Case-Finding Instruments for Depression

Two Questions Are as Good as Many

Mary A. Whooley, MD, Andrew L. Avins, MD, MPH, Jeanne Miranda, PhD, Warren S. Browner, MD, MPH

OBJECTIVE: To determine the validity of a two-question case-finding instrument for depression as compared with six previously validated instruments.

DESIGN: The test characteristics of a two-question case-finding instrument that asks about depressed mood and anhedonia were compared with six common case-finding instruments, using the Quick Diagnostic Interview Schedule as a criterion standard for the diagnosis of major depression.

SETTING: Urgent care clinic at the San Francisco Department of Veterans Affairs Medical Center.

PARTICIPANTS: Five hundred thirty-six consecutive adult patients without mania or schizophrenia.

MEASUREMENTS AND MAIN RESULTS: Measurements were two questions from the Primary Care Evaluation of Mental Disorders patient questionnaire, both the long and short forms of the Center for Epidemiologic Studies Depression Scale, both the long and short forms of the Beck Depression Inventory, the Symptom-Driven Diagnostic System for Primary Care, the Medical Outcomes Study depression measure, and the Quick Diagnostic Interview Schedule. The prevalence of depression, as determined by the standardized interview, was 18% (97 of 536). Overall, the case-finding instruments had sensitivities of 89% to 96% and specificities of 51% to 72% for diagnosing major depression. A positive response to the two-item instrument had a sensitivity of 96% (95% confidence interval [CI], 90–99%) and a specificity of 57% (95% CI 53–62%). Areas under the receiver operating characteristic curves were similar for all of the instruments, with a range of 0.82 to 0.89.

CONCLUSIONS: The two-question case-finding instrument is a useful measure for detecting depression in primary care. It has similar test characteristics to other case-finding instruments and is less time-consuming.

KEY WORDS: depression; case-finding instruments.
J GEN INTERN MED 1997;12:439–445.

Major depression is one of the most common illnesses seen by primary care physicians with a prevalence of 5% to 9% in adult patients.[1] Depressed medical patients have increased disability, health-care utilization, and mortality from suicide and other causes, as well as reduced productivity and health-related quality of life.[2–8] Costs for depression, including patient care, lost productivity, absenteeism, and suicide, were estimated at $43.7 billion in the United States in 1990.[9]

Although primary care physicians manage the majority of patients with major depression, 35% to 50% of cases go unrecognized.[10–12] Several questionnaires have been developed to help providers identify depression in the primary care setting,[13–21] but many practitioners find these measures too cumbersome and time-consuming for routine use.[22,23]

Are there one or two brief questions that could help primary care physicians identify patients with major depression? According to the *Diagnostic and Statistical Manual of Mental Disorders–IV*, the essential feature of a major depressive episode is a period of at least 2 weeks during which there is either depressed mood or the loss of interest or pleasure in nearly all activities.[24] The Primary Care Evaluation of Mental Disorders Procedure (PRIME-MD) includes a 27-item screening questionnaire and follow-up clinician interview designed to facilitate the diagnosis of common mental disorders in primary care. The questionnaire includes two questions about depressed mood and anhedonia: (1) "During the past month, have you often been bothered by feeling down, depressed, or hopeless?" and (2) "During the past month, have you often been bothered by little interest or pleasure in doing things?"[25]

The original PRIME-MD study reported that a "yes" answer to one of these two questions was 86% sensitive and 75% specific compared with a subsequent telephone interview diagnosis of major depressive disorder.[25] However, no study has compared the sensitivity and specificity of these two questions with a simultaneous interview. The goal of our study was to compare the test characteristics of a two-question case-finding instrument with those of six previously validated case-finding instruments. We simultaneously administered a diagnostic interview as the criterion standard for diagnosing major depression.

METHODS

Overview of Design

A consecutive sample of patients visiting the urgent care clinic at the San Francisco Department of Veterans Affairs Medical Center (SFVAMC) between April and November 1995 was asked to participate in the study. Most patients were seeking care for a specific complaint; others sought to establish primary care, to receive subspecialty referral, or to refill medication prescriptions. Of 675 eligible patients, 74 declined to participate, 6 were excluded because they were blind, and 5 were excluded because they were too delusional or intoxicated (as assessed by the interviewer) to understand the interview. Thus, 590 patients were enrolled in the study. The study protocol was approved by the Committee on Human Research at the University of California, San Francisco, and the SFVAMC.

