COMMENTARY ARTICLE

Validation of the patient health questionnaire-9 adaptation in english and swahili for use among adolescents in Kenya and Tanzania

Vinay Pathak

ABSTRACT

OBJECTIVES: In order to be used with teenagers in Kenya, our study sought to verify the Patient Health Questionnaire-9 in both English and Swahili. The Kiddie Schedule of Affective Disorders and Schizophrenia was used in diagnostic interviews conducted by clinicians to assess the validity of the criteria.

METHODS: A total of 250 teenagers between the ages of 10 and 19 were recruited, including 148 girls and 102 boys. All responders received the PHQ-9 in both Swahili and English at the same time. Clinicians then examined teenagers using the Kiddie Schedule of Affective Disorders and Schizophrenia to ascertain whether or not they were currently exhibiting signs of major depressive disorder. Using receiver operating characteristic curves, the sensitivity, specificity, Positive Predictive Value (PPV), negative predictive value, and likelihood ratios for various cut-off scores for the PHQ-9 were examined.

RESULTS: For the PHQ-9, the English version's internal consistency was 0.862, and the Swahili version's was 0.834. On the basis of receiver operating characteristic analysis, the areas under the curve for the Swahili and English versions, respectively, were 0.89 and 0.87. In the English-language version, a cut-off of 9 had a sensitivity of 95.0%, specificity of 73.2, Positive Predictive Value (PPV) of 0.23, and Negative Predictive Value (NPV) of 0.99; in the Swahili version, a cut-off of 9 had a sensitivity of 89.0%, specificity of 70.0, PPV of 0.20, and NPV of 0.90.

CONCLUSION: The PHQ-9's psychometric properties were comparable between its Swahili and English adaptations, indicating that it is a valid and reliable tool for identifying adolescent major depressive disorder and can be used in settings with limited resources to identify young people who need mental health support as early as possible.

Keywords: *Depression in children and adolescents; Developing nations; Psychometrics, screening; Validation*

INTRODUCTION

More than 40% of adolescents in the age range of 10 years to 19 years worldwide suffer from anxiety and depression disorders. Lack of country-level data on adolescent mental health is a major obstacle to prioritizing this crucial public health issue in low-income and middle-income nations. For instance, in Kenya, mental illnesses top the list of factors contributing to years spent living with a disability, and local research has revealed that the incidence of adolescent depression and anxiety in Kenya is a serious cause for concern.

However, methods that have been locally clinically validated and culturally adapted have not been utilized to estimate the burden of mental illnesses in Kenya; as a result, the current prevalence estimates do not take into account cultural characteristics of mental health. This runs the danger of either overestimating or underestimating the burden, which makes it more difficult to promote and expand services and initiatives that support teenage mental health. Implementing mental health interventions for teenagers in Kenya and many other LMICs is constrained by a lack of reliable data and a lack of corresponding policy devoted to child and adolescent mental health.

As a result, clinical validation of a questionnaire for diagnosing adolescent depression was carried out in Kenya in order to determine the psychometric qualities of both the local English dialect and the Swahili versions.

The objective was to gather information that would allow doctors, researchers studying public health, and decision-makers to understand the prevalence of depression at the population level. As an illustration, validation results will be helpful to identify a range of cut-offs to detect depression while minimizing false positives, limiting additional burdens on healthcare providers and potential stigmatization of respondents, as well as false negatives to ensure those in need of services can get them. Additionally, the simultaneous validation in Swahili and English would enable the use of both tools and changes for cut-offs based on language of completion, ensuring that the data are similar.

