Depression screening: a practical strategy

Donald E. Nease, Jr, MD and Jean M. Malouin, MD, MPH
Department of Family Medicine, University of Michigan

Practice recommendations

- A 2-stage strategy, combining an assessment of severity with depression criteria, can help a physician focus on the most severe cases without missing less severe ones that still need treatment (B).

- Because of its brevity, relatively high positive predictive value, and ability to inform the clinician on both depression severity and diagnostic criteria, the PRIME-MD Patient Health Questionnaire (PHQ-9) is the best available depression screening tool for primary care (B).

- One-time screening is cost-effective; physicians may elect to screen more often based on risk factors (A).

Corresponding author: Donald E. Nease, Jr, MD, Department of Family Medicine, University of Michigan, 1018 Fuller St., Ann Arbor, MI 48109-0708. E-mail: dnease@umich.edu.

What is the most efficient and accurate way for a busy primary care physician to screen patients for depression? Many screening tools exist, but they are not equally effective.

A careful review of the literature strongly favors a 2-stage strategy assessing both depression severity and criteria. In this article, we describe this optimal approach against the background of other available resources.

■ HEALTH AND ECONOMIC IMPACT OF DEPRESSION

In the average family practice, around 6 cases of depression go unrecognized each week. This real-world estimate derives from studies that consistently report a 10% prevalence of depression in primary care patients[1] but a rate of recognition by primary care clinicians of only 29% to 35%.[2–4] Depression is a common condition with a large impact on quality of life and productivity, one that indirectly affects other health states, including cardiovascular disease.[5–9] It is responsible for an estimated economic cost in the US of over $40 billion annually. As a result, depression screening has been an active area of research, and a variety of organizations have issued guidelines recommending routine screening for depression in primary care.

TABLE 1

Burdens of screening for cancer, hyperlipidemia, and depression

	Cancer		Hyperlipidemia		Depression	
Burden of performance	Low	Simple test or performance of billable procedure	Low	Blood test	High	Time-intensive administration & scoring
Burden of interpretation	Low	Confirmatory testing often referred to specialists	Low	No confirmatory reference standard testing	High	High false positive rate w/burdensome reference standard
Burden of treatment	Low	Treatment done by specialists	High	Requires activation of patient & frequent monitoring	High	Requires activation of patient & frequent monitoring

■ THE NEED FOR AN EFFICIENT, RELIABLE SCREENING TOOL

Based on a recent review of the evidence on depression screening outcomes in primary care settings,[10] the US Preventive Services Task Force (USPSTF) updated its screening recommendation in 2002 to include an endorsement of depression screening in adults "in clinical practices that have systems in place to assure accurate diagnosis, effective treatment, and follow-up" (strength of recommendation [SOR]=**A**).[11] This endorsement leaves the primary care clinician with no guidance about how or when to screen for depression.

Despite lack of guidance in the USPTF guidelines, we believe depression screening can be done efficiently and reliably in primary care. However, one must begin by understanding that depression screening is different from screening for cancer or cardiovascular risk factors (**Table 1**). The burdens of interpretation of depression screening results are especially noteworthy. For example, the PRIME-MD Patient Health Questionnaire (PHQ) is reported to have a sensitivity of 61% and specificity of 94% for any mood or depressive disorder.[12] This results in a positive predictive value (PPV) of 50% using a reasonable estimate of 10% prevalence for depression in primary care settings.[13]

Put simply, following administration and scoring of the PHQ, the clinician is left with little better odds than a coin toss of identifying a patient that has an active major depressive disorder requiring treatment. If there was no objective help, clinicians would have only their clinical judgment to resolve this, all during an office visit that contains many other competing agendas and demands.[14,15]

We have reviewed the evidence on depression screening instruments with the intent to highlight an instrument that clinicians can efficiently and reliably use to find depressed and impaired patients in their practice whom they might otherwise miss.

■ TWO TYPES OF SCREENING INSTRUMENTS

Depression screening instruments can be grouped into 2 categories:

- **depression assessment scales,** which ask patients to rate the severity or frequency of various symptoms
- **symptom count instruments,** which are based on depression criteria.

