

Validation of the Computerized Adaptive Test for Mental Health in Primary Care

Andrea K. Grabam, PhD^{1*}

Alexa Minc, BA, BFA^{2*}

Erin Staab, MPH²

David G. Beiser, MD, MS³

Robert D. Gibbons, PhD⁴

Neda Laiteerapong, MD, MS²

¹Department of Medical Social Sciences, Northwestern University, Chicago, Illinois

²Department of Medicine, Section of General Internal Medicine, The University of Chicago, Chicago, Illinois

³Department of Medicine, Section of Emergency Medicine, The University of Chicago, Chicago, Illinois

⁴Departments of Medicine, Public Health Sciences, Psychiatry, and Comparative Human Development, The University of Chicago, Chicago, Illinois

*Both authors contributed equally to this work.



Conflicts of interest: Dr Gibbons is a founder of Adaptive Testing Technologies, the company that distributes the CAT-MH. This activity has been reviewed and approved by the University of Chicago. All other authors report none.

CORRESPONDING AUTHOR

Neda Laiteerapong, MD, MS
The University of Chicago
5841 S Maryland Ave, MC 2007
Chicago, IL 60637
nlaiteer@medicine.bs.uchicago.edu

ABSTRACT

PURPOSE The US Preventive Services Task Force recommends screening for depression in the general adult population. Although screening questionnaires for depression and anxiety exist in primary care settings, electronic health tools such as computerized adaptive tests based on item response theory can advance screening practices. This study evaluated the validity of the Computerized Adaptive Test for Mental Health (CAT-MH) for screening for major depressive disorder (MDD) and assessing MDD and anxiety severity among adult primary care patients.

METHODS We approached 402 English-speaking adults for participation from a primary care clinic, of whom 271 adults (71% female, 65% black) participated. Participants completed modules from the CAT-MH (Computerized Adaptive Diagnostic Test for MDD, CAT-Depression Inventory, CAT-Anxiety Inventory); brief paper questionnaires (9-item Patient Health Questionnaire [PHQ-9], 2-item Patient Health Questionnaire [PHQ-2], Generalized Anxiety Disorder 7-item Scale [GAD-7]); and a reference-standard interview, the Structured Clinical Interview for DSM-5 (*Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*) Diagnoses.

RESULTS On the basis of the interview, 31 participants met criteria for MDD and 29 met criteria for GAD. The diagnostic accuracy of the Computerized Adaptive Diagnostic Test for MDD (area under curve [AUC] = 0.85) was similar to that of the PHQ-9 (AUC = 0.84) and higher than that of the PHQ-2 (AUC = 0.76) for MDD screening. Using the interview as the reference standard, the accuracy of the CAT-Anxiety Inventory (AUC = 0.93) was similar to that of the GAD-7 (AUC = 0.97) for assessing anxiety severity. The patient-preferred screening method was assessment via tablet/computer with audio.

CONCLUSIONS Computerized adaptive testing could be a valid and efficient patient-centered screening strategy for depression and anxiety screening in primary care settings.

Ann Fam Med 2019;17:23-30. <https://doi.org/10.1370/afm.2316>.

INTRODUCTION

Major depressive disorder (MDD) and generalized anxiety disorder (GAD) affect nearly 10% of adults¹ and are largely managed in primary care settings.^{2,3} At least one-half of patients with depression in primary care are not recognized or adequately treated, however.⁴⁻⁷ Adequately treating MDD and GAD is imperative, given patients' adverse health outcomes and high health care costs when these conditions go untreated.⁸⁻¹³

A crucial first step to improving depression and anxiety outcomes is adequate screening.¹⁴ The most commonly used screening tools in primary care are paper based and have a limited number of predetermined questions.¹⁵⁻¹⁸ However, nearly 90% of US primary care physicians have electronic health records (EHRs),¹⁹ presenting the opportunity to leverage electronic tools for screening.

Computerized adaptive tests (CATs) are electronic tools that create personalized assessments by adaptively varying the questions administered based on patient responses to previous questions. By design, CATs minimize measurement uncertainty and have greater precision than traditional self-report

assessments. Several CATs for depression and anxiety have been developed,²⁰⁻³⁷ including the Computerized Adaptive Test for Mental Health (CAT-MH). The CAT-MH comprises a suite of assessments, including ones for MDD screening,³⁸ MDD severity,^{39,40} and anxiety severity.⁴¹ It was developed using multidimensional item response theory and random forests to capture the multidimensional nature of psychological disorders.²⁷

The CAT-MH has been validated for adults presenting for outpatient psychiatric treatment,^{38,39,41,42} but has yet to be validated among adult primary care populations. Because the prevalence of depression and anxiety may be comparatively lower among the latter, different questions in the item bank may be more appropriate; thus, validation in this population is warranted. Also, the use of CATs, based on multidimensional item response theory, has great potential in primary care to increase the efficiency of assessing mental and physical health. We therefore evaluated the validity of the CAT-MH for MDD screening and for assessing depression and anxiety severity in adult primary care patients.