This study is a component of continuing efforts to conduct and support cultural adaptation and validation of adolescent mental health assessment tools being coordinated by the United Nations Children's Fund in conjunction with significant institutional and academic partners from Kenya. In order to coordinate efforts across nations, Unicef's Measurement of Mental Health among Adolescents at the Population Level programme created a framework for cultural adaptation and criterion validation utilizing structured diagnostic clinical examinations. The conceptual framework is based on Unicef's drive to support systematic data collecting of adolescent mental health using tools that are culturally customized for use among teenagers. The Mmap initiative was noted as a strong and methodical strategy to gathering and managing mental health data for adolescents globally in the Unicef State of the World Children's report. The Mmap protocol, which devised a method for workable and long-term data collection on adolescent mental health at the population level, provides the theoretical underpinning for the validation study used here. Our validation study shows that utilizing measures that have been culturally adjusted, teenagers who require a mental health examination can be correctly identified. The analytical strategy takes into account the sensitivity and predictive power of these metrics. Through this method, we are producing evidence to support the recommendation of using these tools in a variety of country contexts where psychiatric morbidity is high and an impact has not been widely evaluated at the national level due to a number of factors, including a lack of tools that are feasible and culturally adapted. In Kenya, Swahili, along with English, is the official language of communication in all of the country's regions. We are aware of no study that has approved the use of the Patient Health Questionnaire-9 for teenagers in Kenya in both the local English dialect and Swahili. While Monahan et al. and Mwangi et al. administered and assessed the reliability and validity of the English version of the PHQ-9, the latter employed Swahili version of the PHQ-9 among individuals living with HIV/AIDS in a prior study among head and neck cancer patients conducted in Kenya by Omoro.

Given the foregoing, our report will concentrate on the internal consistency, test-retest reliability, receiver operator characteristics, and discriminant validity of the PHQ-9 in both English and Swahili among adolescents in Kenya aged 10 years to 19 years.

The 2030 Sustainable Development Goals agenda calls for age-specific health data, and the World Health Organization and partners "recommended standard age groups when reporting on health data, especially when reporting on multiple diseases at once, the burden of disease, or a not yet recognized disease or condition." Adolescents are divided into two age groups for the purposes of this study: younger adolescents and older adolescents.

METHOD

Kenya is classified by the World Bank as a lower-middle income country as of 2021. According to the 2019 population census, Kenya is home to 47.5 million people, of which one in four are teenagers between the ages of 10 and 19 and reside in Sub-Saharan Africa. English and Swahili, a Bantu language, are the official languages of Kenya, a multilingual nation. The analysis's research was done in Nairobi, the nation's capital, which is home to 4.4 million people, according to estimates. The Kariobangi and Kangemi health centres, which are managed by the County Government's Nairobi Metropolitan Services, were the sites of the cultural adaptation and validation operations mentioned in this article. 250 teenagers, 148 girls and 102 males, between the ages of 10 and 19, were enlisted. The PHQ-9 was distributed simultaneously to each respondent in both Swahili and English. The Kiddie Schedule of Affective Disorders and Schizophrenia was then used by clinicians to assess teenagers to determine whether they were currently displaying symptoms of severe depressive illness. The sensitivity, specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV), and likelihood ratios for various cut-off scores for the PHQ-9 were evaluated using receiver operating characteristic curves.

Psychiatric evaluation

The English language version of the sample's AUC was 0.886. The AUC for the scale in Swahili was 0.865. For teenagers between the ages of 15 and 19, the Swahili version's AUC was 0.870, whereas the English version's was 0.905. The Cronbach's alpha for the PHQ-9 in both English and Swahili was high, at 0.834 for the English version, 862 for the Swahili version, and 0.835 for participants aged 15 to 19 in both versions. Results in English and Swahili did not significantly differ. Comparing the K-SADS depression scale to the English and Swahili versions of the modified PHQ-9

The appropriate cutoff point for the Swahili and English versions was 9. At this time, the specificity was 73.0%, while the sensitivity was 95.0%. The PPV and NPV values were 23.0% and 99.0%, respectively. Right now, the positive and negative likelihood ratios were 3.53 and 0.77, respectively. The diagnostic OR for the English translation was 49.1, and the accuracy was 74.8%. The sensitivity and specificity of the Swahili version were 89.0% and 70.0%, respectively. The PPV and NPV values were 20.0% and 99.0%, respectively. The odds ratios for the positive and negative outcomes were 2.95 and 0.15, respectively, at this point. The OR was 19.6, and the diagnostic accuracy was 71.2%.

