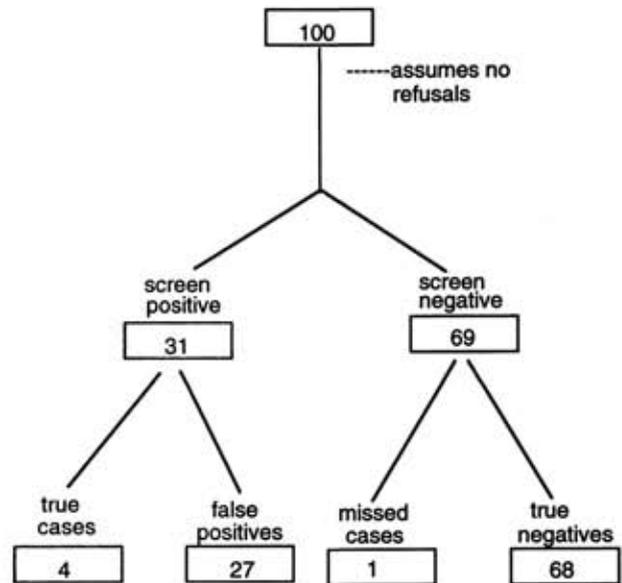
Third, is a case-finding strategy effective in identifying patients with major depression who will receive treatment and who will have improved clinical outcomes? The answer is that we don't know. Although we did not directly address this question, at least 12 studies have assessed the effect of feedback of psychiatric case-finding scale scores on physician practice patterns (42-53). Some of these have shown improved recognition or treatment of depression with such feedback, or both, but none have shown improved clinical outcomes. Most of the studies had significant design limitations. Only 1 incorporated a multifaceted approach to help ensure that clinicians had adequate knowledge with which to make accurate depression diagnoses and treatment recommendations (42). Regardless, the lack of data from randomized trials has led the American Academy of Family Physicians, the Canadian Task Force on the Periodic Health Examination, and the U.S. Preventive Services Task Force to not recommend the routine use of case-finding instruments (54-56). They recommend only that clinicians remain alert for symptoms of depression, particularly in high-risk patients. The recent Agency for Health Care Policy and Research Guideline Panel on Depression noted that case-finding instruments were unlikely to add much if clinicians were already highly attuned to and inquired regularly about depressive symptoms (3). The Panel also noted that the efficiency of case-finding instruments could be increased by targeting high-risk patients.

Unfortunately, recent data suggest that primary care clinicians still miss as many as 35% to 65% of patients with major depression (4, 5, 10, 57). Clinician recognition may be unaffected by a patient's sex or level of education,

may be slightly lower among younger patients, and may be higher among patients with the most severe and disabling disease (4, 57, 58). Fewer than 20% of persons with major depression in the primary care setting have severe disease, and a sparse amount of available data suggests that 25% of these persons are still being missed (57). Although case-finding instruments detect a higher proportion of persons with major depression (80%) than primary care clinicians do (35% to 65%), it is unclear whether they miss the same spectrum of disease most commonly missed by clinicians. Because sensitivity tends to increase with increasing severity of disease, it is likely that case-finding instruments will most often miss persons with mild disease and that they will not miss as many as 25% of persons with severe disease (59).

Targeting high-risk patients with a higher underlying prevalence of depression for case-finding is an intuitively appealing strategy because it offers greater efficiency by improving the positive predictive value of the case-finding instrument. However, gains in efficiency are related to the magnitude of the risk factor and the prevalence of that risk factor in the patient population (60, 61). For example, the strongest demographic risk factor for depression is female sex (relative risk, 2) (62, 63). In primary care practices, women patients outnumber men by a ratio of 60:40 (64). If only women were screened with a case-finding instrument that was 80% sensitive and 72% specific, the positive predictive value would increase only marginally, from 13% to 16%. This minimal gain would be associated with a 20% decrease in sensitivity for the case-finding strategy and a potential for missing 40% of depressed patients. A more potent (relative risk, 3) and less common (10%) risk factor is known previous depressive illness (65). If the "highly attuned" clinician limited case-finding to patients with this risk factor, he or she would increase the positive predictive value of the case-finding strategy to 29%, but overall sensitivity would be greatly reduced. Seventy-five percent of patients with major depression and no past known depressive illness would not be screened.

