Is This Patient Clinically Depressed?

John W. Williams, Jr, MD, MHS
Polly Hitchcock Noël, PhD
Jeffrey A. Cordes, MD
Gilbert Ramirez, DrPH
Michael Pignone, MD, MPH

Context Depressive disorders are highly prevalent in the general population, but recognition and accurate diagnosis are made difficult by the lack of a simple confirmatory test.

Objective To review the accuracy and precision of depression questionnaires and the clinical examination for diagnosing clinical depression.

Data Sources We searched the English-language literature from 1970 through July 2000 using MEDLINE, a specialized registry of depression trials, and bibliographies of selected articles.

Study Selection Case-finding studies were included if they used depression questionnaires with easy to average literacy requirements, evaluated at least 100 primary care patients, and compared questionnaire results with accepted diagnostic criteria for major depression. Eleven questionnaires, ranging in length from 1 to 30 questions, were assessed in 28 published studies. Reliability studies for the clinical examination required criterion-based diagnoses made by at least 2 clinicians who interviewed the patient or reviewed a taped examination. Fourteen studies evaluated interrater reliability.

Data Extraction Pairs of authors independently reviewed articles. For case-finding studies, quality assessment addressed sample size and whether patients were selected consecutively or randomly, the criterion standard was administered and interpreted independently of and blind to the results of the case-finding instrument, and the proportion of persons receiving the criterion standard assessment was less than or more than 50% of those approached for criterion standard assessment. For reliability studies, quality assessment addressed whether key patient characteristics were described, the interviewers collected clinical history independently, and diagnoses were made blinded to other clinicians' evaluations.

Data Synthesis In case-finding studies, average questionnaire administration times ranged from less than 1 minute to 5 minutes. The median likelihood ratio positive for major depression was 3.3 (range, 2.3-12.2) and the median likelihood ratio negative was 0.19 (range, 0.14-0.35). No significant differences between questionnaires were found. For mental health care professionals using a semistructured interview, agreement was substantial to almost perfect for major depression (κ=0.64-0.93). Nonstandardized interviews yielded somewhat lower agreement (κ=0.55-0.74). A single study showed that primary care clinicians using a semistructured interview have high agreement with mental health care professionals (κ=0.71).

Conclusions Multiple, practical questionnaires with reasonable performance characteristics are available to help clinicians identify and diagnose patients with major depression. Diagnostic confirmation by mental health care professionals using a clinical interview or by primary care physicians using a semistructured interview can be made with high reliability.

JAMA. 2002;287:1160-1170 www.jama.com

CLINICAL SCENARIO

Mr P is a 52-year-old small-business owner with a 5-year history of controlled hypertension, for which he takes a β-blocker. Otherwise, he is in good health. He presents for routine follow-up and notes a 1-month history of mild to moderate, bitemporal headaches, and feeling fatigued. The headaches occur about twice a week and are relieved by acetaminophen. He denies chest pain or dyspnea on exertion. He notes wryly that the "new economy" has left him feeling a bit "frazzled."

You wonder if the headache and fatigue are stress related, a somatic presentation of depression. Alternatively, is the fatigue related to his antihypertensive medication or a physical illness? What is the most effective and efficient method for diagnosing depression? How does one distinguish between somatic symptoms related to depression vs those related to coexisting physical illness?

Why Is This an Important Question to Answer With a Clinical Examination?

Depressive disorders are prevalent, cause marked personal suffering, and are associated with increased mortality. In primary care settings, the prevalence of major depression ranges from 4.8% to 8.6%, and dysthymia ranges from 2.1% to 3.7%.[1] The World Health Organization estimates that major depression alone was the fourth leading

Author Affiliations: The South Texas Veterans Health Care System, Audie Murphy Division, San Antonio, Tex (Drs Williams, Noël, and Cordes); San Antonio Evidence-based Practice Center (Dr Ramirez) and Department of Psychiatry (Dr Cordes), University of Texas Health Science Center at San Antonio; and Department of Medicine, University of North Carolina and RTI-UNC Evidence-based Practice Center, Chapel Hill (Dr Pignone). Dr Williams is now with the Center for Health Services Research in Primary Care, HSR&D, Department of Veterans Affairs Medical Center, and Duke University Medical Center, Durham, NC.

Corresponding Author and Reprints: John W. Williams, Jr, MD, MHS, Department of Veterans Affairs Medical Center, 508 Fulton St, HSR&D, Bldg 6, Durham, NC 27705 (e-mail: jw.williams@duke.edu).

The Rational Clinical Examination Section Editors: David L. Simel, MD, MHS, Durham Veterans Affairs Medical Center and Duke University Medical Center, Durham, NC; Drummond Rennie, MD, Deputy Editor, *JAMA*.

Table 1. Diagnostic Criteria and Questions to Assess Major Depression*

Symptom	*DSM-IV* Diagnostic Criteria	Suggested Questions
Depressed mood	Depressed mood most of the day, nearly every day	How has your mood been lately? OR Do you ever feel down, depressed, or blue? How often does that happen? How long does it last?
Anhedonia	Markedly diminished interest or pleasure in almost all activities most of the day, nearly every day	Have you lost interest in your usual activities? Do you get less pleasure in things you used to enjoy?
Sleep disturbance	Insomnia or hypersomnia nearly every day	How have you been sleeping? How does that compare to your normal sleep?
Appetite or weight change	Substantial change in appetite nearly every day or unintentional weight loss or gain (eg, ≥5% of body weight in 1 month)	Has there been any change in your appetite or weight?
Decreased energy	Fatigue or loss of energy nearly every day	Have you noticed a decrease in your energy level?
Increased or decreased psychomotor activity	Psychomotor agitation or retardation nearly every day	Have you been feeling fidgety or had problems sitting still? Have you felt slowed down, like you were moving in slow motion or stuck in mud?
Decreased concentration	Diminished ability to think or concentrate, or indecisiveness nearly every day	Have you been having trouble concentrating? Is it harder to make decisions than before?
Guilt or feelings of worthlessness	Feelings of worthlessness or excessive guilt nearly every day	Are you feeling guilty or blaming yourself for things? How would you describe yourself to someone who had never met you before?
Suicidal ideation	Recurrent thoughts of death or suicide	Have you felt that life is not worth living or that you'd be better off dead? Sometimes when a person feels down or depressed, they might think about dying. Have you been having any thoughts like that?

*The diagnosis of major depression requires 5 or more symptoms, including depressed mood or anhedonia, which have been present during the same 2-week period and cause clinically significant distress or impairment in social, occupational, or other important areas of functioning. Adapted from the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)*.[11]

cause of disability worldwide in 1990 and will soon be second only to heart disease as a cause of disability.[2] Antidepressants and depression-specific psychological treatments are clearly effective for depression, improving both depressive symptoms and functional status.[3,4] Many patients can be treated effectively in primary care settings. Quality improvement initiatives[5] and disease management models[6-10] are cost-effective compared with usual care and improve patient outcomes in primary care settings. Until effective prevention strategies are developed, high-quality depression care begins with recognition and accurate diagnosis. This evidence-based review will discuss case-finding and clinical interview strategies for depression diagnosis.

Table 2. Diagnostic Categories for Depressive Disorders

Diagnostic Category	*DSM-IV* Criteria*	Symptom Duration
Major depression	≥5 depressive symptoms, including depressed mood or anhedonia, causing significant impairment in social, occupational, or other important areas of functiong	≥2 wk
Minor depression†	2 to 4 depressive symptoms, including depressed mood or anhedonia, causing significant impairment in social, occupational, or other important areas of functioning	≥2 wk
Dysthymia	3 or 4 dysthymic symptoms, including depressed mood, causing significant impairment in social, occupational, or other important areas of functioning‡	≥2 y

**DSM-IV* indicates *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)*.
†Minor depression is included in *DSM-IV* as a research criteria diagnosis that requires further evaluation.
‡Dysthymic symptoms are depressed mood, poor appetite or overeating, insomnia or hypersomnia, low energy, low self-esteem, poor concentration or indecisiveness, and hopelessness.