Recruiters and participants

Between June and July 2021, individuals were recruited, and assessments were given. For comparison with "controls" deemed to be free of the disease, the validation sampling was created to be "enriched" for persons likely to exhibit signs of depression. The required ratio between those who were most likely to have depression and those who were most likely to not was 2:1 among the chosen individuals. This strategy was utilized to make sure that the sample size for depressed participants was big enough to support the

execution of a comprehensive range of psychometric analyses for clinical diagnosis validation.

A focused two-stage recruitment process was used. The pregnant and parenting adolescent subsample was identified from an ongoing study under which this work was embedded. Participants were identified by community health workers from households with family issues, marriage conflicts among parents were invited as likely to have depression, and participants from these households were identified by community health workers using the community case detection tool for children and adolescents.

Instruments

The Diagnostic and Statistical Manual of Mental Disorders' nine criteria for the diagnosis of depression are included in the PHQ-9, a nine-item self-report measure. This scale was created to help primary care practitioners identify the most prevalent mental disorders. It is based on the primary care evaluation of mental diseases. One of the instruments used in the MMAP procedure to evaluate teenage depressive symptoms is the nine-item PHQ-9 scale. In both clinical and research contexts, the test was created to screen for and evaluate the severity of depression symptoms. In a sizable primary care sample, the PHQ-9's initial validation studies revealed good reliability and criterion, concept, and external validity. The psychometric qualities of the PHQ-9 have since been examined in a variety of patient samples, including additional primary care samples, psychiatric patients, pregnant women, particular groups of medical patients, as well as diverse multicultural communities. Validation of the PHQ-9 scale was done among university students from Nigeria, China, and Korea in a classroom setting.

Based on the transcultural translation and modification method that had previously been adapted for use with children and adolescents, the PHQ-9 was translated into Swahili and modified for English as spoken in Kenya.

Schedule for adolescents with schizophrenia and affective disorders

For kids aged 6 to 18 years old, the Kiddie Schedule for Affective Disorders and Schizophrenia (K-SADS) is a semi-structured diagnostic clinical interview. To administer the K-SADS, clinic staff receive specialized training. The most recent K-SADS version offers criteria-based algorithms to determine whether existing mental health illnesses meet DSM-5 criteria. We utilized both the depression additional module and the screening module for this study. The Kenyan English and Swahili versions of the PHQ-9's psychometric validity were assessed using the K-SADS diagnosis of major depressive episodes as the benchmark.

Prior to completing the other mental health modules of the MMAP tool, participants first finished the modified PHQ-9. Then, over the course of the following 48 hours, skilled clinicians evaluated the subjects using the K-SADS. The PHQ-9 results for the individual were kept a secret from the clinicians giving the K-SADS.

Training

Community health volunteers have received training to administer the PHQ-9 as part of the MMAP instrument. Community health assistants received supervision training. The K-SADS instrument was taught to twenty clinicians, including clinical officers, psychiatric nurses, psychologists, and psychiatrists. To make sure that participants were being rated by the doctors consistently and

equitably, training included pair evaluations and inter-rater reliability computation. 12 doctors out of these were invited to undertake the diagnostic evaluation using this tool and met the inter-rater reliability (IRR) requirements (IRR of 0.7 and greater). All of our enumerators, clinicians, and respondents were fluent in both English and Swahili, and the data collection was done in those two languages.

Statistic evaluation

In order to translate, adapt, and validate the tool, the required sample size was chosen with the aim of getting a sample size large enough to estimate the receiver operating characteristic curve and area under the curve. A sample size of 100 is typically sufficient to conduct a qualitative evaluation of the utility of a test, despite the fact that a priori sample size determinations for AUC are highly vulnerable to assumptions about the performance of the test. It is crucial to over recruit the desired conditions. In other words, we need methods to spot young people who are most likely to have the relevant mental health issues. In light of this context, a sample of 150+ persons was taken into consideration for recruitment to test positive for depression or anxiety using a community case detection method, and a sample of 100 participants was taken into consideration to test negative. In order to better understand local contextual factors, we added a sample of 50 pregnant adolescents.

Data evaluation

The sociodemographic variables' item means, standard deviations (SDs), frequencies, and percentages were computed. According to the K-SADS, independent samples t-tests were used to compare the mean PHQ-9 scores between the depressed and non-depressed participant groups.

