

Validation of a Brief Screening Instrument for Identifying Psychiatric Disorders Among Newly Incarcerated Adults

Julian D. Ford, Ph.D.

Robert L. Trestman, Ph.D., M.D.

Valerie H. Wiesbrock, M.A.

Wanli Zhang, Ph.D.

Objective: This study sought to cross-validate a brief mental health screening instrument for newly incarcerated adults with undetected psychiatric impairment.

Methods: Women (N=360) and men (N=630) aged 18 to 64 years with no institutionally identified mental health condition were randomly recruited upon entry to jail. Nonclinical research staff administered the Correctional Mental Health Screen for women (CMHS-F) or men (CMHS-M). A randomly selected subsample (100 women and 106 men) subsequently completed research clinical diagnostic interviews for standardized clinical cross-validation.

Results: Accuracy for the CMHS-F and CMHS-M (73%–80%) for the identification of current axis I or II psychiatric disorders replicated results of an instrument development study and exceeded the accuracy levels reported for alternative brief mental health screening measures with incarcerated adults. Cutoff points were identified that maximized positive or negative predictive power.

Conclusions: This cross-validation demonstrates that the CMHS-F and CMHS-M are efficient, accurate screens to identify newly incarcerated adults with undetected psychiatric disorders. (*Psychiatric Services* 60:842–846, 2009)

Correctional mental health clinicians, administrators, and researchers have called for evidence-based protocols for accurately identifying individuals with psychiatric disorders in a timely manner on entry to jail (1,2). Lifetime prevalence estimates of psychiatric disorders for adults in jails and prisons are two to four times higher than those for non-incarcerated adults. Although prevalence estimates are lower for jail detainees than for prison inmates, more than half of adult jail detainees meet criteria for lifetime affective, anxiety, or psychotic mental disorders (1,2). More than 30% of men detained in jail have a current severe psychiatric or substance use disorder (1). Almost one in ten women detained in jail meet criteria for a current psychiatric disorder (2).

Within the past decade most states' departments of correction have adopted policies mandating mental health screening for newly admitted inmates (3). By 2000 almost 70% of state prisons had formal policies mandating the screening of inmates at intake, with almost two-thirds conducting psychiatric assessments and 90% providing mental health services for inmates. One in eight state pris-

oners are reported to be receiving mental health therapy or counseling, primarily pharmacotherapy. However, mental health screening is less common on entry to jail. Compared with prison inmates, jail inmates with mental illness are 50% less likely to receive any mental health service and almost 200% less likely to receive counseling or therapy (3). However, jail inmates with mental illness are 50% more likely than jail inmates without mental illness to have serious disciplinary problems (3). Thus mental health screening in jails may be a key first step toward rehabilitation and treatment.

Screening measures that assess both mental competence and the attitudes or personality characteristics that may lead to disciplinary problems (especially risk of violence) have been developed for adult correctional populations, but fewer psychiatric screening measures have been developed. The 14-item Referral Decision Scale (RDS) (4,5) has shown evidence of moderate sensitivity and negative predictive power but low specificity and positive predictive power (.15–.35) among male prison inmates and jail detainees. A refinement of the RDS designed to identify *DSM-IV* axis I mood or psychotic disorders—the eight-item Brief Jail Mental Health Screen (BJMHS) (6)—correctly classified 74% of males and 62% of females who met research diagnostic criteria in a sample of newly admitted detainees. The BJMHS and RDS do not address anxiety or

Dr. Ford, Dr. Trestman, and Dr. Zhang are affiliated with the Department of Psychiatry, University of Connecticut School of Medicine, MC1410, 263 Farmington Ave., Farmington, CT 06030 (e-mail: ford@psychiatry.uchc.edu). Ms. Wiesbrock is with the Center for Clinical Research and Technology, University Hospitals Case Medical Center, Cleveland, Ohio.

axis II disorders. The BJMHS's 35% level of false negatives among women, although an improvement over the 63% rate reported in the only prior screening study with incarcerated women (5), is problematic. The BJMHS validation cohort also oversampled inmates with mental health needs, rather than using a representative sample of detainees with undetected mental health needs. Low specificity or low positive predictive power, which was potentially related to overreporting of mental health problems, limited testing with female detainees (5,6), and an absence of attention to anxiety or personality disorders limits the utility of extant brief mental health screens for correctional populations.

