

Identifying Depressed Older Adults in Primary Care: A Secondary Analysis of a Multisite Randomized Controlled Trial

Corrine I. Voils, Ph.D.; Maren K. Olsen, Ph.D.;
and John W. Williams, Jr., M.D., M.H.S., for the IMPACT Study Investigators

Objective: To determine whether a subset of depressive symptoms could be identified to facilitate diagnosis of depression in older adults in primary care.

Method: Secondary analysis was conducted on 898 participants aged 60 years or older with major depressive disorder and/or dysthymic disorder (according to DSM-IV criteria) who participated in the Improving Mood—Promoting Access to Collaborative Treatment (IMPACT) study, a multisite, randomized trial of collaborative care for depression (recruitment from July 1999 to August 2001). Linear regression was used to identify a core subset of depressive symptoms associated with decreased social, physical, and mental functioning. The sensitivity and specificity, adjusting for selection bias, were evaluated for these symptoms. The sensitivity and specificity of a second subset of 4 depressive symptoms previously validated in a midlife sample was also evaluated.

Results: Psychomotor changes, fatigue, and suicidal ideation were associated with decreased functioning and served as the core set of symptoms. Adjusting for selection bias, the sensitivity of these 3 symptoms was 0.012 and specificity 0.994. The sensitivity of the 4 symptoms previously validated in a midlife sample was 0.019 and specificity was 0.997.

Conclusion: We identified 3 depression symptoms that were highly specific for major depressive disorder in older adults. However, these symptoms and a previously identified subset were too insensitive for accurate diagnosis. Therefore, we recommend a full assessment of DSM-IV depression criteria for accurate diagnosis.

(Prim Care Companion J Clin Psychiatry 2008;10:9–14)

Received Oct. 1, 2007; accepted Nov. 26, 2007. From the Center for Health Services Research in Primary Care, Durham Veterans Affairs Medical Center (Drs. Voils, Olsen, and Williams), and the Department of Medicine, Division of General Internal Medicine (Drs. Voils and Williams); Department of Biostatistics and Bioinformatics (Dr. Olsen); and Department of Psychiatry and Behavioral Sciences (Dr. Williams), Duke University Medical Center, Durham, North Carolina.

This study was supported by grants from The John A. Hartford Foundation, the California Healthcare Foundation, the Hogg Foundation, and the Robert Wood Johnson Foundation.

Acknowledgements appear at the end of this article.

The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

The authors report no additional financial or other relationships relevant to the subject of this article.

Corresponding author and reprints: Corrine I. Voils, Ph.D., Health Services Research and Development, Durham Veterans Affairs Medical Center (152), 508 Fulton St., Durham, NC 27705 (e-mail: voils001@mc.duke.edu).

Primarily care physicians are at the front lines of depression care, prescribing approximately 70% of antidepressant medications.^{1,2} Primary care physicians identify depression management as one of the most challenging aspects of practice, and indeed, health services research identifies consistent gaps in the quality of care.^{3–5} In response, the National Institutes of Health, several foundations, and other groups launched national programs to improve the recognition of depressed patients and close the quality gaps.⁶ These programs, together with the introduction of newer, easier-to-use antidepressants and direct-to-consumer marketing, have contributed to a 3-fold increase in antidepressant prescriptions.

Although it appears that a depressed patient today is more likely to be identified and treated than one decade ago, there is increasing concern that much of the growth in antidepressant prescriptions is misspent on patients who are unlikely to benefit from active treatment.⁷ Both direct and indirect evidence supports this thesis. Among patients referred by primary care physicians for antidepressant treatment, only about one half meet formal criteria for major depressive disorder. In addition, surveys show that only about 50% of primary care physicians can cite at least 5 depression criterion symptoms, and only 16% report using formal diagnostic criteria, citing insufficient time as an important barrier.^{8,9} Thus, diagnostic imprecision may contribute to inappropriate antidepressant prescribing and rising health care costs.