Reliability

Cronbach's alpha reliability coefficients were determined to look at the PHQ-9's repeatability and consistency.

Specificity and sensitivity

Sensitivity and specificity were calculated for each PHQ-9 cut-off point. In this instance, sensitivity is determined by the percentage of people who met the K-SADS criteria for current depressive episodes and were correctly identified by PHQ-9, whereas specificity is determined by the percentage of people who met the standard-SADS criteria for depression and were correctly identified by PHQ-9 as such.

There were reported positive and negative likelihood ratios as well as 95% confidence intervals for each of these measures. In contrast, LR was defined as the likelihood that a person without the condition would have a negative test, while +LR was defined as the likelihood that a person without the condition would have a positive test. To create the ROC curve, sensitivity and specificity were placed on a curve. The ROC curve represents a depiction of the test's sensitivity against all potential threshold values.

The area under the ROC curve, or AUC, was calculated. AUC is a crucial psychometric that estimates the likelihood that an individual with the target condition will be randomly selected and receive a higher score on the instrument than an individual without the condition.

Calculations were made for both positive and negative predictive values. Odds ratio for diagnostics Calculated in terms of sensitivity,

specificity, and positive and negative probability ratios, the odds of a positive test in those with disease in comparison to the odds of a positive test in those without disease. As a result, this single measure, which tends to stay fairly stable despite diagnostic threshold, contains information about both sensitivity and specificity. For the PHQ-9 test, the threshold value for which the value of is maximized can be selected using one form of criterion called Youden's index. The area under the ROC curve was used to measure the PHQ-9's accuracy.

By calculating means and SDs and comparing using t-tests, conducted separately for the overall sample and older adolescents aged 15 to 19 years with Bonferroni correction for multi-testing, it was possible to analyses the discriminant ability of the adapted PHQ-9 items for adolescents with and without diagnosis as per K-SADS.

Results for respondents aged 15 years to 19 years old and for the entire sample were evaluated independently. Swahili and English replies were also assessed separately. R-Version 4.1.2 and Stata version 14.0 software were used for all analyses.

Ethical endorsement

The Kenyatta National Hospital/University of Nairobi Ethical Review Committee gave its clearance for the study. A permission from the Kenyan National Commission for Science, Technology, and Innovation was later acquired when Nairobi County Health gave its approval. Participants were referred to the Kenyatta National Hospital's Department of Mental Health where three clinicians were asked to review their case and provide management for those who were at a high risk of depressive disorder, defined as those who were experiencing high distress or those who were suicidal on the PHQ-9 and/or were assessed by the clinician to be warranting further assessment or treatment. Further family- and community-based follow-ups were conducted when necessary by the community health professionals integrated into the MMAP project.

Training education

The PHQ-9, a component of the MMAP instrument, has been trained to be administered by community health volunteers. The supervisory skills of community health aides were trained. Twenty clinicians, including clinical officers, psychiatric nurses, psychologists, and psychiatrists, received training on the K-SADS instrument. Training included pair assessments and inter-rater reliability computation to ensure that participants were being evaluated by the doctors fairly and consistently. Out of them, 12 doctors who met the criteria for inter-rater reliability (IRR) were invited to participate in the diagnostic evaluation utilizing this instrument IRR of 0.7 and greater. The data collection was conducted in English and Swahili, and all of our enumerators, clinicians, and respondents were bilingual in both languages.

Material

The PHQ-9 is a nine-item self-report questionnaire that includes the nine criteria for the diagnosis of depression from the Diagnostic and Statistical Manual of Mental Disorders. To aid primary care physicians in recognizing the most common mental diseases, this scale was developed. It is based on a primary care physician's assessment of mental illnesses. The nine-item PHQ-9 scale is one of the tools used

in the MMAP process to assess teenage depressive symptoms. The test was developed to screen for and gauge the severity of depressive symptoms in both clinical and research settings. The PHQ-9's early validation studies in a significant primary care population showed strong reliability and criteria, concept, and external validity.