Depression assessment scales preceded symptom count instruments, and many were developed prior to the establishment of formal diagnostic criteria within the *Diagnostic and Statistical Manual of*

TABLE 2

Accuracy and ease of administration of commonly available screening instruments

Instrument	Time and scoring	LR+ (95% CI)	LR– (95% CI)	PPV (95% CI)	Web source
Assessment scale					
Beck Depression Inventory (BDI)[32]	2–5 min; simple	4.2 (1.2–13.6)	0.17 (0.1–0.3)	29.6% (10.7–57.6)	www.psychcorpcenter.com /content/bdi-II.htm
Center for Epidemiologic Studies Depression Scale (CES-D)[34]	2–5 min; simple	3.3 (2.5–4.4)	0.24 (0.2–0.3)	24.8% (20–30.6)	http://www.mhhe.com/ hper/health/personal health/labs/Stress/ activ2-2.html
Geriatric Depression Scale (GDS)[35]	2–5 min; simple	3.3 (2.4–4.7)	0.16 (0.1–0.3)	24.8% (19.4–32)	http://www.stanford.edu /~yesavage/ GDS.html
Hospital Anxiety and Depression Scale* (HADS)[20]	2–5 min; simple	7.0 (2.9–11.2)	0.3 (0.3–0.4)	41.3% (22.6–52.8)	www.clinical- supervision.com/ hads.htm
Zung Self Assessment Depression Scale (Zung SDS)[33]	2–5 min; simple	3.3 (1.3–8.1)	0.35 (0.2–0.8)	24.8% (11.5–44.8)	http://fpinfo. medicine.uiowa. edu/calculat.htm
Symptom count					
Primary Care Evaluation of Mental Disorders[†] (PRIME-MD)[27]	2 min; complex	2.7 (2.0–3.7)	0.14 (0.1–0.3)	21.3% (16.7–27)	Available upon request to Robert Spitzer, MD: RLS8@columbia.edu
PRIME-MD Patient Health Questionnaire (PHQ)	5–7 min; simple	10.2[‡] (6.5–17.5)	0.4[‡] (0.3–0.5)	50.4% (39.4–63.6)	fpinfo.medicine. uiowa.edu/calculat.htm
Symptom-Driven Diagnostic System for Primary Care[†] (SDDS-PC)	2 min; complex	3.5 (2.4–5.1)	0.2 (0.1–0.4)	25.9% (19.4–33.8)	No website available
PRIME-MD Patient Health Questionnaire (PHQ-9)	2–5 min; simple	12.2 (8.4–18)	0.28 (0.2–0.5)	55% (45.7–64.3)	www.depression- primarycare.org/ ap1.html

* Unless noted by (*), adapted from Williams et al.[18]
† Values reflect the initial brief screening portion of these instruments.
‡ PHQ vaues obtained from original position and reflect diagnosis of "any mood disorders."
LR+, positive likelihood ratio; LR–, negative likelihood ratio; PPV, positive predictive value; CI, confidence interval

Mental Disorders (DSM) system.[16] **Table 2** lists available examples of depression assessment scales and symptom count instruments, along with websites where you may access further information and the instruments themselves.

Pros and cons of assessment scales

The advantages of using a scale are due to the manner in which patients experience depressive symptoms, along a continuum of mild to severe. A scale is able to represent these gradations in severity and may be helpful in guiding the need for treatment and treatment adjustments.

Unfortunately, this ability to measure the dimensional nature of depression is also a weakness, as a threshold must be identified above which the patient is classified as warranting further investigation. Ideally, these thresholds should be established in a representative primary care sample and predict functional status as well as likelihood of meeting *DSM-IV* diagnostic criteria. The ability of a scale to accurately identify patients in need of attention depends directly on the threshold.

Pros and cons of symptom counts

Instruments based on depression criteria are a relatively new innovation, appearing since the establishment of *DSM-IV* criteria that define reference symptoms, a minimum number of which must be present to diagnose depression. Depression criteria–based instruments have the advantage of not being dependent on a threshold of symptom severity.

