# Improving the Evaluation of Adult Mental Disorders in the Criminal Justice System With Computerized Adaptive Testing

Robert D. Gibbons, Ph.D., Justin D. Smith, Ph.D., C. Hendricks Brown, Ph.D., Mary Sajdak, M.S., Nneka Jones Tapia, Ph.D., Andrew Kulik, M.D., Matthew W. Epperson, Ph.D., John Csernansky, M.D.

**Objective:** The authors sought to develop and validate a suite of dimensional measures of psychiatric syndromes for use in a criminal justice population.

**Methods:** The previously validated Computerized Adaptive Test–Mental Health (CAT-MH) was administered to a sample of 475 defendants in the Cook County Bond Court. Itemlevel data were used to determine which test items exhibited differential item functioning in this population compared with the population used for the original calibration.

**Results:** After removal of nine items that exhibited differential item functioning from the CAT-MH, correlations

On any given day, 300,000 to 400,000 people with mental illness are incarcerated in jails and prisons across the United States, and an additional 500,000 are under correctional supervision in the community (1). An analysis of individuals incarcerated in jails in two states found that the rates of severe mental illness were 14.5% for men and 31% for women (2). Rates of less severe mental illness (e.g., some anxiety disorders) were 35% for men and 27% for women (3). The prevalence of posttraumatic stress disorder (PTSD) and suicide are at least three times higher in jails and prisons (4) than in the community, and the rate of substance use disorder is seven times higher (5).

Those confined to correctional facilities in the United States are legally entitled to adequate mental health care, which requires effective screening and identification of people with mental health needs. In a recent systematic review of 22 mental health screening tools, only one instrument was found to have low risk of bias and low concerns regarding applicability, and only a handful of screening tools have undergone replication studies (6). Some instruments, such as the Jail Screening Assessment Tool, offer promising results, but can take up to 30 minutes to complete and require trained clinical interviewers (7). The Brief Jail between scores based on the original calibration from a nonjustice-involved population and the newly computed scores based on a sample of bond court defendants showed a correlation coefficient of r=0.96 to r=0.99.

**Conclusions:** With a slight modification of the original CAT-MH, the tool was successfully used to measure severity of depression, anxiety, mania and/or hypomania, suicidality, and substance use disorder in an English- and Spanishspeaking criminal justice population.

Psychiatric Services 2019; 70:1040-1043; doi: 10.1176/appi.ps.201900038

Mental Health Screen (BJMHS) can be completed in less than 3 minutes, but sensitivity is low (70% for men and 61% for women) (8). The Correctional Mental Health Screen (CMHS) has accuracy rates of up to 80% (9).

#### HIGHLIGHTS

- The authors used differential item functioning to study whether previously validated computerized adaptive tests for mental health (CAT-MH) were valid in the criminal justice population.
- Nine items, out of a bank of over 1,000, exhibited biased responses (i.e., differential item functioning) in a sample of 475 bond court defendants. After removal of these items, the correlation between scores based on the original validated CAT-MH item calibration and the new calibration based on the bond court population ranged from r=0.96 to 0.99.
- These findings demonstrate that with a slight modification of the original CAT-MH, the severity of depression, anxiety, mania and/or hypomania, suicidality, and substance use disorder can be validly measured in an English- or Spanish-speaking criminal justice population.

Mental health measurement (including the BJMHS and CMHS) is based almost exclusively on subjective judgment and classical test theory. In this approach, the level of impairment is determined by a total score, which requires that all respondents be tested with the same set of symptom items and that all items, despite severity level (e.g., "Do you feel sad?" versus "Do you think that you would be better off dead?") are weighted equally. In contrast, computerized adaptive testing (CAT), which is based on multidimensional item response theory, adapts item presentation to the individual's severity and allows different individuals to be tested with different symptom items targeted to their specific impairment level (10). This approach mirrors that of a good clinician and eliminates the need for staff training and test scoring. The duration of testing is shorter (typically 2-10 minutes, depending on the number of domains tested), the results are more precise, and savings are greater than with human-directed assessments. The resulting measures can be used for screening (11) and/or more detailed assessment (12). Because we use multidimensional item response theory, CAT permits adaptive evaluation of complex traits, including depression, anxiety, mania and hypomania, PTSD, psychosis, suicidality, and substance abuse (12). CAT has not yet been tested in a correctional setting, however, which was the goal of this study.