RESULTS

250 adolescents in total, 148 females and 102 boys, ages 10 to 19; mean: 14.76, SD: 2.78; percentage of 10 to 14-year-olds: 45.2%; percentage of 15 to 19-year-olds: 54.8%. 90 girls and 47 boys aged 15 to 19 were present. The participants all spoke Swahili and English very well. Regarding K-SADS results for clinical diagnoses, 19 adolescents met the K-SADS DSM-5 depression criteria, with five adolescents belonging to the 10–14 year old group and fourteen adolescents to the 15–19 year old group. The internal consistency of the PHQ-9 was 0.862 for the English version and 0.834 for the Swahili version. The areas under the curve for the Swahili and English versions, based on receiver operating characteristic analysis, were 0.89 and 0.87, respectively. In the Swahili version, a cut-off of 9 had a sensitivity of 89.0%, specificity of 70.0, PPV of 0.20, and NPV of 0.90; in the English version, a cut-off of 9 had a sensitivity of 95.0%, specificity of 73.2, positive predictive value (PPV) of 0.23, and negative predictive value (NPV) of 0.99.

Psychological testing

The scale for sample's English language version's AUC was 0.886. The Swahili translation of the scale's AUC was 0.865. The Swahili version's AUC was 0.870 while the English version's was 0.905 for adolescents between the ages of 15 and 19. For both the English and Swahili versions of the PHQ-9, the Cronbach's alpha was high 0.834 for the English version, 862 for the Swahili version, and 0.835 for 15–19-year-olds in both versions. Results in Swahili and English did not differ noticeably. Comparing the English and Swahili versions of the adapted PHQ-9 to the K-SADS depression scale

For both the English and Swahili versions, the proper cut-off point was 9. The sensitivity was 95.0% at this time, and the specificity was 73.0%. The respective PPV and NPV values were 23.0% and 99.0%. At this moment, the positive and negative likelihood ratios were respectively 3.53 and 0.77. For the English version, the diagnostic OR was 49.1 and the accuracy was 74.8%. The Swahili version's sensitivity was 89.0% and its specificity was 70.0%. The respective PPV and NPV values were 20.0% and 99.0%. At this moment, the odds ratios for the favorable and negative outcomes were 2.95 and 0.15, respectively. Diagnostic accuracy was 71.2%, and the OR was 19.6.

Item evaluations

Comparing adolescents with depressive diagnoses to those without diagnoses, item means were determined. It demonstrates the products' capacity for discrimination. Significant disparities between respondents with and without diagnoses in all of the PHQ-9 items show that the majority of the Swahili PHQ-9 items functioned well. Some items, however, did not accurately distinguish between people with and without K-SADS depression diagnosis. There were no appreciable variations in these items based on depression state when

utilizing the complete sample, although PHQ Items #1, #3 in English, and #7 in Swahili performed badly. With the exception of Item #1, which showed little interest or enjoyment in English, there were notable disparities among the older teenagers in each of the items. However, the combined score for both language versions showed a noticeable difference in their overall capacity for discrimination. The mean among the adolescents with depression was higher than the mean among the adolescents without depression for the overall sampling PHQ-9 score in English.

The means for each of the nine PHQ items by K-SADS diagnostic status, contrasting a diagnosis of depression with an undiagnosed condition, are shown. The most often endorsed symptom among diagnosed teens was item number 6, Feeling awful about you. The most often endorsed symptom among non-diagnosed participants was item number 2, which is feeling sad or depressed. Additionally, among all participants—diagnosed and undiagnosed—this was the Swahili scale item that received the greatest endorsements. The capacity of each item to distinguish between the groups was good.

Application of validation results: population prevalence adjustments

Policymakers can increase the precision of population estimates of adolescent depression with the help of the validation results. using sensitivity-specificity-balanced cut-offs. uses the psychometric features discovered in this study to present visualization for calculating real depression prevalence based on prevalence collected from the PHQ-9 for the Kenyan English and Swahili versions. The PHQ-9-identified prevalence rates for depression and anxiety can be adjusted to approximate what the true prevalence may be in the population by controlling for false positive and false negative results. For instance, the estimated genuine prevalence is most likely closer to 20% if a prevalence of 39% for depression among 15-year–19-year-old teenagers is found using the English language version. A detected prevalence rate of 45% would be appropriate for an actual prevalence of 20%. Due to the contribution of false positive vs false negative results to the estimates made, the degree of adjustment varies depending on prevalence. Individual adolescent's PPV and NPV differ according on prevalence rates. When allocating resources and creating initiatives, policymakers can make adjusted prevalence estimates using algorithms or data like these. This is significant since the PHQ-9 significantly overestimates population prevalence at typical genuine prevalence rates of 10%–20%.

