Improved detection of depression in primary care through severity evaluation

DONALD E. NEASE, JR, MD; MICHAEL S. KLINKMAN, MD, MS; AND ROBERT J. VOLK, PhD
Ann Arbor, Michigan, and Houston, Texas

KEY POINTS FOR CLINICIANS

- Existing instruments designed to improve primary care detection of depression carry significant associated burdens that may make their use difficult to sustain in routine practice.

- A brief instrument designed to assess symptom severity can effectively target severely symptomatic patients for evaluation with *Diagnostic and Statistical Manual of Mental Disorders* (DSM) criteria for depression.

- A strategy of initial assessment of symptom severity, followed by assessment for DSM depression criteria in the most symptomatic patients, can decrease the burden on primary care clinicians by accurately identifying depressed patients most in need of treatment.

■ OBJECTIVES To determine whether the use of a symptom severity measure to augment an existing *Diagnostic and Statistical Manual of Mental Disorders–Third Edition, Revised* (DSM-III-R) criteria–based depression screener (PRIME-MD) would decrease the difficulties associated with depression screening in primary care by filtering out patients with minimal impairment.

■ STUDY DESIGN The study design was secondary data analysis.

■ POPULATION The study sample comprised 1317 patients, with intentional oversampling by ethnicity and sex, presenting for routine care at a university family practice center in Galveston, Texas.

■ OUTCOMES MEASURED The primary outcomes were cross-sectional, health-related quality-of-life outcomes of subjects who met symptom severity criteria as well as criteria for a DSM-III-R mood disorder. Health care utilization outcomes were examined as secondary outcomes.

■ RESULTS The combination of a 6-item depression severity instrument and the PRIME-MD resulted in 71% of depressed subjects being categorized as severely symptomatic and 29% as minimally symptomatic. Severely symptomatic subjects had significantly worse SF-36 Mental Health Component Summary scale (MCS) scores than did minimally symptomatic subjects (32.8 vs 43.5, $P < .05$). Minimally symptomatic subjects had MCS scores similar to those of a third group of subjects who did not meet DSM-III-R "threshold" criteria for mood disorder but who were severely symptomatic. Adjusted health care utilization was higher for the initial 3-month charge period in the severely symptomatic depressed subjects compared with minimally symptomatic depressed subjects ($679.20 vs $462.38, $P < .05$).

■ CONCLUSIONS The 6-item depression severity measure effectively separated patients meeting DSM-III-R "threshold" depression criteria into 2 groups: one presenting with severe symptoms and impairment and the other presenting with mild symptoms and significantly less impairment. A strategy of initial screening using a brief depression severity instrument, followed with a DSM criteria–based instrument, could decrease the immediate clinician workload by one third and focus treatment on those most likely to benefit.

■ KEY WORDS Depression; health care utilization; predictive value of tests; primary health care; screening. (*J Fam Pract 2002; 51:1065–1070*)

Numerous efforts have been directed toward improving primary care clinicians' detection of depression since the report of early findings that depressive disorders are common yet often unrecognized in primary care.[1,2] Despite the recent release of a new United States Preventive Task Force recommendation,[3] controversy exists about the benefits and cost-effectiveness of routine screening.[4-7]

Despite the controversies around depression screening, it is clear that there is significant room for improvement in detection of and treatment outcomes for depression in primary care. Additionally, there is ample evidence from clinical trials that depressed patients with higher severity of illness receive the highest benefit from pharmacological treatment. Therefore, it makes sense to target these highly impaired, depressed patients for detection and treatment.

TABLE 1

Brief Depression Rating*

Over the LAST 2 WEEKS, how often have you experienced any of the following?*

- Feeling sad.
- Having no interest in being with other people.
- Feeling like a failure as a person.
- Having trouble making decisions.
- Feeling so down that nothing could cheer you up.
- Feeling depressed.

*Responses to questions are on a 5-point Likert scale ranging from "none of the time" to "all of the time."