This project was designed to address these limitations by developing and validating a brief mental health screen for jail detainees. An eight-item Correctional Mental Health Screen for women (CMHS-F) and a 12-item screen for men (CMHS-M) (7) were empirically derived (8) from a 55-item composite screen (including RDS and BJMHS items). These screens have shown accuracy of at least 70%–80%, relatively low rates of false positives and negatives versus other correctional screens, and incremental utility versus the RDS and BJMHS (8). This study replicates the first CMHS study (7) with an independent sample of newly incarcerated jail detainees used the CMHS as a freestanding screen conducted by nonclinical research staff. Within five days of completing the screening, a subsample of randomly selected participants, stratified by gender, were invited to participate in a follow-up structured diagnostic interview for standardized clinical cross-validation.

Methods

Study participants were recruited from February 2003 through September 2003 in five correctional facilities that serve as the jails for all adults incarcerated in Connecticut. Persons were eligible for the study if they entered jail 24 to 72 hours before recruitment, were 18 years or older, were able to speak English, were not "high bond" security risks (these persons could not be interviewed with-

out a custody officer present), were not admitted to the medical unit for immediate care because of wounds or injuries or acute substance intoxication or detoxification, and were not admitted to the medical unit for acute psychosis, mania, suicidality, or delirium or a history of psychiatric treatment. A total of 990 adults participated (360 women, 36%, and 630 men, 64%). Age of participants ranged from 18 to 64 years (mean \pm SD=32.0 \pm 9.7 years). Self-reported race or ethnicity of participants was white (N=397, 40%), black (N=358, 36%), Hispanic (N=218, 22%), and other (N=17, 2%). Education levels ranged from three to 16 years of school (mean=11.1 \pm 1.2 years). Demographic characteristics and legal charges of the sample were comparable with those of the overall jail population except for underenrollment of Hispanics, who comprised 25% of jail inmates, according to census data. Of the 1,094 detainees invited to participate, 104 (10%) declined. Gender, age, and race or ethnicity were unrelated to likelihood of refusal to participate in screening; however, black women (N=27 of 175, 15%) were more likely than white women (N=19 of 237, 8%) or Hispanic women (N=6 of 114, 5%) to refuse ($\chi^2=13.0$, $df=2$, $p=.002$).

Criminal charge data were obtained from Department of Correction records for 90% of the 990 participants (N=882); 108 participants did not consent to release this information to the study. Gender, age, and ethnicity did not differ for participants regardless of whether they consented to release of their records. Primary legal charges included drug offenses (possession or distribution) (N=168, 17%); fraud, burglary, or larceny (N=131, 13%); minor nonviolent offenses (conspiracy, criminal attempt, or contempt of court) (N=187, 19%); and violation of probation (N=112, 11%). Most (N=626, 63%) charge offenses were for nonviolent crimes, 18% (N=179) were for violent crimes (that is, use of a weapon, physical or sexual assault, manslaughter, or murder), and 19% (N=185) could not be classified because the charges were for nonspecific offenses (for example, probation violation). All pro-

cedures were approved by the institutional review boards of the University of Connecticut Health Center and the State Department of Correction. Screens were administered by bachelor's-level research staff (three white women, one Latina, and one white man).

After completing the screening, 223 randomly selected participants, stratified by gender, were invited to participate in a follow-up structured diagnostic interview for standardized clinical cross-validation; 17 (8%) declined, leaving 206 to participate. There was no difference between those interviewed and other participants in race, age, marital status, or education. Men (N=14 of 120, 12%) were more likely than women (N=2 of 102, 2%) to refuse the interview ($\chi^2=4.5$, $df=1$, $p<.05$). Bachelor's-, master's-, or M.D.-level clinical research assessors (three white women and one white man) who were blind to screening results gathered a second consent from interviewees and conducted follow-up interviews within five days of the initial screening.