METHODS

Participants

Participants were adults aged 18 years or older presenting to the internal medicine clinic in an urban, academic medical center. Individuals were eligible if they spoke English, they could see and hear study directions, they screened negative for dementia,^{43,44} and their physician assented to approaching them for recruitment.

Procedure

This study was approved by the medical center's institutional review board and monitored by a data safety monitoring board. Patients were approached for participation while waiting to meet with their physician. All participants provided informed consent. Study activities took place in private clinic rooms with only the participant and assessor present.

We rotated the order of assessments (ie, interview, CAT-MH, brief questionnaires) weekly to reduce bias due to testing order. If a participant expressed active suicidal or homicidal ideation, plan, or intent, the assessor notified the patient's physician and study principal investigator (N.L.). Safety assessments were conducted, with follow-up as necessary. At the end of the study, participants received a referral list of mental health resources and \$10 gift card.

Measures

Demographics

Participants reported their age, sex, race, ethnicity, education, income level, and medical history.

Clinical Interview

Trained assessors (A.K.G. and A.M.) administered the SCID—Structured Clinical Interview for *DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition)*, Research Version—a semistructured interview, to assess for MDD and GAD.⁴⁵ The SCID is generally considered the reference-standard psychiatric assessment and has been used in past validation studies.^{42,46} Interviewers were not aware of the results of the other assessments.

Computerized Adaptive Test for Mental Health

Participants completed the CAT-MH without assistance using a tablet computer. They could both read and listen to the questions, and used the tablet touchscreen to provide responses; headphones were offered for privacy. The CAT-MH was delivered via a secure, Health Insurance Portability and Accountability Act (HIPAA)-compliant server.

The Computerized Adaptive Diagnostic Test for MDD (CAD-MDD) was administered to screen for MDD,³⁸ and the CAT-Depression Inventory (CAT-DI) was administered to assess depression severity among patients who screened positive for MDD on the former.³⁹ The CAD-MDD and CAT-DI select questions from an item bank of 389 possible questions. Questions are adaptively administered until a precise symptom severity estimate is achieved. The first question is selected randomly from the middle of the severity range; additional items are selected based on their information content conditional on the current severity score determined by the items already administered. CAT-DI scores range from 0 to 100 and are grouped as normal (less than 50), mild symptoms (50 to 65), moderate symptoms (66 to 75), and severe symptoms (higher than 75).³⁹

The CAT-Anxiety Inventory (CAT-ANX) assesses anxiety severity based on 431 possible questions. Scores range from 0 to 100 and are grouped as normal (less than 35), mild symptoms (35 to 50), moderate symptoms (51 to 65), and severe symptoms (higher than 65).⁴¹

The CAT-MH is distributed by Adaptive Testing Technologies, of which author R.D.G. is a founder. He contributed to the study design and writing, but was not responsible for data acquisition or analysis. Information regarding use of the CAT-MH is available from the company (<https://adaptivetestingtechnologies.com/>).

Brief Questionnaires

Participants self-administered brief questionnaires using paper and pen.

The 9-item Patient Health Questionnaire (PHQ-9) assesses MDD symptoms over the past 2 weeks and has been validated for screening and severity assessment in primary care.^{18,47,48} Scores range from 0 to 27; scores of 10 and higher indicate likely MDD.¹⁸ Scores

are grouped as normal or minimal symptoms (0 to 4), mild symptoms (5 to 9), moderate symptoms (10 to 14), moderately severe symptoms (15 to 19), and severe symptoms (20 and higher).¹⁸

The 2-item Patient Health Questionnaire (PHQ-2) contains the first 2 questions of the PHQ-9, which assess depressed mood and anhedonia over the past 2 weeks.^{17,47} The PHQ-2 has been validated in primary care, with a sensitivity of 0.61 and specificity of 0.92.^{17,49} Scores range from 0 to 6, and a score of 3 or higher is a common cutoff for indicating likely MDD.

The Generalized Anxiety Disorder 7-item Scale (GAD-7) assesses GAD symptoms, with good internal and test-retest reliability for detecting GAD in primary care.¹⁶ Scores range from 0 to 21; scores of 10 and higher indicate likely GAD.¹⁶ Scores are grouped as mild symptoms (5 to 9), moderate symptoms (10-14), and severe symptoms (15 and higher).¹⁶

We used a self-administered paper questionnaire to assess participant preference for screening delivery method. Participants were asked, "When answering questions about your mood, what format did you like best?" and "What format did you like least?" Response options to both questions were online (CAT-MH; on the tablet/computer), interview (SCID; with an assessor), and paper and pencil (questionnaires).