Defining Clinical Depression

Clinical depression is a syndromal diagnosis based on patient history and the exclusion of competing diagnoses. Depressive symptoms are evaluated along several continuums: intensity, duration, and impact on daily functioning. Using these elements, symptoms can range from low mood lasting hours or a few days to major depression, characterized by multiple symptoms and substantial impact on daily functioning, based on criteria from the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (*DSM-IV*) (TABLE 1).[11] A diagnostic nomenclature that helps guide treatment is "major depression," "dysthymia," and "depression not otherwise specified." Major depression is defined by depressed mood or loss of interest in nearly all activities for at least 2 weeks accompanied by a minimum of 3 to 4 (for a total of 5) psychological (eg, decreased concentration) or somatic symptoms (eg, insomnia).[11] Dysthymia is a characterized by fewer symptoms than major depression (<5) and a chronic course lasting at least 2 years (TABLE 2). Depression not otherwise specified includes syndromes without a sufficient number of symptoms (<5) or duration (<2 weeks) to meet major depression criteria. Within this category, minor depression, an unofficial diagnosis that has been nominated for further

study, is an example with an insufficient number of symptoms.[11,12]

How to Evaluate Patients for Clinical Depression

There are 2 recommended approaches to recognizing and diagnosing depression. One approach, endorsed by the US Preventive Services Task Force, is to cue physicians to possible clinical depression by asking patients to complete a depression questionnaire during a routine appointment, an approach known as case-finding.[13] Patients who score above a specified threshold are evaluated more carefully for depression. A second approach is to evaluate patients for depression only when the clinical presentation triggers the suspicion of depression.[1] Chronic medical illness, chronic pain syndromes, recent life changes or stressors, fair or poor self-rated health, and unexplained physical symptoms are associated with depression.[14] The likelihood of a depressive disorder increases by approximately 1.5 to 3.5 times if any of these factors is present.[15]

For either approach, a clinical interview is used to make a definitive diagnosis. Experts recommend proceeding from open to narrowly focused questions.[16] In patients such as Mr P who present with somatic symptoms, a transition is recommended from inquiry about these symptoms to questions about emotional health. Many experts find a question such as "How are things at home?" or "How are things at work?" a useful transition. More narrowly focused questions should follow (Table 1), with priority given to questions about mood and anhedonia (a loss of interest or decreased pleasure in activities) since at least 1 of these 2 cardinal symptoms is required to diagnose clinically significant depression. Because successive generations use different synonyms for depressed mood, several alternatives should be offered in the question. For example, it may be helpful to ask, "Have you been feeling sad, down, depressed, or blue?" If answers to questions about mood and anhedonia are "no," clinically significant depression is unlikely and alternative diagnoses should be considered more strongly.

Patients admitting to either depressed mood or anhedonia should be asked additional questions to determine if there are sufficient symptoms to warrant a diagnosis of clinical depression. Assessing the impact of depressive symptoms on functioning and suicide risk are critical elements in the initial treatment decision. A helpful question to assess functioning is, "Have these symptoms of [fill in patient's symptoms] affected your home or work life?" Suicide assessment is more complex. Because patients rarely volunteer thoughts of suicide or their intentions to their physicians, it is important to ask directly. There is no evidence to suggest that asking about suicide precipitates suicidal thinking or acts.[17] One useful screening question is, "Have you been feeling that life is not worth living or that you would be better off dead?"[18] Another approach is to say, "Sometimes when a person feels down or depressed, they might think about dying. Have you been having any thoughts like that?" For patients with suicidal ideation, the next step is to ask, "Do you have a plan?" If a patient answers "yes," inquire about the plan and determine if he or she has assembled the materials required, set a time, and if there are any factors that may precipitate or keep the patient from carrying out the plan. Major risk factors for suicide include hopelessness, substance abuse, and prior suicide attempts. Patients at high risk of suicide should be referred for psychiatric evaluation; those at imminent risk should be evaluated emergently.[19]

Expert guidelines recommend a careful review of systems to detect general medical conditions that may masquerade as depression or complicate its treatment.[1] Physical conditions, such as hypothyroidism or Cushing disease, may cause depression and some experts recommend a thyrotropin measurement in women older than 50 years because of the increased prevalence of hypothyroidism.[1,20] Because these physical conditions are etiologic, treatment is directed at the underlying condition rather than the depressive symptoms. Similarly, medications such as glucocorticoids, anabolic steroids, high-dose reserpine, and cocaine or amphetamine withdrawal may cause depression.[21-23] Other medical conditions such as malignancies, diabetes mellitus, autoimmune disorders, and coronary heart disease are highly associated but not causative for depression, and treatment is directed simultaneously at the clinical depression and the associated physical illness.[1,4,24-27] Diagnostic testing for these disorders is only indicated when clinical symptoms suggest the condition. For example, patients with weight loss out of proportion to the depression should be evaluated for malignancy or other systemic disorders associated with weight loss. Psychiatric illnesses such as alcohol abuse are common in primary care settings and often cooccur with depression.[28] The combination is difficult to treat, often requiring mental health specialty care. The CAGE questions (*C* Have you ever felt the need to *c*ut down on your drinking? *A* Have you ever felt *a*nnoyed by criticism of your drinking? *G* Have you ever felt guilty about your drinking? *E* Have you ever taken a drink [*e*ye opener] first thing in the morning?) are a pragmatic and effective screen for alcohol abuse.[29]

Once depression is diagnosed, additional history should be elicited about factors that may affect treatment. First, explore the patient's understanding and acceptance of the diagnosis. Stigmatizing beliefs about depression or outright rejection of the diagnosis may interfere with treatment adherence. Second, elicit the patient's treatment preferences and information on response to therapy for previous episodes of depression. This is particularly important for pharmacotherapy, as antidepressant agents that have been used successfully for past depressive episodes are likely to be effective and well tolerated for the current episode.[4] Finally, assess the number of prior episodes. The risk of relapse, and hence the need for longer-term treatment, increases with the number of prior episodes.[1,4]

Criterion Standard Diagnosis

Clinical depression is a syndromal diagnosis. There is no physiological or laboratory test, radiological examination, or

tissue diagnosis to definitively establish the diagnosis. Instead, a trained interviewer conducts a clinical interview to determine if the patient meets established criteria. The most commonly used criteria, which are updated periodically, are the *DSM-IV* or the *International Classification of Diseases, Ninth Revision, Clinical Modification*.[11,30]

METHODS

Search Strategy and Inclusion/Exclusion Criteria

We conducted separate searches of MEDLINE and a specialized registry of depression trials[31] for English-language medical literature published from 1970 through July 2000 for studies evaluating the performance of case-finding instruments in primary care settings and the reliability of the clinical interview. All searches included the terms *depressive disorder or depression* and additional terms as appropriate for the specific search. Unpublished data were not sought.

For case-finding, we modified inclusion criteria used in our prior literature synthesis[32] to select instruments that are most readily used in clinical situations. Studies were included if they were conducted in a primary care setting, administered a case-finding instrument, and used a standard interview such as the Structured Clinical Interview (SCID) to make a criterion-based diagnosis (eg, *DSM-III*, *DSM-III-R*, *DSM-IV*) of depression. Further, we specified that the case-finding instrument have easy to average literacy requirements,[33] be scored without a calculator, have a depression-specific component, and be evaluated in at least 1 study with at least 100 subjects. Of 1766 articles identified by the search strategy, 379 potentially eligible studies were reviewed. Twenty-eight studies, involving 11 case-finding instruments, met all inclusion criteria.[34-61]

For reliability studies, we required criterion-based diagnoses made by 2 or more clinicians who interviewed the same patient or reviewed an audiotape or videotape interview. Clinicians evaluated patients with known or suspected psychiatric illness who were recruited from inpatient or outpatient settings in both mental health and general medical settings. Studies using nonclinician interviewers were excluded. Among studies using semistructured interviews, we only included those using the SCID, a commonly used research instrument for diagnosing psychiatric illness.[62] The search yielded 6103 potentially eligible articles, of which 14 met all inclusion/exclusion criteria.[63-76]

Data Abstraction and Statistical Methods

Two independent reviewers abstracted articles. For case-finding studies, quality assessment addressed (1) sample size greater than 100, (2) whether patients were selected consecutively or randomly, (3) whether the criterion standard was administered and interpreted independently of and blind to the results of the case-finding instrument, and (4) whether the proportion of persons receiving the criterion standard assessment was less than or more than 50% of those approached for criterion standard assessment. For reliability studies, quality assessments addressed (1) whether key patient characteristics were described (eg, depression severity), (2) whether the interviewers collected clinical history independently, and (3) whether diagnoses were made blinded to other clinicians' evaluations.