After informed consent was obtained, each patient was asked to complete a self-report questionnaire. Three psychology graduate students were trained to administer a 20-minute diagnostic interview. They were blinded to the results of the case-finding instruments while conducting the interview. Each subject was offered $5 for participating in the study.

Information on age, gender, ethnicity, education, income, employment status, marital status, current substance abuse, previous diagnosis of depression, previous or current therapy for depression, and previous or current medical illness was obtained by interview or questionnaire. The self-report questionnaire consisted of several demographic items followed by the two-question instrument: (1) "During the past month, have you often been bothered by feeling down, depressed, or hopeless?" and (2) "During the past month, have you often been bothered by little interest or pleasure in doing things?"

After the two questions were completed, six common case-finding instruments were administered,[13–16,26,27] using one of two versions of a questionnaire in which the instruments were arranged in different orders. Following the case-finding instruments, subjects were asked the four CAGE questions for alcoholism[28] and "How many times have you used street drugs in the past year?" to detect illicit drug use. Substance abuse was defined as two or more positive answers to the CAGE questions or having used "street drugs" 12 or more times in the previous year. The questionnaires and interview took approximately 45 minutes to complete.

Case-Finding Instruments

The Center for Epidemiologic Studies Depression Scale (CES-D) is a 20-item self-report instrument (range 0–60) that covers the number and duration of depressive symptoms.[14] We used a standard cutpoint of 16 to diagnose depression. We also tested a 10-item short form of the CES-D (range 0–30) using a cutpoint of 10.[26] The Beck Depression Inventory (BDI) is a 21-item scale (range

0–60).[13] A standard cutpoint of 10 was used to diagnose depression. We also tested a 13-item short form of the BDI (range 0–39) using a cutpoint of 5.[27] Test characteristics for other cutpoints were evaluated by use of receiver operating characteristic curves.

The Medical Outcomes Study (MOS) depression measure was developed for use in the National Study of Medical Care Outcomes.[15] This 8-item instrument (range 0.001–0.9) incorporates two items from the Diagnostic Interview Schedule and six items from the CES-D. We used a standard cutpoint of 0.060 to diagnose depression. The Symptom-Driven Diagnostic System for Primary Care (SDDS-PC) is an instrument designed to assess multiple mental disorders in the primary care setting.[16,23] It includes a 5-item case-finding measure (range 0–4) for depression. We used a standard cutpoint of 2 to diagnose depression.

Criterion Standard

The National Institute of Mental Health Diagnostic Interview Schedule (DIS) is a highly structured interview designed to be administered by lay interviewers to yield psychiatric diagnoses according to the *Diagnostic and Statistical Manual of Mental Disorders* criteria. The DIS has a sensitivity of 80% and specificity of 84% compared with *DSM-III* criteria for depression and has been used extensively to study the epidemiology of depression.[3,29]

We used modules for major depressive episode, manic episode, and schizophrenia from the Quick DIS-III-R (QDIS), a computerized version of the DIS, as a criterion standard for diagnosing major depression in this study. The depression module was modified to detect major depression in the past year rather than during the subject's lifetime. The QDIS has demonstrated good test-retest reliability and agreement with the standard-format DIS for diagnosing depression ($\kappa = 0.76$), mania ($\kappa = 0.75$), and schizophrenia ($\kappa = 0.87$).[30] In our study, the average interrater reliability on a subset of patients ($n = 20$) interviewed by all three interviewers was excellent ($\kappa = 0.88$).

Recognition

The percentage of depression cases recognized by urgent care providers was determined by blinded chart review. Patients in the urgent care clinic were evaluated by attending physicians, resident physicians, or both. To qualify as "recognized," the provider was required to have noted the term *depression* or *depressed* on the visit record or to have referred the patient to a psychiatrist for further evaluation of depressive symptoms.

Statistical Analysis

Patients with concurrent mania or schizophrenia (found on QDIS modules) were excluded from analysis. The precision of the estimates of the prevalence of depres-

sion and the proportion of cases recognized by health care providers were determined with exact binomial 95% confidence intervals (CIs).[31] For the two-question instrument, a "yes" answer to either of the following two questions was considered a positive test: (1) "During the past month, have you often been bothered by feeling down, depressed, or hopeless?" or (2) "During the past month, have you often been bothered by little interest or pleasure in doing things?" For the other six instruments, standard cutpoints for diagnosis of depression were obtained from published literature.