Statistical Analysis

To assess performance of the CAT-MH compared with that of the brief questionnaires, we needed to recruit 270 participants to achieve a 90% area under the curve (AUC) with a 5% margin of error.⁵⁰

We performed descriptive analyses, including tests of associations between SCID diagnoses and participant self-reported histories of MDD and GAD. The diagnostic performance of the CAT-MH and questionnaires was compared with that of the SCID using receiver operating characteristic curve analysis.⁵¹ Agreement between the CAT-MH and brief questionnaires for MDD and anxiety severity were compared using κ statistics.⁵² As κ statistics require an equal number of categories between variables, we calculated 2 κ values for MDD severity, by collapsing the fourth PHQ-9 severity category (moderately severe) into the third (moderate) or fifth (severe) category. These groupings were selected to merge clinically similar categories (eg, moderately severe and severe), not based on an empirical distribution of the responses.

We used logistic regression models to compare anxiety severity scores with SCID diagnoses and generate predicted probabilities of GAD for specific CAT-ANX scores. The Kruskal-Wallis H test and χ^2 test were used to assess associations between patients' preferred screening delivery method and demographics.

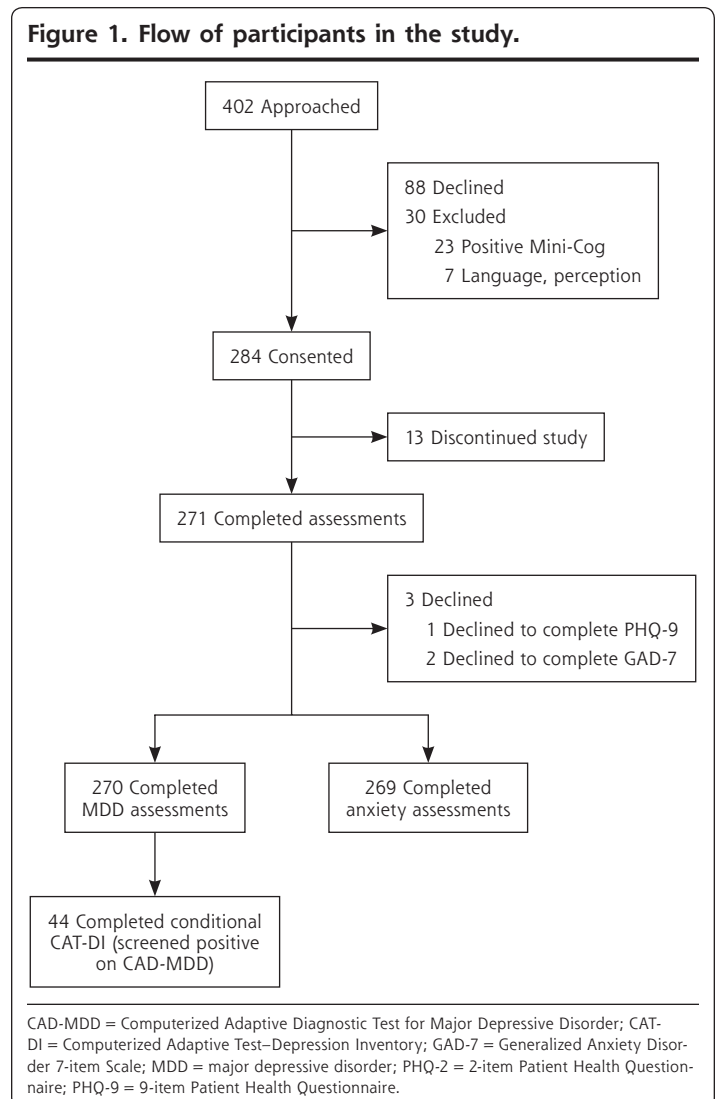
Analyses were conducted using SPSS version 22 (IBM Corp). Statistical significance was set at a 2-sided P value $< .05$.

RESULTS

Participant Characteristics

Figure 1 presents the study flow diagram. Of 402 patients approached, 271 (67%) completed the study assessments. Table 1 shows their sociodemographic and clinical characteristics, which reflect those of the overall clinic population.

Figure 1. Flow of participants in the study.



On the basis of the SCID, 31 participants met criteria for MDD and 29 met criteria for GAD. SCID-diagnosed MDD was associated with self-reported depression (odds ratio [OR]= 19.2; 95% CI, 7.8-46.7) and anxiety (OR= 14.3; 95% CI, 6.2-33.2). SCID-diagnosed GAD was associated with self-reported depression (OR= 9.1; 95% CI, 3.9-21.2) and anxiety (OR= 13.6; 95% CI, 5.7-32.5).

Screening for MDD

The CAD-MDD identified 42 participants as screening positive for likely MDD, whereas the PHQ-9 identified 37 participants. The CAD-MDD administered 4.2 (SD 0.5) questions on average (range, 4 to 6), and the median time to completion was 42 seconds (interquartile range [IQR]= 34 to 60).

With the SCID as the reference, accuracy of the CAD-MDD (AUC= 0.85) was similar to that of the PHQ-9 (AUC= 0.84) and higher than that of the PHQ-2 (AUC= 0.76) (Table 2). Agreement for MDD screening between the CAD-MDD and PHQ-9 ($\kappa= 0.66 \pm 0.07$; $P < .001$) was higher than agreement between the CAD-MDD and PHQ-2 ($\kappa= 0.45 \pm 0.08$; $P < .001$).