Established cutpoints for case-finding instruments were used except for short versions of original instruments that had proportionally lower thresholds[35,43,46,60] and 1 study that used a higher threshold than originally recommended.[44] Two-by-two tables were used to categorize the number of screened-positive and screened-negative persons who did and did not meet criterion standard diagnosis for major depression. When appropriate, we adjusted for verification bias.[77] Of 11 authors contacted for additional information, 10 responded with the needed data. The average likelihood ratio positive and likelihood ratio negative, weighted by study precision and corrected for 2-stage assessment techniques when indicated, were computed for each case-finding instrument.[78-80] A scattergram plotting true-positive against false-positive rates was constructed to visually evaluate variability among studies. To provide a visual reference for the consistency of study results, we modeled a summary receiver operating characteristic curve based on the logit transformation of the true-positive and false-positive rates. The effectiveness score was used to evaluate overall performance and study heterogeneity.[81] Studies of reliability were not combined quantitatively because of marked heterogeneity in study design.

RESULTS

Accuracy of Case-Finding Questionnaires for Depression

Eleven questionnaires, ranging from 1 to 30 items, met all inclusion criteria (TABLE 3). Six are depression-specific (Beck Depression Inventory [BDI], Center for Epidemiologic Studies Depression Screen [CES-D], Depression Scale [DEPS], Geriatric Depression Scale [GDS], Zung Self-Assessment Depression Scale [SDS], and Single Question [SQ]), 1 addresses depression and anxiety (Duke Anxiety and Depression Scale [DADS]), and 4 are multicomponent (Hopkins Symptom Check List [HSCL]), Primary Care Evaluation of Mental Disorders [PRIME-MD], PRIME-MD Patient Health Questionnaire [PHQ], and Symptom Driven Diagnostic System-Primary Care [SDDS-PC]). All of the questionnnaires can be self-administered in less than 5 minutes and include specific questions aimed at assessing depressed mood, and except for the SQ instrument, all assess anhedonia. Resources to obtain the full instruments are listed in the BOX.

The BDI, the CES-D, and the SDS 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. These 3 instruments are among the most thoroughly evaluated in primary care and can be used to rate the severity of depression and monitor response to therapy. Shortened ver-

sions of the BDI and the CES-D have been tested recently in primary care.[35,43] The GDS exists in both 30- and 15-item versions and uses a "yes" or "no" response format that simplifies telephone administration. It has been evaluated only in populations 60 years old and older. DADS, DEPS, and SQ ("Have you felt depressed or sad much of the time in the past year?") are newer, brief instruments that have been evaluated in single studies.

The PRIME-MD and SDDS-PC instruments are multidimensional questionnaires. Each has screening questions arranged in several categories (eg, depression, anxiety, alcohol abuse) that are used to trigger more extensive diagnostic interviewing sections for specific *DSM* diagnoses. Recently, the PHQ, a completely self-administered version of the PRIME-MD, has been evaluated. It scores each *DSM-IV* depression symptom as present or absent to diagnose depression and can also be scored continuously to monitor response. The HSCL screens for general psychiatric illness and has a specific category for depression.

These instruments, encompassing 37 evaluations in 28 published studies studies,[34-61] involved 25 550 screened patients, of whom 9218 were administered an acceptable criterion standard for diagnosing depression (TABLE 4). Nine of the 28 studies had potential major selection biases because more than 50% of persons selected did not receive a criterion standard assessment, either because they refused the assessment or failed to keep an appointment.[36,37,41,44,51-53,56,58] Considering independent and blind administration of the criterion standard, major selection bias, and sample size, 15 (54%) of the 28 studies were of reasonably high quality for diagnostic test evaluations.

The FIGURE plots the study results for true-positive and false-positive rates for case-finding instruments used to detect major depression. Standard cutpoints were used for these calculations (Table 3) except for 1 study using higher than recommended thresholds for the CES-D.[44] The cutpoint for mild depression was used for the 2 scales with 3

Table 3. Characteristics of Depression Case-Finding Instuments Validated in Primary Care Settings

Instrument*	No. of Items†	Scope	Response Format	Time Frame of Questions	Score Range	Usual Cutpoint‡	Literacy Level§	Administration Time, min	Monitor Severity or Response
BDI	21, 13, 7	Depression-specific (multiple versions)	4 Statements of symptom severity per item	Today	0-63	10-19 mild 20-29 moderate ≥30 severe	Easy	2-5	Yes
CES-D	20, 10	Depression-specific (2 versions)	4 Frequency ratings: "less than 1 day" to "most or all (5-7 days)"	Past week	0-60	≥16	Easy	2-5	Yes
DEPS	10	Depression-specific	4 Frequency ratings: "not at all" to "extremely"	Last month	0-30	≥9	Average	≤2	Unknown
DADS	7	For anxiety and depression	3 Frequency ratings: "yes, somewhat, no" for 3 items; "none, some, a lot" for 4 items	Past week	0-100	>30	Average	≤2	Unknown
GDS	30, 15	Depression-specific (2 versions)	Yes or no	Past week	0-30	≥11	Easy	2-5	Yes
HSCL	25, 13	Multiple versions and multiple components with depression category	4 Frequency ratings: "not at all" to "much more than usual"	Past week	25-100	≥43	Average	2-5	Yes
PRIME-MD	2	Multiple components with depression	Yes or no	Past month	0-2	≥1	Average	<1	No
PRIME-MD (PHQ)	9	Multiple components with depression	4 Frequency ratings: "not at all" to "nearly every day"	Past 2 weeks	0-9 for diagnosis; 0-27 for response	For diagnosis: 5 symptoms For severity: 0-4 none 5-9 mild 10-14 moderate 15-19 major 20-27 severe	Average	<2	Yes
SDDS-PC	5	Multiple components with depression	Yes or no	Past month	0-5	≥2	Easy	<2	Unknown
SDS	20	Depression-specific	4 Frequency ratings: "little of the time" to "most of the time"	Recently	25-110 (sum/80 × 100)	50-59 mild 60-69 moderate ≥70 severe	Easy	2-5	Yes
SQ	1	Depression-specific	Yes or no	Past year	0-1	1	Easy	<1	No

*See "Results" for expansions of case-finding instrument abbreviations.
†Numbers refer to different versions of the same instrument and are listed from most to least number of items. Item numbers for the DADS, PRIME-MD, PRIME-MD (PHQ), and SDDS-PC refer to depression questions only; item numbers for the HSCL refer to depression plus anxiety questions.
‡Cutpoint is given for the instrument version with the most number of items.
§Easy indicates third- to fifth-grade reading level; average, sixth- to ninth-grade reading level.[33]

listed cutpoints (BDI and SDS); the study by Raft et al[54] only had information corresponding to moderate depression for the SDS. Two studies were identified as outliers.[36,54] The study by Raft et al[54] used the higher cutpoint for the SDS scale and had an unusually low sensitivity (31%; 95% confidence interval [CI], 16%-51%). The study by Perez-Stable et al[36] had an unusually low specificity of 40% for the BDI. They studied patients with high levels of medical comorbidity and high ethnic minority representation, factors that may have decreased specificity.

