Comparing the predictive utility of two screening tools for mental disorder among probationers

Jennifer Eno Louden, Ph.D.
The University of Texas at El Paso

Jennifer L. Skeem, Ph.D.
University of California, Irvine

Alishia Blevins, M.A.
Lewis & Clark College

Jenni Dillman, M.A.
University of Southern California

©American Psychological Association, 2012. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. Please do not copy or cite without author's permission. The final article is available, upon publication, at: doi: 10.1037/a0031213
Author Note

Jennifer Eno Louden, Department of Psychology, The University of Texas at El Paso; Jennifer L. Skeem, Department of Psychology and Social Behavior, University of California, Irvine; Alishia Blevins, Department of Counseling Psychology, Lewis and Clark College; Jenni Dillman, School of Social Work, University of Southern California.

This research was supported by grants from the American Psychology-Law Society (AP-LS), the Society for the Psychological Study of Social Issues (SPSSI), and the University of California, Irvine. The authors thank administrators at the County of San Bernardino Probation Department for their assistance with this research. Portions of this research were presented at the 2008 conference of the American Psychology-Law Society.

Correspondence regarding this article should be addressed to Jennifer Eno Louden, Psychology Department, The University of Texas at El Paso, 500 W. University Ave., El Paso, TX, 79968. E-mail: jlenolouden@utep.edu
Abstract

To increase the likelihood that probationers with serious mental disorder can be identified by probation agencies, we tested the utility of two promising mental health screening tools, the K-6 and the Brief Jail Mental Health Screen (BJMHS), in identifying probationers with DSM-IV Axis I mental disorders. In this study, 4,670 probationers completed the screening tools as part of routine intake procedures at a probation agency. We interviewed a subset of 149 probationers using a structured clinical interview to determine whether they met criteria for an Axis I anxiety, mood, or psychotic disorder at any point during their lifetime. The BJMHS correctly identified 77% of probationers with mental disorder overall, and contrary to our hypotheses, the measure was as sensitive with women as it was with men. Adding items assessing PTSD provided some incremental utility to the measure. The K-6 demonstrated similar levels of sensitivity to the BJMHS when using a lower cutoff score informed by prior research with offenders. Contrary to prior research, we found that the BJMHS was sensitive to mental disorder among female probationers. We suggest that probation agencies implement either the BJMHS or K-6 as part of their intake procedures to identify probationers with serious mental disorder so the complex needs of this group can be better addressed.

Keywords: screening, psychopathology, criminal justice
The rate of serious mental disorders in correctional populations—major depression, bipolar disorder, and schizophrenia—is 3 times higher for men and 2 times higher for women than in the general population (Teplin, 1990; Teplin, Abram, & McClelland, 1996). Overall, approximately 14.5% of male offenders and 31.0% of female offenders have a serious mental disorder (Steadman, Osher, Robbins, Case, & Samuels, 2009). Because offenders with mental disorder have special needs and often experience relatively poor criminal justice outcomes, a number of experts have recommended that criminal justice agencies systematically “screen” offenders at intake to help identify those with mental disorder so that the challenges associated with their supervision can be better addressed (see Chandler, Peters, Field, & Juliano-Bult, 2004; National Institute of Justice, 2007; Taxman, 1998). These recommendations have focused primarily on prisons and jails, which are required to provide mental health services to inmates who need them (see National Institute of Justice, 2007). The notion of screening offenders in probation agencies has received less attention, which is unfortunate, given that (a) these agencies supervise the majority (58%) of all offenders in the United States (Glaze & Bonczar, 2011), and (b) offenders with mental disorder are more likely to fail community supervision than their relatively healthy counterparts, which often results in a return to custody (Eno Louden & Skeem, 2011; Porporino & Motiuk, 1995).

Screening probationers is a first step towards addressing their complex needs. Many probation agencies across the United States have implemented specialty mental health caseloads, where probationers with mental disorder are supervised by officers with specialized training and smaller than average caseloads (Skeem, Emke-Francis, & Eno Louden, 2006). Screening probationers at intake would identify those who may be eligible for such caseloads. Even in
agencies without specialty caseloads, probationers identified with mental disorder may be referred to local treatment providers for mental health treatment.

Unfortunately, there does not appear to be any evidence on the validity of specific mental health screens in probation populations. A few mental health screening tools have shown promise for institutional settings (i.e., jail and prison; see National Institute of Justice, 2007), but as described below, these may not be ideal for use in probation agencies. Thus, identifying a mental health screen for use in probation agencies is imperative.

**Characteristics of the Ideal Mental Health Screen for Probation Agencies**

The ideal mental health screen for probation agencies would be easily administered, be sensitive to serious mental disorder, and have predictive utility for both men and women (see Council of State Governments, 2009). First, as noted above, probation agencies process much higher volumes of offenders relative to jails and prisons (Glaze & Bonczar, 2011), so brevity and ease of use is crucial in this context. Some well-studied screens are not appropriate for use in probation agencies, given the amount of time needed to administer them (e.g., the 53-item Brief Symptom Inventory, Derogatis & Melisaratos, 1983; the 67-item Defendant and Offender Screening Tool, Ferguson & Negy, 2006). A short, self-report measure would be ideal, so that multiple offenders could be processed simultaneously.

Second, the tool should accurately identify mental disorder among probationers. Commonly used indices of measurement accuracy include sensitivity (rate of true positives), specificity (rate of true negatives), positive predictive value (PPV, likelihood that an individual has a disorder when the test indicates he does) and negative predictive value (NPV, likelihood that an individual has no disorder when the test indicates he has no disorder; Zhou, McClish, & Obuchowski, 2002). There are no set standard optimal values for these indices; rather, ideal
values should reflect the consequences of incorrect classifications made by the measure (Zhou et al., 2002). In this case, we must weigh the consequences of falsely identifying a probationer with a mental disorder versus letting a probationer who does have a mental disorder go undetected. For screening tools, sensitivity is more important than specificity, given that failure to identify a probationer with a serious mental disorder would result in the probationer not receiving needed services, whereas false positives could be weeded out with further evaluation. Further, high NPV is desirable, as this indicates that probationers who are identified as having no disorder by the measure are unlikely to actually have a disorder and thus go undetected.