However, in primary care settings this can also be a weakness because the presence of depression criteria alone may not be a reliable indicator of depression-related impairment.[17] Instruments that can be used in both a diagnostic criteria and scale modes have a particular advantage in that the weaknesses of each are offset.

■ CHARACTERISTICS OF SELECTED SCREENING INSTRUMENTS

We searched MEDLINE and the Cochrane databases for reviews of depression screening, with partic-

ular attention to reviews of primary care-based trials. Forty-one papers emerged, 3 of which were systematic reviews. For this paper, we focused on the review published by Williams and colleagues,[18] which summarizes primary care data on the depression screening instruments most widely used. They examined 379 studies that compared the primary care performance of these instruments with a reference standard diagnostic interview, such as the Structured Clinical Interview for *DSM-IV* (SCID).[19] Twenty-eight studies met their criteria and were included in the systematic review.

In **Table 2** we have adapted the information from Williams's review and added a calculation of PPV based on a 10% prevalence estimate for depression in primary care populations. We chose to exclude information on the Single Question (SQ) screen because of its very low PPV and the Hopkins Symptom Checklist (HSCL) because of its length (25 questions). In addition, we chose to add the Hospital Anxiety and Depression Scale (HADS), using operating characteristic information from 2 studies,[20,21] because of its purported advantages in medically ill populations.

Beyond the SQ, it is useful to comment on "2-question screening" as suggested by the USPSTF. We are unable to find justification for this in the paper by Pingone and colleagues, which served as background for the recommendations.[10] Although Pingone et al did cite the report of Wells and colleagues as using a 2-item screener, their study used not only 2 questions on mood and anhedonia but also other criteria in screening their population.[22] Therefore, it is not appropriate as a source for 2-item screening performance characteristics.

Comparison of the operating characteristics of the selected instruments reveals that most yield PPV values in the 20% to 30% range, with the exception of the HADS, the PHQ, and the PHQ-9, which yield PPV values of 41.3%, 50%, and 55%, respectively.

The PHQ-9 (included in the **Appendix**) offers a further advantage over the HADS and other instruments listed in that within a 9-item instrument both the presence of diagnostic criteria and severity may be assessed. Kroenke and colleagues have exam-

ined the use of the PHQ-9 as a severity instrument and found it to be a reliable and valid measure of depression severity when compared with the Medical Outcomes Study Short Form (SF-20).[23]

We purposely have not examined negative predictive values (NPV) for the listed instruments. NPV is useful when screening using biomedical markers where a negative result allows extrapolation into the future due to a known, predictable time course for development of the screened-for condition. For example, a negative screening colonoscopy has value not just because of its current predictive value, but because we know something about how long it may take to develop precancerous polyps in a negative screened patient. However, this is not the case with depression. A patient that fails to meet criteria for depression today could fully meet criteria in 2 weeks and be quite depressed. Therefore we have chosen to focus on PPV in comparing depression screening instruments.

■ SELECTION AND USE OF A SCREENING INSTRUMENT

How should a busy clinician select a depression screening instrument? Ease of administration and interpretation are key. Ideally, a depression screen should function similarly to a vital sign, providing an easy-to-assess yet reliable marker of the need to address a patient's depression. It is not enough to know that formal depression criteria are met; it is also important to know whether a patient's functioning is impaired. Research indicates that it is difficult in primary care to "clinically" assess functioning in the face of numerous competing demands,[15] even when clinicians know from a screening test that a patient meets criteria for depression.[24] For this reason, even watchful waiting for the "positive screening/low impairment" patients[25] may be difficult to put into practice.

Two-stage strategy to assess impairment

Use of a 2-stage strategy, combining an assessment of severity with an assessment of depression criteria, appears to answer this dilemma. One study[26] has attempted to assess whether this strategy could identify the appropriate patients for clinician attention, using an existing data set that included the PRIME-MD[27] and 6 items identified from the original data via factor analyses that assess depression severity.

The results suggest that a combined assessment of depression severity and criteria could help clinicians focus on the most severely depressed patients without missing less severely impaired patients that need treatment (SOR=**B**).