DISCUSSION

Lack of tools that can be verified and culturally adapted has made it difficult to collect national-level statistics on adolescent mental health that can guide investments in introducing or expanding services. This work adds to the proof that appropriate and verified tools are accessible in LMICs like Kenya. The Swahili and Kenyan English versions of the culturally modified PHQ-9 for use with teenagers contain terminology that is well understood by teenagers, making it simple to administer the survey and communicate with respondents, The translation and adaption process was driven by stakeholders. This study's PHQ-9 internal consistency value, which ranged from 0.824 to 0.869 in previous studies, was comparable to those values. The sensitivity at the cut-off values of 10 for the English version and 9

for the Swahili version or greater was 93.0% for teenagers aged 15 to 19, while the specificity ranged from 67.0% to 74.0%. Although those studies presented results of people older than 18 years old, these values, in particular the sensitivity, are higher than those published in two meta-analyses utilizing PHQ-9. These cutoffs, which range from 8 to 11, were determined by meta-analyses to be the best ones for detecting major depressive episodes in Kenyan teenagers.

The validation findings among teenagers showed psychometric qualities equivalent to those found among adults and in high-income environments. The psychometric features created by a cut-off score of > 8 were comparable to those reported in the literature for other LMICs including South Africa, Argentina, Thailand, Malaysia, and Malaysia.

Limitations

This study is one of the first to use population-level mental health measurements in Kenya to validate treatment techniques with adolescents as young as 10 years old. We were unable to publish results for this group because there were so few adolescents aged 10 to 14 who underwent screening for depression (N = 5). The COVID-19 pandemic's limits caused difficulties, but it was nevertheless possible to complete the work, albeit for a longer time than anticipated. Only Kenyan English and Swahili, Kenya's two official working languages, were addressed in the validation. We did not validate the tool for additional Bantu languages that are extensively spoken in Kenya, such as Kikuyu, Kamba, and Luhya. The work was also done in two ethnically diverse periurban informal settlements in Nairobi, albeit it was not done in rural areas. Therefore, caution should be used while extrapolating the tool to rural areas. Despite these drawbacks, this research helps Kenyan teenagers by adopting a technique that was previously only used among adults and has been culturally adjusted and validated for use with them.

CONCLUSION

The PHQ-9's performance in Kenya's validation studies for identifying adolescent depression was satisfactory. By using tools that can be administered by trained interviewers in the context of population-based data collection, we are increasing the feasibility of data collection and strengthening the evidence for the availability of culturally and clinically valid measurement tools to estimate the prevalence of depression at the population level.

The variety of cut-off scores offered by this analysis can also help practitioners, such as physicians, policymakers, and researchers, to decide which cut-off score is most appropriate for the intended usage in Kenyan settings. The PHQ-9's psychometric qualities were similar between its Swahili and English adaptations, indicating that it is a valid and reliable tool for diagnosing adolescent major depressive disorder and can be used in settings with limited resources to find children and adolescents who need mental health support as soon as possible.

Support for ethics

The study was approved by the Kenyatta National Hospital/ University of Nairobi Ethical Review Committee. Later, after receiving approval from Nairobi County Health, the Kenyan National Commission for Science, Technology, and Innovation granted its permission. Participants were referred to the Kenyatta National Hospital's Department of Mental Health, where three clinicians were

asked to review each participant's case and provide management for those who were at a high risk of depressive disorder, defined as those who were exhibiting high distress or who scored suicidal on the PHQ-9 and/or were judged by the clinician to be deserving of further evaluation or treatment. The MMAP project's community health specialists who were included in the family and community-based follow-ups carried out additional ones when needed.