

PHQ-9: global uptake of a depression scale

Depression is the most prevalent mental disorder, a greater cause of disability than any other disease, and a major contributor to direct and indirect health care costs¹. In the absence of a laboratory or imaging test, eliciting patient symptoms by clinical interview or a self-report scale is the principal way to detect depression and monitor its response to treatment.

First published in 2001, the Patient Health Questionnaire 9-item depression scale (PHQ-9) has had global dissemination, with over 11,000 scientific citations and translations into more than 100 languages. It has been used in hundreds of clinical and population-based studies, incorporated into numerous depression guidelines, and implemented in many clinical practice settings. Depression screening is far from universal; however, where it occurs, the PHQ-9 is a leading scale².

The international spread of the PHQ-9 is likely due to multiple factors³. Its nine items comprise the DSM criteria for depressive disorders, making it both a severity and potentially diagnostic measure. The total score is a simple summation of item scores, and cut-points are easy to memorize: 5, 10, 15, and 20 represent thresholds for mild, moderate, moderately severe, and severe depressive symptoms, respectively. Unlike some depression scales, the PHQ-9 is free to use as a public domain measure, and the many translations make it accessible to populations around the world.

The PHQ family includes several abbreviated versions and companion scales⁴. The PHQ-2 is an ultra-brief screener that comprises the first two items (depressed mood and anhedonia), which are core criteria for depressive disorders. The PHQ-8 omits the ninth item that asks about thoughts of “being better off dead or of hurting yourself in some way”. Although conventionally considered a screening question for suicidal ideation, most positive responses represent endorsement of the first part of this compound item (i.e., being better off dead) rather than active thoughts of self-harm^{5,6}. Because the ninth item is the least frequently endorsed one, PHQ-8 and PHQ-9 scores are nearly identical, as are severity cut-points⁷.

The PHQ-8 is sometimes used in studies where depression is a secondary outcome and not the focus of the investigation, in population-based studies where interviews are administered by non-mental health professionals, or in clinical settings where patient-reported outcomes (PROs) are captured outside of an office visit, causing delays in clarifying positive responses to the ninth item.

Companion scales evaluate common fellow travelers of depression. The P4 is a 4-item measure that evaluates suicidal ideation in individuals who endorse the ninth item of the PHQ-9⁶. The Generalized Anxiety Disorder 7-item (GAD-7) measures anxiety symptoms that co-occur in a third to half of patients with depression. Although initially developed for generalized anxiety disorder, the GAD-7 is also an effective screener for panic, social anxiety, and post-traumatic stress disorders⁴. The PHQ-15 and its abbreviated version (the Somatic Symptom Scale-8, SSS-8) assess the presence and severity of physical symptoms that are the com-

plaints with which depressed patients most frequently present, and may denote concurrent somatic symptom disorder and other somatizing conditions⁸. Finally, the PHQ-4 consists of the PHQ-2 and the GAD-2 (abbreviated version of the GAD-7) and serves as an ultra-brief screener for depression and anxiety as well as general psychological distress. The PHQ family of scales, including many translations, are available at www.phqscreeners.com.

Practical issues still constrain use of depression and other PRO measures in some clinical settings. Routine administration by the clinician or ancillary staff and manual entry of scores into the health records require time that is typically unreimbursed. The PHQ-9 and other PROs generally do not require an interview but rather can be self-administered using a variety of modes (e.g., paper or web-based forms, iPads, apps) before an office visit or while at home. Completed PROs can then be electronically imported or scanned into the records.

Whereas universal depression screening is advocated by some guidelines, the optimal frequency of screening is not established. One approach is to screen all new patients and then annually in established patients. Because screening every patient at every visit is excessive, reminders of when screening is due are required.

Another key role of depression measures is to monitor outcomes in response to active treatment of depression or, in some cases, watchful waiting. Again, flagging which patients require follow-up PHQ-9 administration must be operationalized.

One critique of scales like the PHQ-9 is that depression is not simply a number. Certainly, a depression score alone should not generate a reflexive depression diagnosis or antidepressant prescription, but requires clinical evaluation to determine if the threshold for clinical action has been reached. The length of time symptoms have been present, the degree of functional impairment, and patient treatment preferences, combined with the severity of symptoms as denoted by the depression score, collectively inform treatment decisions, be it psychotherapy, medications, or watchful waiting.

When following depression longitudinally, it is useful to couple the PHQ-9 score with a question about global change: “Are your symptoms the same, better, or worse?”. Discordance between the depression score and global impression of change may have several explanations, including residual somatic symptoms such as insomnia or fatigue; co-occurring symptoms such as anxiety or pain; other medical or psychiatric comorbidity; stress or interpersonal factors; or a lag in functional improvement.

Is a universal depression measure necessary? The PHQ-9 has generally been shown to be similar or superior in performance to competing depression scales, including in special populations such as older adults, adolescents, pregnant or postpartum women, diverse racial/ethnic groups, patients with various medical and psychiatric diseases, and across clinical settings. Nonetheless, a number of depression scales are available and have their proponents, and methods for cross-walking depression scores

across measures are increasingly available⁹.

Incorporating PROs into practice is less about the specific measure (presuming it is well validated) than the act of measuring; it is more about the verb than the noun. On the other hand, using a common measure may facilitate communication across clinical settings and avoid the Tower of Babel phenomenon wherein different “languages” (i.e., metrics) are used for the same condition.

Uptake of the PHQ-9 in the past two decades has paralleled the increasing recognition of depression as an international public health priority, and the discovery that measurement is the first step towards detection and improved management. In the words of M. Chan, former director of the World Health Organization, “accountability means counting; what gets measured gets done”. Ditto for depression.

Kurt Kroenke

Indiana University School of Medicine and Regenstrief Institute, Indianapolis, IN, USA

1. Kroenke K, Unützer J. *J Gen Intern Med* 2017;32:404-10.
2. Mitchell AJ, Yadegarfar M, Gill J et al. *BJPsych Open* 2016;2:127-38.
3. Kroenke K, Monahan PO, Kean J. *J Clin Epidemiol* 2015;68:1085-92.
4. Kroenke K, Spitzer RL, Williams JB et al. *Gen Hosp Psychiatry* 2010;32:345-59.
5. Simon GE, Coleman KJ, Rossom RC et al. *J Clin Psychiatry* 2016;77:221-7.
6. Dube P, Kroenke K, Bair MJ et al. *Prim Care Companion J Clin Psychiatry* 2010;12:e1-8.
7. Wu Y, Levis B, Riehm KE et al. *Psychol Med* 2020;50:1368-80.
8. Toussaint A, Kroenke K, Baye F et al. *J Psychosom Res* 2017;101:44-50.
9. Wahl I, Lowe B, Bjorner JB et al. *J Clin Epidemiol* 2014;67:73-86.

DOI:10.1002/wps.20821