We suggest the PHQ-9 as the instrument of choice for primary care depression screening because it measures both depression criteria and severity. The PHQ-9 provides a simple way to assess both diagnostic criteria and severity with a single, well-validated instrument. While its PPV is not appreciably greater than 50%, this reflects use in a purely "diagnostic mode," ie, a cut-point of 10.

A well done, primary care evaluation of the PHQ-9 suggests that a score of 15 or greater reliably indicates both satisfaction of *DSM-IV* depression criteria and a moderate to severe level of impairment (SOR=**A**).[28] Patients screening positive at this level should be targeted by their physician for a discussion of their symptoms and a recommendation for treatment (SOR=**B**). Patients with a score of 10–14 meet diagnostic criteria for depression but at a lower level of severity; these patients could be candidates for a strategy of repeat testing or watchful waiting (SOR=**B**).

Before leaving the topic, a comment is warranted regarding 2-stage screening using an initial 1- or 2-question screen followed by a more lengthy instrument. This type of strategy was embodied in the original PRIME-MD with its 2-question Patient Questionnaire (PQ).[27] The intent is to reduce the burden of applying a full diagnostic instrument to an entire practice population. By giving the full instrument only to patients that are positive on the initial 2-question screen, the screening performance burden (as identified in **Table 1**) is reduced. Use of a brief instrument

such as the PHQ-9, which requires only 2 to 5 minutes to fully complete, makes it possible to accurately assess both diagnostic criteria and depression severity in an entire patient population, with little administration burden.

When to screen

Once a decision is made to screen, and an instrument is selected, an interval for screening must be determined. Suggested ranges vary greatly from one-time to annual screening. The recent USPSTF recommendations provide little guidance, stating simply, "the optimal interval for screening is unknown."[11]

Regular intervals. One-time screening was found to be cost-effective by Valenstein and colleagues,[13] suggesting that, at a minimum, screening should occur when a new patient enters a practice (SOR=**A**). If a more frequent schedule of screening is desired, depression screening should be linked to other periodic preventive services provided in a practice, such as routine Pap smears or health maintenance exams, to ensure that screening occurs in a systematic fashion (SOR=**C**).

Risk factors. A practice may also elect to screen based on risk factors (SOR=**D**). Important risk factors to consider include prior history of treated depression, family history of depression, postpartum status, and any history of substance abuse.

Patients with chronic diseases known to have a high rate of comorbidity with depression—ie, diabetes, congestive heart failure, myocardial infarction—should also be considered as having risk factors for depression.

■ EASE OF IMPLEMENTATION

The depression screening instruments reviewed in this paper may all be completed by a patient with a sixth- to ninth-grade reading level, and can therefore be given to patients to complete in an exam room while they wait for their physician. Scoring may be then quickly completed either by the patient or by the physician.

Positive screens should prompt the physician to engage the patient in a discussion of their symptoms, the need for treatment, and a quick assessment for the presence of any suicidal ideation.

Finally, when depression is identified by screening, the potential presence of other psychiatric disorders should be noted. Anxiety disorders are frequently diagnosable in depressed patients, although it is unclear whether comorbid anxiety necessitates a change in treatment plans.[29] In contrast, a comorbid substance abuse should be recognized and addressed. Similarly, coexisting dysthymia may contribute to depressed patients' functional impairment.[30]

■ PHQ-9 REASONABLE FOR MONITORING TREATMENT

It is important to note that the USPSTF recommendation specifies screening "in clinical practices that have systems in place to assure accurate diagnosis, effective treatment, and follow-up." Routine, periodic monitoring is an important aspect of a systems approach to depression care. The PHQ-9, when scored as an assessment scale, and the depression assessment scales listed in **Table 2** should be considered for periodic monitoring of patients being treated for depression (SOR=**B**). Active monitoring may alert the clinician to improvement in symptoms or to a need for treatment adjustment when symptoms do not improve.