Incorporating case-finding instruments for depression into practice has its pitfalls. Routine screening for depressive disorders may lead clinicians to focus only on depression and ignore other common psychiatric disorders. Patients with anxiety disorders and alcohol abuse may have positive scores on depression instruments. Unless these disorders are ruled out, patients may be mistakenly labeled as depressed or may not be recognized as having depression comorbid with another psychiatric condition. Some case-finding instruments, such as the General Health Questionnaire, are designed to detect any psychiatric disorder and may be less likely to bias the clinician toward focusing only on depression. Two other recently developed instruments, the Primary Care Evaluation of Mental Disorders and the Symptom Driven Diagnostic System, are multidimensional questionnaires that assess several specific psychiatric disorders (24, 25). For example, the Primary Care Evaluation of Mental Disorders includes questions related to depression, anxiety, somatization disorder, and alcohol abuse. Positive responses to any of the four sections trigger specific structured interview questions that are used to make definitive diagnoses. Such multidimensional instruments may minimize undue



**Figure 4. Yield of case-finding instrument strategy.** Hypothetical group of 100 undiagnosed primary care clinic patients with 5% prevalence of major depression. Sensitivity of instrument is set at 80% and specificity at 72%.

emphasis on depression and identify comorbid psychiatric disorders.

Several limitations of our review merit mention. As with any review, ours represents a retrospective compilation of data, some of which were originally collected for reasons other than evaluating the accuracy of a case-finding instrument. In the studies, many fewer patients received criterion standards than were screened with case-finding instruments. Whether selection biases were present and what their potential effects may have been could not be evaluated. It is possible that persons with severe depression more often refused or were less likely to volunteer for criterion standard assessment. This could have resulted in a relatively healthy sample without adequate representation of persons with severe depression; this, in turn, would lead to an underestimation of sensitivity and an overestimation of specificity. On the other hand, if persons with severe depression were more likely to volunteer for criterion standard assessment, sensitivity would be overestimated and specificity underestimated. Although a formal criterion standard for the diagnosis of depression was used in all studies, these standards sometimes varied in required duration and in numbers and types of symptoms counted toward the diagnosis. However, stratified analyses did not show differences in the performance characteristics of different criterion standards. Lastly, the ability of instruments to identify related and clinically relevant diagnoses other than major depression or dysthymia, such as minor depression or anxiety, was not evaluated.

Despite these limitations, salient lessons that can be used to guide current decisions and frame future research are apparent. Nine different instruments have been used feasibly in primary care settings and have shown reasonable and consistent accuracy. Moreover, there is no need to develop new instruments aimed at identifying primary care patients with major depression or dysthymia. However, an important remaining question is whether any of

the instruments currently available should be routinely incorporated into practice. They could be self-administered in waiting rooms or interviewer-administered during a review of systems. But would they be any more acceptable, accurate, and efficient in diagnosing depression than the clinician who routinely ascertains depressed mood and anhedonia in his patients and asks additional questions only in patients with at least one of these two symptoms? (Such a strategy would theoretically identify most patients with depressive disorders, because most diagnostic schemes for these disorders require the presence of either depressed mood or anhedonia.) And a final question that continues to deserve study is, even if case-finding instruments are more acceptable, accurate, and efficient than routine physician questions during office visits, which strategy leads to better treatment and outcomes?

An informed, rational clinician can incorporate several feasible, reasonably accurate case-finding instruments for major depression into practice. Whether the use of such instruments ultimately will lead to improved outcomes is unknown. Choices about whether to use an instrument should take into account estimates of the resources required (for example, questionnaire forms, staff time for scoring, and for interviewing patients unable to complete self-administered forms) compared with potential resources saved (for example, diagnostic interviewing time, potentially improved outcomes, and decreased health care utilization if improved diagnosis is linked with treatment). If the clinician opts to use a case-finding instrument, several are available and have similar operating characteristics. Choice of a particular instrument should take into account ease of use and the time required for administration as well as whether the clinician wants to use the instrument for other purposes, such as screening for other psychiatric disorders or monitoring treatment response. Finally, regardless of whether primary care clinicians routinely incorporate case-finding instruments into their practices, all of them should know how to diagnose major depression and should aim to link accurate diagnosis with effective treatment strategies.

