

## EDITORIAL

# Close, but not close enough: Brazilian norms for the Patient Health Questionnaire (PHQ-9)

Yuan-Pang Wang,<sup>1</sup>  Clarice Gorenstein<sup>1,2</sup> 

<sup>1</sup>Instituto e Departamento de Psiquiatria, Faculdade de Medicina, Universidade de São Paulo, São Paulo, SP, Brazil. <sup>2</sup>Instituto de Ciências Biomédicas, Departamento de Farmacologia, Universidade de São Paulo, São Paulo, SP, Brazil.

Determining where to set the boundary between normal and abnormal behavior is an important focus of psychiatry. Damiano et al.<sup>1</sup> just proposed normative cut-offs for the Patient Health Questionnaire-9 (PHQ-9) for the general population, having derived four severity categories of depressive symptoms from sophisticated psychometric analyses. This is an important update regarding depressive morbidities among Brazilian adults. Nevertheless, some of these recommendations should be refined after careful consideration.

Depression is a major burden for the healthcare system worldwide, and general practitioners provide most care for people with depression. Although the self-reported PHQ-9 uses agreed-upon criteria to measure depressive symptom severity in primary care, we should consider several points before implementing it in clinical practice.

First, different levels of depressive symptomatology could be used to categorize severity and indicate treatment type. PHQ-9 scores were originally divided into the following user-friendly categories of increasing severity: 0-4, 5-9, 10-14, 15-19, and  $\geq 20$  (i.e., regular intervals of 5 points).<sup>2</sup> These categories were chosen for pragmatic and empirical reasons. These categories are a mnemonic aid for clinicians and present associations between increasing severity and measures of validity. However, cut-offs should be based on score distributions in the sample, comparing the patient's score to national averages. Most studies consider scores  $\geq 10$  as the intervention threshold.<sup>3</sup> Additional discriminant validity studies could determine the utility of 4-category systems over the traditional dichotomic threshold.

Second, this one-size-fits-all framework has its shortcomings in terms of applicability. Although the raw scores were from the 2019 Brazilian National Health Survey, a large non-institutionalized sample of adults ( $n = 90,846$ ), they should be weighted for the demographic features of the national population to adjust for potential confounders, sampling error, and differential probability of participation. The fact that the data were collected at the respondents' residence might explain the lower scores. Further limitations should be noted regarding people not included in the

survey (e.g., homeless, hospitalized, institutionalized, or incarcerated people and rural dwellers).

In addition to sampling issues, local prevalence variations also require further examination. A normative cut-off point is subject to social, demographic, and cultural heterogeneity in Brazil due to the country's large national territory. For example, according to the  $\geq 10$  cut-off point, the prevalence of "depression" was as high as 19.1% in a 2013-2014 survey in the Amazonian region<sup>4</sup> and as low as 4.1% in the 2013 National Health Survey.<sup>5</sup> Thus, the PHQ-9's ability to identify or rule out major depression seems to vary considerably according to the context.

Third, using rank percentile methods, such as T- and D-scores, to generate categories is also subject to criticism. The width between item categories is not the same, as in the original PHQ-9 scoring intervals. Thus, averaging raw scores into means and dispersion indicators is not recommended. Uneven intervals among categories can lead to misinterpretation of their severity, jeopardizing the accuracy and variability of the results, since the difference between different percentiles may not be equivalent. Rank percentiles fail to account for underlying distributional characteristics of the data, such as skewness or kurtosis. Misrepresentation of non-normal data distribution is a common limitation of percentile ranks, which are insensitive to the true distribution. Typically, population-based data are left-skewed. Alternate methods, such as using cut-off scores based on established criteria, or clustering techniques could generate meaningful categories that are accurate for data interpretation.

Fourth, a test's accuracy is crucial in any screening program. The lack of criterion validity for sensitivity and specificity estimates is a major shortcoming. The PHQ-9 has been widely validated in two-stage screening processes. Nevertheless, in samples where the PHQ-9 is highly specific at the standard cut-off, the results tend to be sub-optimal. When the PHQ-9 is excessively sensitive, authors often prefer results based on higher cut-offs. Consequently, sensitivity would increase with higher cut-off points. In other words, if the purpose is to identify only truly depressive participants, a higher specific cut-off

should be used. Most researchers regard flexible cut-offs and longitudinal studies as necessary to provide evidence of long-term screening effectiveness, rather than one-off assessments or medical records.

Practitioners should be mindful of the meaning of a cut-off when choosing a threshold for primary care. General practitioners who use these norms should compare their patient's PHQ-9 score to that of the general population: a reference point for symptom severity. Many complaints are merely transient manifestations of life dissatisfaction, grief, unemployment, or interpersonal conflict. While most complaints will resolve themselves, some become aggravated and require specialized attention. With a lower cutoff (e.g.,  $\geq 7$ ), fewer truly depressed individuals will be missed (false-negatives), although more individuals without depression will screen positive (false-positives). This trade-off should guide public health decisions.

There has been great interest in finding depression cases, particularly in primary care, since many patients do not complain of depressive symptoms to clinicians. Analysis of pooled data from 50,371 patients in 41 primary care studies revealed that general practitioners were capable of ruling out depression in most people who are not depressed,<sup>3</sup> with substantial misidentification outnumbering missed cases. Experts disagree with routine screening for depression in all settings. Their skepticism is due to a lack of evidence from well-conducted randomized controlled trials that it provides any benefit, as well as concerns about high false-positive rates, overdiagnosis, resource use, and the adverse

impact on patients who are screened and treated but do not improve.

Considering that these categorical cut-off points have not been fully validated in the Brazilian population, clinicians should use them with caution in the absence of sufficient empirical data. Due to overclassification, psychiatric professionals could be accused of labeling people with a disease they can't treat. The proposed severity categories for PHQ-9 await fair scrutiny to avoid an overflow of treatment-seeking in primary care.

## Disclosure

The authors report no conflicts of interest.

## References

- 1 Damiano RF, Hoffmann MS, Gosmann NP, Pan PM, Miguel EC, Salum GA. Translating measurement into practice: Brazilian norms for the Patient Health Questionnaire (PHQ-9) for assessing depressive symptoms. *Braz J Psychiatry*. 2023;45:310-17.
- 2 Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16:606-13.
- 3 Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. *Lancet*. 2009;374:609-19.
- 4 Santos ER, Huang H, Menezes PR, Sczufca M. Prevalence of depression and depression care for populations registered in primary care in two remote cities in the Brazilian Amazon. *PLoS One*. 2016;11:e0150046.
- 5 Munhoz TN, Nunes BP, Wehrmeister FC, Santos IS, Matijasevich A. A nationwide population-based study of depression in Brazil. *J Affect Disord*. 2016;192:226-33.