

Validation of Computerized Adaptive Testing in an Outpatient Nonacademic Setting: The VOCATIONS Trial

Eric D. Achtyes, M.D., M.S., Scott Halstead, Ph.D., LeAnn Smart, B.A., Tara Moore, M.A., M.P.H., Ellen Frank, Ph.D., David J. Kupfer, M.D., Robert Gibbons, Ph.D.

Objective: Computerized adaptive testing (CAT) provides an alternative to fixed-length assessments. The study validated a suite of computerized adaptive tests for mental health (CAT-MH) in a community psychiatric sample.

Methods: A total of 145 adults from a community outpatient clinic, including 19 with no history of a mental disorder (control group), were prospectively evaluated with CAT for depression (CAD-MDD and CAT-DI), mania (CAT-MANIA), and anxiety symptoms (CAT-ANX). Ratings were compared with gold-standard psychiatric assessments, including the Structured Clinical Interview for DSM-IV-TR (SCID), Hamilton Rating Scale for Depression (HAM-D-25), Patient Health Questionnaire (PHQ-9), Center for Epidemiologic Studies Depression Scale (CES-D), and Global Assessment of Functioning (GAF).

Results: Sensitivity and specificity for CAD-MDD were .96 and .64, respectively (.96 and 1.00 for major depression

versus the control group). CAT for depression severity (CAT-DI) correlated well with the HAM-D-25 ($r=.79$), PHQ-9 ($r=.90$), and CES-D ($r=.90$) and had an odds ratio (OR) of 27.88 across its range for current SCID major depressive disorder. CAT-ANX correlated with the HAM-D-25 ($r=.73$), PHQ-9 ($r=.78$), and CES-D ($r=.81$) and had an OR of 11.52 across its range for current SCID generalized anxiety disorder. CAT-MANIA did not correlate well with the HAM-D-25 ($r=.31$), PHQ-9 ($r=.37$), and CES-D ($r=.39$), but it had an OR of 11.56 across its range for a current SCID bipolar diagnosis. Participants found the CAT-MH acceptable and easy to use, averaging 51.7 items and 9.4 minutes to complete the full battery.

Conclusions: Compared with gold-standard diagnostic and assessment measures, CAT-MH provided an effective, rapidly administered assessment of psychiatric symptoms.

Psychiatric Services 2015; 66:1091–1096; doi: 10.1176/appi.ps.201400390

With expansion of Medicaid eligibility and passage of the Affordable Care Act, there is additional pressure on the mental health care system to efficiently and effectively provide mental health assessment and treatment for millions of additional people seeking care. As measurement-based care becomes the standard for assessment of illness severity and improvement with treatment, well-validated, affordable, and quick measures are needed to help busy clinicians treat patients rapidly and effectively.

Computerized adaptive diagnosis (CAD) and computerized adaptive testing (CAT) have the potential to provide rapid, systematic testing on a population level (1,2). The paradigm shift between traditional fixed-length tests and adaptive tests is that traditional tests fix the items and allow the measurement precision to vary, whereas adaptive tests fix measurement precision and allow the items to vary. The net result is that it is possible to extract the relevant information contained in a bank of hundreds of symptom-related questions by using only a small number of optimal items for each person. Depending on the application, the degree of required precision can be selected a priori, so that national screening programs can use less precision than clinic screening, which

in turn may require less precision than a randomized clinical trial.

