LETTER TO THE EDITOR

PHQ-9 and PHQ-2 in Western Kenya

The Authors Reply: Coyne and colleagues cite one meta-analysis of diagnostic validation studies but proffer a more negative interpretation than the authors who concluded that “the PHQ-9 has good diagnostic properties, and was able to correctly diagnose major depression (sensitivity 92%) while being able to exclude this condition with some certainty (specificity 80%)”¹. A second PHQ-9 meta-analysis found 77% sensitivity and 94% specificity². In a review of case-finding instruments in primary care, the PHQ-9 had operating characteristics comparable to longer depression measures³. Other features of the PHQ-9 have contributed to its popularity, including its brevity, its focus on the nine core symptoms of DSM-IV depressive disorders, its sensitivity to change, its robust performance across different race/ethnic groups, its translation into more than 70 languages, and its nonproprietary nature.

We agree with the authors on several points. First, by using the PHQ-9 diagnostic algorithm to determine sensitivity and specificity, the operating characteristics of the PHQ-2 in our particular study may have been inflated. It should be noted, however, that agreement even among mental health professionals using criterion standard diagnostic interviews is only modest⁴. Second, we agree that the use of the PHQ-9 or any depression measure must be coupled with a clinical interview to confirm diagnosis of a depressive disorder and to assess severity, duration, and functional impairment. Third, we agree that depression detection is warranted only if systems are in place to assure effective treatment and follow-up⁵. However, this is no different than any other chronic disease: measuring serum glucose, blood pressure, or lipids in the absence of effective diagnosis, treatment and follow-up would be a fruitless endeavor. Our AMPATH clinics in western Kenya provide comprehensive services. Fourth, we concur that ultra-brief measures like the PHQ-2 should not be used in isolation: a high score on the PHQ-2 should trigger completion of the full PHQ-9 as well as a clinical interview.

While PHQ-9 test-retest reliability (0.59) was moderate, we stated this “may be acceptable” because two major events occurring during the interval could have slightly affected “true”

depression scores for some participants—a tumultuous national two-day holiday to allow travel to a national election with several reports of civil unrest—and some (not all) participants attended a support group therapy session.

Finally, we acknowledged that future psychometric studies are needed in this population, including comparisons of PHQ-9 and PHQ-2 with gold standard diagnostic interviews. However, given that our study showed construct validity with general health ratings, content validity in which focus groups indicated the PHQ-9 was generally well understood but might benefit from two potential minor modifications to its instructions, high factor loadings, a pattern of item means similar to US validation samples, high internal consistency, and moderate test-retest reliability, we believe our conclusion that the PHQ-9 and PHQ-2 “appear to be” valid and reliable is warranted.

Patrick O. Monahan, PhD, *Indiana University School of Medicine, Indianapolis, IN 46202, USA (e-mail: pmonahan@iupui.edu); Kurt Kroenke, MD*, *Regenstrief Institute for Health Care and Indiana University School of Medicine, Indianapolis, IN 46202, USA (e-mail: kkroenke@iupui.edu).*

REFERENCES

1. **Gilbody S, Richards D, Brealey S, Hewitt C.** Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. *J Gen Intern Med.* 2007;22:1596-602.
2. **Wittkamp KA, Naeije L, Schene AH, Huyser J, van Weert HC.** Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. *Gen Hosp Psychiatry.* 2007;29:388-95.
3. **Williams JW Jr., Pignone M, Ramirez G, Perez Stellato C.** Identifying depression in primary care: a literature synthesis of case-finding instruments. *Gen Hosp Psychiatry.* 2002;24:225-37.
4. **Williams JBW, Gibbon M, First MB, et al.** The Structured Clinical Interview for DSM-III-R (SCID). II. Multisite test-retest reliability. *Arch Gen Psychiatry.* 1992;49:630-6.
5. **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.

J Gen Intern Med 24(7):891

DOI: 10.1007/s11606-009-0986-9

© Society of General Internal Medicine 2009

Published online April 25, 2009