OPEN

Accuracy of Alternative PHQ-9 Scoring Algorithms to Screen for Depression in People Living With HIV in Sub-Saharan Africa

Charlotte Bernard, PhD,[a] Hélène Font, PhD,[a] Natalia Zotova, PhD,[b] Kara Wools-Kaloustian, MD,[c] Suzanne Goodrich, MD,[c] Edith Kamaru Kwobah, PhD,[d] Ajeh Rogers Awoh, PhD,[e] Guy Calvin Nko'o Mbongo'o, MD,[f,g] Dominique Mahambu Nsonde, MD,[h] Paul Gandou, MD,[h] Albert Minga, MD,[i] Judicaël Malick Tine, MD,[j] Ibrahima Ndiaye, MD,[k] François Dabis, MD, PhD,[a] Moussa Seydi, MD, PhD,[j] Nathalie de Rekeneire, MD, PhD,[l] Marcel Yotebieng, PhD,[b] and Antoine Jaquet, MD, PhD,[a] on behalf of the IeDEA Cohort Collaboration

Received for publication February 29, 2024; accepted August 27, 2024.

From the [a]University of Bordeaux, National Institute for Health and Medical Research (INSERM) UMR 1219, Research Institute for Sustainable Development (IRD) EMR 271, Bordeaux Population Health Centre, Bordeaux, France; [b]Department of Medicine, Albert Einstein College of Medicine, Bronx, NY; [c]Division of Infectious Diseases, Department of Medicine, Indiana University School of Medicine, Indianapolis, IN; [d]Department of Mental Health, Moi Teaching and Referral Hospital and Moi University, Eldoret, Kenya; [e]Clinical Research Education and Consultancy, Yaoundé, Cameroon; [f]Department of Psychiatry, Jamot Hospital, Yaoundé, Cameroon; [g]Faculty of Medicine and Pharmaceutical Sciences of Dschang, Dschang, Cameroon; [h]Centre de Traitement Ambulatoire, Brazzaville, Republic of Congo; [i]Centre Médical de Suivi des Donneurs de Sang (CNTS-CI), Abidjan, Côte D'Ivoire; [j]Service de Maladies Infectieuses et Tropicales, CHNU de Fann, Dakar, Senegal; [k]Service de Psychiatrie, CHNU de Fann, Dakar, Senegal; and [l]Institut Pasteur du Cambodge, Phnom Penh, Cambodia.

The International Epidemiology Databases to Evaluate AIDS (IeDEA) are supported by the US National Institutes of Health's National Institute of Allergy and Infectious Diseases, the Eunice Kennedy Shriver National Institute of Child Health and Human Development, the National Cancer Institute, the National Institute of Mental Health, the National Institute on Drug Abuse, the National Heart, Lung, and Blood Institute, the National Institute on Alcohol Abuse and Alcoholism, the National Institute of Diabetes and Digestive and Kidney Diseases, and the Fogarty International Center: Central Africa, U01AI096299; East Africa, U01AI069911; West Africa, U01AI069919. Informatics resources are supported by the Harmonist Project, R24AI24872.

Meetings: 25th International Workshop on HIV and Hepatitis Observational Databases (IWHOD), Athènes, Grèce, 23–25 Mars 2023; 12th IAS Conference, Brisbane, Australia, 23–26 Juillet 2023.

The authors have no conflicts of interest to disclose.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.jaids.com).

Correspondence to: Charlotte Bernard, PhD, Global Health in the Global South Team, University of Bordeaux, National Institute for Health and Medical Research (INSERM) UMR 1219, Research Institute for Sustainable Development (IRD) EMR 271, Bordeaux Population Health, 146 rue Léo Saignat, 33076 BORDEAUX Cedex, France (e-mail: charlotte.bernard@u-bordeaux.fr).

Background: Screening for depression remains a priority for people living with HIV (PLWH) accessing care. The 9-item Patient Health Questionnaire (PHQ-9) is a widely used depression screening tool, but has limited accuracy when applied across various cultural contexts. We aimed to evaluate the performance of alternative PHQ-9 scoring algorithms in sub-Saharan African PLWH.

Setting: Five HIV programs in Cameroon, Côte d'Ivoire, Kenya, Senegal, and the Republic of Congo.

Methods: Adult PLWH were screened for depression during the 2018–2022 period. Diagnosis confirmation was done by psychiatrist blinded clinical evaluation (gold standard). Diagnostic performances, including sensitivity and area under the curve (AUC) of the traditional PHQ-9 scoring (positive screening − score ≥ 10), were compared to alternative scoring algorithms including (1) the presence of ≥1 mood symptom (PHQ-9 items 1 and 2) combined with ≥2 other symptoms listed in the PHQ-9, and (2) a simplified recoding of each 4-response item into 2 categories (absence/presence).