The operating characteristics of each of the seven case-finding instruments were compared with the diagnosis of depression as determined by the QDIS (criterion standard). Sensitivity, specificity, and likelihood ratios (LRs) were determined by standard formulae[32]; exact binomial 95% CIs were calculated for the sensitivity and specificity. Each of the case-finding instruments was then converted to its continuous or ordinal scale and receiver operating characteristic (ROC) curves were generated; areas under these curves were calculated by the trapezoidal rule.[31] Ninety-five percent CIs for the area under the ROC curves were determined by bootstrapping methods. Areas under the ROC curves were compared using the method of Hanley and McNeil.[33] All analyses were performed using Stata statistical software, version 5.0 (College Station, Tex., 1996).[31]

RESULTS

A total of 590 subjects participated in the study. Of these, 47 were excluded because of concurrent mania, schizophrenia, or both, and 7 were excluded because of missing data, leaving 536 subjects for the analysis. Participants were mainly middle-aged men ranging from 21 to 89 years of age (Table 1). Only 14% had not finished high school.

The prevalence of major depression as determined by the QDIS interview was 18.1% (97 of 536; 95% CI 15–21%). Of those determined to be depressed in the past year, 78% (76 of 97) stated they had experienced a depressive episode in the past month. Overall, the case-finding instruments had sensitivities of 89 to 96% and specificities of 51 to 72% (Table 2). Likelihood ratios for positive tests ranged from 2.0 to 3.3; LRs for negative tests ranged from 0.07 to 0.17.

The two-question instrument was 96% sensitive (95% CI 90–99%) with a negative LR of 0.07 and a negative predictive value of 98% in this sample. Its specificity was 57% (95% CI 53–62%) with a positive LR of 2.2 and a positive predictive value of 33%. These values were stable across different age groups (Table 3). When subjects who answered yes to the question "Have you ever been diagnosed with or treated for depression?" (*n* = 101) were excluded from the analysis, the two-question instrument was 98% sensitive (95% CI 89–100%) and 59% specific (95% CI 54–64%).

Table 1. Characteristics of Study Population (*n* = 536)

Characteristic	Number (%)
Age, years	53 ± 14*
Male	522 (97)
Ethnicity	
African American	158 (29)
Asian or Pacific Islander	32 (6)
Latino	41 (8)
White	294 (55)
Other	11 (2)
High school education	463 (86)
Homeless	44 (8)
Annual income	
<$10,000	284 (53)
$10,000–30,000	214 (40)
≥$30,000	38 (7)
Employment	
Full-time	92 (17)
Part-time	63 (12)
Not working	381 (71)
Marital status	
Married	143 (27)
Widowed	33 (6)
Separated/divorced	215 (40)
Never married	145 (27)

*Mean ± SD.

Areas under the ROC curves were similar for all of the instruments, with a range of 0.82 to 0.89. The CES-D, CES-D-short, and MOS instruments had greater areas under the ROC curve than the two-question instrument, but areas under the ROC curves were not statistically different for the two-question instrument compared with the other three instruments. The two-question instrument took less than 1 minute to complete and score, while each of the other instruments took longer to complete (range 2–5 minutes) and was more cumbersome to score.

When the 175 participants with concurrent substance abuse were excluded from the analysis (121 for alcohol, 36 for illicit drug use, and 46 for both), the sensitivity of the two-question instrument was 96% (95% CI 86–99%) with a negative predictive value of 99% (Table 4). The specificity increased to 66% (95% CI 60–71%), but the positive predictive value slightly decreased to 30% because the prevalence of depression was lower (13.5%) in this group. Excluding subjects with substance abuse, areas under the ROC curves were similar for all of the instruments, with a range of 0.84 to 0.91. With the exception of the MOS instrument, which had a greater area under the ROC curve than the two-question instrument (*p* = .02), areas under the ROC curves were not statistically different for the two-question instrument compared with the other six instruments.