In previous studies exploring the relationships between symptom severity and diagnostic criteria in a large sample of primary care patients, we found that (1) the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria for major depression were nonspecific at low levels of impairment but more accurate at high levels, and (2) mood symptom severity assessment performed better than DSM criteria as an independent predictor of impairment and utilization.[8,9] These findings lend support to the notion that case-finding methods incorporating severity in addition to criteria can improve the efficiency of screening in primary care. This study represents our initial exploration of the potential impact of severity-enhanced screening for depression.

We used a retrospective cohort design to answer the following study questions: (1) Can the administration of a symptom severity scale effectively "filter out" a group of patients who meet diagnostic criteria for "threshold" depression but who have less impairment (and may therefore not need treatment)? (2) Does this filtering strategy inappropriately "filter out" patients who are in need of treatment?

METHODS

Population and setting

Our sample consisted of 1317 patients presenting for routine care in a university-based family medicine center at the University of Texas Medical Branch (UTMB) in Galveston. The sample, originally recruited for a National Institute for Alcohol Abuse and Alcoholism–funded study of primary care alcohol screening, has been previously described.[10] The study methods and additional data collection methods were reviewed and approved by the UTMB Institutional Review Board.

Evaluation measures

We used the Medical Outcomes Study SF-36 subscales and component summary scale[11,12] to assess health-related quality of life (HRQOL) in all sub-

jects. Medical comorbidity was assessed using electronic medical record review as described previously.[8] We also examined health care utilization using charge data from the billing system of UTMB. As previously described,[8] we obtained all inpatient and outpatient charge data for a 15-month period beginning 3 months before the visit at which each subject was surveyed. Outpatient pharmacy data were not included. The results of the charge data subanalysis are presented online in Figure W1, at www.jfponline.com.

Analytic strategy

All subjects were screened with the Clinician Evaluation Guide mood module from the Primary Care Evaluation of Mental Disorders (PRIME-MD).[13] A "DSM criteria positive" screen included major depressive disorder (MDD), dysthymia, and partial remission of MDD. Symptom severity was assessed using a 6-item Brief Depression Rating (BDR) scale (Table 1) derived from a principal components analysis of 15 mood and anxiety symptom severity questions used in the original study and our subsequent investigations.[8] Factor analysis of the 6 BDR items confirmed that they occupy a domain distinct from the somatic symptoms included as PRIME-MD depression criteria.

Cronbach's alpha for the BDR in our sample was 0.8911. Because the distribution of subjects was skewed toward lower severity (median = 9, mean = 10.47, skewness = 1.415), we chose the 75th percentile score[13] as our cut point for a "positive" BDR. This choice reflected a more conservative definition of severity than the use of a standard cut point of 1 standard deviation above the mean (in this case, a score of 15).

We "filtered out" low-severity patients by matching BDR scores and DSM criteria to create 4 groups for comparison: "low severity and DSM negative," "high severity only," "DSM positive only," and "high severity and DSM positive."

Data analysis

We used analysis of variance to compare the 4 groups on demographic and outcome measures of interest. We made adjustments where demographic variables or medical comorbidity contributed significantly to the differences between groups by using analysis of covariance (ANCOVA). We examined interaction effects between the covariates and the severity/DSM groups. Where possible and appropriate, we used Bonferroni or Games-Howell adjustments for multiple comparisons between groups.

RESULTS

Size and demographic comparisons

The distribution of the 1317 subjects available for analysis is depicted in Table 2. Fully 75% of the total sample fell below the BDR severity threshold. The

TABLE 2

Group demographics

Characteristic	PRIME-MD criteria (−)		PRIME-MD criteria (+)	
	BDR severity (−)	BDR severity (+)	BDR severity (−)	BDR severity (+)
Subjects, n	893	119	91	214
Female subjects, %	66.2	72.3	74.4	84.1
Race, %				
White	38.3	35.3	41.7	40.2
African American	34.5	41.2	28.6	36.9
Hispanic	27.2	23.5	29.7	22.9
Mean age, y	43.9*	43	42.5	40.0*

Chi-square is significant for sex ($P < .001$) but not for racial distributions ($P = .500$).
*Significant differences exist for mean age by analysis of variance using Bonferroni adjustment ($P = .012$).
BDR, Brief Depression Rating; Prime-MD, Primary Care Evaluation of Mental Diseases.