Results: A total of 735 participants were included [54% women, median age 42 years (interquartile range 34–50)]. Depression was diagnosed by a psychiatrist in 95 (13%) participants. Alternative scoring sensitivities (0.59–0.74) were higher than that of the traditional score's (0.39). Compared to traditional scoring, AUC was significantly higher for PHQ-9 alternative scoring. Across settings, alternative scoring algorithms increased sensitivity and reduced variability.

Conclusions: As a primary screening test, new scoring algorithms seemed to improve the PHQ-9 sensitivity in identifying depression and reducing heterogeneity across settings. This alternative might be considered to identify PLWH in need of referral for further diagnostic evaluations.

Key Words: HIV, depression, screening tool, PHQ-9, Sub-Saharan Africa

(*J Acquir Immune Defic Syndr* 2025;98:143–149)

INTRODUCTION

Depression is the most common neuropsychiatric disorder in people living with HIV (PLWH).[1] According to

recent meta-analyses in sub-Saharan Africa (SSA), its prevalence is high (13%–14%).[2,3] Depression has harmful consequences on the daily life and health of patients,[4,5] especially compromising antiretroviral (ART) adherence and the achievement of the 95-95-95 targets toward ending the HIV epidemic.[6] Screening and access to care for depression must be prioritized in the HIV/AIDS response in SSA.[1,6] As of today, systematic screening for depression, a critical first step in accessing diagnosis and treatment, is limited in PLWH and faces several challenges, including the lack of mental health specialists and time constraints reported by primary care providers.[7,8] There is, therefore, a need for validated rapid screening tools that can be easily administered by nonmental health professionals. Several brief or full depression screening scales have been validated in low- and middle-income countries (LMICs), including in HIV care settings in SSA, with varying levels of reliability and validity and high heterogeneity among studies, as reported in meta-analyses.[2,9,10]

The 9-item Patient Health Questionnaire (PHQ-9) is widely used in the context of HIV disease, and has a design that makes it easier to use in routine settings.[11,12] It consists of 9 items assessing key symptoms of depression occurring in the past 2 weeks, according to the DSM-IV criteria.[12] Each symptom frequency is rated from 0 to 3 in the past 2 weeks, with a total score ranging from 0 to 27. Kroenke et al recommended a score of 10 or higher to indicate moderate or severe depressive symptoms,[12] and this cutoff has been commonly used in PLWH in SSA to date.[11,13–17] Kroenke et al[12] also reported a categorical algorithm to screen for major depressive disorder (MDD). According to this algorithm, MDD is considered present if at least 5 symptoms are present on "more than half of the days" in the previous 2 weeks; one of these symptoms is a mood symptom (either depressed mood or anhedonia); and PHQ-9–item 9 ("Thoughts that you would be better off dead or of hurting yourself in some way") is considered regardless of its frequency. Although this categorical algorithm has been used in various populations, no data has evaluated its performance in the context of HIV infection and in SSA.[18]

Previous studies have assessed the PHQ-9 performance using the traditional cutoff score of 10 in PLWH in SSA with mixed results and using different gold standard. Sensitivity varied from 0.27 to 0.92, while specificity varied from 0.71 to 0.94.[11,13–17] Finally, for the categorical algorithm in the non-HIV population, performance is lower and heterogeneous depending on the reference standard used (sensitivity: 0.35 to 0.57 and specificity: 0.95 to 0.97).[18]

To date, the reasons for the observed variability in PHQ-9 performance among PLWH in SSA have been poorly investigated. One reason may be differences in the understanding and expression of mental health problems, including depression, from one culture to another.[19–22] Another reason, more specific to chronic diseases such as HIV infection, may be the overlap of symptoms between the 2 conditions, because people tend to attribute their depressive symptoms to the symptoms of their underlying chronic disease and its progression.[23,24] As this PHQ-9 remains an acceptable and widely used tool in routine clinical practice, we considered an alternative scoring algorithm to improve its accuracy while preserving the content of the questionnaire itself.

We proposed 2 adaptations to improve PHQ-9 performances. First, we hypothesized there may be difficulties in understanding the response categories to a given item. Although many studies of depression scale adaptation in Africa have found that some items should be dropped because they are poorly understood, the impact of number of response options has been poorly explored.[25,26] A recent qualitative study conducted in Cameroon based on in-depth and cognitive interviews reported challenges to understand and sequence the 4 available answer categories.[23] So recoding scores into 2 categories (presence or absence of symptoms) might improve test performance. This approach is also in line with the one used for diagnosis (eg, in the MhGAP depression module).