Third, the tool should be able to identify lifetime history of serious mental disorder, rather than only symptoms present at intake. Although current symptoms may be a pressing issue for officers to address given immediate treatment needs and potential crisis situations, cases of mental disorder tend to persist, and probationers with histories of serious disorders such as schizophrenia could decline in functioning over time (Teplin, Abram, & McClellan, 1997; Teplin & Swartz, 1989). Serious psychotic and mood disorders are episodic in nature (American Psychiatric Association, 2004), so if a measure identifies lifetime presence of disorder, agencies will know both (a) which probationers currently need services, and (b) which probationers may need services in the future.

Fourth, the tool should effectively identify mental disorder in both men and women. The proportion of women on community corrections supervision is steadily increasing (Glaze & Bonczar, 2011). These women have particularly high rates of mental disorder compared to men, and women generally present with different disorders than men do. Often, mental health screens developed in correctional samples are developed with predominantly male samples and may not accurately detect mental disorder among women (e.g., Steadman, Scott, Osher, Agnese, &
Robbins, 2005). One noteworthy example is post-traumatic stress disorder (PTSD), which is common among incarcerated women (Teplin 1996) and is associated with a history of trauma (Messina & Grella, 2006). An effective screen should be sensitive to the mental disorders present in both men and women in correctional samples, including major depression, bipolar disorder, psychotic disorders, and post-traumatic stress disorder (Teplin, 1990; Teplin et al., 1996).

Promising Candidate Screens

Two existing screening tools hold promise for use in probation agencies across the criteria mentioned above: the K-6 (Kessler, Andrews, Colpe, Hiripi, Mroczek, Normand, Walters, & Zaslavsky, 2002) and the Brief Jail Mental Health Screen (BJMHS; Steadman et al., 2005; 2007). We discuss the relative strengths and weaknesses of these screens next.

The K-6. The K-6 (Kessler et al., 2002) is a 6-item self-report screening tool designed to discern mental disorder from general distress. The K-6 asks respondents to rate six symptoms using the prompt “During the past 30 days, about how often did you feel…” The items are listed in Table 1. Respondents are asked to rate each item on a 5-point scale (where 0 = “None of the time” and 4 = “All of the time.” Scores on these items are summed to produce a total score that may range from 0 to 24, and the traditional cutoff score for community samples is 13.

The K-6 was developed for, and extensively studied in, general community populations in large epidemiological studies (Kessler, Barker, Colpe, Epstein, Gfroerer, Hiripi, Howes, et al., 2003; Patel, Araya, Chowdhary, King, Kirkwood, et al., 2008; Swartz & Lurigio, 2006). Among 155 non-offenders assessed as part of the U.S. National Health Interview Survey, the Area Under the Curve (AUC) value assessing precision for the K-6 at its traditional cutoff score at identifying any DSM-IV axis I mental disorder (not substance abuse or dependence) and a
Global Assessment of Functioning score below 60 was 0.87. This AUC value equates to an effect size of $d = 1.22$, a large effect (see Cohen, 1992; Rice & Harris, 2005). This indicates that a person with mental disorder and a person without mental disorder, chosen randomly, would be correctly categorized based on their scores on the K-6 87% of the time. Although this is strong evidence for the accuracy of the K-6, one potential drawback is that its strength is primarily in its high specificity (0.96) rather than its sensitivity (0.36; Kessler et al., 2003). However, this is evidence that the K-6 is meeting its intended goal: it was developed to screen community samples to identify adults with “serious mental illness” who would qualify for services under a specific program (Kessler et al., 2003).

Although the K-6 is brief and is sensitive to mental disorder in the general population, a question can be raised about its generalizability to correctional populations, as it was designed to screen non-referred community samples. Questions assessing affect or anxiety may be over-endorsed among individuals who have been recently incarcerated or placed on probation supervision, because they are distressed over their legal situation rather than exhibiting symptoms of a mental disorder. We could locate only one study of the K-6 with a correctional sample. Here, based on a sample of 515 jailed women, Kubiak, Beeble, and Bybee (2009) found an AUC value of 0.92 for the K-6 in detecting self-reported major depression and PTSD, two of the most common disorders among incarcerated women. More specifically, the sensitivity of the measure was 0.67 for depression and 0.59 for PTSD when using the traditional cutoff score of 13. Although this is promising evidence for the use of accuracy of the K-6 in correctional sample, it is only one study with limitations—data are only available for women (not men), and all measures were self-report, meaning the relationship between the K-6 and other symptoms could be inflated due to monomethod bias (Shadish, Cook, & Campbell, 2002).
Another potential limitation of the K-6 is that its cutoff score was developed for use in the general community, not in offenders. Kessler et al. (2002) recommended adjusting the cutoff score in populations where the prevalence of mental disorder was different than that of the general community, so altering the cutoff score in correctional samples is appropriate. Kubiak et al. (2009) noted that the K-6’s sensitivity could be improved by lowering the standard cutoff score, and, based on their analyses, suggested an offender cutoff score of 6, rather than the general population cutoff of 13. However, this lower cutoff score has yet to be examined in other samples.

The BJMHS. The Brief Jail Mental Health Screen (BJMHS; Steadman et al., 2005) is an 8-item interview scale that focuses on symptoms of serious mental disorder. The BJMHS items are presented in Table 1, and consist of six items assessing current symptoms and two additional items assessing current or past treatment; all items are answered “yes” or “no.” Steadman et al. (2005) report that a person screens in on the instrument if they either (a) answer yes to two or more of the symptom items, or (b) acknowledge currently taking medication or having a history of psychiatric hospitalization. Although the BJMHS was developed as an interview, the straightforward “yes or no” format of the items makes it a good candidate for use as a self-report measure, which would lend itself to efficient administration in probation agencies, where there may be need to process multiple offenders simultaneously. Further, a self-report format may be perceived by probationers as more private than an interview with a corrections officer.

In its initial validation study, where 10,330 pretrial detainees were screened and 357 were evaluated with structured mental health interviews, the BJMHS demonstrated sensitivity of 0.66 and specificity of 0.77 (Steadman et al., 2005). This initial study suggested that the BJMHS performed poorly among women, where 35% of women with a mental disorder were falsely
determined to be non-disordered (Steadman et al., 2005); in other words, the BJMHS “missed” a sizeable number of women who had a mental disorder. However, a second study was conducted in four jails, where a total of 464 detainees were administered clinical interviews. This study suggested that the measure’s performance varied as a function of the facility where it was applied (Steadman, Robbins, Islam, & Osher, 2007). In this second study, the authors found comparable levels of sensitivity (0.61) and specificity (0.75) in women as they had in men in the initial study (Steadman et al., 2007).