In post hoc analyses, the first question alone ("During the past month, have you often been bothered by feeling down, depressed, or hopeless?") was 93% sensitive (95% CI 86–97%) with a negative LR of 0.11 and a negative pre-

Table 2. Test Characteristics for Seven Depression Case-Finding Instruments in 536 Subjects

Instrument	Sensitivity, % (95%CI)	Specificity, % (95% CI)	Likelihood Ratio		Area Under ROC Curve (95% CI)
			Positive	Negative	
Two questions[25]	96 (90–99)	57 (53–62)	2.2	0.07	.82 (.78–.86)
SDDS-PC[16]	96 (90–99)	51 (47–56)	2.0	0.08	.86 (.82–.89)
MOS[15]	93 (86–97)	72 (68–76)	3.3	0.10	.89 (.85–.91)*
CES-D[14]	93 (85–97)	69 (65–74)	3.0	0.10	.89 (.85–.92)*
CES-D-short[26]	90 (82–95)	72 (67–76)	3.2	0.14	.87 (.83–.91)*
BDI[13]	89 (81–95)	64 (59–68)	2.5	0.17	.87 (.82–.91)
BDI-short[27]	92 (85–97)	61 (56–66)	2.4	0.13	.86 (.82–.90)

Area under ROC curve was greater for the CES-D (p = .006), CES-D-short (p = .034), and MOS (p = .004) compared with the two questions; areas under ROC curves for the other three instruments were not statistically different (p > .05) than for the two questions.

dictive value of 98%. It was 62% specific (95% CI 58–67%) with a positive LR of 2.4 and a positive predictive value of 35%. These values were stable across different age groups. When subjects who answered yes to the question "Have you ever been diagnosed with or treated for depression?" (n = 100) were excluded from the analysis, the first question was 94% sensitive (95% CI 83–99%) and 65% specific (95% CI 60–70%). The area under the ROC curve for the first question alone (0.78; 95% CI 0.74, 0.81) was less than for all of the other instruments, including the two-question instrument. The second question alone ("During the past month, have you often been bothered by little interest of pleasure in doing things?") was 79% sensitive (95% CI 70–87%) and 72% specific (95% CI 68–76%).

Excluding the 175 participants with substance abuse, the first question alone was 90% sensitive (95% CI 78–97%) with a negative LR of 0.14 and a negative predictive value of 98%, and 69% specific (95% CI 64–74%) with a positive LR of 2.9 and a positive predictive value of 31% in this sample. The second question alone was 71% sensitive (95% CI 57–83%) and 80% specific (95% CI 75–84%).

Complete charts and visit records were available for 429 of 536 subjects. Of these, only 8.8% of subjects with depression (6 of 68 subjects; 95% CI 3–18%) were recognized as being depressed by the health care provider. Excluding subjects with substance abuse, 6.7% of subjects with depression (2 of 30 subjects; 95% CI 1–22%) were recognized as being depressed by the health care provider.

DISCUSSION

A two-question case-finding instrument was an effective means of identifying subjects with major depression.

A "no" response to both of two questions made depression highly unlikely, with a LR of 0.07 and a posterior probability of 2%. The test characteristics of the two-question instrument were similar to those of six common case-finding instruments for major depression and were stable across different age groups.

This is the first study to compare seven case-finding instruments for depression with a simultaneous criterion standard interview. Our results are consistent with those of a meta-analysis that combined the results of 18 studies and found that nine case-finding instruments for depression had similar test characteristics.[34]

Sensitivity should be maximized in choosing a case-finding instrument, so that cases of depression are not missed. Thus, it is most important to compare the ROC curves when sensitivity is high (e.g., 80–100%). The slightly greater areas under the curves for the CES-D, CES-D-short, and MOS instruments are due to greater specificity when sensitivity is low (Figure 1). The 20-item CES-D and the 10-item CES-D-short are much longer than the two-question instrument. The MOS requires multiplication of scored answers to each of eight questions by their individual regression coefficients to obtain a sum value that can be compared with the standard cutpoint for depression, and thus is impractical for use in primary care settings.