Second, we considered that not all items of the PHQ-9 should have similar weight as diagnostic criteria for depression. According to the most recent version of the Diagnostic and Statistical Manual of Mental Disorders-5 (DSM-5),[27] major depressive disorder (MDD) is characterized by 2 primary diagnostic criteria: depressed mood and loss of interest or pleasure in activities (anhedonia) and other secondary diagnostic criteria. For this reason, we believe that retaining the presence of at least 1 mood-related symptom (depressed mood and/or loss of interest in everyday things) to detect depression should be used as a central element to improve test performance. With the alternative scoring algorithm, we have sought to respect and reproduce the hierarchy of symptoms to be taken into account when calculating the score. Based on these proposed adaptations, the main objective of this study was to evaluate the performances of the PHQ-9 using alternative scoring guidelines, to screen for depression in PLWH across 5 SSA countries.

METHODS

Design, Setting, and Population

A cross-sectional design was used to assess sensitivity, specificity, and predictive values of various PHQ-9 scoring algorithms, pooling data from 2 similar PHQ-9 validation studies conducted in 5 countries in the west, central, and east African regions contributing to the International Epidemiology Databases to Evaluate AIDS (IeDEA) consortium https://www.iedea.org/. Data were collected from January 2018 to July 2022 across 5 sites. In west Africa, patients were enrolled at the National Blood Transfusion Center (Centre National de Transfusion Sanguine Côte d'Ivoire) (CNTS-CI) in Abidjan, Côte d'Ivoire, and at the Infectious and Tropical Diseases Department of the National University Hospital Center (CHNU) of Fann in Dakar, Senegal. In central Africa, patients were enrolled at the Outpatient Treatment Centre, in the Republic of Congo and at the Yaoundé Jamot Hospital, in Cameroon. In east Africa, patients were enrolled in the Academic Model Providing Access to Healthcare (AMPATH) clinic at Moi Teaching and Referral Hospital in Eldoret, Kenya. These contributing sites were all located in urban areas, with nearby mental health services. Depression screening was systematically proposed during routine visits.

Each site has a large cohort of actively followed PLWH, who were included through their routine HIV visits if they were 18 years of age or older. Exclusion criteria were not being fluent in any of the languages in which the questionnaire was translated (English, French, Lingala, Kituba, Luo, and Swahili), being unable to give consent, or refusing to participate in the study. Each participant was compensated for their time and transport to the clinic.

Study Procedures

During their routine HIV clinical follow-up visit, patients were consecutively recruited during the study period and offered screening for depression. Trained research assistants or social workers administered the PHQ-9. After completing the questionnaire, patients were referred to a psychiatrist. Psychiatrists, masked to the PHQ-9 responses, independently evaluated patients for depressive disorders. Most assessments were completed on the same day. Psychiatrists were encouraged to use the MINI to guide the interview. Psychiatrists were provided with a separate diagnosis form to complete at the end of the participant visit, which was returned to the study team.

PHQ-9

The PHQ-9 is based on the diagnostic criteria of the DSM-IV for depressive disorders. It consists of 9 questions evaluating the severity of depressive symptoms. The first 2 items focus on mood symptoms (anhedonia and depressed mood). The last item focuses on suicidality. For each patient, the frequency of each symptom is evaluated for the past 2 weeks and then rated on a 4-point Likert scale including the following responses: "Not at all," "Several days," "More than half the days," and "Nearly every day". The PHQ-9 was administered in English, Canadian French, Swahili, Lingala, or Kituba depending on the country. Versions were retrieved from the Pfizer Patient Health Questionnaires website,[28] except for Lingala and Kituba that were available locally.[29] Definitions of traditional scores and alternative scorings are presented in Table 1.

Each PHQ-9 item score ranges from 0 ("Not at all") to 3 ("Nearly every day"). The total score, obtained by summing up the scores on the individual items, can range from 0 to 27 and indicates the severity of depressive symptoms. Depressive symptoms are generally defined as mild (PHQ-9 scores: 5 to 9), moderate (10–14), moderately severe (15–19), or severe (≥20). With the categorical algorithm, the depression screening is positive if (1) 5 or more of the 9 depressive symptom criteria are endorsed as having been present at least "more than half the days" and (2) one of the endorsed symptoms was anhedonia or depressed mood. If the response to the last item on suicidality is positive at all, the depression screening was considered positive.

At the end of the PHQ-9, after asking the 9 items, the impact of the severity of depressive symptoms is evaluated with a specific question (hereafter "difficulties"): "if checked off any problems, how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?". The patient rated the level of difficulties on a 4-point Likert scale (from no difficulty to extremely difficult).

PHQ-9 Alternative Scoring

We considered 2 alternatives, after recoding each item into 2 categories (absence or presence). Depression screening was considered as positive when:

- For the alternative scoring 1: at least 5 depressive symptom criteria were present, of which at least 1 was depressed mood or anhedonia (PHQ-9 item 1 or item 2);
- For the alternative scoring 2: at least 1 of the 2 first symptoms was present depressed mood or anhedonia and at least 2 additional items.

