Brief Screening Tool for Stepped-Care Management of Mental and Substance Use Disorders

Kathryn L. Lovero, PhD[1], **Cale Basaraba, MPH**[1,2], **Saida Khan, MA**[3,4], **Antonio Suleman, MD**[5,6], **Dirceu Mabunda, MD**[3,4], **Paulino Feliciano, BS**[5,6], **Palmira dos Santos, PhD**[4,7], **Wilza Fumo, MD**[4,7], **Flavio Mandlate, MD**[4,7], **M. Claire Greene, PhD**[1], **Andre Fiks Salem, BS**[1], **Jennifer J. Mootz, PhD**, **Ana Olga Mocumbi, MD**[4,8], **Cristiane S. Duarte, PhD**[1], **Lidia Gouveia, MD**[4,7], **Maria A. Oquendo, MD**[9], **Melanie M. Wall, PhD**[1,2], **Milton L. Wainberg, MD**[1]

[1]Department of Psychiatry, New York State Psychiatric Institute and Columbia University Vagelos College of Physicians and Surgeons, New York, New York

[2]Department of Biostatistics, Columbia University Mailman School of Public Health, New York, New York

[3]Health Directorate of Maputo City, Ministry of Health, Maputo, Mozambique

[4]Universidade Eduardo Mondlane School of Medicine, Maputo, Mozambique

[5]Health Directorate of Nampula Province, Ministry of Health, Nampula, Mozambique

[6]Nampula Psychiatric Hospital, Nampula, Mozambique

[7]Department of Mental Health, Ministry of Health, Maputo, Mozambique

[8]National Institute of Health, Marracuene, Mozambique

[9]Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania

Abstract

Background—Widespread implementation and sustainability of stepped mental healthcare requires a rapid method for non-specialists to detect illness. This study aimed to develop and validate a brief instrument, the Mental Wellness Tool (mwTool), for identification and classification of mental disorders.

Methods—Cross-sectional development and validation samples included adults at six health facilities in Mozambique. Mini International Neuropsychiatric Interview diagnoses served as criterion standard. Nine mental disorder and functioning assessments comprised the battery of candidate items. For mwTool development, regression modeling and expert consultation determined best items for identifying any mental disorder and classification of positives into disorder categories (severe mental disorder, common mental disorder, substance use disorder, and

Corresponding Author Kathryn L. Lovero, PhD, Division of Translational Epidemiology, Department of Psychiatry, New York State Psychiatric Institute and Columbia University Vagelos College of Physicians and Surgeons, New York, NY, kate.lovero@nyspi.columbia.edu.

suicide risk). For validation, sensitivity and specificity were calculated for any mental disorder (index and proxy respondents) and disorder categories (index).

Results—Development (911 participants, mean±SD age 32±11 years; 63% female): From the 99-item battery administered, 13 items were selected for the mwTool, three with 0.83 sensitivity (95% CI=0.79–0.86) for any mental disorder and 10 additional items classifying participants with specificity of 0.72 (severe mental disorder) to 0.90 (suicide risk). Validation (480 participants, age 31±11 years; 59% female): Sensitivity for any mental disorder was 0.94 (0.89–0.97) using index and 0.73 (0.58–0.85) using family proxy respondents. Specificity for disorder categories was 0.47 (severe mental disorder) to 0.93 (suicide risk). Removing one item increased severe mental disorder specificity to 0.63 (0.58–0.68).

Conclusions—The mwTool performs well for identification of any mental disorder using index and proxy responses to 3 items and for classification of positives into treatment categories using index responses to an additional 9 items.

Introduction

Mental and substance use disorders, henceforth mental disorders, are the largest contributor to global burden of disease[1], yet the majority of people living with mental disorders in low- and middle-income countries (LMIC) do not have access to care owing to scarce funding and human resources[2]. A stepped-care approach, in which non-specialists manage detection of mental disorders and provide treatment or referral to specialists, is an efficient method to close the treatment gap in LMIC[3–5].

In stepped-care, only severe mental disorders require consultation with a mental health specialist, whereas non-severe mental disorders can be managed by primary-care providers and lay workers[6]. Thus, key to implementation and sustainability of a comprehensive stepped-care management of mental disorders is a rapid, reliable method for minimally-trained providers to identify presence and type of mental disorder. Many screening tools for mental disorder detection have been validated in high-income and LMIC settings[7]. However, these tools, comprised of five to more than 20 items and sometimes with cost per use, are designed to detect one disorder at a time (e.g., depression[8]) or symptoms common to some disorders (e.g., psychological distress[9]). Using a combination of these screens for all mental disorders is unfeasible in low-resource health systems.