Application of CAT differs from standard assessments of symptom severity in several important ways. First, traditional scales may be hampered by a “practice effect,” which results from retaking the same measure repeatedly over time. Because CAT adapts to the current severity level of a patient, these practice effects are eliminated because the patient receives different items each time the test is administered. Second, for repeated assessments, traditional tests make no use of the information contained in the preceding test administrations. By contrast, in CAT, the last CAT-based severity measure can be used to start the next CAT, selecting the next most informative item conditional on the estimated severity level from previous sessions. Third, traditional measurement provides a score (typically the sum of the item scores) but no estimate of uncertainty in the score for a given patient. The standard approach of computing a total score also adds potential bias because items with different numbers of response categories (for example, the Hamilton Rating Scale for Depression [HAM-D-25])

are weighted differently when computing a total score (that is, an item with two categories receives less weight than an item with five categories). Because CAT is based on an underlying statistical model of measurement (item response theory [IRT]), the number of categories no longer differentially weights the importance of the item in computing the severity score, and each estimated score has a corresponding uncertainty estimate. IRT produces the estimate of uncertainty, and CAT mandates that all patients are tested until they achieve a desired level of uncertainty; hence all patients are tested with the same level of precision. Traditional tests lack this desirable statistical property. [An online supplement to this article presents further explanation of CAT and IRT principles.]

It is also important to note that severity measurement and diagnosis are two very different operations. In severity measurement, we seek to maximize information surrounding the symptom severity of the patient. In diagnosis, we seek to maximize information at the threshold above which the probability of the diagnosis exceeds 50%. Gibbons and colleagues (3) have developed a computerized adaptive diagnostic screener for depression (CAD-MDD). They found that the CAD-MDD could ascertain a diagnosis of major depressive disorder with sensitivity of .95 and specificity of .87 by using an average of four questions and taking less than one minute to administer (mean of 46 ± 29 seconds), making it an exceedingly rapid and effective screener.

If shown to be valid across a wide variety of patient populations, these tools could fill a key void, allowing automated testing of millions of people with a quick, easily administered online tool. Standard scales such as the nine-item Patient Health Questionnaire (PHQ-9) have been validated in a wide range of treatment settings. The CAT for depression severity (CAT-DI), the CAD-MDD (depression diagnostic screener), and CAT-ANX (anxiety severity) have been validated previously in both an academic and a nonpsychiatric community hospital. To assess the validity and potential impact of these tests on general outpatient community psychiatric practice, as well as to provide initial validation of the CAT-MANIA (mania severity), we sought in this study to validate the utility of the CAT-MH (mental health) suite of tests in a nonacademic, community sample of adult psychiatric outpatients.

METHODS

Item Bank and Original Calibration Sample

The original studies developed a 1,008-item question bank consisting of 452 depression items, 467 anxiety items, and 89 mania items (1–4). Separate CATs were developed for each of these three primary domains. The items were selected on the basis of a review of more than 100 existing depression or depression-related rating scales, with most items modified to refer to the previous two-week time period and self-rated on a 5-point ordinal scale. These tools and methods have been described in detail elsewhere (1–11) and have been previously validated in an academic center (University of Pittsburgh psychiatric clinics) and a nonpsychiatric community general medical hospital (DuBois Regional Medical Center).

Validation Sample

The VOCATIONS trial (Validation of Computerized Adaptive Testing in an Outpatient Nonacademic Setting) was designed as a prospective cross-sectional validation study of the CAT-MH suite of tests and was conducted between April 18, 2012, and March 29, 2013, at the outpatient clinics of Pine Rest Christian Mental Health Services, located in Grand Rapids, Michigan. Pine Rest is a large, not-for-profit, free-standing psychiatric system with a spectrum of comprehensive psychiatric services ranging from inpatient to partial hospitalization, including a network of outpatient clinics in the surrounding community. In the population served by Pine Rest outpatient clinics, 64% of patients have a commercial insurance plan, 12% are self-pay, 12% are covered by Medicare, and 12% are covered by community mental health contracts (that is, uninsured) or by Medicaid. This study was conducted in compliance with the ethical principles of the Declaration of Helsinki, the U.S. Food and Drug Administration guidelines, and the International Conference on Harmonization's Good Clinical Practices Guidelines. The Human Participants Review Board at Mercy Health Saint Mary's approved the study, and individuals signed a written informed consent form prior to initiation of any study procedures.