We also assessed the performance of 2 other alternative scoring, named "Alternative scoring 1 + difficulties" and "Alternative scoring 2 + difficulties" evaluating the impact of taking the "difficulties" item into account on PHQ-9 performance. For that, we added the presence of the item "difficulties" to alternative scorings 1 and 2, after recoding the responses into 2 categories (absence or presence).

Statistical Analyses

Patient characteristics were described using medians and their interquartile range (IQR) for numerical variables, and frequency (percentage) for categorical variables. The description was done for the combined data and by country.

TABLE 1. Traditional and Alternative PHQ-9 Scores Used in the Present Study

	PHQ-9 Scores
Traditional scores	
Cutoff ≥10	Total score ≥10
Algorithm	At least 5 items with a score ≥ 2 of which at least 1 is item 1 or 2. Item 9 is counted as present as soon as its value is > 0
Alternative scoring	
Alternative scoring 1	At least 5 items present of which at least 1 was item 1 or 2
Alternative scoring 2	At least 1 of the 2 first symptoms is present and at least 2 additional items
Alternative scoring 1 + item "difficulty	At least 5 items present of which at least 1 was item 1 or 2 + presence of difficulties
Alternative scoring 2 + item "difficulty"	At least 1 of the 2 first symptoms is present and at least 2 additional items + presence of difficulties

Country differences were investigated using Pearson χ^2 test for categorical variables and the Kruskal–Wallis rank sum test for numerical variables. The 9 items were summarized using a Likert plot. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated for the different scores using the psychiatrist's diagnosis as the gold standard. Results were presented overall and by countries. Receiver operating characteristic (ROC) curve analysis was performed only for the full sample and the area under the ROC curve (AUC) was assessed. AUC was compared with DeLong test. All estimates were given with 95% confidence intervals. Data analyses were performed using R version 4.3.0.

Ethics

The research was conducted in accordance with the Helsinki Declaration. Ethical approval was obtained from ethics committees from each country participating in the study: the National Ethics Committee for Health and Life Sciences in Côte d'Ivoire (approval number: 013-21/MSHP/CNESVS-km); the National Ethics Committee for Health Research in Senegal (approval number: 00036/MAS/CNERS/SP); the Ohio State University Institutional Review Board (Study 2018H0195); the Regional Ethics Committee for Research on Human Health in Cameroon (approval number: 1846/C'RERSHC/2019); Moi University College of Health Sciences and Moi Teaching and Referral Hospital's Institutional Research and Ethics Committee in Eldoret, Kenya (approval number: IREC/2019/57); and the *Comité d'Ethique de la Recherche* en *Sciences de la Santé* in the Republic of Congo (approval number: 017/CEYFCRM/20 18). All enrolled participants provided written consent.

RESULTS

Participants' Characteristics

A total of 735 participants were enrolled (Cameroon (n = 295), Côte d'Ivoire (n = 148), Kenya (n = 98), Senegal (n = 85), and the Republic of Congo (n = 109)) (Table 2). The median age was 42 years (IQR 35–48) in Cameroon, 48 years (IQR 40–56) in Côte d'Ivoire, 36 years (IQR 29–43) in Kenya, 46 years (IQR 33–55) in Senegal, and 40 years (IQR 32–47) in the Republic of Congo. Half or more of the patients were women in the 5 participating countries. Participants completing primary education or less were 35% (n = 94) in Cameroon, 41% (n = 61) in Côte d'Ivoire, 41% (n = 38) in Kenya, 51% (n = 51) in Senegal, and 18% (n = 19) in the Republic of Congo. Overall, 6.4% (n = 19) in Cameroun, 11% (n = 17) in Côte d'Ivoire, 34% (n = 33) in Kenya, 12% (n = 10) in Senegal, and 15% (n = 16) in the Republic of Congo were diagnosed with depression by a psychiatrist. Among mood symptoms, the most commonly reported symptom by PLWH diagnosed with depression was item 2 "Feeling down, depressed, or hopeless": 72% reported that this symptom occurred several days to nearly every day (compared with 56% for item 1 "Little interest or pleasure in doing things" (see Fig. S1, Supplemental Digital Content, http://links.lww.com/QAI/C380). Among the other symptoms, the 3 less reported were concentration problems (item 7), speech slowing (item 8), and suicidal ideation (item 9) (42%, 38%, and 25%, respectively). For each symptom, most patients screening positive for depression reported their frequency of symptoms as "several days," but less frequently symptoms were reported "more than half of the days" or "nearly every day."

Performance of the Different PHQ-9 Scoring Guidelines Against Psychiatric Diagnosis

Overall, traditional scoring of PHQ-9 resulted in low sensitivity. Using a cutoff score of ≥10, the sensitivity was 0.39 (95% CI: 0.29 to 0.49) and using the traditional algorithm, the sensitivity was 0.26 (95% CI: 0.18 to 0.36). Conversely, traditional scores showed high specificity of 0.95% (95% CI: 0.93 to 0.96) and 0.98 (95% CI: 0.96 to 0.99), respectively (Table3).