Steadman et al. (2007) hypothesized that the BJMHS’s poor performance among women may specifically be due to a lack of criteria assessing post-traumatic stress disorder (PTSD). However, adding an item to the BJMHS to capture PTSD symptoms (along with three additional items capturing depression) did not markedly improve its performance (Steadman et al., 2007). Given that single item scales tend to be unreliable (Lord & Novick, 1968), it is possible that more than one item is needed to capture distress associated with PTSD among incarcerated women.

To our knowledge, no study has directly compared the utility of the K-6 and the BJMHS among correctional samples. One study that approximated this was conducted by Swartz (2007). In this study, he used 2001 and 2002 data from the National Household Surveys on Drug Use and Health, confining his analyses to respondents who indicated having been arrested or on community corrections supervision during the past year (n = 4,884). Here, the sensitivity of the K-6 for men and women was 0.62 and 0.70, respectively, in detecting serious mental disorders based on the Composite International Diagnostic Interview Schedule-Short Form (CIDI-SF). Comparing these figures to Steadman’s (2005) findings, Swartz (2007) concluded that the K-6 was the superior tool. However, this study did not directly compare the measures in the same
sample. Further, this study’s sample—persons living in the community who self-identified as having recent criminal justice involvement—limits the generalizability of its findings. Given the need to detect mental disorder on the front lines in probation agencies, a more ecologically valid comparison of these tools is warranted.

**The Present Study**

Both the K-6 and BJMHS are promising tools that may work well at identifying probationers with mental disorder, but no study to date has directly tested these tools in a probation setting. In the present study, we had one primary aim: to compare the accuracy of the K-6 and BJMHS at identifying probationers with a lifetime diagnosis of mental disorder during routine intake at a probation agency. As noted earlier, the BJMHS is traditionally administered via interview, but given the need for efficiency in probation agencies, we test both measures as self-report. Given prior findings, we assess the accuracy of the K-6 both at its published cutoff score for use with the general population (Kessler et al., 2002; see also Swartz, 2007), and the lower cutoff score suggested by Kubiak et al. (2009). We hypothesize that the BJMHS will be more accurate at identifying mental disorder among male probationers than female probationers, given that it was developed for use in mostly male correctional populations but has mixed support for its accuracy among women (Steadman et al., 2005; 2007). We hypothesize that the K-6 will have similar rates of accuracy for men and women.

Given limitations of each tool identified in prior research, we had two sub-aims, each attempting to further examine prior limitations and identify remedies for them. Sub-aim 1 determines whether the accuracy of the K-6 is moderated by the extent to which distress is caused by recent legal problems. We hypothesized that the extent to which recent legal problems contribute to probationers’ current distress would decrease the effectiveness of the K-6, so
adding an item to capture these problems would improve the measure’s specificity. Sub-aim 2 determines the extent to which adding items capturing symptoms of PTSD would improve the accuracy of the BJMHS. Given Steadman et al.’s (2005, 2007) research, we hypothesized that adding items to capture these symptoms would increase the accuracy of this tool among women.

**Method**

The study was conducted in two stages. In Stage 1, a probation agency administered a screening questionnaire comprised of the K-6, the BJMHS, and the PTSD items from the MINI screen (described below) to probationers during its routine intake process. In Stage 2, a subset of these probationers participated in mental health diagnostic interviews to assess the accuracy of the screening tools.

**Procedure: Stage 1 Screening**

Adult probationers in a large probation agency completed the screens as part of their standard probation intake paperwork. Approximately 10% of probationers did not complete the screens because they could not read English well enough to do so. Probationers who completed the screens were asked to provide their contact information on a separate form if they were interested in potentially completing a later research interview (Stage 2, below). Characteristics of all probationers who completed the screens \((N = 4,670)\) as well as the subset who agreed to be contacted for the Stage 2 interview are listed in Table 2. Although 66.2% of probationers declined to be contacted for the interview \((n = 3,091)\), they did not differ significantly in age or ethnicity from those who agreed to be contacted \((n = 1,579)\), as both groups were approximately 31 years old \((M = 31.3, SD = 10.5\) and \(M = 31.1, SD = 10.1\) respectively; \(t(4611) = .76, p = .45, d = .02\) and were about evenly split across the Caucasian, African-American and Hispanic ethnic groups \((\chi^2(3) = 7.49, p = .06, w = .04)\). Probationers who agreed to be contacted were more
likely to be women than those who refused to be contacted, although the effect size here was small (30.5% versus 23.5%, $\chi^2(1) = 26.1$, $p < .001$, $w = .07$). Further, those who agreed to be contacted scored higher on the screening tools than those who declined to be contacted. Specifically, they had higher K-6 scores ($M = 5.64$ ($SD = 4.97$) versus $M = 4.04$ ($SD = 4.23$); $t(4562) = -11.38$, $p < .001$, $d = .35$), and were more likely to meet the “screen in” criteria for the BJMHS (43.6% versus 27.3%, $\chi^2(1) = 123.13$, $p < .001$, $w = .16$), and the MINI PTSD items (33.0% versus 16.4%, $\chi^2(1) = 160.58$, $p < .001$, $w = .19$).

[INSERT TABLE 2]

Measures

**K-6.** As described earlier, the K-6 (Kessler et al., 2002) consists of six items rated on a scale from 0 to 4 (0 = “None of the time,” 1 = “A little of the time,” 2 = “Some of the time,” 3 = “Most of the time,” and 4 = “All of the time”) and possible scores range from 0 to 24. Scores of 13 or higher indicate that the respondent likely has a mental disorder (defined as any axis I mental disorder with a GAF below 50). The K-6 was internally consistent in our sample, $\alpha = 0.84$. To account for the possibility that the symptoms assessed by the K-6 were due primarily due to the probationer’s recent legal problems rather than indicative of mental disorder, we added an item immediately after the K-6 items: “During the past 30 days, how often have your legal problems been the main cause of these feelings?” This item was scored on a scale of 1 to 5 where 1 = “None of the time” and 5 = “All of the time.”