The brevity of the two-question instrument makes it the most suitable instrument for routine use. Indeed, a post hoc analysis found that one question ("During the past month, have you often been bothered by feeling down, depressed, or hopeless?") was 93% sensitive and 62% specific. A "no" response to this one question also made depression highly unlikely with a LR of 0.11 and a posterior probability of 2%. Although one question might be re-

Table 3. Test Characteristic of Two-Question Case-Finding Instrument, by Age (n = 536)

Age, years	Subjects, n	Depressed, %	Sensitivity, % (95% CI)	Specificity, % (95% CI)
<35	51	14	100 (59–100)	59 (43–74)
35–64	358	23	95 (88–99)	52 (46–58)
≥65	127	5	100 (54–100)	69 (60–77)

Table 4. Test Characteristics for Seven Depression Case-Finding Instruments Excluding 175 Subjects with Substance Abuse

Instrument	Sensitivity, % (95% CI)	Specificity, % (95% CI)	Likelihood Ratio		Area Under ROC Curve (95% CI)
			Positive	Negative	
Two questions[25]	96 (86–99)	66 (60–71)	2.8	0.06	.84 (.80–.89)
SDDS-PC[16]	94 (83–99)	56 (51–62)	2.1	0.11	.88 (.83–.92)
MOS[15]	86 (73–94)	82 (77–86)	4.8	0.17	.91 (.87–.94)*
CES-D[14]	88 (75–95)	74 (69–79)	3.4	0.16	.89 (.85–.94)
CES-D-short[26]	84 (70–93)	75 (70–80)	3.4	0.21	.88 (.83–.92)
BDI[13]	84 (71–94)	71 (66–76)	2.9	0.23	.88 (.81–.93)
BDI-short[27]	87 (74–95)	67 (62–73)	2.6	0.19	.86 (.79–.92)

Area under ROC curve was greater for the MOS than the two questions (p = .02); areas under ROC curves for the other five instruments were not statistically different (p > .05) than for the two questions.

membered more easily by primary care providers, we believe the added sensitivity and greater area under the ROC curve make it worthwhile to ask two questions.

We found 18% of the subjects enrolled in this study had major depression by the criterion standard interview. This is consistent with the high prevalence of depression found in previous studies of VA outpatients.[35,36] Recognition of depression by health care providers in our study was lower than in previous reports.[11,12,37] This was not surprising as the role of providers in an urgent care setting differs from that in a primary care environment, and examination of the medical record is an imperfect way of assessing physician recognition.

Several limitations of this study deserve comment. First, test characteristics found in an urban VA population may not generalize to populations with a lower prevalence of depression or to other practice settings, especially those with a greater proportion of women patients. Second, the case-finding instruments were administered on a self-report questionnaire; test characteristics might differ if they were asked verbally by a health-care provider. Third, although 78% of the depressed subjects in this study

stated they had experienced a depressive episode in the past month, the criterion standard with which the case-finding instruments were compared tested for depression in the past year. Finally, the high sensitivity of the two-question instrument may reflect the fact that both the two-question instrument and the QDIS criterion standard require endorsing depression or anhedonia as primary criteria for diagnosis. As these are necessary criteria for *DSM-IV* diagnosis of depression, however, it is surprising that we have not adopted such a simple case-finding instrument sooner.

Adopting the two-question instrument for routine case-finding in populations with a prevalence of depression similar to that in our study (18%) would result in 17 true positives and only 1 false negative for every 100 patients. However, it also would result in 35 false positives. This low positive predictive value means that approximately two of three patients who test positive for depression would have to undergo a diagnostic interview that ultimately would determine they were not depressed. In populations with a lower prevalence of depression, the false-positive rate would be even higher. Routine use of any of the case-finding instruments would result in similar numbers of false positives because specificity was low. Administering a diagnostic interview to all patients who test positive for depression might not be viewed favorably by primary care physicians who already have many competing demands on their time.