Participants were a convenience sample of women and men, ages 18–70, who presented to Pine Rest Christian Mental Health Services clinics seeking care and a control sample of adults with no current or past history of a mental disorder. Participants were recruited using institutional review board–approved advertisements in clinic waiting rooms and on the Pine Rest Web site. Patients had to be willing and able to provide written informed consent in order to participate. Exclusion criteria were schizophrenia, schizoaffective disorder, or other psychotic disorder; organic mood disorder due to a general medical condition or a substance use disorder; drug or alcohol dependence in the prior three months; severity of illness sufficient to require inpatient hospitalization because of suicide risk or psychosis; and Alzheimer's or Parkinson's disease.

Upon signing informed consent, participants were administered the following assessments by trained raters blinded to the patients' clinical diagnoses prior to evaluation: Structured Clinical Interview for DSM-IV-TR (SCID) (12), the HAM-D-25 (13), PHQ-9 (14), Center for Epidemiologic Studies Depression Scale (CES-D) (15), Global Assessment of Functioning (GAF) (16), a questionnaire about demographic characteristics, and a study participation evaluation. Participants also took the most recent version of the CAT-MH, which contains the depression, anxiety, and mania-hypomania components of the entire 1,008-item bank, including the CAD-MDD for current depression diagnosis, CAT-DI for current depression severity, CAT-ANX for current anxiety severity, and CAT-MANIA for current manic-hypomanic symptom severity. CAT-MH depression, anxiety, and mania scores were correlated with SCID, HAM-D-25, CES-D, and PHQ-9 scores and with *DSM-IV-TR* cases of depression, anxiety, and bipolar disorders.

Statistical Methods

Sample size computations were conducted to determine the ability to find significant differences in sensitivity and specificity between

the original findings for the CAD-MDD and the results of this validation study. Assuming a type I error rate of 5% and power of 80%, $N=150$ permits detection of approximately 10% differences in sensitivity (.95 versus .86) and specificity (.87 versus .75).

Data analysis was performed by the senior author (RG) at the University of Chicago. The goal was to test the reproducibility of previous analyses of sensitivity, specificity, and correlation with gold-standard symptom severity scales (HAM-D, CES-D, and PHQ-9) in this community sample. Logistic regression was used to examine relationships between severity scores and the presence or absence of *DSM-IV-TR* diagnoses.

RESULTS

Participants

A total of 150 patients provided written informed consent. Four did not meet inclusion criteria, and one withdrew consent. A total of 145 patients completed all testing and were included in the analysis. [A CONSORT diagram in the online supplement provides additional details on sample recruitment.]

Patient Demographic Characteristics

Of the 145 adult patients in the sample, 79% were female, 10% were Hispanic, 90% were Caucasian, 5% were African American, 3% were Asian, and 3% indicated other race. In addition, 58% were married, 24% were never married, 5% were living with a partner, and the remainder were divorced (10%), separated (2%), or widowed (<1%). In terms of education, 40% had a college degree or higher, 42% had some college, and 16% had graduated from high school or had a GED (Table 1).

Diagnoses

In terms of current *DSM-IV-TR* diagnoses, 27 of the 145 patients had major depressive disorder, 27 had generalized anxiety disorder, 13 had bipolar I disorder, 11 had bipolar II disorder, 15 had dysthymic disorder, and 16 had panic disorder. Other diagnoses are shown in Table 2. Many patients had comorbid disorders, which explains why the sum of diagnoses exceeds the sample size. Nineteen of the 145 participants had no current or past history of a *DSM-IV-TR* diagnosis (control group).