**Brief Jail Mental Health Screen (BJMHS).** The BJMHS (Steadman et al., 2005) contains eight “yes” or “no” questions that are organized in two sections. As described earlier, the first section consists of six items that assess the occurrence of mental health symptoms in the last six months. For the majority of analyses, we scored the BJMHS in the manner described by
Steadman et al. (2005) where a person screens in on the instrument if they either (a) answer yes to two or more of the symptom items, or (b) acknowledge currently taking medication or having a history of psychiatric hospitalization. For some analyses, we computed the BJMHS score as the number of “yes” responses (which could range from 0 to 8) so that the presence of a linear relationship between BJMHS items endorsed and diagnosis could be assessed—these analyses are identified below. Unless identified as using a continuous score, we conducted analyses using the binary screen in/out status for the BJMHS, as this follows its intended scoring scheme suggested by its authors (see Steadman et al., 2005). Although this measure was developed as an interview, it was administered as a self-report, since the goal of this study was to identify a measure that could be used by probation agencies as a self-report. The reliability coefficient of the BJMHS in our sample was $\alpha = 0.63$.

**MINI Screen: PTSD items.** The MINI International Neuropsychiatric Interview (Sheehan, Janavs, Baker, Harnett-Sheehan, Knapp, & Sheehan, 2004) offers a 21-item screen that includes three items assessing PTSD. These items address both the presence of a traumatic event (e.g., physical or sexual assault) and responses typical to persons with PTSD (i.e., responding to the event with intense fear, helplessness or horror; re-experiencing the event in distressing ways such as flashbacks). The three items from the MINI Screen (Sheehan et al., 2004) that assess PTSD were added to determine whether they moderated the sensitivity of the BJMHS with women. These items ask respondents to indicate (yes or no) whether they have ever experienced a traumatic event (such as sexual or physical assault) and if so, whether they (a) responded to the trauma with intense fear, helplessness, or horror, and (b) re-experienced the event in a distressing way (such as flashbacks or physical reactions) within the past 30 days. These screening items are designed to inform administration of the MINI diagnostic interview,
where a positive response to any one of the items elicits further questioning, but the full PTSD module of the interview is only administered if the interviewee endorses all three PTSD screening items. Because our goal was to identify probationers who may meet criteria for PTSD, we scored this PTSD screen as “positive” if respondents indicated they experienced a traumatic event and endorsed at least one of the items reflecting post-traumatic symptoms. Internal consistency of scores on the MINI PTSD items in our sample was $\alpha = 0.70$.

**Procedure: Stage 2 Structured Clinical Interview**

As noted in the previous section, 1,579 probationers agreed to be contacted to participate in a research interview. Our target sample for structured clinical interviews was 150; a power analysis suggested would provide sufficient power to detect at least medium differences in the utility of the screens in predicting whether probationers receive diagnoses of major mental disorders ($SE = 0.04$ with a ROC approximately 0.75; see Hanley & McNeil, 1982). To maximize statistical power, it was important to recruit balanced subsamples of participants with- and without- major mental disorder (i.e., $n = 75$ per group; see Kessler, 2003; Metz, 1978). Because the K-6 was most likely to work with male and female probationers, we used it as the method for categorizing interviewees with- and without- mental disorder.

As shown in Figure 1, according to the K-6 at its community cutoff score, 174 (11%) of the 1,579 probationers who agreed to be contacted for an interview “screened in” as having a mental disorder. The remaining 1,405 (89%) “screened out.” We randomly selected probationers in each group to recruit for the interview. As shown in Figure 1, we recruited 123 probationers who screened in and 132 who screened out, with 41% and 42%, respectively, actively or passively refusing to participate. This resulted in 149 probationers agreeing to be interviewed, 72 of whom screened in and 77 who screened out.
Compared to probationers who were successfully recruited, those who were not able to be recruited were not different in terms of gender ($\chi^2(1) = 0.24, p = .62, w = .03$), ethnicity ($\chi^2(4) = 6.27, p = .18, w = .16$), or mean age ($t(252) = -1.78, p = .08, d = .22$). Further, those who were interviewed had similar scores on the K-6 to those who were not successfully recruited ($M = 9.81, SD = 6.7$ versus $M = 9.82, SD = 6.5$, $t(252) = .01, p = .99, d = .00$) and endorsed a similar number of items on the BJMHS ($M = 1.68, SD = 1.6$ versus $M = 1.46, SD = 1.6$, $t(251) = -1.01, p = .28, d = .14$). Because we used screening status on the K-6 at its community cutoff as criteria for sorting probationers into screened in/screened out groups for the purposes of ensuring enough variability in screening status among probationer who were interviewed, there was the possibility that there would be limited range on the BJMHS. However, as shown in Table 2, the percentage of probationers who screened in on the BJMHS was the same as the percentage who screened in on the K-6 at its lower cutoff. Further, the range of each measure in the Stage 2 sample represented the full range of values for each measure, with values on the K-6 ranging from 0 to 24 and values on the BJMHS ranging from 0 to 6.

Interviews lasted approximately 60 to 90 minutes and were conducted in the probationer’s home, a quiet public place (such as a coffee shop), or in a private room in the probation agency. On average, these interviews occurred 86.8 ($SD = 75.4$) days after the probationer completed the screening questionnaire at the probation agency. Probationers were assured strict confidentiality of their participation and responses, and at the end of the interview were paid $20 for their time.

[INSERT FIGURE 1]
Structured Clinical Interview for DSM-IV (SCID). The Structured Clinical Interview for DSM-IV (SCID), a semi-structured clinical interview protocol that is specifically designed to conform to diagnostic criteria of the DSM-IV (First, Spitzer, Gibbon, & Williams, 2002), was used as the criterion for determining whether a probationer had a current or lifetime diagnosis of a mental disorder. For the purposes of the present study, we used the Mood Disorders, Psychotic Disorders, and Anxiety Disorders modules, because they include the most serious mental disorders found in correctional populations (see Fazel & Danesh, 2002; Teplin et al., 1996).

Interviews were administered by advanced research assistants who participated in at least 16 hours of didactic training on psychopathology, clinical interviewing skills, and administration of the SCID. This training was informed by the Training and Quality Assurance Program for SCID interviewers developed for the UCLA Research Center for Major Mental Illnesses (Ventura, Liberman, Green, Shaner, & Mintz, 1998). This was administered by a doctoral-level clinical psychologist (the second author), and included watching training videotapes, studying the user guide distributed by the authors of the SCID, participating in role-playing cases, and observing live interviews. As suggested by Ventura et al. (1998), we computed inter-rater reliability at the symptom level of the SCID using weighted kappa. We did so for the last four (out of six) training cases for each interviewer, where each interviewer’s rating was compared to a criterion rating, and the average value of weighted kappa was 0.82. To protect against rater drift over time, interviewers rated 3 ongoing interviews in pairs. The mean weighted kappa value for these live interviews was 0.82.