*Acknowledgments:* The authors thank the many investigators who provided either raw data or reanalysis of data or who made unpublished information available.

*Grant Support:* In part by a Robert Wood Johnson Generalist Physician Faculty Award and the Mexican American Medical Treatment Effectiveness Research Center funded by the Agency for Health Care Policy and Research.

*Requests for Reprints:* Cynthia D. Mulrow, MD, MSc, Audie L. Murphy Memorial Veterans Hospital (11C6), 7400 Merton Minter Boulevard, San Antonio, TX 78284.

*Current Author Addresses:* Drs. Mulrow, Williams, Gerety, Ramirez, Montiel, and Kerber: Audie L. Murphy Memorial Veterans Hospital (11C6), 7400 Merton Minter Boulevard, San Antonio, TX 78284.

## References

- Katon W, Berg AO, Robins AJ, Risse S. Depression—medical utilization and somatization. *West J Med.* 1986;144:564-8.
- Greenberg PE, Stiglin LE, Finkelstein SN, Berndt ER. The economic burden in depression in 1990. *J Clin Psychiatry.* 1993;54:405-18.
- Volume 2: treatment of major depression. In: *Depression in Primary Care.* Rockville, MD: U.S. Dept of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research; 1993; AHCPH publication no. 93-0551.
- Simon GE, VonKorff M. Recognition, management, and outcomes of depression in primary care. *Arch Fam Med.* 1995;4:99-105.
- Gerber PD, Barrett J, Barrett J, Manheimer E, Whiting R, Smith R. Recognition of depression by internists in primary care: a comparison