The CMHS consists of dichotomous items—eight items for women (CMHS-F) and 12 items for men (CMHS-M)—derived from psychometric analyses (7) of a 55-item composite screen, including items from the Structured Clinical Interview for DSM-IV–Patient Version (SCID-P) (9), the Primary Care PTSD [post-traumatic stress disorder] Screen (PC-PTSD) (10), the RDS (4), and the Iowa Personality Disorder Screen (IPDS) (11).

The SCID-P and SCID-II (Structured Clinical Interview for DSM-IV Axis II Disorders) (9) were used to assess axis I and II disorders, except for PTSD. In order to use the best validated structured interview for PTSD, we used the Clinician Administered Posttraumatic Stress Disorder Scale (CAPS) (12) instead of using the SCID-P PTSD module. Lifetime research diagnoses were identified by using the SCID-P, SCID-II, and CAPS in the interview sample. Prevalence estimates were as follows: psychotic disorders (N=5, 2%), affective disorders (N=84, 41%), anxiety disorders other than PTSD (N=83, 40%), PTSD (N=71, 34%), cluster A axis II

disorders (N=21, 10%), cluster B axis II disorders (N=99, 48%—including N=86, 42%, antisocial personality disorder and N=41, 20%, borderline personality disorder), and cluster C axis II disorders (N=46, 22%). Current research diagnoses were identified in the sample for psychotic disorders (N=4, 2%), affective disorders (N=49, 24%), anxiety disorders other than PTSD (N=65, 32%), and PTSD (N=46, 22%).

The internal consistency, interrater concordance, and retest reliability of the CMHS-F and CMHS-M were examined first. Next, the predictive validity of the CMHS-F and CMHS-M was tested for the identification of four composite diagnostic categories derived from the SCID and CAPS structured interviews (7): first, any current axis I or II psychiatric disorder; second, any current axis I or II disorder excluding antisocial personality disorder; third, current depressive disorder (unipolar major depression or dysthymia); and fourth, current anxiety disorder (generalized anxiety, social anxiety, panic with or without agoraphobia, obsessive-compulsive disorder, and PTSD).

Analyses of receiver operating characteristics were conducted separately for men and women and provided estimates of CMHS sensitivity (ability to identify respondents who meet criteria for the diagnosis), specificity (ability to identify those who do not meet criteria for the diagnosis), positive predictive power (probability that someone identified by the screen as having a disorder meets diagnostic criteria), and negative predictive power (the probability that someone identified by the screen as not having a disorder does not meet diagnostic criteria) (8). The area under the curve for each diagnostic category was calculated for cutoff points identified with the Youden index (J), a measure of biomarker effectiveness calculated to maximize sensitivity and specificity (13). CMHS criterion validity was assessed with the F test (for continuous variables) or the chi square statistic (for dichotomous variables) in comparisons of true and false positives and true and false negatives on variables expected to be associated with mental health status: Global Assess-

ment of Functioning (GAF) from the SCID-P, number of SCID-P current axis I diagnoses, number of SCID-II borderline personality disorder symptoms, number of SCID-II antisocial personality disorder symptoms, correctional staff ratings of mental health needs (ranging from 1, no mental health needs, to 5, immediate need for inpatient psychiatric care), and receipt of psychiatric evaluation within 14 days of incarceration. To test discriminant validity, true- and false-positive and true- and false-negative subgroups were compared on correctional staff ratings that should be relatively unrelated to mental health status: medical needs (ranging from 1, none, to 5, immediate need for inpatient medical care), behavioral risk (ranging from 1, minimal behavioral risk, to 5, high security management), and history of substance abuse treatment.

Results

Internal consistency first was confirmed for the CMHS-F and CMHS-M (Cronbach's α = .76 and .78, respectively). Retest reliability for each also was acceptable (Pearson r = .82 and .84 and Kendall's τ -C = .70 and .73, respectively). Interrater agreement was 100% for all tests.