Should the two-question case-finding instrument be adopted for routine use in primary care settings, or should it be used only in selected (e.g., high-risk) patients? Case-finding instruments enhance detection of major depression, and treatment of major depression improves outcomes.[38,39] However, studies conflict as to whether early detection and treatment of depression leads to improved outcomes compared with usual care given at the time symptoms are first recognized.[11,40–45] Neither the U.S. Preventive Services Task Force nor the Canadian Task Force on the Periodic Health Examination has found sufficient evidence to recommend for or against the routine use of case-finding questionnaires for depression in primary care patients.[46,47] Ongoing studies of improving
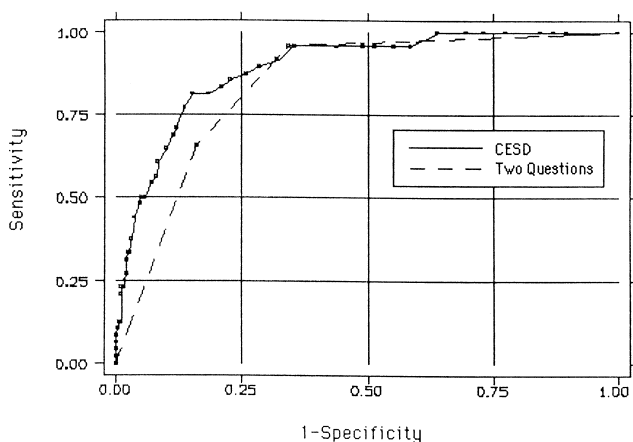

FIGURE 1. Receiver operating characteristic curves for two-question case-finding instrument compared with Center for Epidemiologic Studies Depression Scale (CES-D).

quality of care for depression in primary care patients may clarify when these instruments are most beneficial.

We believe that two questions, "During the past month, have you often been bothered by feeling down, depressed, or hopeless?" and "During the past month, have you often been bothered by little interest or pleasure in doing things?" should be used by primary care providers to improve diagnosis of major depression in patients who are at high risk or who present with symptoms suggestive of depression. A negative response to both questions makes depression highly unlikely. For patients who answer yes to either of the two questions, other symptoms such as fatigue, restlessness, guilt, poor concentration, suicidal ideation, and change in sleep or appetite should be elicited to confirm the diagnosis of depression. Mania and psychosis must be ruled out before initiating therapy. Whether ascertaining depressed mood and anhedonia in unselected patients will improve patient outcomes remains to be determined.