CAD-MDD: A Diagnostic Screen for Major Depression

Given the high degree of pathology and comorbidity in the sample, it was expected that the high sensitivity seen in other studies would be replicated, but with lower specificity. This was found in the overall sample, where sensitivity was .96 (.95 in the original CAD-MDD study) and specificity was .64 (.87 in the original CAD-MDD study (3), which included a much greater number and proportion of individuals with no current or past *DSM-IV-TR* diagnoses). However, when the sample was restricted to patients meeting *DSM-IV-TR* criteria for major depressive disorder in the past month and individuals with no current or past *DSM-IV-TR* diagnoses, sensitivity remained at .96, but specificity increased to 1.00 (that is, there were no false positives and only one false negative in a total of 46 patients). These results are consistent with what would be

TABLE 1. Demographic characteristics of 145 adults in a community outpatient psychiatric sample

| Characteristic | N | % |
|----------------------------|-----|----|
| Age group | | |
| 18–29 | 24 | 16 |
| 30–39 | 25 | 17 |
| 40–49 | 49 | 34 |
| 50–59 | 33 | 23 |
| 60–70 | 14 | 10 |
| Gender | | |
| Male | 31 | 21 |
| Female | 114 | 79 |
| Race | | |
| Caucasian | 130 | 90 |
| African American | 7 | 5 |
| Asian | 4 | 3 |
| Other | 4 | 3 |
| Hispanic | 14 | 10 |
| Marital status | | |
| Married | 84 | 58 |
| Never married | 34 | 24 |
| Living with partner | 8 | 5 |
| Divorced | 15 | 10 |
| Separated | 3 | 2 |
| Widowed | 1 | <1 |
| Education | | |
| College degree or more | 57 | 40 |
| Some college | 61 | 42 |
| High school diploma or GED | 24 | 16 |
| Some high school | 3 | 2 |

expected in a primary care setting, where the majority of patients would not meet criteria for a *DSM-IV-TR* major depressive disorder (17,18). These results were achieved with an average of 4.1 questions, which took 36.1 seconds to complete.

CAT-DI: Depression Severity Measure

The dimensional measure of depressive severity (CAT-DI) demonstrated correlations with traditional scales, such as the HAM-D-25 ($r=.79$), PHQ-9 ($r=.90$), CES-D ($r=.90$), and GAF ($r=-.70$) (Table 3). The CAT-DI correlated highly with the CAT-ANX ($r=.82$) but less so with the CAT-MANIA ($r=.38$). In terms of its relationship with current *DSM-IV-TR* major depressive disorder diagnosis, the CAT-DI had an odds ratio (OR) of 6.97 ($p<.001$). This means that for every unit increase in CAT-DI score, the likelihood of a current *DSM-IV-TR* major depressive disorder diagnosis increased sevenfold. Given that the range of scores on the CAT-DI is from -2 to 2 , the actual span gives an OR of 27.88, a 28-fold increase in probability of major depressive disorder from the low to the high end of the CAT-DI scale. This scale took an average of 16.8 items and 3.4 minutes to complete.

CAT-ANX: Anxiety Severity Measure

The dimensional measure of anxiety severity (CAT-ANX) demonstrated correlations with traditional scales, such as the HAM-D-25 ($r=.73$), PHQ-9 ($r=.78$), CES-D ($r=.81$), and GAF ($r=-.68$) (Table 3). These results indicate that depression and anxiety have considerable overlap, which is known to be true neurobiologically and is also observed clinically

TABLE 2. Diagnostic prevalence (N of patients) among 145 adults in a community outpatient psychiatric sample^a

| Diagnosis | Total current and lifetime diagnoses | Current | | |
|---|--------------------------------------|---------------------|----------------------|-------------------------------|
| | | Meets full criteria | In partial remission | Asymptomatic (full remission) |
| Bipolar I disorder | 20 | 13 | 3 | 4 |
| Bipolar II disorder | 20 | 11 | 7 | 2 |
| Bipolar disorder, not otherwise specified (NOS) | 3 | 1 | 0 | 2 |
| Major depressive disorder | 62 | 27 | 15 | 20 |
| Dysthymic disorder | 15 | 15 | 0 | 0 |
| Depressive disorder NOS | 6 | 2 | 0 | 4 |
| Generalized anxiety disorder | 27 | 27 | 0 | 0 |
| Panic disorder | 34 | 16 | 12 | 6 |
| Agoraphobia | 8 | 6 | 2 | 0 |
| Social phobia | 17 | 13 | 4 | 0 |
| Obsessive-compulsive disorder | 15 | 11 | 1 | 3 |
| Specific phobia | 16 | 9 | 5 | 2 |
| Posttraumatic stress disorder | 33 | 12 | 17 | 4 |
| Anxiety disorder NOS | 19 | 15 | 0 | 4 |