Assessing Depression Severity

The CAT-DI administered 7.6 (SD 1.9) questions on average (range, 5 to 15), and the median time to completion was 71 seconds (IQR= 52 to 93). CAT-DI scores strongly correlated with PHQ-9 scores ($r= 0.76$; $P < .001$). The CAT-DI and PHQ-9 severity levels had fair agreement, regardless of whether participants classified as having moderately severe symptoms by the PHQ-9 were grouped with those having moderate symptoms ($\kappa= 0.26 \pm 0.09$; $P = .001$) or with those having severe symptoms ($\kappa= 0.22 \pm 0.09$; $P = .008$).

Assessing Anxiety Severity

The CAT-ANX administered 11.8 (SD 4.1) questions on average (range, 5 to 22), and the median time to completion was 94 seconds (IQR= 67 to 150). Compared with the SCID, the CAT-ANX and the GAD-7 performed similarly well (AUC= 0.93 and 0.97, respectively) (Table 2). Participants' odds of SCID-diagnosed GAD increased with each 1-unit increase in CAT-ANX score (measured on a 100-point scale) (OR= 1.10;

95% CI, 1.07-1.13) and each 1-category increase in severity (normal, mild, moderate, severe) (OR= 6.4; 95% CI, 3.7-10.9). The CAT-ANX scores correlated with GAD-7 scores ($\rho= 0.74$; $P < .001$). There was fair agreement between severity classifications on the CAT-ANX and GAD-7 ($\kappa= 0.40 \pm 0.06$; $P < .001$). Figure 2 shows the probability that a patient would meet criteria for GAD based on the SCID given various CAT-ANX scores, as predicted by logistic regression analysis.

Preferred Screening Delivery Method

Participants preferred using the tablet computer most often (53%), followed by the interview (33%), and,

Table 1. Participant Sociodemographic and Clinical Characteristics

Characteristic	Total (N = 271)	Participants With MDD ^a (n = 31)	Participants With GAD ^a (n = 29)
Age, mean (SD), y	57 (17)	53 (12)	46 (14)
Female, No. (%)	191 (71)	26 (84)	24 (83)
Ethnicity, No. (%)			
Hispanic	11 (4)	0	0
Non-Hispanic	254 (94)	31 (100)	29 (100)
Prefer not to answer	6 (2)	0	0
Race, No. (%)			
Black	176 (65)	27 (87)	19 (66)
White	73 (27)	3 (10)	10 (34)
Other	8 (3)	0	0
Prefer not to answer	13 (5)	1 (3)	0
Level of education, No. (%)			
College degree or higher	136 (50)	7 (22)	11 (39)
Some college or junior college	83 (31)	12 (39)	8 (27)
High school graduate/GED	34 (12)	8 (26)	8 (27)
Some high school (grades 9-12) or less	18 (7)	4 (13)	2 (7)
Household income, No. (%)			
≥\$100,001	44 (16)	1 (4)	3 (10)
\$50,001-\$100,000	84 (31)	6 (20)	10 (35)
\$25,001-\$50,000	60 (22)	10 (33)	6 (21)
≤\$25,000	43 (16)	10 (33)	7 (24)
Prefer not to answer	40 (15)	3 (10)	3 (10)
Self-reported medical diagnoses, No. (%)			
Depression	54 (20)	23 (74)	17 (59)
Anxiety	39 (14)	18 (58)	16 (55)
Diabetes	62 (23)	8 (26)	7 (12)
Heart disease	42 (15)	5 (16)	4 (14)
Kidney disease	18 (7)	3 (10)	0
Liver disease	14 (5)	2 (6)	1 (3)
Stroke	14 (5)	5 (16)	4 (14)
Chronic pain	43 (16)	11 (35)	9 (31)
Non-skin cancer	11 (4)	0	0

DSM-5 = Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; GAD = generalized anxiety disorder; GED = general equivalency diploma; MDD = major depressive disorder.

^a Based on diagnoses generated from the Structured Clinical Interview for DSM-5 Disorders (SCID).

Table 2. Performance of the Computerized Adaptive Tests and Brief Questionnaires Relative to the Reference-Standard Clinical Interview

Measure	Test and Value(s)		
	CAD-MDD	PHQ-9	PHQ-2
Screening for MDD			
Sensitivity	0.77	0.75	0.58
Specificity	0.93	0.94	0.93
Positive predictive value	0.57	0.62	0.52
Negative predictive value	0.97	0.97	0.95
AUC (95% CI)	0.85 (0.76-0.94) ^a	0.84 (0.75-0.94) ^a	0.76 (0.65-0.87) ^a
Assessing anxiety severity	CAT-ANX	GAD-7	–
AUC (95% CI)	0.93 (0.90-0.97) ^a	0.97 (0.96-0.99) ^a	–
OR (95% CI) for 1-point increase in score	1.10 (1.07-1.13) ^a	1.58 (1.40-1.80) ^a	–
OR (95% CI) for 1-category increase in severity	6.37 (3.72-10.91) ^a	11.48 (5.76-22.88) ^a	–

AUC = area under the curve; CAD-MDD = Computerized Adaptive Diagnostic Test for Major Depressive Disorder; CAT-ANX = Computerized Adaptive Test–Anxiety Inventory; GAD-7 = Generalized Anxiety Disorder 7-item Scale; MDD = major depressive disorder; OR = odds ratio; PHQ-2 = 2-item Patient Health Questionnaire; PHQ-9 = 9-item Patient Health Questionnaire.