The Hamilton Rating Scale for Depression (HAM-D) is often used as a reference standard for monitoring of outcomes in clinical trials, but it is administered by trained interviewers and is therefore impractical to administer in a routine patient care setting. The Beck Depression Inventory (BDI) and Zung Self-rating Depression Scale (SDS) have been used as outcome measures as well, but they are not as sensitive to change over time as the HAM-D.[31]

The sensitivity to change over time of the PHQ-9 has not yet been formally compared to the HAM-D, but it still represents a reasonable option until the results of such a comparison are available.

REFERENCES

1. Katon W, Schulberg H. Epidemiology of depression in primary care. *Gen Hosp Psychiatry* 1992; 14:237–47.
2. 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.
3. Coyne JC, Schwenk TL, Fechner-Bates S. Nondetection of depression by primary care physicians reconsidered. *Gen Hosp Psychiatry* 1995; 17:3–12.
4. Williams JW, Mulrow CD, Kroenke K, et al. Case-finding for depression in primary care: a randomized trial. *Am J Med* 1999; 106:36–43.
5. Greenberg PE, Stiglin LE, Finkelstein SN, Berndt ER. The economic burden of depression in 1990. *J Clin Psychiatry* 1993; 54:405–18.
6. Katon W, Von Korff M, Lin E, et al. Distressed high utilizers of medical care. DSM-III-R diagnoses and treatment needs. *Gen Hosp Psychiatry* 1990; 12:355–62.
7. Von Korff M, Ormel J, Katon W, Lin EH. Disability and depression among high utilizers of health care. A longitudinal analysis. *Arch Gen Psychiatry* 1992; 49:91–100.
8. 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.
9. Ford DE, Mead LA, Chang PP, Cooper-Patrick L, Wang NY, Klag MJ. Depression is a risk factor for coronary artery disease in men: the precursors study. *Arch Intern Med* 1998; 158:1422–6.
10. Pignone MP, Gaynes BN, Rushton JL, et al. Screening for depression in adults: a summary of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Med* 2002; 136:765–76.
11. U.S. Preventive Services Task Force. Screening for depression: recommendations and rationale. *Ann Intern Med* 2002; 136:760–4.
12. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA* 1999; 282:1737–44.
13. Valenstein M, Vijan S, Zeber JE, Boehm K, Buttar A. The cost-utility of screening for depression in primary care. *Ann Intern Med* 2001; 134:345–60.
14. Jaen CR, Stange KC, Nutting PA. Competing demands of primary care: a model for the delivery of clinical preventive services. *J Fam Pract* 1994; 38:166–71.
15. Klinkman MS. Competing demands in psychosocial care. A model for the identification and treatment of depressive disorders in primary care. *Gen Hosp Psychiatry* 1997; 19:98–111.
16. American Psychiatric Association, American Psychiatric Association, Task Force on DSM-IV. *Diagnostic and statistical manual of mental disorders: DSM-IV-TR.* 4th ed. Washington, DC: American Psychiatric Association; 2000.
17. Schwenk TL, Coyne JC, Fechner-Bates S. Differences between detected and undetected patients in primary care and depressed psychiatric patients. *Gen Hosp Psychiatry* 1996; 18:407–15.
18. Williams JW Jr., Noel PH, Cordes JA, Ramirez G, Pignone M. Is this patient clinically depressed? *JAMA* 2002; 287:1160–70.
19. Spitzer RL, Williams JB, Gibbon M, First MB. The Structured Clinical Interview for DSM-III-R (SCID). I: History, rationale, and description. *Arch Gen Psychiatry* 1992; 49:624–9.
20. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983; 67:361–70.
21. Silverstone PH. Poor efficacy of the Hospital Anxiety and Depression Scale in the diagnosis of major depressive disorder in both medical and psychiatric patients. *J Psychosom Res* 1994; 38:441–50.
22. 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; 283:212–20.
23. Stewart AL, Hays RD, Ware JE Jr. The MOS short-form general health survey. Reliability and validity in a patient population. *Med Care* 1988; 26:724–35.
24. Rost K, Nutting P, Smith J, Coyne JC, Cooper-Patrick L, Rubenstein L. The role of competing demands in the treatment provided primary care patients with major depression. *Arch Fam Med* 2000; 9:150–4.
25. Leon AC, Portera L, Olfson M, et al. False positive results: a challenge for psychiatric screening in primary care. *Am J Psychiatry* 1997; 154:1462–4.
26. Nease DE, Jr., Klinkman MA, Volk RJ. Improved detection of depression in primary care through severity detection. *J Fam Pract* 2002; 51:1065–70.
27. Spitzer RL, Williams J, Kroenke K, Linzer M, deGruy FV, Hann SR, et al. Utility of a new procedure for diagnosing mental disorders in primary care: the PRIME-MD 1000 study. *JAMA* 1994; 272:1749–56.
28. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001; 16:606–13.
29. Coyne JC, Fechner-Bates S, Schwenk TL. Prevalence, nature, and comorbidity of depressive disorders in primary care. *Gen Hosp Psychiatry* 1994; 16:267–76.
30. Wells K, Burnam M, Rogers W, Hays R, Camp P. The course of depression in adult outpatients: results from the Medical Outcomes Study. *Arch Gen Psychiatry* 1992; 49:788–94.
31. Lambert MJ, Hatch DR, Kingston MD, Edwards BC. Zung, Beck, and Hamilton Rating Scales as measures of treatment outcome: a meta-analytic comparison. *J Consult Clin Psychol* 1986; 54:54–9.
32. Beck A, Ward C, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry* 1961; 4:561–71.
33. Zung WW, Richards CB, Short MJ. Self-rating depression scale in an outpatient clinic. Further validation of the SDS. *Arch Gen Psychiatry* 1965; 13:508–15.
34. Radloff LS. The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement* 1977; 1:385–401.
35. Sheikh JI, Yesavage JA. Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. In: *Clinical Gerontology: A Guide to Assessment and Intervention.* New York: Haworth Press; 1986:165–73.