The sensitivity of the alternative scores was higher than the traditional scoring algorithm. For the alternative scoring 1, the sensitivity was 0.59 (95% CI: 0.48 to 0.69) and for the

TABLE 2. Participants Characteristics (n = 735)

Characteristics*	Overall	Cameroon	Côte d'Ivoire	Kenya	Senegal	The Republic of Congo	
N	735	295	148	98	85	109	*P*†
Age	42 (34, 50)	42 (35, 48)	48 (40, 56)	36 (29, 43)	46 (33, 55)	40 (32, 47)	<0.001
Unknown	1	0	0	0	1	0	
Gender							0.002
Female	394 (54%)	146 (49%)	74 (50%)	64 (66%)	40 (47%)	70 (66%)	
Unknown	4	0	0	1	0	3	
Formal education							
Primary or less	263 (38%)	94 (35%)	61 (41%)	38 (41%)	51 (60%)	19 (18%)	
Secondary	313 (45%)	141 (53%)	54 (36%)	38 (41%)	26 (31%)	54 (52%)	
Superior	120 (17%)	32 (12%)	33 (22%)	17 (18%)	8 (9.4%)	30 (29%)	
Unknown	39	28	0	5	0	6	

*n (%); Median (IQR).
†Pearson χ^2 test; Kruskal–Wallis rank sum test.

alternative scoring 2, the sensitivity was 0.74 (95% CI: 0.64 to 0.82). Specificity was lower than specificity of the traditional scoring algorithm [0.87 (95% CI: 0.84 to 0.90) and 0.71 (95% CI: 0.67 to 0.75) respectively]. For the 2 alternative scorings, PPV was low [0.41 (95% CI: 32% to 49%) and 0.27 (95% CI: 0.22 to 0.33)], whereas NPV was high [0.93 (95% CI: 0.91 to 0.95) and 0.95 (95% CI: 0.92 to 0.97)], respectively. The AUC was significantly higher for the alternative scorings than the traditional PHQ-9 score based on a cutoff score ≥10 [AUC alternative scoring 1: 0.73 (95% CI: 0.68 to 0.78); AUC alternative scoring 2: 0.72 (95% CI: 0.68 to 0.77); AUC cutoff score ≥10: 0.67 (95% CI: 0.62 to 0.72)] (Fig. 1).

Using the item "difficulties," the last question of the PHQ-9 did not improve PHQ-9 alternative scorings 1 and 2 sensitivity [0.52 (95% CI: 0.41 to 0.62) and 0.64 (95% CI: 0.54 to 0.74), respectively]. Across all settings, alternative scoring increased sensitivity and decreased heterogeneity. Indeed, for the traditional cutoff score ≥10, sensitivity varied between 0.10 and 0.71 across sites, whereas for alternative scoring 1, it varied between 0.47 and 0.76, and for alternative scoring 2, it varied between 0.60 and 0.88 (Fig. 2).

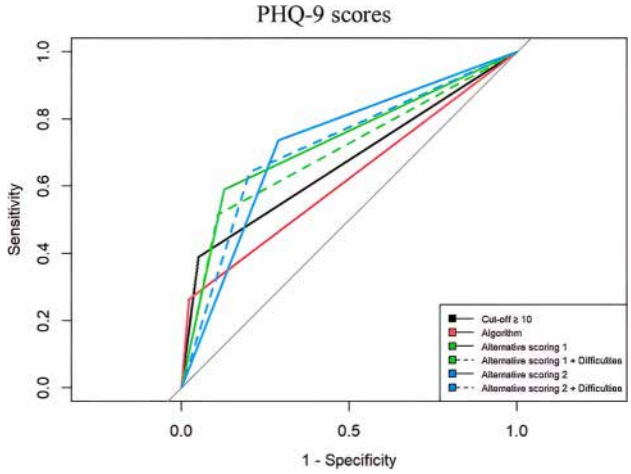

FIGURE 1. ROC curves for the PHQ-9 different scorings.

DISCUSSION

Using valid tools to screen for depression is urgently needed in HIV primary care facilities to assist in the identification of patients suffering from depression. In this multicountry study, we observed that alternative scoring algorithms improved the sensitivity of the PHQ-9 in identifying depression and seemed to increase reproducibility across different sites. Recoding should also facilitate administration of the PHQ-9 by simplifying the response scale.