^a $P < .001$.

lastly, paper-and-pencil questionnaires (14%) (Supplemental Figure 1A, available at <http://www.annfammed.org/content/17/1/23/suppl/DC1/>). The majority of participants (64%) rated paper-and-pencil questionnaires as their least preferred screening method (Supplemental Figure 1B, available at <http://www.annfammed.org/content/17/1/23/suppl/DC1/>).

Screening by tablet was the preferred method among black individuals compared with nonblack individuals ($\chi^2 = 7.8$; $P = .02$). There was no association between preferred screening method and age, sex, education, income, self-reported depression or anxiety, or SCID-diagnosed depression or anxiety.

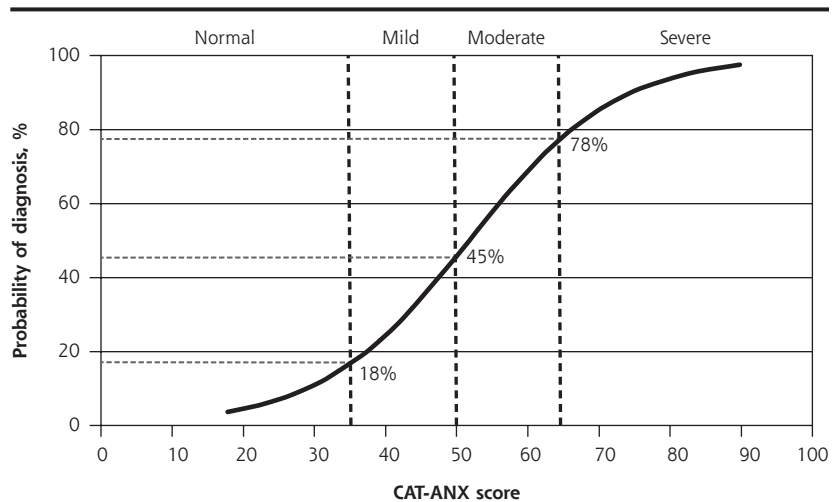
DISCUSSION

To improve depression and anxiety detection and management in primary care, efficient and accurate screening tools are essential. We evaluated CATs among adult primary care patients and demonstrated that the CAT-MH is a valid instrument for screening for MDD and assessing depression and anxiety severity compared with reference-standard interviews. Also, the CAT-MH had

higher accuracy than the commonly used PHQ-2 for depression screening.^{15,49} Participants preferred delivery by tablet computer over interview and paper-based questionnaires, highlighting the acceptability of this screening approach.

The CAD-MDD performance was comparable to that of the PHQ-9 for MDD screening and administered fewer questions on average (4 vs 9). The CAT-ANX performance also was comparable to that of the GAD-7 for assessing anxiety but required more questions (12 vs 7). The CAD-MDD outperformed the PHQ-2 for screening for MDD and required only 2 additional questions on average. As the CAD-MDD median completion time was 42

seconds, efficiency was not sacrificed. Compared with past CAD-MDD validation studies,^{38,42} this study showed lower sensitivity but higher specificity, which may be due to a lower prevalence of MDD in primary care than in psychiatric clinics. Our results also may differ because the clinic population has a much higher proportion of black individuals (65%) compared with past CAT-MH study populations (eg, 10%³⁸ and 5%⁴²). Future work to tai-

Figure 2. Predicted probability of a GAD diagnosis given CAT-ANX score.

CAT-ANX = Computerized Adaptive Test–Anxiety Inventory; GAD = Generalized Anxiety Disorder; SCID = Structured Clinical Interview for DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition) Disorders, Research Version.

Note: Figure shows the probability of a GAD diagnosis on the SCID for given cutoff scores on the CAT-ANX, as predicted using logistic regression modeling. For example, a patient with moderate anxiety on the CAT-ANX (a score of 50) has a 45% probability of receiving a GAD diagnosis via the SCID.

lor the item bank questions, the algorithm, or both for primary care patients may improve the accuracy and efficiency of the CAT-MH.