The CMHS-F cutoff point derived with the J index for identifying any current axis I or II psychiatric diagnosis (4) produced a significantly larger area under the curve (.80) than cutoff points maximizing positive (7) or negative (2) predictive power (.61 and .67, respectively). The optimal CMHS-M cutoff point (5) for identifying any current axis I or II psychiatric disorder also yielded a significantly larger area under the curve (.73) than cutoff points maximizing positive (8) or negative (1) predictive power (.66 and .62, respectively). At the optimal cutoff points, the area under the curve for men and women for identification of psychiatric disorders other than antisocial personality disorder and depressive disorders was .77–.82. [Tables showing areas under the curve, sensitivity, specificity, and positive and negative predictive power findings in detail are available as an online supplement at ps.psychiatryonline.org.] The CMHS-M produced a comparably

robust area under the curve for anxiety disorders, but the CMHS-F did not (.69). The CMHS-M was more accurate in identifying cases with any axis I or II disorder when antisocial personality disorder was excluded (area under the curve = .79) because of greater sensitivity (.84 versus .64 when antisocial personality disorder was included).

At cutoff points selected to maximize negative predictive power, sensitivity increased to >.90. Likewise, with cutoff points selected to maximize positive predictive power and sensitivity, specificity reached 1.00 in every case. Overall screening accuracy was statistically significantly lower at these extreme cutoff points than at the J-based cutoff points.

With regard to criterion validity, participants with true-positive ratings on the CMHS had significantly lower scores on the GAF (indicating worse psychosocial functioning) than other participants, and those with true-negative ratings had significantly higher GAF scores than those with false-negative ratings (Table 1). Compared with other participants, those with true-positive ratings also had significantly more psychiatric diagnoses and borderline personality disorder symptoms and higher mental health needs scores, although this was not statistically significant (Table 1). Those with true-positive ratings and those with false-negative ratings also were most likely to receive psychiatric evaluations (respectively, 16% and 7%) than those with false-positive ratings (0%) or true-negative ratings (5%), although this finding was not statistically significant. Consistent with discriminant validity, persons with true- and false-positive ratings and those with true- and false-negative ratings did not differ on variables expected to be unrelated to the CMHS, including medical need and behavioral risk ratings (Table 1) and past substance use disorder (true positive=46%, false positive=42%, false negative=61%, and true negative=36%).

Discussion

When administered by nonclinical research staff, the CMHS-F and CMHS-M showed evidence of reliability, criterion and discriminant va-

Table 1Validation of the Correctional Mental Health Screen with a structured diagnostic interview among 206 jail detainees^a

Morbidity indicator	True positive (N= 94)		False positive (N=12)		False negative (N=41)		True negative (N=59)		F ^b	p
	M	SD	M	SD	M	SD	M	SD		
GAF ^c	52.4 [†]	13.5	70.7 ^{†,§}	15.9	64.5 [†]	15.0	75.6 [§]	15.5	35.1	<.001
Structured Clinical Interview for DSM-IV–Patient Version										
Number of current axis I disorder diagnosis categories ^d	1.4 [†]	1.3	.0 [†]		.6 [†]	.8	.0 [†]		27.9	<.001
Structured Clinical Interview for DSM-IV–Axis II Version										
Number of borderline person- ality disorder symptoms ^e	3.7 [†]	.9	.8 [†]	.6	1.6 [†]	2.2	.4 [†]	.8	29.4	<.001
Number of antisocial personality disorder symptoms ^f	3.1 [†]	2.6	.0 [†]		3.4 [†]	2.5	.0 [†]		36.2	<.001
Mental health needs ^g	1.9 [†]	.8	1.6 [†]	.7	1.4 [†]	.7	1.4 [†]	.6	6.2	<.001
Medical needs ^g	2.0	.7	2.0	.8	1.9	.7	1.9	.7	.5	.658
Behavioral risk ^h	3.1	.7	2.8	.7	3.1	.7	2.8	.9	2.3	.078