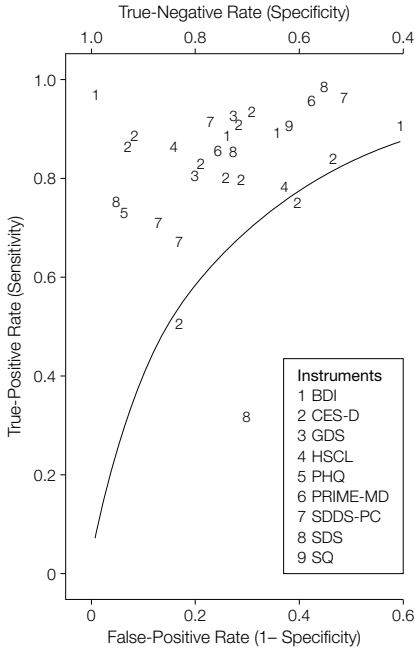
The median likelihood ratio positive for all studies was 3.3 (range: 2.3-12.2), meaning that a positive depression screen is 3.3 times more likely to be seen in someone with depression than in someone without the illness. The median likelihood ratio negative was 0.19 (range: 0.14-0.35), meaning that a negative depression screen was 0.2 times as likely to be seen in someone with depression than in someone without the illness. Performance did not differ significantly between instruments. Using the effectiveness score as a measure of overall performance, there was statistically significant heterogeneity for the BDI, CES-D, HSCL, and SDS ($P<.05$), indicating that the instruments performed variably across the individual studies. The variability may be due to differences in the patient populations or study design. Finally, a subset of studies reported instrument performance for major depression and separately for the combined category of major depression or dysthymia.[55-61] Performance characteristics for detecting this combined category were not statistically significantly different from those for detecting major depression alone.

Given the similar performance, case-finding instruments should be selected based on brevity, response format (particularly if telephone administration is planned), the desire to screen for other psychiatric illnesses, and the need to monitor response. The PHQ best meets these criteria with only 9 items for depression, modules for other psychiatric illness, and a simple response format that is sensitive to change. For clinicians who wish to screen only for depression, the SQ is an attractive alternative that could be asked during preventive medicine evaluations or in response to triggers that increase the likelihood of depression. Positive responses would need to be explored by a more careful clinical interview. In a clinic with an 8% prevalence of major depression or dysthymia, a clinician seeing 100 patients per week can expect that 30 will screen positive for depression, of whom 7 would meet criteria for clinical depression after a more careful clinical interview. Among the 70 patients who screen negative for depression, 1 would have clinical depression. If case-finding were used only in selected high-risk patients (eg, those with chronic pain), a positive screen would more likely be a true positive, but more patients with clinical depression would be missed.

Box. Web Sites for Case-Finding Instruments

Beck Depression Inventory (BDI): http://www.uea.ac.uk/~wp316/depression.pdf
Center for Epidemiologic Studies Depression (CES-D): http://chipts.ucla.edu/Assessment_Instruments/asmt_dp2.html
Depression Scale (DEPS): No Web site available
Duke Anxiety-Depression Scale (DADS): http://healthmeasures.mc.duke.edu/
Geriatric Depression Scale (GDS) - Long Version: http://www.stanford.edu/~yesavage/GDS.english.long.html
GDS - Short Version: http://www.stanford.edu/~yesavage/GDS.english.short.score.html
Hopkins Symptom Checklist (HSCL): No Web site available
Primary Care Evaluation of Mental Disorders (PRIME-MD): No Web site available (pdf file available from Robert Spitzer, MD, at the following e-mail address: RLS8@columbia.edu)
PRIME-MD Patient Health Questionnaire (PHQ): http://www.cmecenter.com/primemdtoday/
Symptom Driven Diagnostic System-Primary Care (SDDS-PC): No Web site available
The Zung Self-Rating Depression Scale (SDS): http://www.wellbutrin-sr.com/hcp/depression/zung.html

Figure. Plot of True-Positive Rate Against False-Positive Rate for Case-Finding Instruments to Detect Major Depression



The numbered points represent individual studies. The curve represents the summary receiver-operating curve through all data points. Standard cutpoints (see Table 3) were used for these calculations except for 1 study[44] that used higher than recommended thresholds for the CES-D. See "Results" for expansion of case-finding instrument abbreviations.

Accuracy and Reliability of the Clinical Interview for Depression

Because the criterion standard diagnosis is based on a clinical interview, there is no simple method for establishing its accuracy. However, we identified relevant studies comparing the diagnostic agreement between 2 mental health

professionals, between primary care physicians and a mental health professional, and the effects of coexisting medical illness on reliability.

We identified 7 studies using the SCID, which evaluated inter-rater reliability for major depression (TABLE 5).[63-69] The SCID is a widely used research instrument that uses a semistructured interview to elicit symptoms that are applied to the current *DSM* criteria to establish a diagnosis.[62] It is designed in part to decrease variability related to the range of symptoms explored and the manner in which a clinical interviewer presents questions. Study design varied considerably, ranging from multiple clinicians viewing a videotaped interview, to paired interviewers conducting sequential interviews. Examiners' training and experience ranged from psychology trainees to practicing psychiatrists with a special interest in mood disorders. All were conducted in

Table 4. Case-Finding Instrument Performance in Primary Care Settings

Source, y	Instrument*	Sample Size†	Population‡	Likelihood Ratio Positive (95% CI)§	Likelihood Ratio Negative (95% CI)‖	Quality Score¶
			Major Depressive Disorder			
Whooley et al,[34] 1997	BDI	536	Veterans Affairs	2.46	0.18	1
Steer et al,[35] 1999	BDI, 7 items	120	Academic	96.55	0.03	3
Perez-Stable et al,[36] 1990	BDI	265	Mixed	1.50	0.25	4
Zich et al,[37] 1990	BDI	31	Mixed	3.37	0.17	4
Summary				4.2 (1.2-13.6)	0.17 (0.1-0.3)	
Kirmayer et al,[38] 1993	CES-D	685	Academic	3.05	0.28	1
Whooley et al,[34] 1997	CES-D	536	Veterans Affairs	2.99	0.10	1
Williams et al,[39] 1999	CES-D	296	Mixed	3.25	0.13	1
Fechner-Bates et al,[40] 1994	CES-D	425	Community	2.73	0.29	2
Hendrie et al,[41] 1995	CES-D	125	Academic (age ≥60 y)	2.94	0.60	2
Hough et al,[42] 1983	CES-D	525	Health maintenance organization	3.91	0.23	2
Irwin et al,[43] 1999	CES-D, 10 items	68	Academic (age ≥60 y)	10.94	0.14	3
Lyness et al,[44] 1997	CES-D	130	Community (age ≥60 y)	12.24	0.15	4
Perez-Stable et al,[36] 1990	CES-D	214	Mixed	1.94	0.40	4
Zich et al,[37] 1990	CES-D	34	Mixed	1.77	0.31	4
Summary				3.3 (2.5-4.4)	0.24 (0.2-0.3)	
Neal and Baldwin,[45] 1994	GDS	45	Academic (age >65 y)	4.00	0.25	2
D'Ath et al,[46] 1994	GDS, 15 items	120	Community (age ≥65 y)	3.26	0.12	4
Summary				3.3 (2.4-4.7)	0.16 (0.1-0.3)	
Schmitz et al,[47] 1999	HSCL	421	Community	2.02	0.37	1
Hough et al,[42] 1983	HSCL	525	Health maintenance organization	5.36	0.17	2
Summary				3.2 (1.7-6.2)	0.24 (0.1-0.5)	
Spitzer et al,[48] 1999	PHQ	585	Mixed	12.2 (8.4-18)	0.28 (0.2-0.5)	1
Spitzer et al,[49] 1994	PRIME-MD	431	Mixed	3.42	0.19	1
Whooley et al,[34] 1997	PRIME-MD	536	Veterans Affairs	2.23	0.07	1
Summary				2.7 (2.0-3.7)	0.14 (0.1-0.3)	
Leon et al,[50] 1996	SDDS-PC	501	Community	5.43	0.34	1
Whooley et al,[34] 1997	SDDS-PC	536	Veterans Affairs	1.96	0.08	1
Broadhead et al,[51] 1995#	SDDS-PC	388	Community	3.95	0.12	3
Broadhead et al,[51] 1995	SDDS-PC	257	Mixed	3.92	0.40	3
Summary				3.5 (2.4-5.1)	0.22 (0.1-0.4)	
Spitzer et al,[49] 1994	SDS	337	Mixed	3.31	0.19	1
Okimoto et al,[52] 1982	SDS	55	Veterans Affairs (age ≥60 y)	2.16	0.05	3
Magruder-Habib et al,[53] 1990	SDS	206	Veterans Affairs	14.86	0.27	4
Raft et al,[54] 1977	SDS	69	Academic	1.03	0.97	4
Summary				3.3 (1.3-8.1)	0.35 (0.2-0.8)	
Williams et al,[39] 1999	SQ	291	Mixed	2.3 (1.8-2.9)	0.16 (0.0-0.6)	1
Median performance for all instruments				3.3	0.19	