#### METHODS

The study took place (July 2017 through February 2018) in the Cook County Bond Court in northeastern Illinois. The bond court is connected to the Cook County Jail, the largest single-site jail in the United States. Every person arrested and detained (either in the Cook County Jail or a police precinct) for a felony charge in the City of Chicago goes through Bond Court, typically within 48 hours of arrest. At Bond Court, a judge determines whether the person may be released on bond and, if so, the dollar amount of that bond. If the person is not released on bond, they go to the Cook County Jail. On December 12, 2017, 95% of the people in the Cook County Jail were incarcerated pretrial, meaning they were either not given a bond or could not pay the bond amount (chicagodatacollaborative.org). Since 2012, the Cook County Jail has required mental health assessments (conducted by psychologists and social workers), which inform the housing location, level of treatment, and medication schedule of the detainee during incarceration. These screenings also provide judges and public defenders with information on the health status of all defendants, including those who may be released on bond.

During the course of each detainee's health screening in bond court, we provided the detainee with a tablet computer and invited him or her to take the Computerized Adaptive Test–Mental Health (CAT-MH) (12). We provided no further details to the detainee. The CAT-MH reads the questions to the subject through headphones, helping to overcome any issues related to literacy. We used six validated modules of the current CAT-MH system to conduct the screening: major depressive disorder (computerized adaptive diagnosis–MDD), depressive severity (CAT–Depression Inventory [CAT-DI]), severity of anxiety (CAT-ANX), severity of mania and/or hypomania (CAT-MANIA), suicidality (CAT-SS), and severity of substance misuse (CAT-SA) (12). We sequentially recruited 475 defendants for the study, and all agreed to participate. Two percent took the tests in Spanish (13). Ninety-six percent completed the CAT assessment. The 4% who did not complete the CAT-MH were called to court during the assessment. Eighty-one percent of the defendants were male, 61% were black, and 17% were Hispanic.

To assess differential item functioning, the itemresponse patterns from the bond court sample were used to estimate a new bifactor model (10) based solely on this sample. CAT-MH items that had factor loadings on the primary dimension of less than 0.3 were identified as having poor discrimination in this criminal justice population and were eliminated from further analysis and scoring. Using the remaining item parameter estimates, we then scored the response patterns of the 475 defendants for each of the five domains (depression, anxiety, mania and/or hypomania, suicidality, and substance use disorder). Scores were also computed on the basis of the original bifactor model calibration developed with a sample of psychiatric patients and a control group of healthy individuals (10). The scores for the new bifactor model calibration and the original calibration were then tested for agreement by using a correlation coefficient. Data were plotted on the original underlying normal scale, which has a range of scores from -3 to 3, scaled to have a mean of 0 and variance of 1 for both calibrations to adjust for differences in severity between the bond court sample and the original sample. In the bond court sample, items exhibiting differential item functioning were ones that no longer differentiated between high and low levels of the underlying disorder, presumably because they were produced by the experience of incarceration and no longer correlate with the other symptoms shown to be related to the disorder. To provide an analogy, in perinatal depression, the somatic symptom of fatigue is not a good discriminator, because fatigue affects most pregnant and postpartum women whether or not they are depressed (14).

This study was approved by the institutional review board of the Cook County Health and Hospital System.

## RESULTS

The median time required to complete the entire battery of six adaptive tests (five domains and the major depressive disorder screener) was 9:45 minutes, with an interquartile range of 7:50–12:03 minutes.

Plots of the correlations across the score spectrum for each scale and a table of the score distributions are available in the online supplement. For depression, there was no indication of differential item functioning except for a single item ("In the past 2 weeks, I felt that everything that I did was an effort") that exhibited differential item functioning in the bond court sample compared with the original nonjustice-involved psychiatric population (a mixture of psychiatric patients with mood disorders and a control group of healthy individuals). Removal of that item revealed a correlation of r=0.99 between the bond court calibration and the original calibration. Plots of the correlations showed close agreement throughout the severity score range, with a small amount of bias at the low end of the scale, where the bond court calibration yielded slightly higher scores on depression severity.