- of internist and gold standard psychiatric assessments. *J Gen Intern Med.* 1989;4:7-13.
- Feightner JW, Worrall G. Early detection of depression by primary care physicians. *Can Med Assoc J.* 1990;142:1215-20.
- Leserman J, Koch G. Review of self-report depression and anxiety measures. *Drug Inform J.* 1993;27:537-48.
- Coulehan JL, Schulberg HC, Block MR. The efficiency of depression questionnaires for case finding in primary medical care. *J Gen Intern Med.* 1989;4:541-7.
- Attkisson CC, Zich JM, eds. *Depression in Primary Care: Screening and Detection.* New York: Routledge; 1990.
- Pérez-Stable EJ, Miranda J, Muñoz RF, Ying YW. Depression in medical outpatients. Underrecognition and misdiagnosis. *Arch Intern Med.* 1990;150:1083-8.
- Zich JM, Attkisson CC, Greenfield TK. Screening for depression in primary care clinics: the CES-D and the BDI. *Int J Psychiatry Med.* 1990;20:259-77.
- Kirmayer LJ, Robbins JM, Dworkind M, Yaffe MJ. Somatization and the recognition of depression and anxiety in primary care. *Am J Psychiatry.* 1993;150:734-41.
- Schulberg HC, Saul M, McClelland M, Ganguli M, Christy W, Frank R. Assessing depression in primary medical and psychiatric practices. *Arch Gen Psychiatry.* 1985;42:1164-70.
- Hendrie HC, Callahan CM, Levitt EE, Hui SL, Musick B, Austrom MG, et al. Prevalence rates of major depressive disorders: the effects of varying the diagnostic criteria in an older primary care population. *Am J Ger Psych.* 1995;2:119-31.
- Hough R, Landsverk J, Stone J, et al. Comparison of psychiatric screening questionnaires for primary care patients. Final report for NIMH Contract No. 278-81-0036 (DB), 1983.
- Fechner-Bates S, Coyne JC, Schwenk TL. The relationship of self-reported distress to depressive disorders and other psychopathology. *J Consult Clin Psychol.* 1994;62:550-9.
- Hoepfer EW, Nycz GR, Cleary PD, Regier DA, Goldberg ID. Estimated prevalence of RDC mental disorder in primary care. *Int J Ment Health.* 1979;8:6-15.
- Von Korff M, Shapiro S, Burke JD, Teitelbaum M, Skinner EA, German P, et al. Anxiety and depression in a primary care clinic. Comparison of Diagnostic Interview Schedule, General Health Questionnaire, and practitioner assessments. *Arch Gen Psychiatry.* 1987;44:152-6.
- Goldberg DP, Blackwell B. Psychiatric illness in general practice. A detailed study using a new method of case identification. *Br Med J.* 1970;2:439-43.
- Finlay-Jones RA, Murphy E. Severity of psychiatric disorder and the 30-item general health questionnaire. *Br J Psychiatry.* 1979;134:609-16.
- Okimoto JT, Barnes RF, Veith RC, Raskind MA, Inui TS, Carter WB. Screening for depression in geriatric medical patients. *Am J Psychiatry.* 1982;139:799-802.
- Williams JW, Kerber CA, Mulrow CD, Medina A, Aguilar C. Depressive disorders in primary care: prevalence, functional disability, and identification. *J Gen Intern Med.* 1995;10:7-12.
- Burnam MA, Wells KB, Leake B, Landsverk J. Development of a brief screening instrument for detecting depressive disorders. *Med Care.* 1988;26:775-89.
- Broadhead WE, Leon AC, Weissman MM, Barrett GE, Blacklow RS, Gilbert TT, et al. Development and validation of the SDDS-PC screen for multiple mental disorders in primary care. *Arch Fam Med.* 1995;4:211-9.
- Spitzer RL, Williams JB, Kroenke K, Linzer M, deGruy FV 3d, Hahn SR, et al. Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 study. *JAMA.* 1994;272:1749-56.
- Magruder-Habib K, Zung WW, Feussner JR, Alling WC, Saunders WB, Stevens HA. Management of general medical patients with symptoms of depression. *Gen Hosp Psychiatry.* 1989;11:201-7.
- Raft D, Spencer RF, Toomey T, Brogan D. Depression in medical outpatients: use of the Zung scale. *Dis Nerv Syst.* 1977;38:999-1004.
- Ormel J, Koeter MW, van den Brink W, van de Willige G. Recognition, management, and course of anxiety and depression in general practice. *Arch Gen Psychiatry.* 1991;48:700-6.
- Ring JM, Marquis P. Depression in a Latino immigrant medical population: an exploratory screening and diagnosis. *Am J Orthopsychiatry.* 1991;61:298-302.
- Hinton WL, Du N, Chen YC, Tran CG, Newman TB, Lu FG. Screening for major depression in Vietnamese refugees: a validation and comparison of two instruments in a health screening population. *J Gen Intern Med.* 1994;9:202-6.
- Barrett JE, Barrett JA, Oxman TE, Gerber PD. The prevalence of psychiatric disorders in a primary care practice. *Arch Gen Psychiatry.* 1988;45:1100-6.
- Kraemer HC. *Evaluating Medical Tests: Objective and Quantitative Guidelines.* Newbury Park, CA: Sage; 1992.
- Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics.* 1983;39:207-15.
- Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC, et