(continued)

mental health specialty settings. Diagnoses were made blind to the other raters' diagnosis in 6 studies, patient histories were elicited independently in only 1 study, and no study described depression severity. Despite the variability in study design and examiner training, interrater agreement corrected for chance was substantial to almost perfect (κ=0.64-0.93). These studies show that major depression can be diagnosed reliably by a mental health professional who uses a careful, semistructured interview.

Studies that use nonstandardized interviews to make *DSM*-based diagnoses may better simulate clinical practice. Seven studies, involving psychiatry trainees to practicing psychiatrists, evaluated interrater agreement using this approach.[70-76] Diagnoses were based typically on paired interviewers conducting joint or sequential interviews; 1 study used a videotaped interview.[71] Diagnoses were made blind to the other raters' diagnosis in 5 studies, patient histories were elicited independently for most patients in 3 studies, and no study described depression severity. Interrater agreement corrected for chance was moderate to substantial (κ=0.55-0.74). Compared with semistructured interviews, agreement was somewhat lower for nonstandardized interviews. However, chance-corrected agreement remained good compared with many other clinical findings.[82-84] Less is known about the reliability of depression diagnoses made by primary care physicians. We were able to identify only 1 study that compared a primary care clinician's diagnoses based on *DSM* criteria to that of a mental health professional using the same criteria. Spitzer et al[49] compared primary care clinicians' diagnoses using a semistructured instrument to mental health professionals' diagnoses using an SCID-based *DSM* measure of depression. This study found good agreement (simple agreement, 88%; κ=0.71). It is unknown how well primary care physicians using a nonstandardized interview would agree with diagnoses made by mental health professionals.

These studies have a number of design limitations. The severity of major depression and spectrum of competing medical and psychological illnesses that may make diagnosis more difficult were not typically described. In addition, studies using joint interviews and videotape review may overestimate interrater reliability because both interviewers hear identical information. Two of the studies compared diagnoses made by emergency department psychiatrists to those made by the patient's inpatient treating physician and were thus not blinded evaluations, again potentially biasing these studies toward higher agreement.[73,76] Finally, only 1 study reported 95% CIs for the estimate of interrater agreement.[69]

Effect of Physical Illness on Diagnosis

Because the psychological and physical symptoms of depression may overlap with other physical illness, diagnosing depression in patients with severe or multiple chronic medical illnesses can be especially challenging.[85] If symptoms due to the physical illness (eg, fatigue related to congestive heart failure) are attributed to depression, then patients may receive unnecessary treatment. Conversely, if depressive symptoms are misattributed to a concurrent physical illness, then effective depression treatment may be withheld. A number of different strategies have been proposed in an attempt to improve the accuracy and reliability of diagnosis in physically ill patients. The "inclusive" approach counts depressive symptoms toward the diagnosis of depression, regardless of whether the clinician judges that the symptom is due to medical or psychological causes. The *DSM-IV* criteria use an "etiologic" approach that counts symptoms toward a major depression diagnosis unless the symptom is "clearly and fully accounted for by a general medical condition."[11] Because clinicians must make a judgment about the cause of individual symp-

Table 4. Case-Finding Instrument Performance in Primary Care Settings (cont)

Source, y	Instrument*	Sample Size†	Population‡	Likelihood Ratio Positive (95% CI)§	Likelihood Ratio Negative (95% CI)‖	Quality Score¶
Major Depressive Disorder or Dysthymia						
Klinkman et al,[55] 1997	CES-D	425	Mixed	2.88	0.27	2
Schulberg et al,[56] 1985	CES-D	294	Community	5.19	0.26	4
Salokangas et al,[57] 1995	DEPS	436	Community	4.93	0.31	2
Parkerson and Broadhead,[58] 1997	DADS	481	Academic	2.27	0.28	3
Van Marwijk et al,[59] 1995	GDS, 30 items	586	Community (age ≥65 y)	3.90	0.53	1
Arthur et al,[60] 1999	GDS, 15 items	201	Community (age ≥75 y)	3.43	0.05	3
Nettelbladt et al,[61] 1993	HSCL	186	Community	2.83	0.36	2
Median performance				3.9	0.30	

*See "Results" for expansions of case-finding instrument abbreviations.
†The sample size refers to the actual number who received the criterion standard.
‡Mixed-community and university-affiliated clinics, academic university–affiliated clinics, and community–private practice clinics.
§The likelihood ratio positive describes how much more likely a positive depression screen would be seen in an individual with depression than in someone without depression. It is calculated as sensitivity/(1−specificity). Summary is a weighted average across all studies. CI indicates confidence interval.
‖The likelihood ratio negative describes how much more likely a negative depression screen would be seen in an individual with depression than in someone without depression. It is calculated as (1−sensitivity)/(specificity). Summary is a weighted average across all studies.
¶Lower scores indicate higher quality.
#The study by Broadhead et al is listed twice for the same instrument because it included both an initial test set of patients and a validation set of patients.

toms, this approach may be less reliable than the inclusive approach. A third strategy, called the "substitutive" approach, replaces depression criterion symptoms that are most likely to be confused with medical illness (eg, loss of energy, weight loss, impaired concentration) with psychological symptoms.[86] This approach was developed in an attempt to better discriminate between patients with depression and physical illness and those with only physical illness. Koenig et al[87] evaluated these strategies in a consecutive series of elderly, hospitalized patients. The prevalence of major depression was 20.7% using the inclusive approach, 16.5% using the etiologic approach, and 15.0% using the substitutive approach. Measures of depression severity and the need for treatment did not differ significantly across the 3 diagnostic approaches. For minor depression, both the prevalence and measures of severity varied more significantly. In a related study, interobserver agreement among mental health professionals was slightly higher for the inclusive approach ($\kappa = 1.0$) than for the etiologic approach ($\kappa = 0.88$).[88] Two other studies have shown high levels of agreement between the etiologic and substitutive approaches.[89,90] Although the data are limited, these studies show high concordance between the different approaches and high interobserver agreement in physically ill patients. Because the substitutive approach requires learning new criteria and does not offer a clear advantage, we recommend the inclusive or etiologic approaches.

How Can I Improve My Skills for Diagnosing Depression?