We aimed to develop and validate a brief questionnaire, the Mental Wellness Tool (mwTool), to screen for mental disorders and classify individuals into disorder categories that facilitate comprehensive stepped-care management of mental disorders. Through novel application of a variable selection technique (Least Absolute Shrinkage and Selection Operator, LASSO)[10], we sought to identify a small set of items, selected from widely-validated screening measures for individual mental disorders, with high sensitivity for identification of any mental disorder and high specificity for classification of severe mental disorder, common mental disorder, substance use disorder, and suicide risk. In a separate validation sample, we assessed performance of the mwTool in identifying and classifying mental disorders. Specifically, we examined performance using index as well as proxy responses,

which facilitate epidemiological research and community-based care wherein it is not feasible or possible to interview all index cases.

Methods

Participants provided written informed consent as approved by the NYSPI Institutional Review Board (#7479) and the Eduardo Mondlane University Institutional Health Bioethics Council (CIBS FM & HCM/54/2017). Study analyses and reporting follow the Standards for Reporting of Diagnostic Accuracy Studies (STARD)[11].

Study Setting

Development data were collected at two primary care clinics and one hospital in Maputo City, Mozambique from May 16th-June 8th, 2018. These facilities provide primary care, emergency, and outpatient mental health services. The hospital also provides services for victims of interpersonal violence and inpatient health and psychiatric services. Validation data were collected from December 5th–12th, 2018 at three primary care clinics in Nampula, Mozambique. These facilities provide primary care and emergency services. People with mental disorders are referred to Nampula's provincial psychiatric hospital.

Study Population

Adults (patients and accompaniers) in health facility waiting rooms were invited to participate. All volunteers were taken to a private area for eligibility assessments and informed consent. Potential participants were excluded if they were less than 18 years old and/or were unable to sufficiently communicate in Portuguese, determined by interviewers asking potential participants to repeat study objectives in their own words. For the development sample, we planned to enroll ~400 people with at least one psychiatric diagnosis and ~400 without any psychiatric diagnosis to ensure ±5% margins of error for sensitivity and specificity estimates. For diversity of psychiatric diagnoses, we aimed to obtain ~40 gender-balanced participants with each specific diagnosis (detailed below). For the validation sample, we aimed to obtain ~40 gender-balanced index participants (those providing responses about their own mental health) with each specific diagnosis (described below), of whom ~200 would provide proxy responses (regarding the mental health of another index participant with whom they were attending the health facility), allowing enough precision for ±7% margins of error for sensitivity and specificity.

Measures

For all instruments except the Psychosis Screening Questionnaire (PSQ) and Primary Care Post-Traumatic Stress Disorder Screen (PC-PTSD), we used existing Brazilian or Portuguese translations and local research team members made minor adjustments for the Mozambican context (e.g., local terms for substances). The PSQ and PC-PTSD were translated from English to Portuguese by the local research team, back-translated by a native English speaker fluent in Portuguese, and reviewed for translation accuracy by a measurement specialist at Columbia University (unassociated with the present study). All measures were pre-tested while training interviewers (Mozambican mental health

specialists) and underwent final review using cognitive interviews with 10 Mozambican adults attending primary care.

Mental disorder diagnosis and classification—Mental disorder diagnoses were made using the Brazilian version of The Mini International Neuropsychiatric Interview (MINI) Plus[12,13], a structured diagnostic interview that has been widely-used as a reference standard across many contexts[7]. Based on MINI diagnoses, we classified participants into the following categories corresponding to different stepped-care pathways: 1) severe mental disorder for diagnoses of mania, psychosis, or the presence of psychotic symptoms associated with another disorder (e.g., depression); 2) common mental disorder for major depressive episode, panic disorder, PTSD, anxiety, and/or somatization; 3) substance use disorder for alcohol abuse or dependence and/or substance abuse or dependence; 4) suicide risk if they were scored as having moderate or high suicide risk (i.e. 6 points or higher, indicative of past-month active ideation, planning, and/or attempt). All diagnoses were for current disorder, except psychosis, for which even lifetime diagnoses were considered since patients with history of psychosis require referral to specialists.