PRIME-MD Patient Health Questionnaire (PHQ-9)

Patient Name:_____

Date:_____

1. Over the <u>last 2 weeks</u>, how often have you been bothered by any of the following problems?

	Not at all	Several days	More than half the days	Nearly every day
	0	**1**	**2**	**3**
a. Little interest or pleasure in doing things	☐	☐	☐	☐
b. Feeling down, depressed, or hopeless	☐	☐	☐	☐
c. Trouble falling/staying asleep, sleeping too much	☐	☐	☐	☐
d. Feeling tired or having little energy	☐	☐	☐	☐
e. Poor appetite or overeating	☐	☐	☐	☐
f. Feeling bad about yourself— or that you are a failure or have let yourself or your family down	☐	☐	☐	☐
g. Trouble concentrating on things, such as reading the newspaper or watching television	☐	☐	☐	☐
h. Moving or speaking so slowly that other people have noticed. Or the opposite—being so fidgety or restless that you have been moving around a lot more than usual	☐	☐	☐	☐
i. Thoughts that you would be better off dead or of hurting yourself in some way	☐	☐	☐	☐

2. If you checked off <u>any</u> problem on this questionnaire so far, how <u>difficult</u> have these problems made it for you to do your work, take care of things at home, or get along with other people?

Not difficult at all	**Somewhat difficult**	**Very difficult**	**Extremely difficult**
☐	☐	☐	☐

This questionnaire may be photocopied for use in the physician office. Copyright Pfizer.

Instructions—How to score PHQ-9
Major Depressive Syndrome is suggested if:
• Of the 9 items, 5 or more are checked as at least "More than half the days"
• Either item #1 or #2 is positive, that is, at least "More than half the days"

Other Depressive Syndrome is suggested if:
• Of the 9 items, 2, 3, or 4 are checked as at least "More than half the days"
• Either item #1 or #2 is positive, that is, at least "More than half the days"

Guide for Interpreting PHQ-9 Scores

Score	Action
≤4	The score suggests the patient may not need depression treatment.
≥5-14	Physician uses clinical judgment about treatment, based on patient's duration of symptoms and functional impairment.
≥15	Warrants treatment for depression, using antidepressant, psychotherapy, or a combination of treatment.