CATs using multidimensional item-response theory and cloud-based assessments offer potential advantages over traditional written assessments for use in medicine. Ideally, the growing integration of EHRs in primary care¹⁹ could enable implementation of CATs in clinical practice.^{22,23,25,53-56} Patients' test responses could be added to EHRs in real time and to searchable forms to automate development of disease-specific population registries. The online format can incorporate modules for additional mental health concerns (eg, suicidality⁵⁷) and can be modified in real time. Using cloud-based assessments may enhance possibilities for patients to self-administer these tools, including outside of the clinic.⁵⁸ Because clinicians can immediately access patients' responses, physicians can monitor symptom changes without necessarily requiring in-person visits. Further, as the same questions on the CAT-MH are not repeatedly administered, patients can be routinely assessed in or out of the clinic without producing response bias due to repeated administration of the same questions using traditional instruments.⁴⁰ The CAT-MH was developed using multidimensional item response theory, which permits measurement of complex traits such as depression and anxiety, and allows for much larger item banks than CATs based on unidimensional item response theory. These features offer advantages over other electronic tools that have been tested in primary care, such as the PsyScan e-tool,⁵⁹ Patient-Reported Outcome Measure Information System (PROMIS) symptom measures (for which CATs are available),^{54,60,61} and the Adaptive Pediatric Symptom Checklist.⁶²

The potential impact of health-related technologies for improving mental health assessment may be tempered by their limited integration in EHRs, however, as EHRs in many health care systems are not yet capable of integrating stand-alone programs. Also, in general, the impact of screening tools on mental health outcomes is limited because clinician assessment is necessary for diagnosis.

Limitations of this study should be noted. Our sample was disproportionately female, black, and well educated (with one-half having a college degree or higher). Replication in other populations and non-US samples is warranted and would allow for detailed analyses of severity scores. We recruited a small number of participants in accordance with our sample size calculation, which precludes analyzing findings among subpopulations. Lastly, we compared the CAT-MH with paper questionnaires, rather than with electronic questionnaires, because the former are commonly used in clinical practice.

Participant preferences may have been influenced by delivery mode differences, however.

In conclusion, CATs could be a valid, highly efficient, and patient-centered approach for depression and anxiety screening and assessment in primary care patients. In this first validation study in primary care, the CAT-MH had similar diagnostic accuracy and correlated well with the PHQ-9 and GAD-7, and was delivered in a way that patients preferred.

To read or post commentaries in response to this article, see it online at <http://www.AnnFamMed.org/content/17/1/23>.

Key words: screening; depression; anxiety; mental health; symptom assessment; surveys and questionnaires; health informatics; electronic health records; vulnerable populations; primary care; practice-based research

Submitted December 7, 2017; submitted, revised, August 21, 2018; accepted September 10, 2018.

Funding support: This work was supported by the National Institutes of Health (grants K23 DK097283, R01 MH100155, R01 MH66302, and F32 HD089586), the Agency for Healthcare Research & Quality (grant T32 HS000078), and an Innovation Award from the University of Chicago Medicine.

Previous presentations: Portions of this work have been presented at the Society for General Internal Medicine Annual Meeting; April 19-22, 2017; Washington, DC; and the Midwest Society for General Internal Medicine Regional Meeting; September 14-15, 2017; Chicago, Illinois.

Supplemental Materials: Available at <http://www.AnnFamMed.org/content/17/1/23/suppl/DC1>.

References

1. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005;62(6):617-627.
2. Wang PS, Lane M, Olfson M, Pincus HA, Wells KB, Kessler RC. Twelve-month use of mental health services in the United States: results from the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005;62(6):629-640.
3. Wang PS, Demler O, Olfson M, Pincus HA, Wells KB, Kessler RC. Changing profiles of service sectors used for mental health care in the United States. *Am J Psychiatry*. 2006;163(7):1187-1198.
4. Cepoiu M, McCusker J, Cole MG, Sewitch M, Belzile E, Ciampi A. Recognition of depression by non-psychiatric physicians—a systematic literature review and meta-analysis. *J Gen Intern Med*. 2008; 23(1):25-36.
5. Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. *Lancet*. 2009;374(9690):609-619.
6. Cunningham PJ. Beyond parity: primary care physicians' perspectives on access to mental health care. *Health Aff (Millwood)*. 2009; 28(3):w490-w501.
7. Trude S, Stoddard JJ. Referral gridlock: primary care physicians and mental health services. *J Gen Intern Med*. 2003;18(6):442-449.
8. Loeppke R, Taitel M, Haufler V, Parry T, Kessler RC, Jinnett K. Health and productivity as a business strategy: a multiemployer study. *J Occup Environ Med*. 2009;51(4):411-428.