To increase diagnostic precision, knowledge and time barriers must be addressed. One possibility is to use self-administered diagnostic assessment tools such as the Patient Health Questionnaire depression screener (PHQ-9).¹⁰ These diagnostic tools perform well, but, despite robust educational campaigns, they are used routinely by fewer than 5% of primary care physicians. Another strategy is to better focus the diagnostic interview by identifying a reduced set of core symptoms. Brody et al.¹¹ identified 4 core symptoms (SALsA: sleep disturbance, anhedonia, low self-esteem, and decreased appetite) that accounted for a significant proportion of variance in functional status and well-being in a predominately midlife sample. These symptoms had a sensitivity of 65% and specificity of 99% for major depression, and, thus, when present, “rule in” clinical depression. Because older adults may have a somewhat different symptom pattern and high rates of medical comorbidity,¹² it is uncertain if these core symptoms would perform as well in this population.

We performed a secondary data analysis from a large study of primary care depression to (1) identify a subset of depressive symptoms that was associated with decreased social, physical, and mental functioning, as impairment would indicate significant debilitation and signal the need for intervention; (2) determine the diagnostic performance of these symptoms; and (3) test the SALsA symptoms in our older adult population.

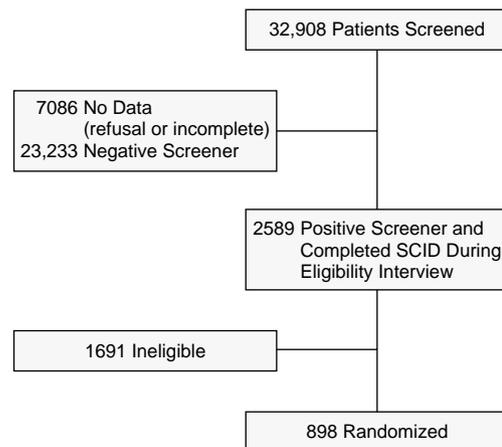
METHOD

Participants and Procedure

Data for the current study were part of the Improving Mood–Promoting Access to Collaborative Treatment (IMPACT) study, a multisite randomized controlled trial of a primary care–based collaborative care management program for late-life depression (recruitment from July 1999 to August 2001).¹³ All trial sites were approved by their respective institutional review boards and adhered to the regulatory procedures concerning informed consent.

Details of the method are described elsewhere.¹³ Briefly, patients were approached systematically in a primary care clinic, or they were recruited via healthcare staff referral or self-referral. Approached patients completed the Primary Care Evaluation of Mental Disorders (PRIME-MD) 2-item depression screener.¹⁴ Patients who responded affirmatively to the PRIME-MD screening question or who were referred to the study were asked to complete an eligibility interview, which included the Structured Clinical Interview for DSM-IV (SCID).¹⁵ Patients were eligible for the study if they met SCID criteria for major depressive disorder and/or dysthymic disorder, were aged 60 years or older, and planned to use one of the 18 participating primary care clinics as their main source of general medical care in the following year. Exclusion

Figure 1. Flowchart of IMPACT Study Participants Included in the Current Analyses



Abbreviations: IMPACT = Improving Mood–Promoting Access to Collaborative Treatment, SCID = Structured Clinical Interview for DSM-IV.

criteria included current drinking problem, history of bipolar disorder or psychosis, ongoing treatment with a psychiatrist, moderate to severe cognitive impairment, and suicidal risk requiring immediate psychiatric evaluation. All eligible patients were randomly assigned to either the intervention or usual care. For analyses reported herein, we focused exclusively on patients who completed the PRIME-MD screener (N = 898; see Figure 1) because we needed data from the screener to complete the sensitivity and specificity analyses. Compared to nonscreened patients, screened patients had slightly lower depression severity but did not differ on other clinical characteristics.¹³

Measures

During the baseline interview, a trained interviewer, using a Computer Aided Telephone Interview, elicited SCID symptoms, including depressed mood, anhedonia, appetite change, sleep disturbance, psychomotor changes, fatigue, feeling worthless or guilty, difficulty concentrating, and suicidal ideation.¹⁵ Two measures of impairment were administered as well. The 3-item Sheehan Disability Scale (SDS; $\alpha = .82$) assessed the extent to which emotional symptoms impair family life/home responsibilities, work, and social life.^{16,17} The response scale ranged from 0 (not at all) to 10 (unable to carry on any activities). The Medical Outcomes Study 12-Item Short-Form Health Survey (SF-12) assessed health-related quality of life.¹⁸