For anxiety, two items ("In the past 2 weeks, how much of the time have you had difficulty doing activities involving concentration and thinking?" and "In the past 2 weeks, how much difficulty have you had falling asleep?") exhibited differential item functioning. Removal of those items revealed a correlation of r=0.97. Plots of the correlations indicated close agreement throughout the severity score range, with a small amount of bias at the low end of the scale, where the bond court calibration yielded slightly higher anxiety scores.

For mania and hypomania, the item "In the past 2 weeks, have you had periods of at least 3 days in which you were less sexually active than is usual for you?" exhibited differential item functioning. Removal of that item revealed a correlation of r=0.97. Correlation plots indicated close agreement throughout the severity score range, with no evidence of bias.

For suicidality, the item "In the past 2 weeks, how much have you been distressed by feeling fearful?" exhibited differential item functioning. Removal of that item revealed a correlation of r=0.98. Correlation plots indicated close agreement throughout the severity score range, with slightly increased scores for the bond court calibration compared with the original calibration at the lowest end of the scale.

For substance abuse, four items beginning with "In the past 2 weeks," exhibited differential item functioning ("How often have you been bothered by feeling down, depressed or hopeless?" "Have you had trouble falling asleep, staying asleep, or sleeping too much? "How much of the time have you been feeling distant or cut off from other people?" and "How much of the time have you been feeling lonely?"). Removal of those items revealed a correlation of r=0.96 between the original calibration and the bond court calibration. Correlation plots indicated close agreement throughout the severity score range, with slightly decreased scores for the bond court calibration at the highest end of the scale.

Thirty percent of the defendants screened positive for major depressive disorder, with 9% in the moderate to severe range and 10% in the moderate range. Nine percent were in the severe range for anxiety, and 10% were in the severe range for mania and/or hypomania, suggesting that further assessment was needed for bipolar disorder. Three percent had high risk for suicidality in need of immediate intervention, and 14% were at high risk of having a significant substance use disorder.

## DISCUSSION

The results of this study revealed that after the removal of nine items, the CAT-MH provides the same level of discrimination between high and low levels of severity on the five severity scales in a criminal justice population as it did during a previous validation in a psychiatric population, where results were compared with structured clinical interviews. The deleted items dealt with sleep disturbance, social isolation, decreased sexual activity, and feeling fearful, all of which could plausibly be related to the experience of arrest and incarceration rather than to an underlying psychiatric disorder. Appreciable numbers of defendants had mental health psychopathology, suicide risk, and substance abuse. We found that 10% of the defendants had scores in the severe range for mania and/or hypomania, which would suggest the need for further evaluation to diagnose bipolar disorder. The rate of high risk for suicidality was 3% overall; however, 7% (overall) had both suicidal ideation with intent or a plan or reported recent suicidal behavior in the past month regardless of ideation. This rate is more than double the 3.0% found in a recent study of patients conducted in the University of Chicago emergency department, which is also in Cook County and serves a similar high-risk inner-city population. While 14% of the sample had scores indicating high risk of having a substance use disorder, 22% had scores indicating intermediate risk, for a combined risk estimate of 36%. Thresholds were derived based on 12-month CIDI diagnoses of substance use disorder and self-reported use of alcohol and drugs. In comparison, individuals receiving an intermediate risk score on the CIDI had a positive diagnosis rate for substance use disorder of 22%, and individuals receiving a high risk score on the CIDI had a positive diagnosis rate for substance use disorder of 50%. For self-reports, the rates were 47% and 90%, respectively (unpublished manuscript, Gibbons RD, Alegria M, Markle S, et al., 2019). As such, individuals receiving intermediate and high-risk scores should be considered to have substance misuse.

## CONCLUSIONS

Our results show that the revised version of the CAT-MH can be used to screen and assess a variety of mental health conditions in the criminal justice population. This version can be used to rapidly screen for the presence of one or more serious mental disorders (major depressive disorder, generalized anxiety disorder, bipolar disorder, substance use disorder, and suicidality) and to quantify the severity of illness. With the aid of the CAT-MH, clinicians can be more effectively used to provide treatment and placement into appropriate specialized diversion and criminal justice interventions, rather than to perform routine assessments. For more complex disorders, such as bipolar disorder, the CAT-MH can be used to direct clinicians to individuals who require additional evaluation. We have recently developed CATs for PTSD and psychosis, which will further expand the types of mental disorders that can be rapidly detected in this high-need population. The CAT-MH measures can also be used to monitor the effectiveness of treatment and as a predictor of long-term mental health outcomes when individuals return to their communities.