Mental health screening battery—We administered nine structured instruments commonly used to screen for specific mental disorders and to assess functioning (Table 1, see online supplement for detail)[8,14–21].

Demographic and general health measures—We collected self-reported sociodemographic information (age, gender, marital status, living situation, education, religion, monthly household income, occupation, and ethnicity) and health history (chronic diseases, pregnancy, and parity). For the validation sample, we recorded the relationship between index and proxy respondents.

Responses to all measures were recorded via tablet using the REDCap platform[22].

Procedures

In the development sample, research assistants administered the sociodemographic questionnaire and then the MINI and mental health screening battery in a randomized order. In the validation phase, a research assistant first administered the sociodemographic questionnaire. Then, for participants who were alone or with someone who was not eligible to participate in the study, the research assistant administered the mwTool followed by the MINI to assess the participant's mental health (index response). For participants attending the facility with another participant, research assistants privately asked the items of the mwTool for identification of any mental disorder in regard to their companion's mental health (proxy response) and then administered the complete mwTool and MINI to assess the participant's mental health (index response).

Statistical analysis

We excluded from analysis participants with incomplete responses to the MINI, screening battery, or mwTool. Analyses were performed using R version 3.6.1[23]; the *glmnet* package fit LASSO models[24].

mwTool Development—First, we sought to identify 3–5 items from the screening battery with high sensitivity for the presence of any mental disorder. Second, we sought to identify an additional 6–12 items that provided high specificity for classification into the four disorder categories, to minimize false positives for potentially stigmatizing disorders and undue burden on low-resource systems, while maintaining adequate sensitivity. All items considered for the mwTool are in the online supplement. Ordinal responses were dichotomized for analyses; responses indicating moderate to high symptom strength and frequency were considered positive.

A series of LASSO logistic regression was used to determine the best subset of battery items for presence of any mental disorder and subsequently for each disorder category. LASSO regression incorporates a penalty term based on the sum of the absolute values of all model coefficients. The effect of this term is, at high levels of the shrinkage parameter (λ), coefficient estimates for less important variables shrink to zero and are removed from the model. This allows for variable selection at high levels of λ, which we varied to select the best 3–10 items for predicting each outcome. We confirmed LASSO results only included items with positive coefficients (item presence associated with higher disorder risk). Area under the receiver operating characteristic curve described the accuracy of different best-item subsets. We then selected mwTool items, balancing statistical validity (i.e., empirically best combination of items based on LASSO), feasibility (i.e., fewer total items), and face validity (i.e., items reflecting diagnostic criteria for disorder categories).

Sensitivity and specificity of the mwTool for any mental disorder and the four disorder categories were assessed in the development sample. Because treatment of severe mental disorder takes priority when a person has both severe and common mental disorder in stepped care, participants positive for both by the mwTool were classified as severe mental disorder and not common mental disorder.

mwTool Validation—We calculated sensitivity and specificity of the mwTool for any mental disorder and the four disorder categories using index case responses. We also assessed sensitivity and specificity of mwTool questions for any mental disorder using proxy responses. We excluded proxy responses when proxy and index respondents provided discordant information about their relationship.

Results

mwTool Development

Across the three Maputo sites, 1033 people were screened for eligibility; seven (1%) were under 18 and eight (1%) were not fluent in Portuguese. Twenty-nine (3%) of the 1018 eligible people did not provide informed consent. We excluded from analysis 78 (8%) participants who did not complete the MINI or screening battery (see online supplement). Among the 911 included participants, 570 (63%) were female and the mean±SD age was 32.0±11.3 years. Over half (52%, n=470) had one or more disorder based on MINI diagnoses (Table 2): 29% (n=260) severe mental disorder, 36% (n=330) common mental disorder, 14% (n=124) substance use disorder, and 9% (n=86) suicide risk.

The 3, 5, 8, and 10 screening items that best classified any mental disorder and each of the disorder categories are detailed in the online supplement. In total, 13 screening battery items were selected for inclusion in the mwTool (see diagram in online supplement). Twelve items were chosen from LASSO results. In consultation with expert clinicians and clinical researchers, one additional suicide item was added to capture people with recent attempts, a high-risk group that may lack ideation and thus would not be detected by the LASSO models' best items.[25]

A positive response to any of the first three mwTool items signals for the next 10 items to be asked; a negative response to all three indicates absence of any mental disorder and signals screening completion. When the 10 additional items are asked, a positive response to any item associated with the disorder category indicates presence of that disorder category. Negative responses to all additional 10 items indicate the person should be classified as 1) common mental disorder if they answered positively to PHQ2 or GAD5 or 2) no disorder if they only answered positively to GAD1.