Results

To address the study’s primary aim—testing the predictive utility of the K-6 and the BJMHS—we computed a range of statistics: sensitivity, specificity, positive predictive value
(PPV), and negative predictive value (NPV). Sensitivity is the probability that the test will detect the condition (mental disorder, in this case) when the condition is present (true positives in signal detection theory terminology). Specificity is often presented as 1-specificity, which is the rate of false positives (Zhou et al., 2002). PPV and NPV reflect the ability of the tool to detect true status of the condition (these indicate the condition is present, or absent, respectively).

Because PPV and NPV are affected by the prevalence of the condition in the sample and we oversampled probationers with mental disorder, we adjusted these values for base rates of mental disorder seen in correctional populations (14.5% overall and for men, and 31.0% for women; see Steadman et al., 2009). We did this using formulae provided by Baldessarini, Finkelstein, and Arana (1983).

To provide a single index of accuracy of each measure, we present the overall diagnostic power of each measure, which was computed as the percent of correct classifications (true positives plus true negatives; see Baldessarini et al., 1983). Finally, we computed kappa statistics, which is a measure of chance-corrected agreement and relative performance of screens can be discerned from the degree of overlap of the confidence intervals. We present kappa values as an overall measure of agreement between each screening measure and the SCID so that we could determine if one screen had significantly better agreement with the SCID than the other. Each of these statistics was computed both for the whole sample and separately for men and women.

For sub-aim 1, we computed a logistic regression moderation analysis (see Baron & Kenny, 1986) to determine whether the item capturing distress due to legal problems moderated the performance of the K-6. For sub-aim 2, we computed a regression analysis to examine whether the MINI PTSD items provided incremental utility for the BJMHS.
Preliminary Analyses

Before addressing the aims of the study, we first computed some preliminary analyses to examine (a) intercorrelations among the screening measures, and (b) the base rate of mental disorder in our Stage 2 sample. The intercorrelations among K-6 and BJMHS items are presented in Table 1. Overall, there were low to moderate levels of correlation for items on the K-6 to items on the BJMHS. This is consistent with the notion that the measures capture somewhat different types of psychopathology; the K-6 assesses primarily negative affect whereas the BJMHS assesses symptoms of thought disturbance and physical symptoms associated with serious psychopathology. At the scale level, if we scored the BJMHS as a sum of the number of “yes” responses, it was correlated to total K-6 scores at $r = .52, p < .000$, indicating that although the item content of the scales are largely different, there is quite a bit of overlap in scores at the scale level.

As noted earlier, the criterion for each screen was lifetime diagnosis of any axis I mental disorder, not counting disorders of substance abuse or dependence. Overall, 73.8% of all probationers interviewed had a lifetime history of any axis I mental disorder (i.e. 110 individuals had a diagnosis and 39 had no diagnosis). By gender, 66.3% of male probationers and 89.6% of female probationers had a lifetime history of an axis I disorder. In addition, of probationers who screened in on the K-6 (at the community cutoff), 91.9% had a diagnosis, whereas 56.8% of probationers who screened out on the K-6 had a diagnosis. Given that we used the higher community cutoff score of the K-6 for initial selection into the screened in/screened out groups, it is not surprising that many probationers who screened out on the K-6 had a diagnosis of a mental disorder.

[INSERT TABLE 3]
**Primary Aim: Comparative Accuracy of the K-6 Versus the BJMHS**

The relative accuracy of predicting lifetime mental disorder for each of the screens is presented in the first five rows of Table 3. As shown, there was a great deal of overlap in the accuracy of the BJMHS and K-6 when examining the kappa statistics and their corresponding confidence intervals. We present statistics for the K-6’s general population-based cutoff score of 13 primarily for descriptive purposes; given that (a) there is a different cutoff score suggested for offenders, and (b) the K-6 had low sensitivity for our sample at this cutoff score (0.62). The third column in Table 3 shows the performance of the K-6 at the offender cutoff score of 6 suggested by Kubiak et al. (2009). With this lower cutoff score, the K-6 correctly identified 76% of probationers with mental disorder (sensitivity), which is similar to the BJMHS’s sensitivity of 0.77. Likewise, the measures had similar rates of false positives: the K-6’s 1-specificity rate was 0.34, compared to 0.31 for the BJMHS. This indicates that 34% and 31%, respectively, of probationers identified as having a mental disorder by each screen actually did not have a mental disorder. However, as indicated by the NPV index, the measures were able to accurately identify most probationers who did not have a mental disorder—94% of probationers who screened out on the K-6 and 95% of probationers who screened out on the BJMHS did not have a mental disorder.

In the remaining rows of Table 3, we present the relative accuracy of each screen separately for women and men. As noted earlier, we predicted that the BJMHS would perform well for men but less so for women, and that the K-6 would perform well for both men and women. Contrary to our hypothesis, we found that the BJMHS was sensitive to lifetime disorder for both men and women: its sensitivity was 0.81 for women and 0.75 for men. Using its offender cutoff score, the K-6 was nearly as sensitive as the BJMHS, with sensitivity values of...
0.79 and 0.75 for women and men, respectively. However, at this lower cutoff, the K-6 had a higher rate of false positives among women than the BJMHS (0.20 compared to 0.00). In addition, the K-6’s NPV rate was somewhat lower for women than men, indicating that there is a somewhat higher likelihood that a woman who screens out on the measure actually does have a mental disorder. In terms of overall agreement between the two screens and the SCID, the kappa values were similar (0.36 and 0.42, respectively for the K-6 and the BJMHS), and there was a high degree of overlap in the confidence intervals as shown in Table 3. This indicates that the two measures were similar in the extent to which they “agreed” with diagnostic status on the SCID. In sum, when using the K-6 at the offender cutoff, it demonstrated equivalent sensitivity to the BJMHS in detecting mental disorder among men and women. The measures had similar specificity among men, but the BJMHS was more effective than the K-6 at discriminating mental disorder from non-disorder among women.

Sub-aim 1: Do Legal Problems Moderate the Accuracy of the K-6?