BDR filtered out 29% of those subjects meeting DSM criteria because of low symptom severity. Conversely, 17% of subjects who did not meet DSM criteria had high symptom severity based on the BDR. Although the groups had similar demographic characteristics, subjects in the "high severity and DSM positive" group were significantly younger than subjects in the "low severity and DSM negative" group. The distribution of women in all groups was significantly higher than expected except for the "low severity and DSM negative" group. We found even distributions of subjects by ethnicity.

Mean HRQOL score comparisons

Figure 1 shows mean Mental Health Component Summary (MCS) scores for subjects in the 4 groups, after ANCOVA adjustments for significant covariates (age and African-American ethnicity, $P = .003$ for both). The groups of subjects that scored either positively or negatively on both the BDR and PRIME-MD occupy opposite poles of very low and very high functional status, respectively. The groups of subjects that scored positively on only the BDR or only the PRIME-MD share the middle ground with no significant difference in MCS-related functional status.

A similar pattern was seen for the Physical Component Summary (PCS) scores from the SF-36. PCS score means ranged from 41.60 to 44.17 among the 4 groups after ANCOVA adjustment for significant covariates (income, medical comorbidity, and Hispanic ethnicity, $P < .001$ for each). Only the "low severity and DSM negative" and "high severity and DSM positive" groups differed significantly at either end of this range; however, the absolute difference of 2.57 points carries minimal, if any, clinical significance.

Unadjusted mean values from SF-36 subscale scores across the 4 study groups are shown in Figure 2. Although we saw no differences in the "physical functioning" and "role-physical" subscale scores among the groups, a consistent pattern emerged for the remaining 6 subscales. The "high severity and DSM positive" group had significantly lower mean scores (indicating more impairment) than each of the other 3 groups, whereas the "low severity and DSM negative" group had significantly higher scores than each of the other 3 groups. The other 2 groups' means were in the middle and almost identical across all 8 subscales, indicating that these 2 groups were similar on each SF-36 measure of physical and mental health functioning.

Mean health care charge comparisons

Briefly, adjusted mean health care charges for each group of subjects showed significant charge differences between groups for the period 3 months before the index visit. The adjusted mean health care charges for this period are shown in Figure W1.

DISCUSSION

We believe that the central findings of this study support a severity-targeted screening strategy. The answer to our first study question—Can the addition of a symptom severity scale effectively "filter out" a group of patients who meet diagnostic criteria for "threshold" depression but have less impairment?—is "yes." We were able to separate patients meeting criteria for depression into 2 groups, roughly one third with mild symptom severity and roughly two thirds with moderate to severe symptom severity.

The answer to our second question—Does this filtering strategy filter out patients who are in need of treatment?—appears to be "no." The patterns of HRQOL scores and health care utilization seen for the "filtered-out" patients were indistinguishable from those of a third group of more severely symptomatic patients who did not meet depression criteria at the time of screening and who would not

