

Enhancing the clinical utility of depression screening

Kurt Kroenke MD

See related research article by Manea and colleagues on page E191 and at www.cmaj.ca/lookup/doi/10.1503/cmaj.110829

Screening has a checkered legacy in health care. Too often, clinicians have been advised to screen for a variety of diseases or risk factors in advance of evidence that screening actually leads to improved outcomes. Depression is no exception. A meta-analysis of randomized clinical trials showed that depression screening alone is insufficient to improve outcomes,¹ whereas the addition of coordinated patient follow-up, monitoring of adherence and response to therapy, and adjustment of treatment is beneficial.² A diagnostic meta-analysis of the brief Patient Health Questionnaire (PHQ-9) by Manea and colleagues³ not only helps us determine the optimal cut-off score for one of the more commonly used depression measures but also prompts us to consider what makes any depression measure more clinically useful.

A literature synthesis found that depression measures are more alike than different in their ability to detect major depression.⁴ Therefore, features other than diagnostic performance should factor into selecting a particular measure. From a pragmatic standpoint, the ideal measure is brief, self-administered, multipurpose, free and easy to score. For example, ultra-brief measures (e.g., consisting of two to four items) can perform as well as longer measures for screening purposes.^{5,6} The predecessor of the PHQ-9 was the Primary Care Evaluation of Mental Disorders (PRIME-MD), which coupled a patient-completed screener with follow-up questions that require clinician interview. Conversion of the two-step PRIME-MD into the entirely self-administered PHQ-9 markedly increased uptake into clinical practice.⁶

A multipurpose measure is one that can be used for screening, severity assessment, probable diagnosis and treatment monitoring. A free measure is one that is available at no cost to the clinician or patient and, moreover, is easily accessible on a public-domain website. Easy scoring is exemplified by a measure that provides a single summative score of individual items without the complexity of reverse-scored items, transformation of raw scores into standardized scores, or several subscale scores.

Although there are a number of psychometrically comparable measures for depression screening, the PHQ-9 satisfies the five practical considerations of being brief, self-administered, multipurpose, in the public domain and easy to score. Scores of 5, 10, 15 and 20 on the PHQ-9 represent thresholds of mild, moderate, moderately severe and severe depressive symptoms, respectively. A 5-point change is clinically significant. A score of less than 10 suggests a partial response, and a score of less than 5 represents remission.⁶ The PHQ-9 has been translated into more than 80 languages, and many of these translated versions are freely available (along with other measures in the PHQ family of scales) at www.phqscreeners.com.

Determining an optimal cut-off score and showing sensitivity to change are essential for clinical utility. Although 10 has been the conventional PHQ-9 cut-off score, Manea and colleagues found a higher score (11 or 12) may be preferable in certain settings, and they conclude that “the pooled sensitivity and specificity results show no significant differences in the diagnostic properties of PHQ-9 for cut-off scores between 8 and 11.” Thus, 10 plus or minus 2 points might be viewed as an operational “confidence interval” for the PHQ-9 cut-off score.

The ability of a depression measure to detect valid and clinically meaningful change is especially important because follow-up visits for depression care are often brief, and treatment adjustments are often necessary to optimize out-

Competing interests: None declared.

This article was solicited and has not been peer reviewed.

Correspondence to: Dr. Kurt Kroenke, kkroenke@regenstrief.org

CMAJ 2012. DOI:10.1503/cmaj.112004

KEY POINTS

- Although the conventional PHQ-9 cut-off score for screening for depression has been 10, a wider range (8–12) may be more appropriate, depending on the patient population.
- An ideal measure for screening for depression should be brief, self-administered, all-purpose, in the public domain and easy to score.
- A depression measure that is sensitive to change can be valuable in monitoring response to treatment, a key component of measurement-based care.
- Screening must be combined with patient education, coordinated follow-up, monitoring of adherence and clinical response, and treatment adjustments to achieve optimal outcomes.

comes. Whereas many depression measures are suitable for screening, fewer measures have robust evidence of sensitivity to change; examples include the PHQ-9, the Beck Depression Inventory, the Hospital Anxiety and Depression Scale and the Quick Inventory of Depressive Symptomatology. As long as the responsiveness of a depression measure has been established, I agree with Zimmerman and McGlinchey,⁷ who wrote “More important than which scale clinicians use to measure outcome is that some quantifiable index is used to track the progress of treatment.”