As noted earlier, we computed a moderation analysis as recommended by Baron and Kenny (1986) to determine if the extent to which probationers’ distress was caused by recent legal problems would inflate the number of probationers screening in on the K-6. The mean total score on the K-6 was 15.96 (SD = 6.6) and the mean on the legal problems item was 2.67 (SD = 1.5). We computed a binary logistic regression where the dependent variable was presence of a lifetime axis I diagnosis (dummy coded where 1 identified individuals who had a diagnosis). In the first block, we entered the continuous K-6 total score and continuous legal problems item as predictors. As we expected, the K-6 significantly predicted diagnosis ($B = 0.20, SE = .05, p < .000$), but the legal problems item did not ($B = -.01, SE = .22, p = .98$). In the second block, we entered the interaction between these two variables. If there were a significant interaction, we
would expect that it would indicate that the K-6 was less predictive of mental disorder for probationers with high scores on the legal problems item than for those with low scores on the item. The interaction term was not significant ($B = -0.023$, $SE = .03$, $p = .46$), suggesting that the legal problems item did not moderate the performance of the K-6 in identifying mental disorder. This suggests that the degree to which scores on the K-6 were related to the probationer’s legal problems did not affect the accuracy of the K-6 in predicting lifetime mental disorder. Thus, including the item assessing distress due to legal problems did not markedly improve the measure’s accuracy.

**Sub-aim 2: Do PTSD Items add Incremental Utility to the BJMHS?**

Finally, we examined whether the PTSD items of the MINI screen would add incremental utility to the BJMHS. To do this, we calculated a binary logistic regression analysis where BJMHS screening status (screened in or screened out) was entered as block 1 and MINI scores were entered as block 2 to examine whether the MINI predicted diagnosis, holding constant BJMHS scores. Here, the BJMHS and the MINI both emerged as significant predictors in block 2 ($B = 0.55$, $SE = .20$, $p = .006$, and $B = 0.59$, $SE = .21$, $p = .006$, respectively). This indicates that the MINI offers some incremental utility over the BJMHS.

For descriptive purposes, we present the accuracy statistics for the combined BJMHS and MINI, where a probationer screens in if he or she meets the criteria for either of the two screens. As shown in the last column of Table 3, the addition of these items to the BJMHS increased the measure’s sensitivity slightly to 0.84 for men and women alike. However, the false positive rate also increased with the addition of these items. As shown, the 1-specificity rate changed from 0.31 for the original BJMHS to 0.39 with the addition of the MINI items. False positives were particularly a problem for men, where 1-specificity was 0.41, compared to 0.20. Thus, although
the BJMHS gained a small amount of sensitivity by adding the PTSD items, the gain was at the cost of an increased rate of false positives.

**Discussion**

This study was the first to compare the accuracy of existing mental health screening measures in a probation population. Given the considerable needs of the one in seven men and almost one-third of women in the criminal justice system who have a serious mental disorder, improved identification of psychopathology in this population is imperative. Considering the risk associated with failing to identify probationers who have a mental disorder, we primarily emphasized the measures’ sensitivity to mental disorder in our analyses. After comparing the K-6 and the BJMHS, we had two primary findings. First, the BJMHS was sensitive to lifetime mental disorder among probationers, and contrary to hypotheses informed by prior research, it was as sensitive among women as it was for men. Its sensitivity was improved somewhat by the addition of items assessing PTSD. Second, the K-6 demonstrated comparable levels of accuracy as the BJMHS when using an “offender” cutoff score, and this accuracy was not affected by the degree to which probationers’ distress was accounted for by recent legal problems. Here, we briefly describe the study’s limitations, then discuss our primary findings and conclusions.

**Limitations**

As with any research, this research had some limitations that should be taken into account when interpreting its results. First, our sampling frame was limited by the fact that the majority (66%) of probationers screened declined to be contacted for the research interview, and around 40% of probationers who agreed to be contacted for the research interview either refused participation or were unable to be located. Although those who declined participation were demographically similar to probationers who agreed to be contacted, probationers who agreed to
be contacted for the study had higher scores on the screening instruments than those who refused to be contacted. Ideally, probationers who agreed to be contacted would not differ in any meaningful way from those who refused, but the fact that those who we sampled from had a high likelihood of mental disorder is not discouraging, given that these probationers are most in need of accurate screening. Further, of probationers we attempted to recruit for clinical interviews, those who were successfully recruited did not differ in any meaningful way from those who were not successfully recruited—this indicates that our final sample was representative of probationers who agreed to be contacted. Second, approximately one in ten probationers were not screened because they were not able to read in English. Future research to translate these screening tools into other languages (e.g., Spanish) would partially ameliorate this problem.

**Key Findings**

*The BJMHS accurately identifies lifetime mental disorder among male and female probationers.* Our first primary finding was that the Brief Jail Mental Health Screen (BJMHS; Steadman et al., 2005), a measure developed for jail populations, was generally effective at identifying lifetime psychopathology in probationers. Of the probationers we interviewed, the BJMHS accurately classified three-quarters (75.2%) of them. This is consistent with Steadman et al.’s (2005) initial validation study of the measure, where most (68%) inmates were correctly classified. Further, our findings regarding the accuracy of the BJMHS puts in on par with much longer measures. For example, the 67-item Defendant and Offender Screening Tool, (DOST; Ferguson & Negy, 2006) has a specificity rate of 0.86 at identifying psychosis. Although the DOST is a bit more sensitive, the BJMHS comes close with 59 fewer items. We view this as promising evidence that the BJMHS can be useful at identifying probationers who may have a serious mental disorder and are in need of further assessment.
Further, our results suggest that the BJMHS is accurate with both men and women, which was a concern in prior research with the tool. Although the initial validation study for the BJMHS found that it was not sensitive to mental disorder among women—correctly identifying only 61.6% of women and having a 34.7% false negative rate with this group (Steadman et al., 2005)—our results were consistent with a revalidation study that found acceptable levels of sensitivity for the BJMHS among women (Steadman et al., 2007). Although our sample of women was somewhat small ($n = 48$), the BJMHS correctly classified 83.3% of them.

Another concern raised in prior research with the BJMHS (see Steadman et al., 2005) is that symptoms of post-traumatic stress disorder (PTSD) may affect the accuracy of the BJMHS with women. This concern is related to the fact that there are high rates of PTSD among justice-involved women (Teplin et al., 1996), and existing screens may not be sensitive to these symptoms. We found that adding items assessing PTSD symptoms added a small amount of incremental utility to the BJMHS. However, these items increased false positives for the measure, particularly among men. Our findings suggest that the original BJMHS (Steadman et al., 2005) is accurate at identifying mental disorder among both male and female probationers.