REFERENCES

1. Depression Guideline Panel. Depression in Primary Care, Vol 1: Detection and Diagnosis. Clinical Practice Guideline. Washington, DC: U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research; 1993. AHCPR 93–0550.
2. Johnson J, Weissman MM, Klerman GL. Service utilization and social morbidity associated with depressive symptoms in the community. JAMA. 1992;267:1478–83.
3. Wells KB, Stewart A, Hays RD, et al. The functioning and well-being of depressed patients. Results from the Medical Outcomes Study. JAMA. 1989;262:914–9.
4. Broadhead WE, Blazer DG, George LK, Tse CK. Depression, disability days, and days lost from work in a prospective epidemiologic survey. JAMA. 1990;264:2524–8.
5. Simon GE. Psychiatric disorder and functional somatic symptoms as predictors of health care use. Psychiatr Med. 1992;10:49–59.
6. Katon W, Berg AO, Robins AJ, Risse S. Depression—medical utilization and somatization. West J Med. 1986;144:564–8.
7. Spitzer RL, Kroenke K, Linzer M, et al. Health-related quality of life in primary care patients with mental disorders. [Results from the PRIME-MD 1000 study. JAMA. 1995;274:1511–7.] JAMA 1993; 270:1819–25.
8. Frasure-Smith N, Lesperance F, Talajic M. Depression following myocardial infarction. Impact on 6-month survival. Published erratum appears in JAMA. 1994;271:1082.
9. Greenberg PE, Stiglin LE, Finkelstein SN, Berndt ER. The economic burden of depression in 1990. J Clin Psychiatry. 1993;54:405–18.
10. Shurman RA, Kramer PD, Mitchell JB. The hidden mental health network. Treatment of mental illness by nonpsychiatrist physicians. Arch Gen Psychiatry. 1985;42:89–94.
11. Simon GE, Von Korff M. Recognition, management, and outcomes of depression in primary care. Arch Fam Med. 1995;4:99–105.
12. Von Korff M, Shapiro S, Burke JD, 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.
13. Beck A, Ward C, Mendelson M. An inventory for measuring depression. Arch Gen Psychiatry. 1961;4:53–63.
14. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Measurement. 1977;1:385–401.
15. Burnam MA, Wells KB, Leake B, Landsverk J. Development of a brief screening instrument for detecting depressive disorders. Med Care. 1988;26:775–89.
16. 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;4:211–9.
17. Yesavage JA, Brink TL, Rose TL, et al. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatric Res. 1982;17:37–49.
18. Zung W. A self-rating depression scale. Arch Gen Psychiatry. 1965;12:63–70.
19. Rush AJ, Giles DE, Schlesser MA, Fulton CL, Weissenburger J, Burns C. The Inventory for Depressive Symptomatology (IDS): preliminary findings. Psychiatry Res. 1986;18:65–87.
20. Cleary PD, Goldberg ID, Kessler LG, Nycz GR. Screening for mental disorder among primary care patients. Usefulness of the General Health Questionnaire. Arch Gen Psychiatry. 1982;39:837–40.
21. Rost K, Burnam MA, Smith GR. Development of screeners for depressive disorders and substance disorder history. Med Care. 1993;31:189–200.
22. Andersen SM, Harthorn BH. The recognition, diagnosis, and treatment of mental disorders by primary care physicians. Med Care. 1989;27:869–86.
23. Weissman MM, Olfson M, Leon AC, et al. Brief diagnostic interviews (SDDS-PC) for multiple mental disorders in primary care. A pilot study. Arch Fam Med. 1995;4:220–7.
24. Diagnostic and Statistical Manual of Mental Disorders. 4th ed. Washington, DC: American Psychiatric Association; 1994.
25. 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;272:1749–56.
26. Andersen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). Am J Prev Med. 1994;10:77–84.
27. Beck AT, Beck RW. Screening depressed patients in family practice. A rapid technic. Postgrad Med. 1972;52:81–5.
28. Ewing JA. Detecting alcoholism. The CAGE questionnaire. JAMA. 1984;252:1905–7.
29. Robins LN, Helzer JE, Croughan J, Ratcliff KS. National Institute of Mental Health Diagnostic Interview Schedule. Its history, characteristics, and validity. Arch Gen Psychiatry. 1981;38:381–9.
30. Bucholz KK, Marion SL, Shayka JJ, Marcus SC, Robins LN. A short computer interview for obtaining psychiatric diagnoses. Psychiatr Serv. 1996;47:293–7.
31. Stata Statistical Software: release 5.0. College Station, Tex.: Stata Corp; 1996.
32. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical Epidemiology: A Basic Science For Clinical Medicine. Boston, Mass.: Little, Brown and Co.; 1991.
33. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology. 1983;148:839–43.
34. Mulrow CD, Williams J Jr, Gerety MB, Ramirez G, Montiel OM, Kerber C. Case-finding instruments for depression in primary care settings. Ann Intern Med. 1995;122:913–21.
35. Walker RD, Howard MO, Lambert MD, Suchinsky R. Psychiatric and medical comorbidities of veterans with substance use disorders. Hosp Commun Psychiatry. 1994;45:232–7.
36. Kukull WA, Koepsell TD, Inui TS, et al. Depression and physical illness among elderly general medical clinic patients. J Affect Disord. 1986;10:153–62.

37. Wells KB, Hays RD, Burnam MA, Rogers W, Greenfield S, Ware J Jr. Detection of depressive disorder for patients receiving prepaid or fee-for-service care. Results from the Medical Outcomes Study. JAMA. 1989;262:3298–302.

38. Elkin I, Gibbons RD, Shea MT, et al. Initial severity and differential treatment outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. J Consult Clin Psychol. 1995;63:841–7.

39. 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;28:239–50.

40. Tiemens BG, Ormel J, Simon GE. Occurrence, recognition, and outcome of psychological disorders in primary care. Am J Psychiatry. 1996;153:636–44.

41. 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;42:839–46.

42. Schulberg HC, Block MR, Madonia MJ, et al. Treating major depression in primary care practice. Eight-month clinical outcomes. Arch Gen Psychiatry. 1996;53:913–9.

43. Katon W, Von Korff M, Lin E, et al. Collaborative management to achieve treatment guidelines. Impact on depression in primary care. JAMA. 1995;273:1026–31.

44. Coyne JC, Schwenk TL, Fechner-Bates S. Nondetection of depression by primary care physicians reconsidered. Gen Hosp Psychiatry. 1995;17:3–12.

45. 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.

46. US Preventive Services Task Force, Guide to Clinical Preventive Services. 2nd ed. Baltimore, Md: Williams & Wilkins; 1996.

47. Canadian Task Force on the Periodic Health Examination. Canadian Guide to Clinical Preventive Health Care. Ottawa, Ont.: Canada Communication Group; 1994.