^a Diagnoses based on the Structured Clinical Interview for DSM-IV-TR

(19). The CAT-ANX correlated highly with the CAT-DI ($r=.82$) but less so with the CAT-MANIA ($r=.47$). In terms of its relationship with current *DSM-IV-TR* generalized anxiety disorder diagnosis, the CAT-ANX had an OR of 2.88 ($p<.001$). Given that the range of scores on the CAT-ANX is from -2 to 2 , the actual span gives an OR of 11.52, a 12-fold increase in probability of generalized anxiety disorder from the low to the high end of the scale. This scale took an average of 12.9 items and 2.0 minutes to complete.

CAT-MANIA: Mania Severity Measure

The dimensional measure of the hypomania-manic spectrum (CAT-MANIA) demonstrated relatively low correlations with traditional scales, as expected: HAM-D-25 ($r=.31$), PHQ-9 ($r=.37$), CES-D ($r=.39$), and GAF ($r=-.29$) (Table 3). These results indicate that depression and mania have limited overlap, at least at a single point in time, which has been confirmed clinically: depressive and manic symptoms often co-occur, but true mixed states as defined by *DSM-IV-TR* are uncommon (20,21).

This scale took an average of 17.9 items and 3.4 minutes to complete.

Patient Impressions of Usability of the CAT-MH

Participants took, on average, 51.7 items and 9.4 minutes to complete the entire CAT-MH. As summarized in Table 4, patients found the computerized adaptive tests easy overall and acceptable to use, felt comfortable answering personal questions about themselves, answered them honestly, preferred computerized adaptive tests over a pencil-and-paper test, and felt the test accurately reflected their mood. There was some concern that older patients would not find the computerized test as easy to take. This was not found to be the case; correlations to age ranged from .22 to .35.

DISCUSSION

This was the first prospective, cross-sectional study to validate the CAT-MH suite of tests, including the CAT-MANIA scale, in

TABLE 3. Correlation of CAT-MH tests with each other and with four traditional scales, odds ratios for diagnoses, and number of items and completion time among 145 adults^a

| CAT-MH test | Correlation | | | | | | | Corresponding <i>DSM</i> diagnosis | | | | Average N to complete | |
|-------------|-------------|-------|-------|------|---------|-----------|--------|------------------------------------|-----------------|------------|-------|-----------------------|---------|
| | HAM-D-25 | PHQ-9 | CES-D | GAF | CAT-ANX | CAT-MANIA | CAT-DI | Diagnosis | OR ^b | 95% CI | p | Items | Minutes |
| CAT-DI | .79 | .90 | .90 | -.70 | .82 | .38 | — | Major depressive disorder | 6.97 | 3.14–15.51 | <.001 | 16.8 | 3.4 |
| CAT-ANX | .73 | .78 | .81 | -.68 | — | .47 | .82 | Generalized anxiety disorder | 2.88 | 1.72–4.83 | <.001 | 12.9 | 2.0 |
| CAT-MANIA | .31 | .37 | .39 | -.29 | .47 | — | .38 | Bipolar disorder | 2.89 | 1.47–5.71 | <.002 | 17.9 | 3.4 |

^a CAT-MH, Mental Health CAT (computerized adaptive test); CAT-DI, depression inventory; CAT-ANX, anxiety; CAT-MANIA, mania; HAM-D-25, 25-item Hamilton Rating Scale for Depression; PHQ-9, 9-item Patient Health Questionnaire; CES-D, Center for Epidemiologic Studies Depression Scale; GAF, Global Assessment of Functioning