#### AUTHOR AND ARTICLE INFORMATION

Center for Health Statistics, University of Chicago, Chicago (Gibbons); Department of Psychiatry, Northwestern University, Evanston, Illinois (Smith, Brown, Csernansky); Cook County Health and Hospital System, Chicago (Sajdak, Kulik); Chicago Beyond, Chicago (Tapia); Social Service Administration, University of Chicago, Chicago (Epperson). Send correspondence to Dr. Gibbons (rdg@uchicago.edu).

This study was supported in part by a grant to Dr. Gibbons from the National Institutes of Health for "A New Statistical Paradigm for Measuring Psychopathology Dimensions in Youth" (R01-MH-100155) and a grant from the National Institute on Drug Abuse to Dr. Brown for the Center for Prevention Implementation Methodology for Drug Abuse and HIV (P30DA027828). The authors thank Kayla Morgan and Katherine Vinaitheerthan for collecting the Computerized Adaptive Test–Mental Health (CAT-MH) data.

Dr. Gibbons is a founder of Adaptive Testing Technologies, which distributes the CAT-MH suite of computerized adaptive tests. The terms of this arrangement have been reviewed and approved by the University of Chicago in accordance with its conflict of interest policies. Dr. Csernansky has received consultation fees from Indivior Pharmaceuticals, Inc. The other authors report no financial relationships with commercial interests.

Received January 18, 2019; revisions received March 18 and May 2, 2019; accepted May 16, 2019; published online July 24, 2019.

#### REFERENCES

 Ending an American Tragedy: Addressing the Needs of Justice-Involved People With Mental Illnesses and Co-Occurring Disorders. Washington, DC, National Leadership Forum on Behavioral Health/Criminal Justice Services, 2009. http://www.pacenterofexcellence.pitt.edu/documents/AmericanTragedy.pdf

- Steadman HJ, Osher FC, Robbins PC, et al: Prevalence of serious mental illness among jail inmates. Psychiatr Serv 2009; 60:761–765
- 3. Black DW, Gunter T, Loveless P, et al: Antisocial personality disorder in incarcerated offenders: psychiatric comorbidity and quality of life. Ann Clin Psychiatry 2010; 22:113–120
- Goff A, Rose E, Rose S, et al: Does PTSD occur in sentenced prison populations? A systematic literature review. Crim Behav Ment Health 2007; 17:152–162
- 5. Health, Mental Health, and Substance Use Disorders FAQs. Washington, DC, Justice Center: Council of State Governments. https://csgjusticecenter.org/substance-abuse/faqs/#q2
- 6. Martin MS, Colman I, Simpson AIF, et al: Mental health screening tools in correctional institutions: a systematic review. BMC Psychiatry 2013; 13:275
- Baksheev GN, Ogloff J, Thomas S: Identification of mental illness in police cells: a comparison of police processes, the Brief Jail Mental Health Screen and the Jail Screening Assessment Tool. Psychol Crime Law 2012; 18:529–542
- 8. Steadman HJ, Scott JE, Osher F, et al: Validation of the brief jail mental health screen. Psychiatr Serv 2005; 56:816–822
- Ford JD, Trestman RL, Wiesbrock VH, et al: Validation of a brief screening instrument for identifying psychiatric disorders among newly incarcerated adults. Psychiatr Serv 2009; 60: 842–846
- Gibbons RD, Weiss DJ, Pilkonis PA, et al: Development of a computerized adaptive test for depression. Arch Gen Psychiatry 2012; 69:1104–1112
- Gibbons RD, Hooker G, Finkelman MD, et al: The CAD-MDD: a computerized adaptive diagnostic screening tool for depression. J Clin Psychiatry 2013; 74:669–674
- Gibbons RD, Weiss DJ, Frank E, et al: Computerized adaptive diagnosis and testing of mental health disorders. Annu Rev Clin Psychol 2016; 12:83–104
- Gibbons RD, Alegria M, Cai L, et al: Successful validation of the CAT-MH scales in a sample of Latin American migrants in the United States and Spain. Psychol Assess 2018; 30: 1267–1276.
- 14. Kim JJ, Silver RK, Elue R, et al: The experience of depression, anxiety, and mania among perinatal women. Arch Women Ment Health 2016; 19:883–890