As observed in previous publications, the traditional algorithm, described by Kroenke et al,[12] has poor sensitivity.[18] However, using the categorical algorithm with recoded item response categories (ie, presence/absence) increased the ability to identify cases. A study conducted in the Netherlands also found that the use of a modified categorical algorithm showed a significant improvement in sensitivity (from 28% to 84%).[30] The difficulty of understanding these item response categories supports the relevance of this simplification. Recently, Zotova et al[23] reported that most participants had difficulty understanding fully some of the response categories (ie, "half of the time") and had difficulty mapping their daily experiences and challenges to these categories. In addition, the importance of social desirability may influence the reported responses. For example, patients may tend to under-report the severity of their symptoms for fear of being judged.

The PHQ-9 also includes a question assessing functional impairment. In validation studies, the importance of this item is rarely evaluated, whereas in the DSM-5 diagnostic criteria for depression,[27] functional impairment is part of the diagnosis and may improve accuracy in detecting cases of depression. In this study, the inclusion of functional impairment did not improve accuracy. For the same reasons mentioned above (difficulties in understanding the items response categories and in reporting daily difficulties), the response to this question might be affected.

TABLE 3. Performance of the Different PHQ-9 Scoring Against the Psychiatric Diagnosis

	PHQ-9 Scores					
	Cutoff ≥ 10	**Algorithm**	**Alternative Scoring 1**	**Alternative Scoring 2**	**Alternative Scoring 1 + Item "Difficulties"**	**Alternative Scoring 2 + Item "Difficulties"**
Sensitivity	0.39 (0.29; 0.49)	0.26 (0.18; 0.36)	0.59 (0.48; 0.69)	0.74 (0.64; 0.82)	0.52 (0.41; 0.62)	0.64 (0.54; 0.74)
Specificity	0.95 (0.93; 0.96)	0.98 (0.96; 0.99)	0.87 (0.84; 0.90)	0.71 (0.67; 0.75)	0.89 (0.87; 0.92)	0.80 (0.76; 0.83)
Positive predictive value	0.53 (0.41; 0.65)	0.62 (0.46; 0.77)	0.41 (0.32; 0.49)	0.27 (0.22; 0.33)	0.42 (0.33; 0.51)	0.32 (0.25; 0.39)
Negative predictive value	0.91 (0.89; 0.93)	0.9 (0.87; 0.92)	0.93 (0.91; 0.95)	0.95 (0.92; 0.97)	0.93 (0.90; 0.94)	0.94 (0.91; 0.96)
Area under the curve*	0.67 (0.62; 0.72)	0.62 (0.58; 0.67)†	0.73 (0.68; 0.78)†	0.72 (0.68; 0.77)†	0.70 (0.65; 0.76)	0.72 (0.67; 0.77)

Reference: For PHQ-9 scores: AUC for cutoff ≥ 10.
*Comparison of AUC for the different PHQ-9 scores with a reference using for DeLong test.
†$P < 0.005$.

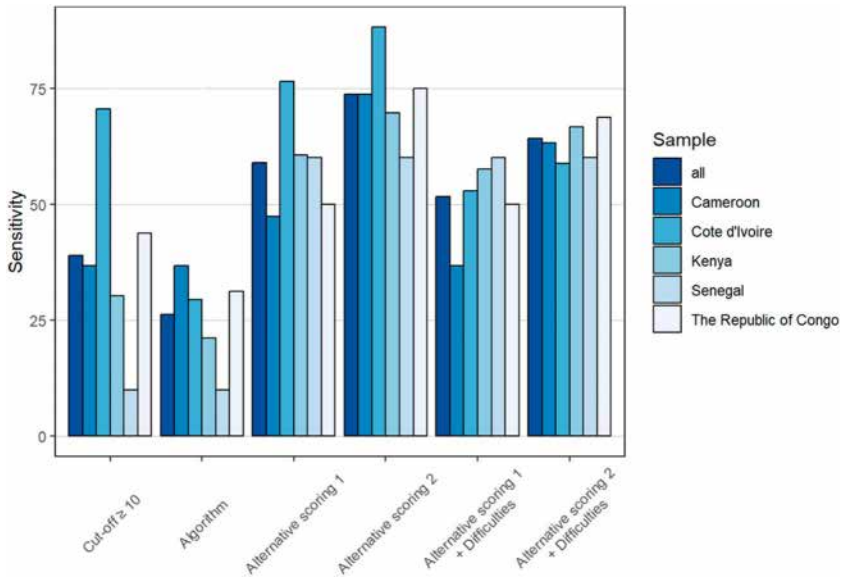

FIGURE 2. Sensitivity of the PHQ-9 different scorings across sites.

Another important aspect observed in the literature is the heterogeneity in the accuracy of the scale from one study to another, even in studies using the same protocol.[2,18,31] We observed that the use of alternative scoring reduced this heterogeneity. Because the PHQ-9 is used in different international consortia to assess the prevalence of depressive symptoms, it seems necessary to work with a tool and an algorithm that reduces this heterogeneity to facilitate comparisons.