Depression is not unlike other chronic medical problems. Diabetes mellitus, hypertension and hypercholesterolemia are also measured as continuous variables, and cut-off scores are operationally defined rather than absolute. Cut-off scores for diagnosis and treatment may vary depending upon medical comorbidity and other risk factors. Simple screening is insufficient unless patients undergo disease-specific education, monitoring of outcomes and therapeutic adjustments. For example, a national survey in the United States revealed that only 59% of patients with hypertension were receiving treatment and only a third had optimally controlled blood pressure.⁸ Clearly, systems-based interventions are as necessary for the adequate treatment of chronic medical disorders as they are for depression.

The optimal frequency of depression screening has not been determined. Indeed, it is possible that an “informed case-finding approach” (e.g., targeting patients who present with chronic pain or unexplained physical complaints, make frequent health care visits, suffer from comorbid medical or psychiatric conditions, report recent stressors or have other risk factors for depression) may be more efficient than universal screening. Also, an “either/or” approach that relies exclusively on diagnostic criteria may be less desirable than weighing additional factors such as severity and duration of depressive symptoms, functional impairment and desire for treatment. For example, the PHQ-9 has an additional item (not counted in the score) that assesses the degree of impairment.⁶ Arroll and colleagues have discovered that adding a single question about desire for treatment (“Is this something with which you would like help?”) improves both the diagnostic specificity and patient-centeredness of depression screening.⁹

Automated assessment using interactive voice-recorded phone calls or the Internet can facilitate home-based depression screening as

well as monitoring.¹⁰ This can further reduce the cost and clinician burden of assessing depression or other conditions (e.g., pain) that are based on patient-reported outcomes. Also, efficient ways of assessing for suicidal ideation are important as depression screening becomes more widespread. In this respect, the four-item P4 screener is a promising tool for stratifying the risk of self-harm.¹¹ Measurement-based care is also being incorporated into psychiatric classification systems such as fifth version of the Diagnostic and Statistical Manual of Mental Disorders (DSM-V) and has been shown to be useful in mental health settings as well as primary care.¹² Finally, the use of ultra-brief measures to assess anxiety, chronic pain and other somatic symptoms⁶ may also be warranted because such conditions commonly co-occur with depression and can adversely affect treatment response.

References

1. Gilbody S, Sheldon T, House A. Screening and case-finding instruments for depression: a meta-analysis. *CMAJ* 2008;178:997-1003.
2. Gilbody S, Bower P, Fletcher J, et al. Collaborative care for depression: a cumulative meta-analysis and review of longer-term outcomes. *Arch Intern Med* 2006;166:2314-21.
3. Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. *CMAJ* 2012;184:E191-6.
4. Williams JW Jr., Pignone M, Ramirez G, et al. Identifying depression in primary care: a literature synthesis of case-finding instruments. *Gen Hosp Psychiatry* 2002;24:225-37.
5. Mitchell AJ, Coyne JC. Do ultra-short screening instruments accurately detect depression in primary care? A pooled analysis and meta-analysis of 22 studies. *Br J Gen Pract* 2007;57:144-51.
6. Kroenke K, Spitzer RL, Williams JB, et al. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. *Gen Hosp Psychiatry* 2010;32:345-59.
7. Zimmerman M, McGlinchey JB. Why don't psychiatrists use scales to measure outcome when treating depressed patients? *J Clin Psychiatry* 2008;69:1916-19.
8. Chobanian AV, Bakris GL, Black HR, et al. Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *Hypertension* 2003;42:1206-52.
9. Arroll B, Goodyear-Smith F, Kerse N, et al. Effect of the addition of a “help” question to two screening questions on specificity for diagnosis of depression in general practice: diagnostic validity study. *BMJ* 2005;331:884.
10. Johns SA, Kroenke K, Theobald D, et al. Telecare management of pain and depression in patients with cancer: patient satisfaction and predictors of use. *J Ambul Care Management* 2011;34:126-139.
11. Dube P, Kroenke K, Bair MJ, et al. The P4 screener: evaluation of a brief measure for assessing potential suicidal risk in 2 randomized effectiveness trials of primary care and oncology patients. *Prim Care Companion J Clin Psychiatry* 2010;12:pii:PCC.10m00978.
12. Duffy FF, Chung H, Trivedi M, et al. Systematic use of patient-rated depression severity monitoring: Is it helpful and feasible in clinical psychiatry? *Psychiatr Serv* 2008;59:1148-54.

Affiliation: Kurt Kroenke is with the Department of Medicine, Indiana University; and the Center of Excellence on Implementing Evidence Based Practice, Department of Veterans Affairs Health Services Research and Development Center, Indianapolis, Ind.