^a Means with different superscripts differ by $p < .05$.^b $df=3$ and 202^c Global Assessment of Functioning. Possible scores range from 0 to 100, with higher scores indicating better psychosocial functioning.^d Possible scores range from 0 to 4.^e Possible scores range from 0 to 9.^f Possible scores range from 0 to 7.^g Possible scores range from 0 to 5, with higher scores indicating need for more intensive services or care.^h Possible scores range from 0 to 5, with higher scores indicating need for more intensive monitoring of potential noncompliance and dangerousness.

lidity, and predictive utility for identifying detainees with undetected psychiatric disorders. CMHS-F accuracy, with the use of optimal cutoff points, was approximately 80% (except for anxiety disorders), comparable with findings for the CMHS-F in the prior development study (7) and exceeding the 62%–72% rate for the BJMHS with incarcerated women (6,14). Compared with the 35% false negative reported initially for the BJMHS (6), CMHS-F's replicated (7) rate of approximately 25% is an improvement and lower than the lowest rate reported previously (6) except for the 15% rate reported for the BJMHS in a recent study (14). The false-positive rating of the CMHS-F was lower still, with 8%–16% false positives (except for detecting anxiety disorders, which were slightly higher), compared with 45%–62% for the BJMHS with women (6,14). CMHS-F accuracy was comparable regardless of whether or not antisocial personality disorder was included, suggesting it has utility for identifying antisocial personality disorder among women. The CMHS was less accurate in identifying anxiety disorders than it was in identifying other disorders, as

a result of false positives and negatives, consistent with prior studies' findings that incarcerated women may both under- and overreport anxiety disorders (2,3).

The CMHS-M also showed evidence of reliability, validity, and predictive utility. CMHS-M accuracy for identifying any axis I or II disorder (73%), particularly when antisocial personality disorder was excluded (79%), was comparable with that found in the CMHS development study (7) and for the BJMHS with incarcerated men (72%–80%) (6,14). The false-negative and false-positive ratings were approximately 20% (range=11%–30%) except when antisocial personality disorder was included (36%), suggesting that the CMHS-M underdetects antisocial personality disorder.

Axis I depressive or anxiety disorders were well detected by the CMHS (albeit somewhat less so for anxiety disorders among women). The CMHS showed promise in identifying comorbid psychiatric disorders: 39% of participants with true positives had at least two axis I disorders, whereas no participants with false positives and 17% of those with false negatives had axis I diagnoses

from two or more of the four composite diagnosis categories (that is, affective disorders, anxiety disorders other than PTSD, PTSD, and psychotic disorders).

Persons considered to have a true-positive rating on the CMHS on average had a mental health need rating close to 2 (reflecting need for further observation and possible mental health evaluation), whereas the false positive and true and false negatives on average were rated between 1 and 2. False positives were the closest on average to the true positives in mental health needs ratings. These differences were statistically significant overall, although they were not statistically significant for post hoc comparisons of subgroups.

Persons who were considered to have a true-positive rating were more likely than other participants to be referred for a mental health evaluation in jail, although this was rare (only 16% of all true positives). Limited statistical power resulting from the relatively small numbers in the false-positive and false-negative cells and from the infrequent provision of mental health evaluations for newly jailed adults may account for the mar-

ginal statistical significance ($p=.08$) of the relationship between screening true positive and independently being selected to receive a mental health evaluation. It appears that the CMHS is generally consistent with institutional ratings of mental health need, but without formal screening, most inmates with a psychiatric disorder are not identified, and few who are identified actually receive psychiatric evaluation. Extrapolated to large systems, the cost of not screening may be substantial because of untreated morbidity and overuse of scarce costly mental health resources.

Study limitations also include underrepresentation of black and Hispanic detainees and not including chronically incarcerated inmates or institutionally identified persons with mental health problems. Additional replication is needed for definitive cutoff points for clinical use (for example, item analyses testing item weights and scoring rules) and to demonstrate acceptability and utility of the CMHS when administered by correctional staff. On the basis of study findings, the CMHS should not be used to identify anxiety disorders for women or antisocial personality disorder for men.