Analyses

The SCID items were dichotomous, with 1 indicating presence of the symptom and 0 indicating absence. Responses to the SDS were averaged to form a composite score that ranged from 0 to 10¹⁷; higher scores correspond

Table 1. Characteristics of Older Adults With Major Depressive Disorder and/or Dysthymic Disorder (N = 898)^a

| Characteristic | Value |
|--|---------------------------------|
| Age (range, 60–93), mean (SD), y | 70.1 (7.3) |
| Income (range, \$0–\$960,000), median (IQR) ^b | \$23,500 (\$12,000 to \$45,000) |
| Chronic diseases (of a list of 10), mean (SD) | 3.9 (2.0) |
| Sheehan Disability Scale score (range, 0–10), mean (SD) | 4.6 (2.6) |
| SF-12 Physical Component Summary score (range, 24–61), mean (SD) | 40.3 (7.6) |
| SF-12 Mental Component Summary score (range, 19–62), mean (SD) | 42.2 (7.7) |
| Female, N (%) | 551 (61.4) |
| Race, N (%) | |
| White | 675 (75.2) |
| Black | 157 (17.5) |
| Hispanic | 51 (5.7) |
| Other | 15 (1.7) |
| Education, N (%) | |
| Less than high school | 207 (23.1) |
| High school graduate | 190 (21.2) |
| Some college | 292 (32.5) |
| College graduate | 209 (23.3) |
| Diagnosis, N (%) | |
| Major depressive disorder | 115 (12.8) |
| Dysthymic disorder | 328 (36.5) |
| Major depressive disorder with dysthymic disorder | 455 (50.7) |
| Mild cognitive impairment, N (%) ^c | 285 (31.7) |
| Depressive symptoms, N (%) | |
| Depressed mood | 613 (68.3) |
| Anhedonia | 580 (64.6) |
| Appetite change | 455 (50.7) |
| Sleep disturbance | 730 (81.3) |
| Psychomotor changes | 336 (37.4) |
| Fatigue | 778 (86.6) |
| Feeling worthless or guilty | 507 (56.5) |
| Difficulty concentrating | 516 (57.5) |
| Suicidal ideation | 264 (29.4) |

^aAll depressive symptoms are from the Structured Clinical Interview for *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition, Text Revision (SCID).

^bDue to outliers, median income is reported.

^cAssessed with a 6-item screener¹³; a score of 3 to 5 indicated mild impairment.

Abbreviation: IQR = interquartile range, SF-12 = Medical Outcomes Study 12-Item Short-Form Health Survey.

to greater impairment. Mental and physical component summary scores (MCS and PCS) were created by summing appropriate SF-12 items and transforming the score to a 0 to 100 scale, with higher scores indicating greater health-related quality of life.¹⁸

The first goal was to identify a subset of symptoms associated with decreased functioning. For each outcome—SDS, MCS, and PCS scores—we conducted a linear regression model, initially entering all SCID symptoms as independent variables. Backwards step-down selection was used to yield a final reduced model.¹⁹ The stopping rule was based on Akaike Information Criterion²⁰ rather than a p value; variables were deleted until the difference in Akaike Information Criterion was significant. Fifty bootstrap samples (sampling with re-

placement) were used to validate the variable selection process. These analyses yielded a set of 3 core symptoms, shown in the Results.

Our second goal was to calculate the sensitivity and specificity of the 3 core symptoms for assessing depression. We could not calculate the sensitivity and specificity using the typical method because our estimates would be biased due to selection bias, also known as verification bias. Selection bias occurs when disease status is verified in a subset of the patients who were tested initially. In this study, we could only verify SCID symptoms in the subset of patients who screened positive with the PRIME-MD and thus were eligible to enroll in the study; we did not have SCID data for patients who screened as negative. We corrected for this bias using the method of Begg and Greenes.²¹ These calculations required an assumption about the true sensitivity of the PRIME-MD screener in order to calculate the probability of not being depressed according to the screener. We used a value of 0.90, which is the mean sensitivity for 4 primary care studies.^{14,22–24} We also conducted sensitivity analyses to determine whether other values would lead to different conclusions.