- al. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med.* 1994;120:667-76.
35. **Greenes RA, Begg CB.** Assessment of diagnostic technologies. Methodology for unbiased estimation from samples of selectively verified patients. *Invest Radiol.* 1985;20:751-6.
  36. **Hedges LV, Olkin I.** *Statistical Methods for Meta-Analysis.* Orlando: Academic Pr; 1985.
  37. **Fleiss JL.** *Statistical Methods for Rates and Proportions.* New York: J Wiley; 1981.
  38. **Streiff LD.** Can clients understand our instructions? *Image J Nurs Sch.* 1986;18:48-52.
  39. **Robins LN, Helzer JE, Croughan J, Ratcliff KS, et al.** National Institute of Mental Health Diagnostic Interview Schedule: Its history, characteristics, and validity. *Arch Gen Psychiatry.* 1981;38:381-9.
  40. **Katon W, Roy-Byrne PP.** Antidepressants in the medically ill: diagnosis and treatment in primary care. *Clin Chem.* 1988;34:829-36.
  41. **Schurman RA, Krooner PD, Mitchell JB.** The hidden mental health network. Treatment of mental illness by nonpsychiatric physicians. *Arch Gen Psychiatry.* 1985;42:89-94.
  42. **Callahan CM, Hendrie HC, Dittus RS, Brater DC, Hui SL, Tierney WM.** Improving treatment of late life depression in primary care: a randomized clinical trial. *J Am Geriatr Soc.* 1994;42:840-5.
  43. **Johnstone A, Goldberg D.** Psychiatric screening in general practice. A controlled trial. *Lancet.* 1976;1:605-8.
  44. **Moore JT, Silimperi DR, Bobula JA.** Recognition of depression by family medicine residents: the impact of screening. *J Fam Pract.* 1978;7:509-13.
  45. **Linn LS, Yager J.** The effect of screening, sensitization, and feedback on notation of depression. *J Med Educ.* 1980;55:942-9.
  46. **Zung WW, Magill M, Moore JT, George DT.** Recognition and treatment of depression in a family medicine practice. *J Clin Psychiatry.* 1983;44:3-6.
  47. **Zung WW, King RE.** Identification and treatment of masked depression in a general medical practice. *J Clin Psychiatry.* 1983;44:365-8.
  48. **Hoepfer EW, Nycz GR, Kessler LG, Burke JD Jr, Pierce WE.** The usefulness of screening for mental illness. *Lancet.* 1984;1:33-5.
  49. **Rucker L, Frye EB, Cygan RW.** Feasibility and usefulness of depression screening in medical outpatients. *Arch Intern Med.* 1986;146:729-31.
  50. **Shapiro S, German PS, Skinner EA, VonKorff M, Turner RW, Klein LE, et al.** An experiment to change detection and management of mental morbidity in primary care. *Med Care.* 1987;25:327-39.
  51. **Rand EH, Badger LW, Coggins DR.** Toward a resolution of contradictions. Utility of feedback from the GHQ. *Gen Hosp Psychiatry.* 1988;10:189-96.
  52. **Magruder-Habib K, Zung WW, Feussner JR.** Improving physicians' recognition and treatment of depression in general medical care. *Med Care.* 1990;28:239-50.
  53. **Brody DS, Lerman CE, Wolfson HG, Caputo GC.** Improvement in physicians' counseling of patients with mental health problems. *Arch Intern Med.* 1990;150:993-8.
  54. **American Academy of Family Physicians, Commission on Public Health and Scientific Affairs.** *Age Charts for Periodic Health Examination.* Kansas City, MO: American Academy of Family Physicians; 1993.
  55. Periodic health examination, 1990 update: 2. Early detection of depression and prevention of suicide. Canadian Task Force on the Periodic Health Examination. *Can Med Assoc J.* 1990;142:1233-8.
  56. **U.S. Preventive Services Task Force.** Screening for depression. In: *Guide to Clinical Preventive Services: An Assessment of the Effectiveness of 169 Interventions.* Report of the U.S. Preventive Services Task Force. Baltimore: Williams & Wilkins; 1989.
  57. **Coyne JC, Schwenk TL, Fechner-Bates S.** Nondetection of depression by primary care physicians reconsidered. *Gen Hosp Psychiatry.* 1995; 17:3-12.
  58. **Ormel J, Koeter MW, van den Brink W, van de Willige G.** Recognition, management, and course of anxiety and depression in general practice. *Arch Gen Psychiatry.* 1991;48:700-6.
  59. **Gerety MB, Williams JW Jr, Mulrow CD, Cornell JE, Kadri AA, Rosenberg J, et al.** Performance of case-finding tools for depression in the nursing home: influence of clinical and functional characteristics and selection of optimal threshold scores. *J Am Geriatr Soc.* 1994;42: 1103-9.
  60. **Baron JA.** The clinical utility of risk factor data. *J Clin Epidemiol.* 1989;42:1013-20.
  61. **Boyko EJ, Alderman BW.** The use of risk factors in medical diagnosis: opportunities and cautions. *J Clin Epidemiol.* 1990;43:851-8.
  62. **Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, et al.** Lifetime and 12-month prevalence of DSM-III-R Psychiatric Disorders in the United States. 1994;51:8-19.
  63. **Regier DA, Boyd JH, Burke JD Jr, Rae DS, Myers JK, Kramer M, et al.** One-month prevalence of mental disorders in the United States. *Arch Gen Psychiatry.* 1988;45:977-86.
  64. **Barker LR, Burton JR, Zieve PD, eds.** *Principles of Ambulatory Medicine.* Baltimore: Williams & Wilkins; 1986.
  65. **Crum RM, Cooper-Patrick L, Ford DE.** Depressive symptoms among general medical patients: prevalence and one-year outcome. *Psychosom Med.* 1994;56:109-17.