^b ORs are for a 1-unit increase in the corresponding CAT, reflecting 25% of the total metric. This means that for every unit increase in CAT-DI score, for example, the likelihood of a current *DSM-IV-TR* major depressive disorder diagnosis increased sevenfold. Given that the range of scores on the CAT-DI is from -2 to 2 , the actual span gives an OR of 27.88, a 28-fold increase in probability of major depressive disorder from the low to the high end of the CAT-DI scale.

a community outpatient psychiatric setting against gold-standard diagnostic and severity measures, including the SCID for DSM-IV-TR, HAM-D-25, CES-D, PHQ-9, and GAF.

Considering the high rate of *DSM-IV-TR* disorders in this clinic sample, the high rate of comorbidity, and the small number of individuals with no current or past *DSM-IV-TR* diagnoses, the CAT-MH performed well. Sensitivity remained at high levels and specificity decreased as expected. However, when the sample was restricted to patients with confirmed major depressive disorder and those with no current or past *DSM-IV-TR* diagnoses, sensitivity for the CAD-MDD was unchanged, but specificity increased to 1.00 (that is, no false positives). Of 46 participants, there was only one misclassification. This bodes well for applications in primary care, where most patients (90% or more) will not have a current *DSM-IV-TR* major depressive disorder.

Even though the sample was of a patient cohort with multiple diagnoses, the three severity tests also performed well. Significant relationships were found to *DSM-IV-TR* diagnoses of major depressive disorder, generalized anxiety disorder, and current bipolar disorders for each of the three dimensional measures (CAT-DI, CAT-ANX, and CAT-MANIA, respectively), and the CAT-DI was strongly related to traditional depression severity measures. In general, patients appeared to have a positive overall impression of the test, were comfortable answering questions using a computer interface, found it easy to use, reported answering honestly, and indicated that the questions accurately reflected their mood. Interestingly, 86% indicated that they preferred the computer interface to a traditional paper-and-pencil test.

The strengths of this study included the prospective nature of the evaluations, the broad inclusion criteria that improved generalizability, and the use of gold-standard diagnostic and symptom severity comparators. Limitations included its cross-sectional design that did not allow for test-retest and longitudinal assessment of improvement over time. Given the adaptive nature of the testing and the large question bank from which to draw unique questions, we would expect that these assessments would be superior to standard assessments for longitudinal follow-up and would avoid the potential bias of the practice effect, but this needs to be demonstrated in future studies.

A further limitation of these assessments was the inability to detect lifetime history of psychiatric disorders. For example, longitudinal data are required for the accurate diagnosis of bipolar disorder, whereas the CAT-MANIA scale is useful only in assessing current manic symptoms. Per the SCID for DSM-IV-TR, there were 13 participants with current manic symptoms that met full criteria indicative of bipolar I disorder, 11 with hypomania indicative of bipolar II disorder, and one with current bipolar disorder NOS. When lifetime episodes of mania or hypomania were taken into account by assessment with the SCID for DSM-IV-TR, a total of 20 patients in this cohort had bipolar I disorder, 20 had bipolar II disorder, and three had bipolar disorder NOS (Table 2). This finding is critical, because if those patients (with mania currently in full remission [N=8] or partial remission [N=10]) were incorrectly diagnosed as having unipolar depression, they may have received inappropriate treatment with antidepressants,

TABLE 4. Ratings of CAT-MH usability by 145 adults in a community outpatient psychiatric sample