The first three mwTool items identified any mental disorder with 0.83 sensitivity (95% CI=0.79–0.86) and 0.49 specificity (95% CI=0.44–0.54) (Table 2), and performed similarly by gender, age, and HIV status (see online supplement). The 10 additional mwTool items classified severe mental disorder with 0.72 specificity (95% CI=0.69–0.76), common mental disorder with 0.79 specificity (95% CI=0.76–0.82), substance use disorder with 0.82 specificity (95% CI=0.79–0.84), and suicide risk with 0.90 specificity (95% CI=0.88–0.92) (Table 2), with little variation across subpopulations (see online supplement). Sensitivity for the specific disorder categories was highest for suicide risk (0.80) and lowest for severe mental disorder (0.62).

mwTool Validation and Final Item Selection

At Nampula sites, 482 people were screened for eligibility; two (<1%) were not fluent in Portuguese. Of the 480 participants, 243 (51%) provided proxy responses to the three initial mwTool items regarding another participant with whom they were attending the health facility. We excluded from analysis 27 (6%) participants who did not complete the MINI or all mwTool items (see online supplement). Among the 453 included participants, 296 (59%) were female. Mean age was 31.1±10.7 years. MINI diagnoses indicated presence of one or more disorder in 39% (n=178) of participants (Table 3): 18% (n=82) severe mental disorder, 30% (n=134) common mental disorder, 6% (n=29) substance use disorder, and 8% (n=35) suicide risk.

Using index respondents, the first 3 items of the mwTool had 0.94 (95% CI=0.89–0.97) sensitivity for identification of any mental disorder. The 10 classification items had specificity of 0.47 (95% CI=0.42–0.53) for severe mental disorder, 0.83 (95% CI=0.78–0.87) for common mental disorder, 0.82 (95% CI=0.78–0.86) for substance use disorder, and 0.93 (95% CI=0.90–0.96) for suicide risk. Sensitivity for disorder categories ranged from 0.77 (suicide risk) to 0.96 (common mental disorder). Family proxy responses had higher sensitivity (0.73; 95% CI=0.58–0.85) than non-family responses (0.62; 95% CI=0.45–0.78), though not significantly different (Table 3). For both family and non-family proxy responses

that led to a positive mwTool screen on the mwTool, 83% of corresponding index responses also led to a positive screen.

While the 13-item mwTool generally showed similar or higher sensitivity and specificity in the validation compared to development samples, the specificity for severe mental disorder was considerably lower. Many participants who were false positives for severe mental disorder (58/195) had endorsed only the first of the additional classification items (GAD-7, item 7). As this item assesses a symptom of anxiety, a common mental disorder, we next evaluated performance of the mwTool excluding this item. Using this 12-item mwTool (see diagram in online supplement), specificity of index responses for classification of severe mental disorder increased to an acceptable level (0.63; 95% CI=0.58–0.68) while the specificity for common mental disorder (which the excluded item also was used to classify in the 13-item mwTool) was reduced but remained good (0.72; 95% CI=0.67–0.77) (Table 4). The 12-item mwTool performed similarly in subpopulations of the validation sample (see online supplement).

Discussion

We employed the novel application of LASSO regression modeling along with expert consultation to select items from mental health screens that can identify and classify mental disorders for stepped-care service provision. We designed a two-step instrument, the mwTool, in which the first three items are asked to all respondents and only those identified as positive for any mental disorder are asked the additional items for classification into disorder categories.

Brevity of screening instruments reduces provider burden and, in turn, promotes adoption in primary care and community settings[26]. Additionally, previous research has shown that screens with fewer items are as accurate as those with more items for individual disorder detection in both high-income countries and LMICs[27–29]. Other analytical techniques, such as item response theory, have been successfully used to shorten screens for common mental disorders and substance use disorders in LMICs[30–32], though no brief instrument exists that provides transdiagnostic mental health assessment. With the LASSO variable selection technique, we reduced 99 items from 9 instruments for 8 different mental disorders to 12 items, or just 12% of the combined screens length, that had acceptable to excellent sensitivity and specificity for all disorder categories.