Mean deviations from standardized SF-36 subscale norms
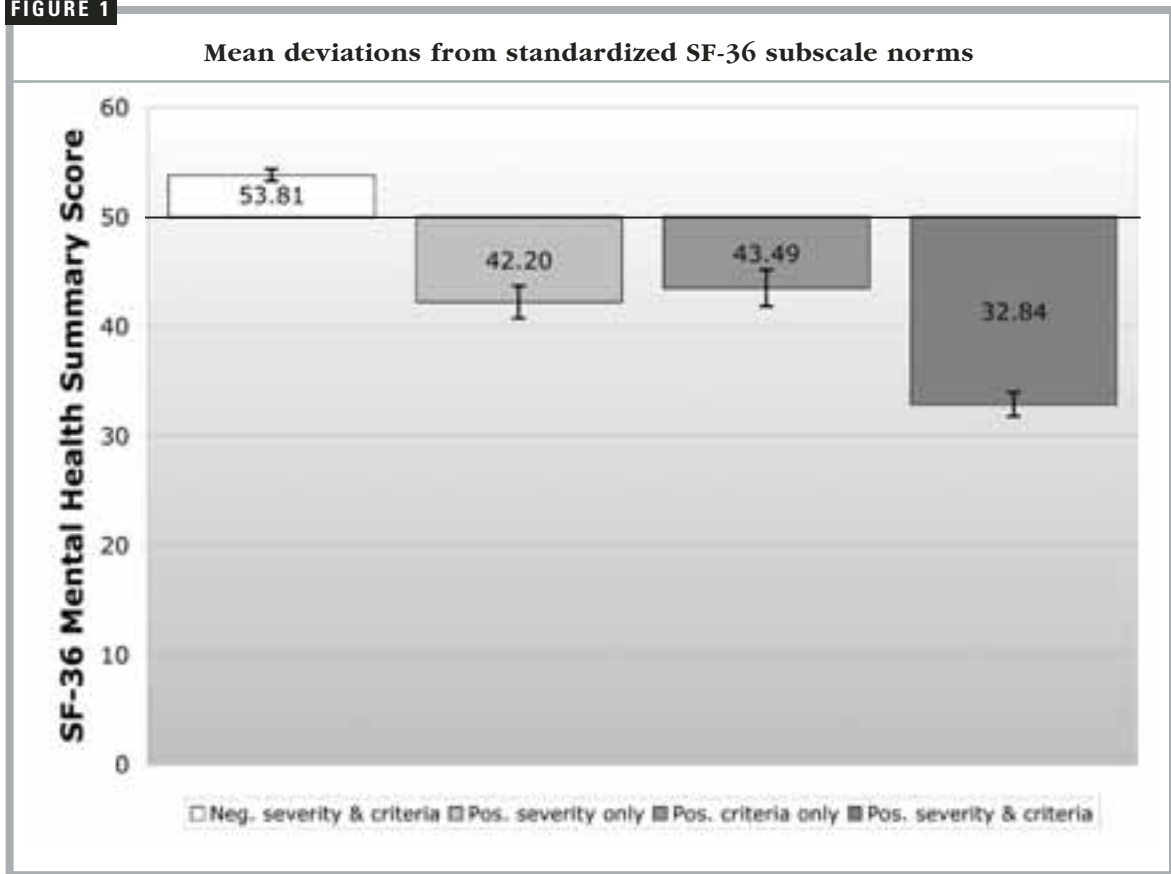
routinely be considered candidates for antidepressant treatment. The presence of a cohort of "middle-ground" patients has been noted in other cross-sectional primary care samples.[14] Whether these patients represent persons with "major depression-in-waiting" or simply distressed and sad individuals is debatable, but there is no evidence to suggest that immediate detection and treatment lead to improved outcomes for these patients. Therefore, in routine clinical practice there would appear to be little risk in failing to identify and treat these patients unless or until their symptom severity increases.
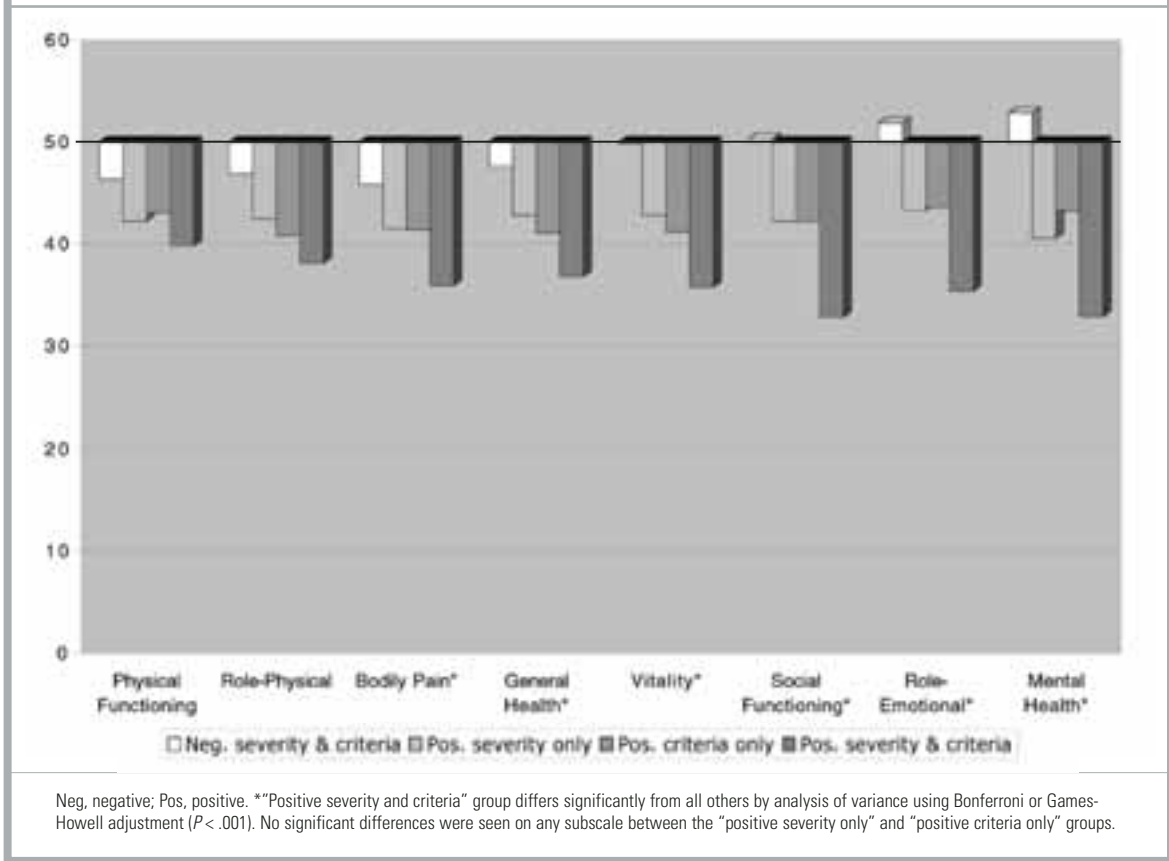
This study does contain some important limitations. First, its cross-sectional nature does not allow us to address important questions about the middle-ground ("high severity only" and "DSM positive only") patients, such as when they might warrant treatment, whether or when rescreening is useful, or whether "watchful waiting" is the appropriate clinical strategy for these 2 groups. Also, our decision to include as "DSM positive" those patients meeting criteria for dysthymia and MDD in remission deserves a brief explanation. Our previous work with this sample suggested that many patients meeting criteria for these 2 syndromes had high levels of distress and might be thought of as "depressed" by clinicians in routine practice. We included them to make our stratification strategy more closely representative of usual primary care