The third goal was to compare our results to those of Brody et al.,¹¹ who evaluated the performance of the SALSA symptoms in a midlife sample. To this end, we calculated the sensitivity and specificity of various combinations of the 4 SALSA symptoms to assess their suitability for identifying depression in older adults. We then compared the performance of the SALSA symptoms to the core subset identified in the current study to determine whether either symptom profile correctly identified depressed and nondepressed older adults.

RESULTS

At baseline, participants ranged in age from 60 to 93 years (Table 1), with a mean of 70 years. More than half of the participants were female, were white, and had attended at least some college. Half of the participants had major depressive disorder with dysthymic disorder, and nearly one third had mild cognitive impairment. The most common SCID items endorsed were sleep disturbance and fatigue, with at least 80% of patients reporting these symptoms. More than half of the patients reported depressed mood, anhedonia, appetite change, feeling worthless or guilty, and difficulty concentrating. The least common depressive symptoms were psychomotor changes and active or passive suicidal ideation, with 37% and 29%, respectively.

To identify a subset of symptoms associated with decreased functioning, we investigated which depressive symptoms were associated with SDS, MCS, and PCS scores. Table 2, which shows the results of the bootstrap regression analyses, shows that no subset of symptoms was associated with decrements across all 3 outcomes.

Table 2. Symptoms Significantly Associated With Disability and Functioning in Older Adults With Major Depressive Disorder and/or Dysthymic Disorder^a

| Dependent Variable and Symptoms | B ^b | SE | p Value |
|---|----------------|------|---------|
| Sheehan Disability Scale (df = 893) | | | |
| Anhedonia | 0.64 | 0.18 | .0004 |
| Psychomotor changes | 0.78 | 0.18 | < .0001 |
| Fatigue | 0.87 | 0.25 | .0005 |
| Suicidal ideation | 0.42 | 0.19 | .0254 |
| SF-12 Physical Component Summary score (df = 895) | | | |
| Psychomotor changes | -2.89 | 0.53 | .0005 |
| Fatigue | -3.11 | 0.54 | < .0001 |
| SF-12 Mental Component Summary score (df = 895) | | | |
| Depressed mood | -1.79 | 0.51 | < .0001 |
| Suicidal ideation | -3.37 | 0.73 | < .0001 |

^aDegrees of freedom vary due to missing data.

^bUnstandardized regression coefficient.

Abbreviation: SF-12 = Medical Outcomes Study 12-Item Short-Form Health Survey.

This is sensible, as the 3 outcome variables were not highly intercorrelated (SDS with PCS, $r = -0.41$; SDS with MCS, $r = -0.25$; and MCS with PCS, $r = -0.18$; all $p < .0001$). Four of the 9 symptoms were associated with SDS scores, accounting for 6% of the variance; 2 were associated with PCS, accounting for 4% of the variance; and 2 were associated with MCS, accounting for 7% of the variance. We selected psychomotor changes, fatigue, and suicidal ideation as our core set of depressive symptoms because each was associated with 2 of the 3 outcomes.

We calculated the sensitivity and specificity of the core set of 3 depressive symptoms adjusting for selection bias (Table 3). The results indicate that these symptoms “rule out” but do not “rule in” depression (i.e., specificity was high, but sensitivity was low). As expected, when any 2 of the 3 symptoms were present, the sensitivity was higher, but it was still inadequate. To examine the possibility that the low sensitivity of the 3 symptoms was due to the assumed high sensitivity of the screener (0.90), we explored the effects of different sensitivities of the screener (from 0.50 to 0.95) on the adjusted sensitivity of the 3 core symptoms. This possibility was not supported; using a sensitivity of 0.50 for the screener, which is substantially lower than any published value, the adjusted sensitivity of the 3 core symptoms was only 0.02.