9. Herrman H, Patrick DL, Diehr P, et al. Longitudinal investigation of depression outcomes in primary care in six countries: the LIDO study. Functional status, health service use and treatment of people with depressive symptoms. *Psychol Med*. 2002;32(5):889-902.
10. Whooley MA, Simon GE. Managing depression in medical outpatients. *N Engl J Med*. 2000;343(26):1942-1950.
11. Rost K, Zhang M, Fortney J, Smith J, Coyne J, Smith GR Jr. Persistently poor outcomes of undetected major depression in primary care. *Gen Hosp Psychiatry*. 1998;20(1):12-20.
12. Merikangas KR, Ames M, Cui L, et al. The impact of comorbidity of mental and physical conditions on role disability in the US adult household population. *Arch Gen Psychiatry*. 2007;64(10):1180-1188.
13. Simon G, Ormel J, VonKorff M, Barlow W. Health care costs associated with depressive and anxiety disorders in primary care. *Am J Psychiatry*. 1995;152(3):352-357.
14. Siu AL, Bibbins-Domingo K, Grossman DC, et al; US Preventive Services Task Force (USPSTF). Screening for depression in adults: US Preventive Services Task Force recommendation statement. *JAMA*. 2016;315(4):380-387.
15. Whooley MA. Screening for depression—a tale of two questions. *JAMA Intern Med*. 2016;176(4):436-438.
16. 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.
17. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care*. 2003;41(11):1284-1292.
18. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606-613.
19. Office of the National Coordinator for Health Information Technology. Office-based physician electronic health record adoption, health IT quick-stat #50. <https://dashboard.healthit.gov/quickstats/pages/physician-ehr-adoption-trends.php>. Published Dec 2016. Accessed May 9, 2017.
20. Gershon RC, Rothrock N, Hanrahan R, Bass M, Cella D. The use of PROMIS and assessment center to deliver patient-reported outcome measures in clinical research. *J Appl Meas*. 2010;11(3):304-314.
21. Cella D, Riley W, Stone A, et al.; PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol*. 2010; 63(11):1179-1194.
22. Pilkonis PA, Yu L, Dodds NE, Johnston KL, Maihoefer CC, Lawrence SM. Validation of the depression item bank from the Patient-Reported Outcomes Measurement Information System (PROMIS) in a three-month observational study. *J Psychiatr Res*. 2014;56:112-119.
23. Wagner LI, Schink J, Bass M, et al. Bringing PROMIS to practice: brief and precise symptom screening in ambulatory cancer care. *Cancer*. 2015;121(6):927-934.
24. Riley WT, Pilkonis P, Cella D. Application of the National Institutes of Health Patient-reported Outcome Measurement Information System (PROMIS) to mental health research. *J Ment Health Policy Econ*. 2011;14(4):201-208.
25. Jensen RE, Rothrock NE, DeWitt EM, et al. The role of technical advances in the adoption and integration of patient-reported outcomes in clinical care. *Med Care*. 2015;53(2):153-159.
26. Vilagut G, Forero CG, Adroher ND, Ollariu E, Cella D, Alonso J; INSAYD investigators. Testing the PROMIS Depression measures for monitoring depression in a clinical sample outside the US. *J Psychiatr Res*. 2015;68:140-150.
27. Gibbons RD, Weiss DJ, Frank E, Kupfer D. Computerized adaptive diagnosis and testing of mental health disorders. *Annu Rev Clin Psychol*. 2016;12:83-104.
28. Sunderland M, Batterham P, Carragher N, Caleare A, Slade T. Developing and validating a computerized adaptive test to measure broad and specific factors of internalizing in a community sample. *Assessment*. <https://doi.org/10.1177/1073191117707817>.
29. Smits N, Cuijpers P, van Straten A. Applying computerized adaptive testing to the CES-D scale: a simulation study. *Psychiatry Res*. 2011; 188(1):147-155.
30. Magnee T, de Beurs DP, Terluin B, Verhaak PF. Applying computerized adaptive testing to the Four-Dimensional Symptom Questionnaire (4DSQ): a simulation study. *JMIR Ment Health*. 2017;4(1):e7.
31. Fliege H, Becker J, Walter OB, Rose M, Bjorner JB, Klapp BF. Evaluation of a computer-adaptive test for the assessment of depression (D-CAT) in clinical application. *Int J Methods Psychiatr Res*. 2009; 18(1):23-36.
32. Flens G, Smits N, Carlier I, van Hemert AM, de Beurs E. Simulating computer adaptive testing with the Mood and Anxiety Symptom Questionnaire. *Psychol Assess*. 2016;28(8):953-962.
33. Devine J, Fliege H, Kocalevent R, Mierke A, Klapp BF, Rose M. Evaluation of computerized adaptive tests (CATs) for longitudinal monitoring of depression, anxiety, and stress reactions. *J Affect Disord*. 2016;190:846-853.
34. Becker J, Fliege H, Kocalevent RD, et al. Functioning and validity of a computerized adaptive test to measure anxiety (A-CAT). *Depress Anxiety*. 2008;25(12):E182-E194.
35. Loe BS, Stillwell D, Gibbons C. Computerized Adaptive Testing provides reliable and efficient depression measurement using the CES-D scale. *J Med Internet Res*. 2017;19(9):e302.
36. Walter OB, Becker J, Bjorner JB, Fliege H, Klapp BF, Rose M. Development and evaluation of a computer adaptive test for 'Anxiety' (Anxiety-CAT). *Qual Life Res*. 2007;16(Suppl 1):143-155.
37. Stochl J, Böhnke JR, Pickett KE, Croudace TJ. An evaluation of computerized adaptive testing for general psychological distress: combining GHQ-12 and Affectometer-2 in an item bank for public mental health research. *BMC Med Res Methodol*. 2016;16:58.
38. Gibbons RD, Hooker G, Finkelman MD, et al. The computerized adaptive diagnostic test for major depressive disorder (CAD-MDD): a screening tool for depression. *J Clin Psychiatry*. 2013;74(7): 669-674.
39. Gibbons RD, Weiss DJ, Pilkonis PA, et al. Development of a computerized adaptive test for depression. *Arch Gen Psychiatry*. 2012;69(11): 1104-1112.
40. Beiser D, Vu M, Gibbons R. Test-retest reliability of a computerized adaptive depression screener. *Psychiatr Serv*. 2016;67(9):1039-1041.
41. Gibbons RD, Weiss DJ, Pilkonis PA, et al. Development of the CAT-ANX: a computerized adaptive test for anxiety. *Am J Psychiatry*. 2014;171(2):187-194.
42. Achtyes ED, Halstead S, Smart L, et al. Validation of computerized adaptive testing in an outpatient nonacademic setting: the VOCATIONS trial. *Psychiatr Serv*. 2015;66(10):1091-1096.
43. Borson S, Scanlan J, Brush M, Vitaliano P, Dokmak A. The mini-cog: a cognitive 'vital signs' measure for dementia screening in multi-lingual elderly. *Int J Geriatr Psychiatry*. 2000;15(11):1021-1027.
44. Borson S, Scanlan JM, Watanabe J, Tu SP, Lessig M. Improving identification of cognitive impairment in primary care. *Int J Geriatr Psychiatry*. 2006;21(4):349-355.
45. First MB, Williams JBW, Karg RS, Spitzer RL. *Structured clinical interview for DSM-5—Research Version (SCID-5 for DSM-5, Research Version; SCID-5-RV)*. Arlington, VA: American Psychiatric Association; 2015.
46. Stuart AL, Pasco JA, Jacka FN, Brennan SL, Berk M, Williams LJ. Comparison of self-report and structured clinical interview in the identification of depression. *Compr Psychiatry*. 2014;55(4):866-869.
47. Kroenke K, Spitzer RL, Williams JB, Löwe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. *Gen Hosp Psychiatry*. 2010;32(4):345-359.