As depression is a major public health problem, having a valid screening tool is an important first step. It could facilitate early identification, provide access to appropriate treatment, and reduce the negative impact of depression. But to effectively reach those who need help, we also need to increase awareness of depression, both among health professionals and patients. Reducing stigma and improving understanding of depression will help patients to open up and feel listened to by the care team.[23,32] The conditions under which depression screening takes place are also important: a place to ensure confidentiality and building patient trust. Training, support, and supervision are also important to ensure fidelity of implementation.[33]

The major strength of this study was the use of a common protocol at all participating sites, including a common reference standard for the diagnosis of depression based on a psychiatric assessment. However, the study also has several limitations. First, item recoding was done retrospectively. It would be interesting to directly test the performance of the PHQ-9 with directly modified answers to questions. Second, selection bias may have been introduced by excluding patients with missing data for the « difficulty » item. However, the 2 samples did not differ in terms of sociodemographic characteristics, and PHQ-9 performance was similar (data not shown). Finally, the PHQ-9 was originally designed to be self-administered. Interviewer administration could influence the reporting of depressive symptoms and their frequency. However, the alternative recoding algorithm may partially compensate for this problem.

CONCLUSIONS

As a short and easy to administer instrument, the PHQ-9 is a widely accepted tool for routine screening for depression symptoms. However, the traditional scoring has low sensitivity and heterogeneous performance across countries, making it difficult to use in international contexts. Simple alternative methods in scoring and interpreting the instrument improved the sensitivity of the PHQ-9 to identify patients in need of further care for depression across various settings in SSA. Recoding might also facilitate PHQ-9 administration by simplifying the response scale. This alternative PHQ-9 scoring approach could be considered to identify PLWH in need of referral for further diagnostic evaluations. Additional studies may be needed to validate our findings and assess the impact of these adjustments on access to care for mental health in PLWH.

ACKNOWLEDGMENTS

The International Epidemiology Databases to Evaluate AIDS (IeDEA) is supported by the US National Institutes of Health's National Institute of Allergy and Infectious Diseases; the Eunice Kennedy Shriver National Institute of Child Health and Human Development; the National Cancer Institute; the National Institute of Mental Health; the National Institute on Drug Abuse; the National Heart, Lung, and Blood Institute; the National Institute on Alcohol Abuse and Alcoholism; the National Institute of Diabetes and Digestive and Kidney Diseases; and the Fogarty International Center: Central Africa, U01AI096299; East Africa, U01AI069911; West Africa, U01AI069919. Informatics resources are supported by the Harmonist project,

R24AI24872. This work is solely the responsibility of the authors and does not necessarily represent the official views of any of the institutions mentioned above.