As an additional sensitivity analysis, we investigated the diagnostic performance of the 5 symptoms that were significant in any 1 of the regression models (depressed mood, anhedonia, psychomotor changes, fatigue, and suicidal ideation; see Table 2) and any combination thereof. The adjusted sensitivity and specificity values for this expanded core, and any combination thereof, were similar to those obtained when using the 3 core symptoms mentioned previously. The sensitivity increased as the number of symptoms decreased, but even with only 2 of the 5 symptoms present, the sensitivity was inadequate (sensi-

Table 3. Sensitivity and Specificity of Core Symptom Profile^a

| Symptom Profile | Sensitivity (95% CI) | Specificity (95% CI) |
|-----------------------------------|------------------------|------------------------|
| 3 of 3 core symptoms ^b | 0.012 (0.010 to 0.015) | 0.994 (0.992 to 0.995) |
| At least 2 core symptoms | 0.076 (0.068 to 0.085) | 0.985 (0.983 to 0.987) |

^aEstimates were adjusted for selection bias using the method of Begg and Greenes.²¹

^bCore symptoms (significant in 2 of the 3 regression models): psychomotor changes, fatigue, and suicidal ideation.

tivity range, 0.008 for all 5 symptoms to 0.303 for at least 2 symptoms; specificity range, 0.998 for all 5 symptoms to 0.976 for at least 2 symptoms). We also examined the performance of the 6 symptoms that were not included in the core subset; these also had low sensitivity (0.011) and high specificity (0.998).

Finally, we compared the performance of the 3 core symptoms identified in this study to the 4 SALSA symptoms identified by Brody et al.¹¹ The adjusted sensitivities and specificities were similar for the 2 symptom profiles, exhibiting high specificity but not sensitivity (sensitivity range, 0.019 for all 4 SALSA symptoms to 0.266 for at least 2 SALSA symptoms; specificity range, 0.997 for all 4 SALSA symptoms to 0.976 for at least 2 SALSA symptoms). Thus, both profiles accurately identified nondepressed older adults but did not accurately identify depressed older adults.

DISCUSSION

We attempted to identify a subset of depressive symptoms that would facilitate primary care physicians' diagnosis of major depressive disorder in older adults. We identified 3 criterion symptoms that were highly specific but insensitive and explained only a small proportion of the variability in social, physical, and mental functional status. Furthermore, the core symptoms identified in our older population had no overlap with the SALSA symptoms identified in a predominately midlife population. Therefore, learning variable core symptoms for differing populations is an unlikely solution to diagnostic imprecision.

We found that our symptom profile and the SALSA profile performed similarly in this older adult population; both had high specificity and low sensitivity.¹¹ The SALSA symptoms were less sensitive (unadjusted sensitivity = 16%, unadjusted specificity = 97%) in this older population compared to the earlier finding of 65% unadjusted sensitivity and 99% unadjusted specificity in a midlife population. We propose 3 possible explanations for these differences. First, our sample consists of older adults, whereas their sample included adults of all ages (age range, 18–90 years; mean = 55 years). Older adults may have a different symptom profile for depression, leading to

fewer patients with the cluster of SALsA symptoms. For example, psychomotor retardation or agitation may be more distinct in older populations.¹² Second, our symptom assessments were obtained using the semistructured SCID interview,¹⁵ whereas Brody et al.¹¹ used the PRIME-MD, which may have led to some differential symptom assessment. Third, the utility of SALsA symptoms has not been replicated in midlife populations and may have been specific to the derivation sample.²⁵

Our study has some limitations. Our exclusive inclusion of depressed patients may have attenuated the proportion of variance in outcomes accounted for by the depressive symptoms. Also, the small number of people who completed the SCID interview (N = 2589) relative to the number who did not (N = 23,233) limited the range of sensitivities that we could obtain. Additionally, the Begg and Greenes²¹ correction for sampling bias relied on literature estimates for the performance of the PRIME-MD screener. However, a sensitivity analysis across a wide range of plausible estimates showed that our findings were not dependent upon these estimates. It should be noted that the Begg and Greenes adjustment for selection bias may have resulted in biased estimates of the sensitivity and specificity compared to other methods.²⁶ Nonetheless, we believe that the performance of the different symptom profiles relative to each other would have been similar had we invoked other correction methods. Our study has many strengths, including careful correction for sampling bias, a large sample size, and representation of diverse practices and patients.