48. Huang FY, Chung H, Kroenke K, Delucchi KL, Spitzer RL. Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. *J Gen Intern Med*. 2006;21(6):547-552.
49. Mitchell AJ, Yadegarfar M, Gill J, Stubbs B. Case finding and screening clinical utility of the Patient Health Questionnaire (PHQ-9 and PHQ-2) for depression in primary care: a diagnostic meta-analysis of 40 studies. *BJPsych Open*. 2016;2(2):127-138.
50. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. *J Biomed Inform*. 2014;48:193-204.
51. McFall RM, Treat TA. Quantifying the information value of clinical assessments with signal detection theory. *Annu Rev Psychol*. 1999; 50:215-241.
52. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med*. 2005;37(5):360-363.
53. Khanna D, Maranian P, Rothrock N, et al. Feasibility and construct validity of PROMIS and "legacy" instruments in an academic scleroderma clinic. *Value Health*. 2012;15(1):128-134.
54. Hinami K, Smith J, Deamant CD, Kee R, Garcia D, Trick WE. Health perceptions and symptom burden in primary care: measuring health using audio computer-assisted self-interviews. *Qual Life Res*. 2015;24(7):1575-1583.
55. Bruckel J, Wagle N, O'Brien C, et al. Feasibility of a tablet computer system to collect patient-reported symptom severity in patients undergoing diagnostic coronary angiography. *Crit Pathw Cardiol*. 2015;14(4):139-145.
56. Schalet BD, Pilkonis PA, Yu L, et al. Clinical validity of PROMIS Depression, Anxiety, and Anger across diverse clinical samples. *J Clin Epidemiol*. 2016;73:119-127.
57. Gibbons RD, Kupfer D, Frank E, Moore T, Beiser DG, Boudreaux ED. Development of a computerized adaptive test suicide scale-The CAT-SS. *J Clin Psychiatry*. 2017;78(9):1376-1382.
58. Marcano-Belisario JS, Gupta AK, O'Donoghue J, Ramchandani P, Morrison C, Car J. Implementation of depression screening in antenatal clinics through tablet computers: results of a feasibility study. *BMC Med Inform Decis Mak*. 2017;17(1):59.
59. Gidding LG, Spigt M, Winkens B, Herijgers O, Dinant GJ. PsyScan e-tool to support diagnosis and management of psychological problems in general practice: a randomised controlled trial. *Br J Gen Pract*. 2018;68(666):e18-e27.
60. Kroenke K, Talib TL, Stump TE, et al. Incorporating PROMIS symptom measures into primary care practice—a randomized clinical trial. *J Gen Intern Med*. 2018;33(8):1245-1252.
61. Bryan S, Davis J, Broesch J, et al. Choosing your partner for the PROM: a review of evidence on patient-reported outcome measures for use in primary and community care. *Healthc Policy*. 2014;10(2): 38-51.
62. Gardner W, Kelleher KJ, Pajer KA. Multidimensional adaptive testing for mental health problems in primary care. *Med Care*. 2002; 40(9):812-823.