Conclusions

Given the infrequency with which the participants with true-positive ratings were independently referred for a mental health evaluation, systematic screening for undetected psychiatric morbidity should be done in two stages (15), beginning with broad-

based screening and followed by selectively administered intensive assessments. The gender-specific CMHS brief mental health screening instruments are feasible, reliable, valid, and accurate measures that may be used efficiently (that is, less than five minutes of administration time by nonclinical staff) for this purpose.

Acknowledgments and disclosures

This research was funded solely through grant 2000-IJ-CX-0044 from the National Institute of Justice (principal investigators, Dr. Ford and Dr. Trestman). The authors acknowledge the collaboration and support of the Connecticut Department of Correction, the University of Connecticut Health Center's Correctional Managed Health Care, and the assistance of Karen Kesten, M.S., Susan Quarti, M.A., Jessie Wagner, M.S., and Kelly Beckom, M.A.

The authors report no competing interests.

References

1. Teplin LA: Psychiatric and substance abuse disorders among male urban jail detainees. *American Journal of Public Health* 84:290-293, 1994
2. Teplin LA, Abram KM, McClelland GM: Prevalence of psychiatric disorders among incarcerated women: I. pretrial jail detainees. *Archives of General Psychiatry* 53:505-512, 1996
3. Beck AJ, Maruschak LM: *Mental Health Treatment in State Prisons, 2000*. Washington, DC, US Department of Justice, Bureau of Justice Statistics, 2001
4. Teplin LA, Swartz J: Screening for severe mental disorder in jails: the development of the Referral Decision Scale. *Law and Human Behavior* 13:1-18, 1989
5. Veysey BM, Steadman HJ, Morrissey JP, et al: Using the Referral Decision Scale to screen mentally ill jail detainees: validity and implementation issues. *Law and Human Behavior* 22:205-215, 1998
6. Steadman HJ, Scott HE, Osher F, et al: Val-

idation of the Brief Jail Mental Health Screen. *Psychiatric Services* 56:816-822, 2005

7. Ford JD, Trestman RL, Wiesbrock V, et al: Development and validation of a brief mental health screening instrument for newly incarcerated adults. *Assessment* 14:279-299, 2007
8. Zimmerman M, Mattia JJ: A self-report scale to help make psychiatric diagnoses: the Psychiatric Diagnostic Screening Questionnaire. *Archives of General Psychiatry* 58:787-794, 2001
9. First MB, Gibbon M, Spitzer RL, et al: *Structured Clinical Interview for DSM-IV Axis I and II Disorders, Patient Version (SCID-P) and Axis II Version (SCID-II)*. Washington, DC, American Psychiatric Publishing, 1996
10. Prins A, Ouimette P, Kimerling R, et al: The Primary Care PTSD Screen (PC-PTSD): development and operating characteristics. *Primary Care Psychiatry* 9:9-14, 2003
11. Langbehn DR, Pfohl BM, Reynolds S, et al: The Iowa Personality Disorder Screen: development and preliminary validation of a brief screening interview. *Journal of Personality Disorders* 13:75-89, 1999
12. Weathers FW, Ruscio AM, Keane TM: Psychometric properties of nine scoring rules for the Clinician Administered Posttraumatic Stress Disorder Scale. *Psychological Assessment* 11:124-133, 1999
13. Perkins NJ, Schisterman EF: The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *American Journal of Epidemiology* 163:670-675, 2006
14. Steadman HJ, Clark Robins P, Islam T, et al: Revalidating the brief jail mental health screen to increase accuracy for women. *Psychiatric Services* 58:1598-1601, 2007
15. Shrout PE, Skodol AE, Dohrenwend BP: A two-stage approach for case identification and diagnosis: first-stage instruments, in *Mental Disorders in the Community: Progress and Challenge*. Edited by Barrett JE, Rose RM. New York, Guilford, 1986