practice. Repeat analyses including only MDD patients as "DSM positive" did not change our primary findings and conclusions, but they did—as expected—decrease the number of subjects in the "positive severity and criteria" group as well as increase the number of subjects in the "high severity only" group.

Despite these limitations, we believe that the results of this study offer hope to practicing physicians trying to cope with the growing depression screening mandate. Primary care physicians seeking to implement depression screening must deal with the fact that depression-screening protocols impose significant burdens on busy clinicians. In the setting of high competing demand[15,16] in primary care, this additional effort—or "cognitive burden"—may render such screening impossible to accomplish in a routine clinical encounter. Several studies support this notion. Rost et al[17,18] found that a screening protocol was not sustainable in primary care, in large part because primary care clinicians were unable to determine which screened patients were most in need of treatment. Dobscha et al[19] found that clinicians failed to adhere to even a limited practice-based screening protocol. Williams et al[20] found no difference in treatment rates or short-term outcomes when comparing brief (1-question) and comprehensive (20-question) case-finding protocols with customary clinical care.

FIGURE 2

Mean deviations from standardized SF-36 subscale norms

Neg, negative; Pos, positive. *"Positive severity and criteria" group differs significantly from all others by analysis of variance using Bonferroni or Games-Howell adjustment ($P < .001$). No significant differences were seen on any subscale between the "positive severity only" and "positive criteria only" groups.

Our results suggest that a simple refinement to a screening protocol—ie, using a brief severity measure to target the patients most appropriate for further DSM diagnostic evaluation—could help clinicians in 2 ways. First, it could decrease the burden of positive screening results by one third according to this study. Second, it could provide a more specific "prompt to act" rather than the "prompt to consider" provided by the use of current DSM criteria–based instruments. The importance of this last point should not be underestimated. Valenstein et al[21] demonstrated that clinicians' perceptions of the value of positive screen results are closely linked to their likelihood to initiate treatment. If we can enhance the value of the positive prompt, we can improve the rate of response to prompting.