REFERENCES

1. Abas M, Ali GC, Nakimuli-Mpungu E, et al. Depression in people living with HIV in sub-Saharan Africa: time to act. *Trop Med Int Health*. 2014; 19:1392–1396.
2. Bernard C, Dabis F, de Rekeneire N. Prevalence and factors associated with depression in people living with HIV in sub-Saharan Africa: a systematic review and meta-analysis. Seedat S, ed. *PLoS One*. 2017;12: e0181960.
3. Bigna JJ, Tounouga DN, Kenne AM, et al. Epidemiology of depressive disorders in people living with HIV in Africa: a systematic review and meta-analysis. *Gen Hosp Psychiatry*. 2019;57:13–22.
4. Mayston R, Kinyanda E, Chishinga N, et al. Mental disorder and the outcome of HIV/AIDS in low-income and middle-income countries: a systematic review. *AIDS*. 2012;26(suppl 2):S117–S135.
5. Nanni MG, Caruso R, Mitchell AJ, et al. Depression in HIV infected patients: a review. *Curr Psychiatry Rep*. 2015;17:530.
6. Godfrey C, Nkengasong J. Prioritizing mental health in the HIV/AIDS response in Africa. *N Engl J Med*. 2023;389:581–583.
7. Parcesepe A, Bernard C, Agler R, et al. Mental health and HIV: research priorities related to the implementation and scale up of "treat all" in sub-Saharan Africa. *J Virus Erad*. 2018;4(suppl 2):16–25.
8. Parcesepe AM, Mugglin C, Nalugoda F, et al. Screening and management of mental health and substance use disorders in HIV treatment settings in low- and middle-income countries within the global IeDEA consortium. *J Int AIDS Soc*. 2018;21:e25101.
9. Tsai AC. Reliability and validity of depression assessment among persons with HIV in sub-saharan Africa: systematic review and meta-analysis. *J Acquired Immune Deficiency Syndromes*. 2014;66:503–511.
10. Akena D, Joska J, Obuku EA, et al. Comparing the accuracy of brief versus long depression screening instruments which have been validated in low and middle income countries: a systematic review. *BMC Psychiatry*. 2012;12:187.
11. Molebatsi K, Motlhatlhedi K, Wambua GN. The validity and reliability of the Patient Health Questionnaire-9 for screening depression in primary health care patients in Botswana. *BMC Psychiatry*. 2020;20:295.
12. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16:606–613.
13. Akena D, Joska J, Obuku EA, et al. Sensitivity and specificity of clinician administered screening instruments in detecting depression among HIV-positive individuals in Uganda. *AIDS Care*. 2013;25:1245–1252.
14. Monahan PO, Shacham E, Reece M, et al. Validity/reliability of PHQ-9 and PHQ-2 depression scales among adults living with HIV/AIDS in western Kenya. *J Gen Intern Med*. 2009;24:189–197.
15. Pence BW, Gaynes BN, Atashili J, et al. Validity of an interviewer-administered patient health questionnaire-9 to screen for depression in HIV-infected patients in Cameroon. *J Affective Disord*. 2012;143:208–213.
16. Cholera R, Gaynes BN, Pence BW, et al. Validity of the patient health questionnaire-9 to screen for depression in a high-HIV burden primary healthcare clinic in Johannesburg, South Africa. *J Affective Disord*. 2014; 167:160–166.
17. Chibanda D, Verhey R, Gibson LJ, et al. Validation of screening tools for depression and anxiety disorders in a primary care population with high HIV prevalence in Zimbabwe. *J Affective Disord*. 2016;198:50–55.
18. He C, Levis B, Riehm KE, et al. The accuracy of the patient health questionnaire-9 algorithm for screening to detect major depression: an individual participant data meta-analysis. *Psychother Psychosom*. 2020; 89:25–37.
19. Huang FY, Chung H, Kroenke K, et al. Using the patient health questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. *J Gen Intern Med*. 2006;21:547–552.
20. Sweetland AC, Belkin GS, Verdeli H. Measuring depression and anxiety in sub-saharan Africa. *Depress Anxiety*. 2014;31:223–232.
21. Ryder AG, Yang J, Zhu X, et al. The cultural shaping of depression: somatic symptoms in China, psychological symptoms in North America? *J Abnormal Psychol*. 2008;117:300–313.
22. Choi SKY, Boyle E, Burchell AN, et al. Validation of six short and ultra-short screening instruments for depression for people living with HIV in ontario: results from the ontario HIV treatment network cohort study. *PLoS One*. 2015;10:e0142706.
23. Zotova N, Watnick D, Awoh AR, et al. Understanding the PHQ-9 items and depression among people living with HIV: a multiple methods study in Yaoundé, Cameroon. *SSM Mental Health*. 2024;6:100353.
24. Bernard C, Mané I, Ziadeh S, et al. Implementation of group interpersonal therapy in two Senegalese HIV care units: a qualitative study. *Front Public Health*. 2024;12:1295181.
25. Bass JK, Ryder RW, Lammers MC, et al. Post-partum depression in Kinshasa, Democratic Republic of Congo: validation of a concept using a mixed-methods cross-cultural approach. *Trop Med Int Health*. 2008;13: 1534–1542.
26. Miller AP, Kintu M, Kiene SM. Challenges in measuring depression among Ugandan fisherfolk: a psychometric assessment of the Luganda version of the Center for Epidemiologic Studies Depression Scale (CES-D). *BMC Psychiatry*. 2020;20:45.
27. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders. DSM-5-TR*. Washington, DC: American Psychiatric Association Publishing; 2022.
28. Pfizer PHQ-9. Available at: https://www.phqscreeners.com/select-screener.
29. Yotebieng KA, Fokong K, Yotebieng M. Depression, retention in care, and uptake of PMTCT service in Kinshasa, the Democratic Republic of Congo: a prospective cohort. *AIDS Care*. 2017;29:285–289.
30. Zuithoff NP, Vergouwe Y, King M, et al. The Patient Health Questionnaire-9 for detection of major depressive disorder in primary care: consequences of current thresholds in a crosssectional study. *BMC Fam Pract*. 2010;11:98.
31. Yotebieng M, Zotova N, Bernard C, et al. Accuracy of PHQ-9 against psychiatric diagnosis for depression among people living with HIV: a cross-sectional study in Cameroon, Cote d'Ivoire, Kenya, Senegal, and the Republic of Congo. *AIDS*. 2024;38:1765–1773.
32. Gelaye B, Williams MA, Lemma S, et al. Validity of the patient health questionnaire-9 for depression screening and diagnosis in East Africa. *Psychiatry Res*. 2013;210:653–661.
33. Pence BW, Stockton MA, Mphonda SM, et al. How faithfully do HIV clinicians administer the PHQ-9 depression screening tool in high-volume, low-resource clinics? Results from a depression treatment integration project in Malawi. *Glob Ment Health (Camb)*. 2019;6:e21.