| Aspect and rating | N | % |
|--|-----|----|
| Overall | | |
| Excellent | 44 | 30 |
| Very good | 61 | 42 |
| Good | 36 | 25 |
| Fair | 4 | 3 |
| Poor | 0 | — |
| Ease of use | | |
| Very easy | 28 | 19 |
| Easy | 71 | 49 |
| Neutral | 42 | 29 |
| Difficult | 4 | 3 |
| Very difficult | 0 | — |
| Comfortable answering personal questions | | |
| Very comfortable | 108 | 75 |
| Comfortable | 32 | 22 |
| Neutral | 1 | .7 |
| Uncomfortable | 1 | <1 |
| Very uncomfortable | 0 | — |
| Missing data | 3 | 2 |
| Answered questions honestly | | |
| Strongly agree | 133 | 92 |
| Agree | 9 | 6 |
| About 50/50 | 0 | — |
| Disagree | 0 | — |
| Strongly disagree | 0 | — |
| Missing data | 3 | 2 |
| Computer versus paper | | |
| Computer | 125 | 86 |
| Paper | 14 | 10 |
| Equivocal | 3 | 2 |
| Missing data | 3 | 2 |
| Questions accurately reflected mood | | |
| A great deal | 129 | 89 |
| Very much | 12 | 8 |
| Somewhat | 4 | 3 |
| Not very much | 0 | — |
| Not at all | 0 | — |

rather than with mood stabilizers; antidepressants may be ineffective for the treatment of bipolar disorder (22–25).

Finally, the sample, in which 79% were women, 90% were Caucasians, 40% had a college degree, and 42% had some college, is not representative of other, more diverse patient populations. Future testing in these populations is required.

CONCLUSIONS

The results of this prospective, cross-sectional validation study suggest that the CAT-MH suite of tests provides a rapidly administered, accurate assessment of depression diagnosis and symptom severity across a broad range of mood and anxiety symptoms in an adult, community outpatient psychiatric population.

AUTHOR AND ARTICLE INFORMATION

Dr. Achtyes, Dr. Halstead, and Ms. Smart are with Pine Rest Christian Mental Health Services, Grand Rapids, Michigan. Dr. Achtyes is also with the Department of Psychiatry and Behavioral Medicine, Michigan State University College of Human Medicine, Grand Rapids (e-mail: achtyes@msu.edu). Ms.

Moore, Dr. Frank, and Dr. Kupfer are with the Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania. Dr. Gibbons is with the Departments of Medicine, Public Health Sciences, and Psychiatry and the Center for Health Statistics, University of Chicago, Illinois. Some findings from this study were presented at the annual meeting of the American Psychiatric Association, New York City, May 3–7, 2014; at the annual meeting of the American Society of Clinical Psychopharmacology, Hollywood, Florida, June 16–18, 2014; and at the Institute on Psychiatric Services, San Francisco, October 30–November 2, 2014.

This project was supported by a grant from the Pine Rest Foundation ("CAT-DI/SCID Assessment Tool") to Dr. Achtyes and Dr. Halstead, by grant MH66302 to Dr. Gibbons from the National Institute of Mental Health, and by in-kind support (computer software) from Michigan State University to Dr. Achtyes. The authors acknowledge the contributions of Robert J. Bielski, M.D., who helped train clinical staff in administering the diagnostic interviews, and Tim Bozung, L.M.S.W., C.T.S., Cheryl Cnossen, L.M.S.W., and Kirk Vanderploeg, L.M.S.W., who interviewed participants and helped gather study data.

Dr. Frank has received royalties or honoraria from the American Psychological Association, Guilford Press, and Lundbeck; has served on an advisory board for Servier International; and has financial interests in Adaptive Testing Technologies (www.adaptivetestingtechnologies.com), through which the CAT-MH tests will be made available, and in HealthRhythms. Dr. Kupfer holds joint ownership of copyright for the PSQI, has received an honorarium from and served on an advisory board for Servier International, and is a stockholder in AliphCom, Adaptive Testing Technologies, and HealthRhythms. Dr. Gibbons has been an expert witness for Merck, Pfizer, and Wyeth and has financial interests in Adaptive Testing Technologies. The other authors report no financial relationships with commercial interests.

Received September 4, 2014; revision received December 20, 2014; accepted February 4, 2015; published online June 1, 2015.