How should these findings be incorporated into recommendations for primary care physicians on depression recognition and diagnosis? The United States Preventive Services Task Force recommends screening when systems are in place to provide high quality care.²⁷ A quick verbal screen for all patients with the 2-item PRIME-MD has logistical advantages and reasonable performance characteristics. This approach could also be used more selectively as a case-finding strategy, when depression is suspected based on the presenting symptoms or nonverbal cues. For practices that can solve the logistics of distributing, collecting, and scoring longer questionnaires, the PHQ-9 is the best validated in primary care.²⁸ Demonstration projects and research studies have implemented systematic screening using Web sites, interactive voice response, hand-held computers, and pencil and paper. Whichever administration method is chosen, patients with positive screens will require additional evaluation to elicit enough symptoms for a DSM-IV diagnosis and to rule out other disorders such as bipolar disorder or substance abuse. On the basis of our failure to validate the SALsA symptoms or to identify a novel symptom set with high sensitivity and specificity in this older population, we cannot recommend a truncated symptom set. Rather, the primary care physician should review at least enough criterion symptoms to

establish or refute a diagnosis of major depressive disorder or dysthymic disorder. For physicians who use the PHQ-9 as a screener, this process can be streamlined because patients self-report criterion symptoms on this instrument. Thus, the physician can use the PHQ-9 as a guide to inquiring about selected symptoms, an inquiry that should always include suicidal ideation because the prevalence in depressed patients is high and the consequences of failure to detect may be dire.

In conclusion, the current findings suggest that primary care providers may need to provide a more comprehensive symptom assessment among patients reporting symptoms of depression. Formal assessment tools, if promoted by educators, insurance plans, and professional societies may be part of the solution to diagnostic imprecision. However, education alone is unlikely to change practice. Innovative solutions, such as depression toolkits, performance indicators, reimbursement for administering depression questionnaires, and linking higher payment to higher quality care via quality indicators among others, need to be evaluated.

Acknowledgments: We acknowledge the contributions and support of patients, primary care providers, and staff at the study coordinating center and at all participating study sites, which include the following: Duke University, Durham, N.C.; South Texas Veterans Health Care System, San Antonio; Central Texas Veterans Health Care System, Austin; San Antonio Preventive and Diagnostic Medicine Clinic, San Antonio, Tex.; Indiana University School of Medicine, Indianapolis; Health and Hospital Corporation of Marion County, Indianapolis, Ind.; Group Health Cooperative of Puget Sound in cooperation with the University of Washington, Seattle; Kaiser Permanente of Northern California, Oakland and Hayward; Kaiser Permanente of Southern California, San Diego; and Desert Medical Group, Palm Springs, Calif. This study is the result of work supported in part with patients, resources, and the use of facilities at the South Texas Veterans Health Care System and the Central Texas Veterans Health Care System.

The IMPACT Investigators include (in alphabetical order) Patricia Arean, Ph.D. (Co-principal investigator [PI]); Thomas R. Belin, Ph.D.; Noreen Bumby, D.O.; Christopher Callahan, M.D. (PI); Paul Ciechanowski, M.D., M.P.H.; Ian Cook, M.D.; Jeffrey Cordes, M.D.; Steven R. Counsell, M.D.; Richard Della Penna, M.D. (Co-PI); Jeanne Dickens, M.D.; Michael Getzell, M.D.; Howard Goldman, M.D., Ph.D.; Lydia Grypma, M.D. (Co-PI); Linda Harpole, M.D., M.P.H. (PI); Mark Hegel, Ph.D.; Hugh Hendrie, M.B., Ch.B., D.Sc. (Co-PI); Polly Hitchcock Noel, Ph.D. (Co-PI); Marc Hoffing, M.D., M.P.H. (PI); Enid M. Hunkeler, M.A. (PI); Wayne Katon, M.D. (PI); Kurt Kroenke M.D.; Stuart Levine, M.D., M.H.A. (Co-PI); Elizabeth H. B. Lin, M.D., M.P.H. (Co-PI); Tonya Marmon, M.S.; Eugene Oddone, M.D., M.H.Sc. (Co-PI); Sabine Oishi, M.S.P.H.; R. Jerome Rauch, M.D.; Michael Sands, M.D.; Michael Schoenbaum, Ph.D.; Rik Smith, M.D.; David C. Steffens, M.D., M.H.S.; Christopher A. Steinmetz, M.D.; Lingqi Tang, Ph.D.; Iva Timmerman, M.D.; Jürgen Unützer, M.D., M.P.H. (PI); John W. Williams Jr., M.D., M.H.S. (PI); Jason Worchel, M.D.; and Mark Zweifach, M.D.