Observational and trial data suggest that specific communication and interviewing skills are related to diagnostic performance. Three studies using "standardized patients," or actors presenting with a scripted set of complaints, suggest that physicians are more likely to recognize or diagnose depression when they ask questions about feelings or psychosocial issues.[91-93] In one of these studies, recognition rates approached 100% if physicians asked about mood and anhedonia.[93]

We did not identify any trials designed to improve the accuracy or reliability of diagnostic interviews for depression. Existing trials focus primarily on improving recognition rates, or sen-

Table 5. Interrrater Reliability for Depressive Disorder Using Semistructured and Nonstructured Interviews*

Source, y	Examiners (N)	No. of Patients Evaluated	No. of Patients With MDD Diagnosis	Setting	Design	Independent Assessment†	Blinding‡	Simple Agreement (κ)
Semistructured Interview								
Fuhrer et al,[63] 1986	Psychiatrists (136)	11	2	Inpatient	Videotape review	No	Yes	NA (0.78)
Riskind et al,[64] 1987	Psychologists (16)	75	25	Outpatient	Videotape review	No	Yes	82% (0.72)
Stukenberg et al,[65] 1990	Psychology trainees (4)	75	NA	Community	Not stated	NS	NS	NA (0.92)
Skre et al,[66] 1991	Psychiatrist (1) Psychologists (4)	54	25	Mixed	Live vs audiotape review	No	Yes	NA (0.93)
Williams et al,[67] 1992	Psychiatrists (14) Psychologists (6) Master's degree (4)	390	121	Mixed	Live, sequential interview	Yes	Yes	NA (0.64)
Segal at el,[68] 1994	Psychology trainees (NS)	33	15	Outpatient	Live vs audiotape or videotape review	No	Yes	85% (0.70)
Keller et al,[69] 1995	Master's degree (NS)	80	68	Mixed	Live vs videotape review	No	Yes	NA (0.72)
Nonstandardized Interview								
Spitzer et al,[70] 1979	Mental health clinicians (274)	281	83	Mixed	Joint or sequential interview	Mixed	Yes	NA (0.70)
Webb et al,[71] 1981	Mental health clinicians (78)	1	1	NA	Videotape review	No	Yes	83% (NA)
Hyler et al,[72] 1982	Psychiatrists (31) Psychologists (3) Social workers (7)	46	14§	Mixed	Joint or sequential interview	Mixed	Yes	NA (0.55)
Lieberman and Baker,[73] 1985	Psychiatrists (NS)	50	6	Emergency department	Sequential interview	NS	No	NA (0.62)
Mellsop et al,[74] 1991	Psychiatrist (5)	60	32§	Inpatient	Joint interview	No	Yes	NA (0.70)
Buchwald and Rudick-Davis,[75] 1993	Psychiatry residents (25) Psychology trainee (1)	43	38	Emergency department	Joint or sequential interview	Mixed	Yes	88% (0.74)
Warner and Peabody,[76] 1995	Psychiatry residents (30) Psychiatrists (6)	190	74	Emergency department	Sequential interview	NS	No	NA (0.64)

*NS indicates not stated; MDD, major depressive disorder; and NA, not available (not reported by authors).
†Yes indicates history obtained independently by 2 or more observers; mixed, history obtained independently for some but not all subjects.
‡Yes indicates diagnosis made without knowledge of other examiners' diagnosis.
§Patients had "affective disorder" rather than the more specific MDD.

sitivity, which is only 1 aspect of diagnostic accuracy. Four randomized controlled trials of continuing medical education programs for physicians (N=329) show generally positive results.[94-97] Three of the trials focused specifically on or included recognizing depression[94-96] and the fourth trial focused more generally on communication skills training designed to address patients' emotional distress.[97] Trained vs untrained physicians were significantly more likely to recognize depression or psychosocial problems in the 2 trials that provided 8-hour training sessions and emphasized communication or interviewing skills[95,96] or in the trial that provided access to an on-site consulting psychiatrist following a shorter training session.[94] These data suggest that motivated physicians can improve their communication skills and sensitivity to emotional distress or depressive disorder. Medical schools and residency programs should consider incorporating similar training in their curricula.

SCENARIO RESOLUTION

You follow up on Mr P's "frazzled" comment and learn that he has been under intense work-related stress. Knowing that recent stress increases the likelihood of clinical depression you ask, "Have you been feeling sad or depressed much of the time?" Mr P has been feeling down nearly every day for several weeks and on further questioning meets criteria for major depression. A focused review of systems and physical examination does not suggest a secondary cause of depression. He does not drink alcohol and has no prior history of depression. From medical school you recall a possible association between depression and β-blocker use. A quick check in a current evidence-based reference book[98] shows the surprising finding that the association is not well supported. With some relief that his antihypertensive medication does not need to be changed, you discuss both antidepressant medication and psychological treatment options.

Author Contributions: *Study concept and design:* Williams, Noël, Cordes, Pignone.
Acquisition of data: Williams, Cordes, Pignone.
Analysis and interpretation of data: Williams, Cordes, Ramirez, Pignone.
Drafting of the manuscript: Williams, Noël, Cordes,
Critical revision of the manuscript for important intellectual content: Williams, Noël, Cordes, Ramirez, Pignone.
Statistical expertise: Williams, Ramirez.
Obtained funding: Williams, Pignone.
Administrative, technical, or material support: Noël.
Funding/Support: This study was supported by the San Antonio Veterans Affairs Health Services Research and Development Center of Excellence and the Agency for Healthcare Quality and Research.
Disclaimer: The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.
Acknowledgment: We thank Ramon Velez, MD, David Steffens, MD, MHSc, and David Simel, MD, MHSc, for their careful manuscript review, and Elizabeth O. Cain, BA, who assisted with manuscript preparation.