We offer one caveat to the adoption of the BJMHS among probation agencies: This was the first research where the BJMHS was administered as a self-report questionnaire. Our findings suggest that probationers are willing to report symptoms assessed by the BJMHS in this format. Further, it is possible that probationers may perceive a self-report questionnaire as more private than directly discussing symptoms and mental health treatment history with a probation officer. Nonetheless, future research should examine the performance of the BJMHS as a self-report measure. In addition, researchers may want to consider revising the items of the BJMHS.
for easier readability. In particular, the item “Have there currently been a few weeks when you felt like you were useless or sinful?” may be confusing to probationers.

*The K-6 accurately identifies mental disorder using an “offender” cutoff.* Our second finding was that the K-6 had comparable accuracy to BJMHS when using a lower cutoff score. Given Kessler et al.’s (2002) recommendation to lower the cutoff score of the K-6 in samples with inflated levels of mental disorder, it is not surprising that the K-6 had low levels of sensitivity when scored with its community cutoff score, as rates of mental disorder in corrections populations are higher than in the community (Teplin, 1990; Teplin et al., 1996). When using the cutoff score of 6 suggested by prior research with incarcerated women (Kubiak et al., 2009), the K-6 matched the BJMHS’s level of sensitivity for both male and female probationers. Importantly, this is the first study we know of that has tested the K-6’s offender cutoff with male offenders, so additional research is needed to replicate this finding. The fact that we found similar levels of accuracy for the K-6 among both male and female probationers is promising.

Although we were concerned that probationers’ recent legal problems would inflate their scores on the K-6, given its focus on negative affect and worry, this was not the case. We added an item assessing the extent to which recent legal problems caused the distress captured by the K-6, but it did not moderate the accuracy of the K-6. This lends support to the notion that the K-6 is able to distinguish psychopathology from transient distress. Overall, our findings suggest that the K-6 is a promising screening tool for use with offender populations.

A concern that agencies may have with the K-6 and BJMHS is that both measures had false-positive rates around 32%, meaning that among probationers who screen in on each measure, a sizeable minority will not actually have a mental disorder. However, false positives
are a price to be paid for high sensitivity for these measures. Some false positives could be weeded out with further assessment by a trained officer before the probationer is referred for an expensive mental health assessment. The alternative would be lowering the sensitivity of the measures, in which case probationers with serious disorders such as schizophrenia may not be identified. Rather than being referred to treatment, these probationers could decompensate and have negative mental health or criminal justice consequences. Thus, the benefit of correctly identifying as many probationers with mental disorder as possible outweighs the consequences of these false positives.

Conclusions

Because offenders have markedly high rates of mental disorder compared to the general population (Teplin, 1990; Teplin et al., 1996), it is imperative that probation agencies—who supervise the bulk of these offenders—have an efficient yet effective tool to screen for mental disorder. Until now, there was no known data on the predictive utility of such tools in a probation setting. We found that both the Brief Jail Mental Health Screen (BJMHS) and the K-6 are accurate at identifying lifetime mental disorder among male and female probationers. These measures have utility in identifying probationers who do have a mental disorder, and are not likely to inaccurately categorize disordered probationers as non-disordered. Both of these measures are brief (6 and 8 items, respectively) and easily scored, and could thus be added to routine intake procedures without being time consuming or cumbersome. Agencies may consider adding the MINI PTSD items to the BJMHS to improve sensitivity, but with the understanding that this will also increase the number of false positives. Whichever measure is used, these tools represent a first step in assessment where probationers who “screen in” should receive a more in-depth assessment to determine whether they meet criteria for a DSM disorder.
Screening probationers will further multiple goals of probation agencies. First, if the agency has specialized services for probationers with mental health problems (e.g., specialty caseloads; Skeem et al., 2006), use of a screening questionnaire such as the BJMHS or the K-6 is a systematic method of identifying probationers who may be eligible for such programming. Given that these caseloads improve both mental health and criminal justice outcomes for probationers (Skeem, Manchak, & Peterson, 2011), increasing the likelihood that probationers are accurately assigned to such caseloads will likely improve outcomes for an increased number of probationers. Second, screening can identify probationers who may be in need of mental health treatment. In this case, agencies could implement procedures involving referrals for services for probationers who “screen in” on the measure. This would increase the likelihood that probationers will receive needed mental health treatment, which would alleviate symptoms and improve their functioning. Further, such treatment may improve probationers’ response to correctional treatment designed to reduce the likelihood that they will reoffend and plunge further into the criminal justice system (see Skeem et al., 2011). We recommend that probation agencies adopt either the BJMHS or the K-6 as part of their routine intake procedures to improve outcomes for probationers with mental disorder.
References


Figure 1

**Sampling Frame and Recruitment**

- **4,670 Screened**
  - 1,579 (34%) eligible
  - 3,091 (66%) refused to be contacted

- **174 (11%) screened in on K-6**
  - 123 recruited
  - 72 interviewed
  - 51 (41%) refused interview

- **1,405 (89%) screened out on K-6**
  - 132 recruited
  - 77 interviewed
  - 55 (42%) refused interview
Table 1