REFERENCES

- Gibbons RD, Weiss DJ, Kupfer DJ, et al: Using computerized adaptive testing to reduce the burden of mental health assessment. *Psychiatric Services* 59:361–368, 2008
- Gibbons RD, Weiss DJ, Pilkonis PA, et al: Development of the CAT-ANX: a computerized adaptive test for anxiety. *American Journal of Psychiatry* 171:187–194, 2014
- Gibbons RD, Hooker G, Finkelman MD, et al: The computerized adaptive diagnostic test for major depressive disorder (CAD-MDD): a screening tool for depression. *Journal of Clinical Psychiatry* 74:669–674, 2013
- Gibbons RD, Weiss DJ, Pilkonis PA, et al: Development of a computerized adaptive test for depression. *Archives of General Psychiatry* 69:1104–1112, 2012
- Gibbons RD, Hedeker DR: Full-information item bifactor analysis. *Psychometrika* 57:423–436, 1992
- Breiman L, Friedman JH, Olshen RA, et al: *Classification and Regression Trees*. Monterey, Calif, Wadsworth and Brooks, 1984
- Quinlan JR: *C4.5: Programs for machine learning*. San Mateo, Calif, Morgan Kaufmann, 1993
- Breiman L: Bagging predictors. *Machine Learning* 24:123–140, 1996
- Freund Y, Schapire RE: Experiments with a new boosting algorithm. *ICML* 96:148–156, 1996
- Breiman L: Random forests. *Machine Learning* 45:5–32, 2001
- Friedman J: Greedy function approximation: a gradient boosting machine. *Annals of Statistics* 29:1189–1232, 2001
- First MB, Spitzer RL, Gibbon M, et al: *Structured Clinical Interview for the DSM-IV-TR Axis I Disorders, Non-Patient Edition (SCID-I/NP, 1/2007 revision)*. New York, New York State Psychiatric Institute, Biometrics Research, 2007
- Hamilton M: A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry* 23:56–62, 1960
- Kroenke K, Spitzer RL, Williams JBW: The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine* 16:606–613, 2001
- Radloff LS: The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement* 1:385–401, 1977
- Endicott J, Spitzer RL, Fleiss JL, et al: The Global Assessment Scale: a procedure for measuring overall severity of psychiatric disturbance. *Archives of General Psychiatry* 33:766–771, 1976
- Wittchen HU, Jacobi F, Rehm J, et al: The size and burden of mental disorders and other disorders of the brain in Europe 2010. *European Neuropsychopharmacology* 21:655–679, 2011
- Whooley MA: Diagnosis and treatment of depression in adults with comorbid medical conditions: a 52-year-old man with depression. *JAMA* 307:1848–1857, 2012
- Gorman JM: Comorbid depression and anxiety spectrum disorders. *Depression and Anxiety* 4:160–168, 1997
- Cassidy F, Yatham LN, Berk M, et al: Pure and mixed manic subtypes: a review of diagnostic classification and validation. *Bipolar Disorders* 10:131–143, 2008
- Vieta E, Morralla C: Prevalence of mixed mania using 3 definitions. *Journal of Affective Disorders* 125:61–73, 2010
- Baldessarini RJ, Leahy L, Arcona S, et al: Patterns of psychotropic drug prescription for US patients with diagnoses of bipolar disorders. *Psychiatric Services* 58:85–91, 2007
- Post RM, Altshuler LL, Leverich GS, et al: Mood switch in bipolar depression: comparison of adjunctive venlafaxine, bupropion and sertraline. *British Journal of Psychiatry* 189:124–131, 2006
- Sachs GS, Nierenberg AA, Calabrese JR, et al: Effectiveness of adjunctive antidepressant treatment for bipolar depression. *New England Journal of Medicine* 356:1711–1722, 2007
- Ghaemi SN, Ostacher MM, El-Mallakh RS, et al: Antidepressant discontinuation in bipolar depression: a Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD) randomized clinical trial of long-term effectiveness and safety. *Journal of Clinical Psychiatry* 71:372–380, 2010