REFERENCES

1. Depression Guideline Panel. *Depression in Primary Care, I: Detection and Diagnosis: Clinical Practice Guideline No. 5.* Rockville, Md: US Dept of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research; 1993. AHCPR Publication 93-0550.
2. Murray CL, Lopez AD. *The Global Burden of Disease.* Cambridge, Mass: Harvard University Press; 1996.
3. Mulrow CD, Williams JW Jr , Trivedi M, et al. *Treatment of Depression: New Pharmacotherapies.* Rockville, Md: Agency for Health Care Policy and Research; 1999. AHCPR Publication 99-E014.
4. Depression Guideline Panel. *Depression in Primary Care: II: Treatment of Major Depression.* Rockville, Md: US Dept of Health and Human Services, Public Health Service; 1993. AHCPR publication 93-0551.
5. Wells KB, Sherbourne C, Schoenbaum M, et al. Impact of disseminating quality improvement programs for depression in managed primary care: a randomized controlled trial. *JAMA.* 2000;283:212-220.
6. Katon W, Von Korff M, Lin E, et al. Collaborative management to achieve treatment guidelines: impact on depression in primary care. *JAMA.* 1995;273:1026-1031.
7. Katon W, von Korff M, Lin E, et al. Collaborative management to achieve depression treatment guidelines. *J Clin Psychiatry.* 1997;58(suppl 1):20-23.
8. Hunkeler EM, Meresman JF, Hargreaves WA, et al. Efficacy of nurse telehealth care and peer support in augmenting treatment of depression in primary care. *Arch Fam Med.* 2000;9:700-708.
9. Rost K, Nutting PA, Smith J, Werner J, Duan N. Designing and implementing a primary care intervention trial to improve the quality and outcome of care for major depression. *Gen Hosp Psychiatry.* 2001;22:66-77.
10. Katzelnick DJ, Simon GE, Pearson SD, et al. Randomized trial of a depression management program in high utilizers of medical care. *Arch Fam Med.* 2000; 9:345-351.
11. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition.* Washington, DC: American Psychiatric Association; 1994.
12. Pincus HA, Davis WW, McQueen LE. "Subthreshold" mental disorders: a review and synthesis of studies on minor depression and other "brand names." *Br J Psychiatry.* 1999;174:288-296.
13. Pignone MP, Gaynes BN, Rushton JL, et al. Screening for depression in adults: a summary of the evidence for the US Preventive Services Task Force. *Ann Intern Med.* In press.
14. Kroenke K, Spitzer RL, Williams JB, et al. Physical symptoms in primary care: predictors of psychiatric disorders and functional impairment. *Arch Fam Med.* 1994;3:774-779.
15. Kroenke K, Jackson JL, Chamberlin J. Depressive and anxiety disorders in patients presenting with physical complaints: clinical predictors and outcome. *Am J Med.* 1997;103:339-347.
16. Lipkin M Jr, Putnal SM, Lazare A, eds. *The Medical Interview: Clinical Care, Education, and Research.* New York, NY: Springer-Verlag; 1995.
17. Zimmerman M, Lish JD, Lush DT, Farber NJ, Plescia G, Kuzma MA. Suicidal ideation among urban medical outpatients. *J Gen Intern Med.* 1995;10:573-576.
18. Zimmerman M. *Diagnosing DSM-IV Psychiatric Disorders in Primary Care Settings: An Interview Guide for the Nonpsychiatrist Physician.* East Greenwich, RI: Psych Products Press; 1994.
19. Hirschfeld RMA, Russell JM. Assessment and treatment of suicidal patients. *N Engl J Med.* 1997;337: 910-915.
20. Whooley MA, Simon GE. Primary care: managing depression in medical outpatients. *N Engl J Med.* 2000;343:1942-1950.
21. Goodwin FK, Bunney WE Jr. Depressions following reserpine: a reevaluation. *Semin Psychiatry.* 1971; 3:435-448.
22. Lewis DA, Smith RE. Steroid-induced psychiatric syndromes: a report of 14 cases and a review of the literature. *J Affect Disord.* 1983;5:319-332.
23. Pope HG Jr, Katz DL. Affective and psychotic symptoms associated with anabolic steroid use. *Am J Psychiatry.* 1988;145:487-490.
24. Bottomley A. Depression in cancer patients: a literature review. *Eur J Cancer Care.* 1998;7:181-191.
25. Gavard JA, Lustman PJ, Clouse RE. Prevalence of depression in adults with diabetes: an epidemiological evaluation. *Diabetes Care.* 1993;16:1167-1178.
26. Fifield J, Tennen H, Reisine S, McQuillan J. Depression and the long-term risk of pain, fatigue, and disability in patients with rheumatoid arthritis. *Arthritis Rheum.* 1998;41:1851-1857.
27. Schleifer SJ, Macari-Hinson MM, Coyle DA, et al. The nature and course of depression following myocardial infarction. *Arch Intern Med.* 1989;149:1785-1789.
28. Ballenger JC, Davidson JR, Lecrubier Y, et al. Consensus statement on the primary care management of depression from the International Consensus Group on Depression and Anxiety. *J Clin Psychiatry.* 1999;60 (suppl 7):54-61.
29. Buchsbaum DG, Buchanan RG, Centor RM, Schnoll SH, Lawton MJ. Screening for alcohol abuse using CAGE scores and likelihood ratios. *Ann Intern Med.* 1991;115:774-777.
30. *International Classification of Diseases, Ninth Revision, Clinical Modification.* Washington, DC: Public Health Service, US Dept of Health and Human Services; 1988.
31. The Cochrane Controlled Trials Register. Oxford, England: Cochrane Library, Update Software; 2000;issue 3.
32. Mulrow CD, Williams JW Jr, Gerety MB, Kerber CA, Ramirez G. Case-finding instruments for depression in primary care settings. *Ann Intern Med.* 1995; 12:913-921.
33. Gunning R. *The Technique of Clear Writing.* New York, NY: McGraw-Hill International Book Co; 1952.
34. Whooley MA, Avins AL, Miranda J, Browner WS. Case-finding instruments for depression: two questions are as good as many. *J Gen Intern Med.* 1997; 12:439-445.
35. Steer RA, Cavalieri TA, Leonard DM, Beck AT. Use of the Beck Depression Inventory for Primary Care to screen for major depression disorders. *Gen Hosp Psychiatry.* 1999;21:106-111.
36. Perez-Stable EJ, Miranda J, Munoz RF, Ying YW. Depression in medical outpatients: underrecognition and misdiagnosis. *Arch Intern Med.* 1990;150:1083-1088.
37. Zich JM, Attkisson CC, Greenfield TK. Screening for depression in primary care clinics: the CES-D and the BDI. *Int J Psychiatry Med.* 1990;20:259-277.