*Intercorrelations Among K-6 and BJMHS Items*

<table>
<thead>
<tr>
<th>Item</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>K-6</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. During the past 30 days, how often did you feel…nervous?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. …hopeless?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. …restless or fidgety?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.61</td>
<td>.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. …so depressed that nothing could cheer you up?</td>
<td></td>
<td></td>
<td></td>
<td>.55</td>
<td>.69</td>
<td>.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. …that everything was an effort?</td>
<td></td>
<td></td>
<td>.27</td>
<td>.27</td>
<td>.31</td>
<td>.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. …worthless?</td>
<td></td>
<td>.50</td>
<td>.72</td>
<td>.55</td>
<td>.69</td>
<td>.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BJMHS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Do you <em>currently</em> believe that someone can control your mind by putting thoughts into your head or taking thoughts out of your head?</td>
<td></td>
<td>.17</td>
<td>.22</td>
<td>.20</td>
<td>.24</td>
<td>.12</td>
<td>.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Do you <em>currently</em> feel that other people know your thoughts and can read your mind?</td>
<td></td>
<td>.16</td>
<td>.22</td>
<td>.19</td>
<td>.22</td>
<td>.12</td>
<td>.22</td>
<td>.39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Have you <em>currently</em> lost or gained as much as two pounds a week for several weeks without even trying?</td>
<td></td>
<td>.25</td>
<td>.25</td>
<td>.26</td>
<td>.28</td>
<td>.17</td>
<td>.25</td>
<td>.16</td>
<td>.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Have you or your family or friends noticed that you are <em>currently</em> much more active than you usually are?</td>
<td></td>
<td>.08</td>
<td>.05</td>
<td>.10</td>
<td>.06</td>
<td>.15</td>
<td>.26</td>
<td>.12</td>
<td>.09</td>
<td>.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Do you <em>currently</em> feel like you have to talk or move more slowly than you usually do?</td>
<td></td>
<td>.24</td>
<td>.27</td>
<td>.28</td>
<td>.27</td>
<td>.16</td>
<td>.26</td>
<td>.20</td>
<td>.19</td>
<td>.23</td>
<td>.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Have there <em>currently</em> been a few weeks when you felt like you were useless or sinful?</td>
<td></td>
<td>.37</td>
<td>.50</td>
<td>.42</td>
<td>.47</td>
<td>.22</td>
<td>.52</td>
<td>.26</td>
<td>.22</td>
<td>.26</td>
<td>.08</td>
<td>.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Are you <em>currently</em> taking any medication prescribed by a physician for any emotional or mental health problems?</td>
<td></td>
<td>.25</td>
<td>.23</td>
<td>.27</td>
<td>.28</td>
<td>.11</td>
<td>.24</td>
<td>.48</td>
<td>.17</td>
<td>.18</td>
<td>.19</td>
<td>.05</td>
<td>.20</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* All correlations significant at $p < .001$. 
Table 2

*Characteristics of Participants*

<table>
<thead>
<tr>
<th></th>
<th>Stage 1 (Screening)</th>
<th>Stage 2 (Interviews)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All ((N = 4,670))</td>
<td>Agreed to be contacted ((N = 1,579))</td>
</tr>
<tr>
<td>Gender (% male)</td>
<td>72.0</td>
<td>69.5</td>
</tr>
<tr>
<td>Mean age ((SD))</td>
<td>31.3 (10.4)</td>
<td>31.1 (10.1)</td>
</tr>
<tr>
<td>Ethnicity: (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>28.9</td>
<td>31.9</td>
</tr>
<tr>
<td>African American</td>
<td>21.9</td>
<td>22.5</td>
</tr>
<tr>
<td>Hispanic/Latino</td>
<td>43.3</td>
<td>42.2</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>1.7</td>
<td>1.1</td>
</tr>
<tr>
<td>Other</td>
<td>5.0</td>
<td>2.3</td>
</tr>
<tr>
<td>Screening results:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean on Legal Problems item ((SD))</td>
<td>1.93 (1.24)</td>
<td>2.15 (1.31)</td>
</tr>
<tr>
<td>Mean K-6 ((SD))</td>
<td>4.59 (4.55)</td>
<td>5.64 (4.97)</td>
</tr>
<tr>
<td>Screened in on K-6 at 6 cutoff (%)</td>
<td>20.3</td>
<td>28.7</td>
</tr>
<tr>
<td>Screened in on BJMHS (%)</td>
<td>31.6</td>
<td>42.8</td>
</tr>
<tr>
<td>Screened in on MINI PTSD items (%)</td>
<td>21.2</td>
<td>33.0</td>
</tr>
</tbody>
</table>

*Note.* The Legal Problems item was scored on a scale of 1 to 5 where 1 = “None of the time” and 5 = “All of the time.”
### Table 3

**Accuracy of Screening Questionnaires in Detecting Lifetime Psychopathology**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Measure</th>
<th>K-6</th>
<th>K-6 with 6 cutoff</th>
<th>BJMHS</th>
<th>BJMHS + Mini</th>
</tr>
</thead>
<tbody>
<tr>
<td>All cases</td>
<td>Sensitivity</td>
<td>0.62</td>
<td>0.76</td>
<td>0.77</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>1-Specificity</td>
<td>0.16</td>
<td>0.34</td>
<td>0.31</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>PPV</td>
<td>0.40</td>
<td>0.27</td>
<td>0.30</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>NPV</td>
<td>0.93</td>
<td>0.94</td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>Kappa (CI)</td>
<td>0.35 (0.20-0.50)</td>
<td>0.38 (0.21-0.55)</td>
<td>0.42 (0.26-0.58)</td>
<td>0.44 (0.27-0.61)</td>
</tr>
<tr>
<td></td>
<td>Dx. power</td>
<td>0.68</td>
<td>0.73</td>
<td>0.75</td>
<td>0.78</td>
</tr>
<tr>
<td>Women only</td>
<td>Sensitivity</td>
<td>0.67</td>
<td>0.79</td>
<td>0.81</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>1-Specificity</td>
<td>0.00</td>
<td>0.20</td>
<td>0.00</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>PPV</td>
<td>1.00</td>
<td>0.64</td>
<td>1.00</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>NPV</td>
<td>0.87</td>
<td>0.89</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>Kappa (CI)</td>
<td>0.30 (0.00-0.61)</td>
<td>0.35 (0.00-0.71)</td>
<td>0.48 (0.16-0.81)</td>
<td>0.42 (0.05-0.79)</td>
</tr>
<tr>
<td></td>
<td>Dx. power</td>
<td>0.71</td>
<td>0.75</td>
<td>0.83</td>
<td>0.83</td>
</tr>
<tr>
<td>Men only</td>
<td>Sensitivity</td>
<td>0.58</td>
<td>0.75</td>
<td>0.75</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>1-Specificity</td>
<td>0.18</td>
<td>0.36</td>
<td>0.35</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>PPV</td>
<td>0.35</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>NPV</td>
<td>0.92</td>
<td>0.94</td>
<td>0.94</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>Kappa (CI)</td>
<td>0.34 (0.16-0.52)</td>
<td>0.37 (0.17-0.56)</td>
<td>0.38 (0.19-0.57)</td>
<td>0.43 (0.24-0.63)</td>
</tr>
<tr>
<td></td>
<td>Dx. power</td>
<td>0.66</td>
<td>0.73</td>
<td>0.72</td>
<td>0.76</td>
</tr>
</tbody>
</table>
Note. For analyses with all cases, $N = 149$ for BJMHS and MINI and $N = 148$ for K-6. For women only, $n = 48$. For men only, $n = 101$ for BJMHS and MINI, and $n = 100$ for K-6. PPV and NPV values have been adjusted for base rates of mental disorder found in correctional populations (14.5% for men and 31.0% for women).