Although we believe that the principle of severity targeting, rather than the specific instrument chosen, will improve screening performance, the instrument must nonetheless be chosen carefully. Kroenke et al[22] examined the utility of using the quantitative score from the Patient Health Questionnaire, 9-item version, (PHQ-9) as a severity measure and found that higher scores correlated with lower functional status scores, greater numbers of sick days, and greater health care uti-

lization. However, their methodology included as "positive" only those patients who met diagnostic criteria for MDD. Our use of an independent severity instrument identified an additional 17% of middle-ground patients who might benefit from close observation ("watchful waiting") without the need for active management.

In summary, we believe that severity-targeted screening represents a promising "next step" in the evolution of office-based screening for depression in primary care. Much more work is needed to determine whether this "prompt to act" will be followed by improved treatment adherence and better treatment outcomes.

REFERENCES
1. Regier DA, Goldberg ID, Taube CA. The de facto US Mental Health Services system: a public health perspective. Arch Gen Psychiatry 1978; 35:685–93.
2. Katon W, Schulberg H. Epidemiology of depression in primary care. Gen Hosp Psychiatry 1992; 14:237–47.
3. Pignone MP, Gaynes BN, Rushton JL, Burchell CM, Orleans CT,

Mulrow CD, 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.

4. Gilbody SM, House AO, Sheldon TA. Routinely administered questionnaires for depression and anxiety: systematic review. Br Med J 2001; 322:406–9.

5. 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.

6. Schoenbaum M, Unutzer J, Sherbourne C, Duan N, Rubenstein LV, Miranda J, et al. Cost-effectiveness of practice-initiated quality improvement for depression: results of a randomized controlled trial. JAMA 2001; 286:1325–30.

7. Simon GE, Manning WG, Katzelnick DJ, Pearson SD, Henk HJ, Helstad CS. Cost-effectiveness of systematic depression treatment for high utilizers of general medical care. Arch Gen Psychiatry 2001; 58:181–7.

8. Nease DE Jr, Volk RJ, Cass AR. Investigation of a severity-based classification of mood and anxiety symptoms in primary care patients. J Am Board Fam Pract 1999; 12:21–31.

9. Nease DE Jr, Volk RJ, Cass AR. Does the severity of mood and anxiety symptoms predict high health care utilization? J Fam Pract 1999; 48:769–77.

10. Volk RJ, Cantor SB, Steinbauer JR, Cass AR. Alcohol use disorders, consumption patterns, and health-related quality of life of primary care patients. Alcohol Clin Exp Res 1997; 21:899–905.

11. Ware JE Jr, Kosinski M, Bayliss MS, McHorney CA, Rogers WH, Raczek A. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the Medical Outcomes Study. Med Care 1995; 33(suppl 4):AS264–79.

12. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992; 30:473–83.

13. 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.

14. Klinkman MS, Coyne JC, Gallo S, Schwenk TL. False positives, false negatives, and the validity of the diagnosis of major depression in primary care. Arch Fam Med 1998; 7:451–61.

15. 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.

16. 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.

17. 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.

18. Rost K, Nutting P, Smith J, Werner J, Duan N. Improving depression outcomes in community primary care practice: a randomized trial of the QuEST intervention. J Gen Intern Med 2001; 16:143–9.

19. Dobscha SK, Gerrity MS, Ward MF. Effectiveness of an intervention to improve primary care provider recognition of depression. Eff Clin Pract 2001; 4:163–71.

20. Williams JW, Mulrow CD, Kroenke K, Dhanda R, Badgett RG, Omori D, et al. Case-finding for depression in primary care: a randomized trial. Am J Med 1999; 106:36–43.

21. Valenstein M, Dalack G, Blow F, Figueroa S, Standiford C, Douglass A. Screening for psychiatric illness with a combined screening and diagnostic instrument. J Gen Intern Med 1997; 12:679–85.

22. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001; 16:606–13.

JFP