REFERENCES

1. Pincus HA, Tanielian TL, Marcus SC, et al. Prescribing trends in psychotropic medications: primary care, psychiatry, and other medical specialties. *JAMA* 1998;279:526-531
2. Harman JS, Crystal S, Walkup J, et al. Trends in elderly patients' office visits for the treatment of depression according to physician specialty: 1985-1999. *J Behav Health Serv Res* 2003;30:332-341
3. Williams ME, Connolly NK. What practicing physicians in North Caro-

- lina rate as their most challenging geriatric medicine concerns. *J Am Geriatr Soc* 1990;38:1230–1234
4. Leclere H, Beaulieu MD, Bordage G, et al. Why are clinical problems difficult? general practitioners' opinions concerning 24 clinical problems. *CMAJ* 1990;143:1305–1315
 5. Glasser M, Gravdal JA. Assessment and treatment of geriatric depression in primary care settings. *Arch Fam Med* 1997;6:433–438
 6. Pincus HA, Pechura C, Keyser D, et al. Depression in primary care: learning lessons in a national quality improvement program. *Adm Policy Ment Health* 2006;33:2–15
 7. Croghan TW, Schoenbaum M, Sherbourne CD, et al. A framework to improve the quality of treatment for depression in primary care. *Psychiatr Serv* 2006;57:623–630
 8. Williams JW Jr, Rost K, Dietrich AJ, et al. Primary care physicians' approach to depressive disorders: effects of physician specialty and practice structure. *Arch Fam Med* 1999;8:58–67
 9. Shao WA, Williams JW Jr, Lee S, et al. Knowledge and attitudes about depression among non-generalists and generalists. *J Fam Pract* 1997;44:161–168
 10. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;16:606–613
 11. Brody DS, Hahn SR, Spitzer RL, et al. Identifying patients with depression in the primary care setting: a more efficient method. *Arch Intern Med* 1998;158(22):2469–2475
 12. Blazer DG. Depression in late life: review and commentary. *J Gerontol A Biol Sci Med Sci* 2003;58:249–265
 13. Unutzer J, Katon W, Callahan CM, et al. Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. *JAMA* 2002;288:2836–2845
 14. 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
 15. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision*. Washington, DC: American Psychiatric Association; 2000
 16. Sheehan DV, Harnett-Sheehan K, Raj BA. The measurement of disability. *Int Clin Psychopharmacol* 1996;11(suppl 3):89–95
 17. Leon AC, Olfson M, Portera L, et al. Assessing psychiatric impairment in primary care with the Sheehan Disability Scale. *Int J Psychiatry Med* 1997;27:93–105
 18. Ware J Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996;34:220–233
 19. Harrel F. *Regression Modeling Strategies*. New York, NY: Springer-Verlag; 2001
 20. Atkinson A. A note on the generalized information criterion for choice of a model. *Biometrika* 1980;67:413–418
 21. Begg C, Greenes R. Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics* 1983;39:207–215
 22. Arroll B, Khin N, Kerse N. Screening for depression in primary care with two verbally asked questions: cross sectional study. *BMJ* 2003;327:1144–1146
 23. Blank K, Gruman C, Robison JT. Case-finding for depression in elderly people: balancing ease of administration with validity in varied treatment settings. *J Gerontol A Biol Sci Med Sci* 2004;59:378–384
 24. Whooley MA, Avins AL, Miranda J, et al. Case-finding instruments for depression. *J Gen Intern Med* 1997;12:439–445
 25. Wasson JH, Sox HC, Neff RK, et al. Clinical prediction rules: applications and methodological standards. *N Engl J Med* 1985;313:793–799
 26. Harel O, Zhou X. Multiple imputation for correcting for verification bias. *Stat Med* 2006;25:3769–3786
 27. 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;136:765–776
 28. Lowe B, Spitzer RL, Grafe K, et al. Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians' diagnoses. *J Affect Disord* 2004;78:131–140