Proxy respondents are common in clinical care and epidemiological research when the index case is unable to self-report, either because they are unavailable, incapable of providing responses, or underage. However, evaluation of proxy responses on other assessment tools has shown them to be less reliable for questions about subjective experiences, like emotions and psychological distress, than objective experiences[33]. Our results demonstrate that the first three items of the mwTool have good sensitivity for identification of any mental disorder using proxy responses from family members. Proxy responses from non-family did not perform as well as family proxy responses, in line with previous research showing friends and healthcare proxy respondents have lower agreement and reliability than family proxies[33]. Future studies are needed to determine in more detail what specific characteristics

of family members—such as cohabitation or relation to the index—promote reliability of proxy responses on the mwTool.

Throughout mwTool development and validation, severe mental disorder was the lowest performing disorder category. In tool development, more questions (5) had to be included for adequate classification of severe mental disorder and sensitivity of the mwTool for severe mental disorder was lower than for any other disorder category. This is unsurprising, as measures for severe mental disorders have routinely been found to have lower performance than for other mental disorders[34]. However, in the validation sample, by removing one question we were able to increase specificity to an acceptable level. We therefore recommend the use of the 12-item mwTool in future assessments.

This study has several limitations. Both development and validation sample participants were recruited from health clinics, and our findings may not be generalizable to other settings. Additionally, because there are no other published data on prevalence of mental disorders in Mozambican healthcare settings and we cannot be certain our data are representative, we did not calculate the positive predictive value or negative predictive value of the mwTool. However, the mwTool did not perform differently by age, gender, and HIV status subgroups. Because we assessed the mwTool in one low-income country and one language, its validity in other settings should be assessed. Moreover, owing to post-hoc elimination of the GAD-7 item from the 13-item mwTool assessed in the validation sample, the 12-item mwTool requires further validation in an independent sample. Finally, our setting had low rates of substance use, and most substance users also used alcohol; therefore, the mwTool includes questions only related to alcohol use. In other settings, it may be necessary to add items for substance use and calibrate dichotomization of the measure according to contextual substance use patterns.

Conclusions

To our knowledge, the mwTool is the first brief screen for non-specialist assessment of common mental disorder, severe mental disorder, substance use disorder, and suicide risk. The mwTool performs well for identification of any mental disorder using index and proxy responses to 3 items and for classification of positives into treatment categories using index responses to an additional 9 items. Though developed in LMIC primary care, the mwTool may have applicability in multiple settings, such as community-based care, emergency situations, and population-based assessments, but further research is required to assess its performance in these settings.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Whiteford HA, Degenhardt L, Rehm J, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. Lancet. 2013;382(9904):1575–1586. [PubMed: 23993280]

2. Saraceno B, van Ommeren M, Batniji R, et al. Barriers to improvement of mental health services in low-income and middle-income countries. Lancet. 2007;370(9593):1164–1174. [PubMed: 17804061]

3. Kakuma R, Minas H, vanGinneken N, et al. Human resources formental healthcare: current situation and strategies for action. Lancet. 2011;378(9803):1654–1663. [PubMed: 22008420]

4. Kohrt BA, Asher L, Bhardwaj A, et al. The role of communities in mental health care in low-and middle-income countries: a meta-review of components and competencies. Int J Environ Res Public Health. 2018;15(6):1279.

5. Patel V, Chisholm D, Parikh R, et al. Addressing the burden of mental, neurological, and substance use disorders: key messages from Disease Control Priorities, 3rd edition. Lancet. 2016;387(10028):1672–1685. [PubMed: 26454360]

6. WHO| mhGAP Intervention Guide-Version 2.0. WHO. 2018.

7. Ali GC, Ryan G, De Silva MJ. Validated Screening Tools for Common Mental Disorders in Low and Middle Income Countries: A Systematic Review. PLoS One. 2016;11(6):e0156939. [PubMed: 27310297]

8. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatr Ann. 2002;32(9):509–515.

9. Kessler RC, Andrews G, Colpe LJ, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med. 2002;32(6):959–976. [PubMed: 12214795]

10. Tibshirani R. Regression shrinkage and selection via the lasso. JR Stat Soc Ser B. 1996;58(1):267–288.

11. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. Radiology. 2015;277(3):826–832. [PubMed: 26509226]

12. Amorim P. Mini International Neuropsychiatric Interview (MINI): validação de entrevista breve para diagnóstico de transtornos mentais. Braz J Psychiatry. 2000;22(3):106–115.

13. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry. 1998.

14. Babor TF, de la Fuente JR, Saunders J, Grant M. The Alcohol Use Disorders Identification Test: Guidelines for use in. Prim Care. 2001.

15. Bebbington P, Nayani T. The psychosis screening questionnaire. Int J Methods Psychiatr Res. 1995.

16. Gierk B, Kohlmann S, Kroenke K, et al. The somatic symptom scale–8 (SSS-8):abrief measure of somatic symptom burden. JAMA Intern Med. 2014;174(3):399–407. [PubMed: 24276929]

17. WHO ASSIST Working Group. The alcohol, smoking and substance involvement screening test (ASSIST): development, reliability and feasibility. Addiction. 2002;97(9):1183–1194. [PubMed: 12199834]

18. Posner K, Brown GK, Stanley B, et al. The Columbia–Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry. 2011;168(12):1266–1277. [PubMed: 22193671]

19. Prins A, Bovin MJ, Smolenski DJ, et al. The primary care PTSD screen for DSM-5(PC-PTSD-5): development and evaluation within a veteran primary care sample. J Gen Intern Med. 2016;31(10):1206–1211. [PubMed: 27170304]

20. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med. 2006;166(10):1092–1097. [PubMed: 16717171]

21. Üstün TB, Kostanjsek N, Chatterji S, Rehm J. Measuring health and disability: Manual for WHO disability assessment schedule WHODAS 2.0. World Health Organization; 2010.

22. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42(2):377–381. [PubMed: 18929686]

23. R. Core Team. R: A language and environment for statistical computing. 2013.

24. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Soft. 2010;33(1):1.

25. Rimkeviciene J, De Leo D. Impulsive suicide attempts: a systematic literature review of definitions, characteristics and risk factors. J Affect Disord. 2015;171:93–104. [PubMed: 25299440]

26. Johnson M, Jackson R, Guillaume L, Meier P, Goyder E. Barriers and facilitators to implementing screening and brief intervention for alcohol misuse: a systematic review of qualitative evidence. J Public Health. 2011;33(3):412–421.

27. Akena D, Joska J, Obuku EA, Amos T, Musisi S, Stein DJ. 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(1):187. [PubMed: 23116126]

28. ChoiS K, Boyle E, Burchell AN, et al. Validation of sixs hort 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(11):e0142706. [PubMed: 26566285]

29. Mitchell AJ, Coyne JC. Doultra-short screening instruments accurately detect depression in primary care?: A pooled analysis and meta-analysis of 22 studies. Br J Gen Pract. 2007;57(535):144–151. [PubMed: 17263931]

30. Betancourt TS, Yang F, Bolton P, Normand SL. Developing an African youth psychosocial assessment: an application of item response theory. Int J Methods Psychiatr Res. 2014;23(2):142–160. [PubMed: 24478113]

31. Doty SB, Haroz EE, Singh NS, et al. Adaptation and testing of an assessment for mental health and alcohol use problems among conflict-affected adults in Ukraine. Conf Health. 2018;12(1):34.

32. Haroz EE, Kane JC, Nguyen AJ, Bass JK, Murray LK, Bolton P. When less is more: reducing redundancy in mental health and psychosocial instruments using Item Response Theory. Glob Mental Health. 2020;7.

33. Andresen EM, Vahle VJ, Lollar D. Proxy reliability: Health-related quality of life (HRQoL) measures for people with disability. Qual Life Res. 2001;10(7):609–619. [PubMed: 11822794]

34. Eaton WW, Hall AL, Macdonald R, Mckibben J. Case identification in psychiatric epidemiology: a review. Int Rev Psychiatry. 2007;19(5):497–507. [PubMed: 17896230]

Table 1.

Description of measures comprising the mental health screening battery used to develop the mwTool.

Disorder	Measure	No. Items
Severe Mental Disorder		
Psychosis/Mania	Psychosis Screening Questionnaire	5[a]
Common Mental Disorder		
Depression	Patient Health Questionnaire	9
Anxiety	Generalized Anxiety Questionnaire	7
PTSD	Primary Care Post Traumatic Stress Disorder Screen for DSM-5	5
Somatization	Somatic Symptom Scale	8
Substance Use Disorder		
Alcohol	Alcohol Use Disorder Identification Test	10
Substances	Alcohol, Smoking & Substance Involvement Screening Test	12[b]
Suicide Risk	Columbia Suicide Severity Rating Scale	7[c]
Functioning	WHO Disability Assessment Schedule 2.0	36
TOTAL		99

Abbreviations: PTSD, post-traumatic stress disorder

[a] Items that lead to positive screen and termination of questionnaire.