38. Kirmayer LJ, Robbins JM, Dworkind M, Yaffe MJ. Somatization and the recognition of depression and anxiety in primary care. *Am J Psychiatry.* 1993;150:734-741.
39. Williams JW Jr, Mulrow CD, Kroenke K, et al. Case-finding for depression in primary care: a randomized trial. *Am J Med.* 1999;106:36-43.
40. Fechner-Bates S, Coyne JC, Schwenk TL. The relationship of self-reported distress to depressive disorders and other psychopathology. *J Consult Clin Psychol.* 1994;62:550-559.
41. Hendrie HC, Callahan CM, Levitt EE, et al. Prevalence rates of major depressive disorders: the effects of varying the diagnostic criteria in an older primary care population. *Am J Geriatr Psychiatry.* 1995;2:119-131.
42. Hough R, Landsverk J, Stone J, et al. *Comparison of Psychiatric Screening Questionnaires for Primary Care Patients.* Bethesda, Md: National Institute of Mental Health; 1983. Final report for NIMH contract 278-81-0036 (DB) edition.
43. Irwin M, Artin KH, Oxman MN. Screening for depression in the older adult: criterion validity of the 10-item Center for Epidemiological Studies Depression Scale (CES-D). *Arch Intern Med.* 1999;159:1701-1704.
44. Lyness JM, Noel TK, Cox C, King DA, Conwell Y, Caine ED. Screening for depression in elderly primary care patients: a comparison of the Center for Epidemiologic Studies-Depression Scale and the Geriatric Depression Scale. *Arch Intern Med.* 1997;157:449-454.
45. Neal RM, Baldwin RC. Screening for anxiety and depression in elderly medical outpatients. *Age Aging.* 1994;23:461-464.
46. D'Ath P, Katona P, Mullan E, Evans S, Katona C. Screening, detection and management of depression in elderly primary care attenders, I: the acceptability and performance of the 15 item Geriatric Depression Scale (GDS15) and the development of short versions. *Fam Pract.* 1994;11:260-266.
47. Schmitz N, Kruse J, Heckrath C, Alberti L, Tress W. Diagnosing mental disorders in primary care: the General Health Questionnaire (GHQ) and the Symptom Check List (SCL-90-R) as screening instruments. *Soc Psychiatry Psychiatr Epidemiol.* 1999;34:360-366.
48. Spitzer RL, Kroenke K, Williams JBW, and the Patient Health Questionnaire Primary Care Study Group. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *JAMA.* 1999;282:1737-1744.
49. Spitzer RL, Williams JB, Kroenke K, et al. Utility of a new procedure for diagnosing mental disorders in primary care: the PRIME-MD 1000 study. *JAMA.* 1994;272:1749-1756.
50. Leon AC, Olfson M, Weissman MM, et al. Brief screens for mental disorders in primary care. *J Gen Intern Med.* 1996;11:426-430.
51. Broadhead WE, Leon AC, Weissman MM, et al. Development and validation of the SDDS-PC screen for multiple mental disorders in primary care. *Arch Fam Med.* 1995;4:211-219.
52. 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.
53. Magruder-Habib K, Zung WW, Feussner JR. Improving physicians' recognition and treatment of depression in general medical care: results from a randomized clinical trial. *Med Care.* 1990;28:239-250.
54. Raft D, Spencer RF, Toomey T, Brogan D. Depression in medical outpatients: use of the Zung scale. *Dis Nerv Syst.* 1977;38:999-1004.
55. Klinkman MS, Coyne JC, Gallo S, Schwenk TL. Can case-finding instruments be used to improve physician detection of depression in primary care? *Arch Fam Med.* 1997;6:567-573.
56. Schulberg HC, Saul M, McClelland M, Ganguli M, Christy W, Frank R. Assessing depression in primary medical and psychiatric practices. *Arch Gen Psychiatry.* 1985;42:1164-1170.
57. Salokangas RK, Poutanen O, Stengard E. Screening for depression in primary care: development and validation of the Depression Scale, a screening instrument for depression. *Acta Psychiatr Scand.* 1995;92:10-16.
58. Parkerson GR Jr, Broadhead WE. Screening for anxiety and depression in primary care with the Duke Anxiety-Depression Scale. *Fam Med.* 1997;29:177-181.
59. van Marwijk HW, Wallace P, de Bock GH, Hermans J, Kaptein AA, Mulder JD. Evaluation of the feasibility, reliability and diagnostic value of shortened versions of the geriatric depression scale. *Br J Gen Pract.* 1995;45:195-199.
60. Arthur A, Jagger C, Lindesay J, Graham C, Clarke M. Using an annual over-75 health check to screen for depression: validation of the short Geriatric Depression Scale (GDS15) within general practice. *Int J Geriatr Psychiatry.* 1999;14:431-439.
61. Nettelbladt P, Hansson L, Stefansson CG, Borgquist L, Nordstrom G. Test characteristics of the Hopkins Symptom Check List-25 (HSCL-25) in Sweden, using the Present State Examination (PSE-9) as a caseness criterion. *Soc Psychiatry Psychiatr Epidemiol.* 1993;28:130-133.
62. Spitzer RL, Williams JB, Gibbon M, First MB. *Structured Clinical Interview for DSM-III-R: Patient Edition.* Washington, DC: American Psychiatric Press; 1991.
63. Fuhrer R, Rouillon F, Lellouch J. Diagnostic reliability among French psychiatrists using DSM-III criteria. *Acta Psychiatr Scand.* 1986;73:12-16.
64. Riskind JH, Beck AT, Berchick RJ, Brown G, Steer RA. Reliability of DSM-III diagnoses for major depression and generalized anxiety disorder using the structured clinical interview for DSM-III. *Arch Gen Psychiatry.* 1987;44:817-820.
65. Stukenberg KW, Dura JR, Kiecolt-Glaser JK. Depression screening scale validation in an elderly, community-dwelling population. *Psychol Assess.* 1990;2:134-138.
66. Skre I, Onstad S, Torgersen S, Kringlen E. High interrater reliability for the Structured Clinical Interview for DSM-III-R Axis I (SCID-I). *Acta Psychiatr Scand.* 1991;84:167-173.
67. Williams JB, Gibbon M, First MB, et al. The Structured Clinical Interview for DSM-III-R (SCID), II: multisite test-retest reliability. *Arch Gen Psychiatry.* 1992;49:630-636.
68. 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.
69. Keller MB, Klein DN, Hirschfeld RM, et al. Results of the DSM-IV mood disorders field trial. *Am J Psychiatry.* 1995;152:843-849.
70. Spitzer RL, Forman JB, Nee J. DSM-III field trials, I: initial interrater diagnostic reliability. *Am J Psychiatry.* 1979;136:815-817.
71. Webb LJ, Gold RS, Johnstone EE, Diclemente CC. Accuracy of DSM-III diagnoses following a training program. *Am J Psychiatry.* 1981;138:376-378.
72. Hyler SE, Williams JB, Spitzer RL. Reliability in the DSM-III field trials: interview v case summary. *Arch Gen Psychiatry.* 1982;39:1275-1278.
73. Lieberman PB, Baker FM. The reliability of psychiatric diagnosis in the emergency room. *Hosp Community Psychiatry.* 1985;36:291-293.
74. Mellsop GW, Thomas CS, Ellis PM, Purdie G, Crawshaw J, Mendis N. Reliability of the draft diagnostic criteria for research of ICD-10 in comparison with ICD-10 and DSM-III-R. *Acta Psychiatr Scand.* 1991;84:332-335.
75. Buchwald AM, Rudick-Davis D. The symptoms of major depression. *J Abnorm Psychol.* 1993;102:197-205.
76. Warner MD, Peabody CA. Reliability of diagnoses made by psychiatric residents in a general emergency department. *Psychiatr Serv.* 1995;46:1284-1286.
77. Greenes RA, Begg CB. Assessment of diagnostic technologies: methodology for unbiased estimation from samples of selectively verified patients. *Invest Radiol.* 1985;20:751-756.
78. Sackett DL. A primer on the precision and accuracy of the clinical examination. *JAMA.* 1992;267:2638-2644.
79. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol.* 1991;44:763-770.
80. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. *J Clin Epidemiol.* 1995;48:119-130.
81. Hasselblad V, Hedges LV. Meta-analysis of screening and diagnostic tests. *Psychol Bull.* 1995;117:167-178.
82. Norden C, Philipps E, Levy P, Kass E. Variation in interpretation of intravenous pyelograms. *Am J Epidemiol.* 1970;91:155-160.
83. Edelman D, Hough DM, Glazebrook KN, Oddone EZ. Prognostic value of the clinical examination of the diabetic foot ulcer. *J Gen Intern Med.* 1997;12:537-543.
84. Deyo RA, McNiesh LM, Cone RO 3d. Observer variability in the interpretation of lumbar spine radiographs. *Arthritis Rheum.* 1985;28:1066-1070.
85. Koenig HG, Cohen HJ, Blazer DG, Krishnan KR, Sibert TE. Profile of depressive symptoms in younger and older medical inpatients with major depression. *J Am Geriatr Soc.* 1993;41:1169-1176.
86. Endicott J. Measurement of depression in patients with cancer. *Cancer.* 1984;53(10 suppl):2243-2249.
87. Koenig HG, George LK, Peterson BL, Pieper CF. Depression in medically ill hospitalized older adults: prevalence, characteristics, and course of symptoms according to six diagnostic schemes. *Am J Psychiatry.* 1997;154:1376-1383.
88. Koenig HG, Pappas P, Holsinger T, Bachar JR. Assessing diagnostic approaches to depression in medically ill older adults: how reliably can mental health professionals make judgments about the cause of symptoms? *J Am Geriatr Soc.* 1995;43:472-478.
89. Rapp SR, Vrana S. Substituting nonsomatic for somatic symptoms in the diagnosis of depression in elderly male medical patients. *Am J Psychiatry.* 1989;146:1197-2000.
90. Kathol RG, Mutgi A, Williams JW Jr, Clamon G, Noyes R. Diagnosis of major depression in cancer patients according to four sets of criteria. *Am J Psychiatry.* 1990;147:1021-1024.
91. Stange KC, Fedirko T, Zyzanski SJ, Jaen CR. How do family physicians prioritize delivery of multiple preventive services? *J Fam Pract.* 1994;38:231-237.
92. Carney PA, Dietrich AJ, Eliassen MS, Owen M, Badger LW. Recognizing and managing depression in primary care: a standardized patient study. *J Fam Pract.* 1999;48:965-972.
93. Carney PA, Eliassen MS, Wolford GL, Owen M, Badger LW, Dietrich AJ. How physician communication influences recognition of depression in primary care. *J Fam Pract.* 1999;48:958-964.
94. Worrall G, Angel J, Chaulk P, Clarke C, Robbins M. Effectiveness of an educational strategy to improve family physicians' detection and management of depression: a randomized controlled trial. *CMAJ.* 1999;161:37-40.
95. Gerrity MS, Cole SA, Dietrich AJ, Barrett JE. Improving the recognition and management of depression: is there a role for physician education? *J Fam Pract.* 1999;48:949-957.
96. Thompson C, Kinmonth AL, Stevens L, et al. Effects of a clinical-practice guideline and practice-based education on detection and outcome of depression in primary care: Hampshire Depression Project randomised controlled trial. *Lancet.* 2000;355:185-191.
97. Roter DL, Hall JA, Kern DE, Barker LR, Cole KA, Roca RP. Improving physicians' interviewing skills and reducing patients' emotional distress: a randomized clinical trial. *Arch Intern Med.* 1995;155:1877-1884.
98. Mulrow CD. *Evidence-based Hypertension.* London, England: BMJ Books; 2001.