[b] Based on inclusion of 6 items regarding cannabis use and 6 regarding cocaine use in past 3 months. No other substance use was reported in our sample.

[c] Final question analyzed as two items, one assessing lifetime and one assessing past 3-month suicidal behavior.

Table 2.

Performance of 13-item mwTool in identification and classification of participants with mental disorders in the development sample.

Disorder	No.	(%)	Sensitivity	95% CI	Specificity	95% CI
Any Disorder[a]	470	52	0.83	0.79–0.86	0.49	0.44–0.54
Severe Mental Disorder	260	29	0.62	0.55–0.67	0.72	0.69–0.76
Psychosis	235	26	0.61	0.55–0.68	0.71	0.67–0.74
Mania	70	8	0.77	0.66–0.86	0.66	0.63–0.69
Common Mental Disorder	330	36	0.83	0.78–0.87	0.79	0.76–0.82
Major Depressive Episode	298	33	0.84	0.80–0.88	0.79	0.76–0.83
Panic Disorder	33	4	0.94	0.80–0.99	0.78	0.76–0.81
PTSD	49	5	0.90	0.78–0.97	0.78	0.75–0.81
Anxiety	65	7	0.91	0.81–0.97	0.79	0.76–0.82
Somatic	13	1	0.85	0.55–0.98	0.79	0.76–0.81
Substance Use Disorder	124	14	0.72	0.63–0.79	0.82	0.79–0.84
Alcohol	115	13	0.73	0.64–0.81	0.81	0.79–0.84
Substance	22	2	0.64	0.41–0.83	0.75	0.73–0.78
Suicide Risk	86	9	0.80	0.70–0.88	0.90	0.88–0.92
Medium Risk	29	3	0.83	0.64–0.94	0.86	0.83–0.88
High Risk	57	6	0.79	0.66–0.89	0.88	0.85–0.90

Abbreviations: PTSD, post-traumatic stress disorder

[a]Calculated based on responses to the initial three mwTool items only.

Table 3.

Performance of 13-item mwTool for index case and proxy respondents in the validation sample.

Index Case	No.	%	Sensitivity	95% CI	Specificity	95% CI
Any Disorder[a]	178	39	0.94	0.89–0.97	0.34	0.28–0.40
Severe Mental Disorder	82	18	0.89	0.80–0.95	0.47	0.42–0.53
Common Mental Disorder	134	30	0.96	0.91–0.98	0.83	0.78–0.87
Substance Use Disorder	29	6	0.86	0.68–0.96	0.82	0.78–0.86
Suicide Risk	35	8	0.77	0.60–0.90	0.93	0.90–0.96
Proxy (Family)	**No.**	**%**	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**
Any Disorder[a]	48	48	0.73	0.58–0.85	0.31	0.19–0.45
Proxy (Non-family)	**No.**	**%**	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**
Any Disorder[a]	37	34	0.62	0.45–0.78	0.51	0.39–0.63

[a]Calculated based on responses to the initial three mwTool items only.

Table 4.

Performance of 12-item mwTool for index case and proxy respondents in the validation sample.

Index Case	No.	%	Sensitivity	95% CI	Specificity	95% CI
Any Disorder[a]	178	39	0.94	0.89–0.97	0.34	0.28–0.40
Severe Mental Disorder	82	18	0.82	0.72–0.89	0.63	0.58–0.68
Common Mental Disorder	134	30	0.93	0.87–0.96	0.72	0.67–0.77
Substance Use Disorder	29	6	0.86	0.68–0.96	0.82	0.78–0.86
Suicide Risk	35	8	0.77	0.60–0.90	0.93	0.90–0.96
Proxy (Family)	**No.**	**%**	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**
Any Disorder[a]	48	48	0.73	0.58–0.85	0.31	0.19–0.45
Proxy (Non-family)	**No.**	**%**	**Sensitivity**	**95% CI**	**Specificity**	**95% CI**
Any Disorder[a]	37	34	0.62	0.45–0.78	0.51	0.39–0.63

[a] Calculated based on responses to the initial three mwTool items only.