



**Criminal Justice
Drug Abuse Treatment Studies**

**The Criminal Justice Co-Occurring
Disorder Screening Instrument (CJ-
CODSI): Study Data Manual**

Study Summary

Lead Center: NDRI Rocky Mountain Research Center

Principal Investigator: Stanley Sacks, Ph.D. & Gerald Melnick, Ph.D.

Collaborating Research Centers:

Texas Christian University (TCU) - Southwest Research Center

University of California, Los Angeles - Pacific Coast Research Center

Lifespan/Rhode Island Hospital - Rhode Island Research Center,

Summary:

Available mental and substance use disorder screening instruments have not been widely validated for use with criminal justice populations and typically do not test for both disorders simultaneously. The Criminal Justice Co-occurring Disorder Screening Instrument (CODSI) study reviewed available screening instruments and determined that while several valid, brief, easy to use screens exist for substance use disorder for criminal justice populations; there were none for mental disorders. The study tested three mental disorder screeners on admissions to prison-based substance abuse treatment programs: The *Global Appraisal of Individual Needs* (GSS, version 1.0); the *Mental Health Screening Form* (MHSF); and the *Modified MINI Screen* (MMS) were tested using the *Structured Clinical Interview for DSM-IV* (SCID) as the criterion for mental disorder. Two additional instruments were created based on items from the three tested screeners in a pilot study of 100 subjects – the CODSI-MD for any mental disorder and the CODSI-SMD for severe mental disorder. A validation study based on an additional 180 subjects found the GSS, MHSF, MMS, CODSI-MD and CODSI-SMD as suitable mental disorder screening instruments for prisoners in substance abuse treatment programs. An Administrative Supplement, the CODSI-MS, was conducted to increase the proportion of minority offenders in the CODSI sample. Results show that the CODSI-MD and CODSI-SMD instruments exhibit no statistically significant differences by race/ethnicity on sensitivity, specificity, and overall accuracy.

Table of Contents

<u>Brief Report.....</u>	<u>3</u>
<u>Study Protocol.....</u>	<u>3</u>
<u>Instruments.....</u>	<u>29</u>
<u>Form Codebooks.....</u>	<u>29</u>
<u>Analysis Codebooks.....</u>	<u>29</u>
<u>Available Datasets.....</u>	<u>30</u>
<u>Publications.....</u>	<u>31</u>
<u>Appendix A:</u>	
CODSI Minority Supplement.....	33

Brief Report

Link:

[Brief Report: Criminal Justice Co-Occurring Disorder Screening Instrument \(CJ-CODSI\).](#)

Study Protocol

Study Objectives

Understanding the presence and extent of the co-occurrence of substance use and mental disorders (COD) is essential to the design of effective treatment that can meet the needs of criminal justice clients, both in prison and in the community. To this end, the investigators conducted a short-term (12 month) study designed to provide a brief screening instrument (to be completed in 20 minutes or less) to identify individuals with COD. The instruments created are appropriate for use by correctional and substance treatment staffs without specialized mental health training, and compatible with the modular set of instruments in the CJ-DATS instrument package.

The instrument has both clinical and research utility and fills a gap in the array of instruments presently available to CJ-DATS researchers. Clinically, the screening instrument provides a method to identify offenders with mental disorders who require a more extensive assessment. Research utility is gained through the provision of a standardized instrument that can be used to survey the rates of mental disorder in substance abusing populations of offenders.

Study Aim

The study aim was to provide a reliable and valid brief screening instrument to identify offenders with COD for use in CJ-DATS and by the criminal justice field. This aim involved (a) identifying and combining existing substance abuse and mental illness screening instruments, and considering them in relation to criteria to be determined by the investigative team working closely with criminal justice stakeholders; (b) constructing additional scales to meet stakeholder needs and adjusting the instrument to criminal justice settings (i.e., degree of risk for suicide, violence, disruptive behavior, and other concerns); and (c) determining the accuracy and other psychometric properties of the identified instruments in prison-based substance abuse treatment setting. The purpose of the screening instrument was to provide scores for substance use (alcohol and drug) and mental disorders, with a small number of additional scales that responded to needs identified by the stakeholder group. Identification of COD was to be based on quantitatively determined cut-off scores; individual items were scored as positive based on the degree to which a symptom interferes with everyday functioning. Criteria for the acceptability of the instrument included meeting appropriate psychometric standards, the ability to detect the presence of COD when it is present (sensitivity), and the ability to identify the absence of COD when it is not present.

Background and Significance

The reported proportion of offenders with serious mental disorders combined with chemical abuse (commonly termed "MICA") disorders in jail and prison populations has ranged from 3% to 16% (Peters & Hills, 1993; Regier, Farmer, Rae et al., 1990; Steadman, Fabisiak, Dvoskin & Holohean, 1987; Teplin, 1990). A recent U.S. Department of Justice special report (Ditton, 1999) estimated that 283,800 offenders with mental illness were

incarcerated across the country; 16% of State prison inmates, 7% of Federal inmates, and 16% of those in local jails cited either a mental condition or an overnight stay in a mental hospital during their lifetime. Offenders in this survey reported a high incidence of drug and alcohol abuse; one third were alcohol dependent, while six in ten were under the influence of alcohol or drugs at the time of offense. About 16% of probationers, or an estimated 547,800 persons, reported a mental condition or an overnight stay in a mental hospital during their lifetime (Ditton, 1999). Clinical assessment data indicate that the incidence of mental disorders within the offender population continues to grow; the Colorado Department of Corrections (DOC) reported an increase from 4% in 1991 to 14% in 2001 (Kleinsasser & Michaud, 2002). Data reported in the last year show that, today, 17% of Colorado inmates have a serious mental disorder (Stommel, 2005); about three quarters of these (or 12.75% of all offenders) also show evidence of a substance use disorder, and can be classified as having COD.

Offenders who have MICA disorders constitute a population of significant interest that presents considerable challenges. In practical terms, their special needs place exceptional demands on the system (e.g., segregated programming, multi-agency discharge planning), and their treatment needs are complex. Studies of non-offender populations have found poorer outcomes for those with COD, particularly those with serious mental disorders (the MICA subset of COD), including higher rates of HIV infection, relapse to substance use, re-hospitalization, depression and suicide risk (Drake et al., 1998; Office of the Surgeon General, 1999), as compared to those with a single disorder. In spite of widespread findings of mental disorders among drug treatment entrants (i.e., COD), and evidence of the negative impact of psychiatric impairment on drug treatment outcomes, the information on COD available to criminal justice program planners and clinical staff is often insufficient for service planning generally or for the development of an individual treatment plan.

In many cases, the instruments available to screen inmates, particularly for mental disorder, within State Departments of Corrections lack standardization and uniformity. The various screening instruments currently in use, either in prison or in community settings, typically focus separately on mental disorders or substance use disorders (alcohol and drug), and fail to examine the extent to which these disorders co-exist (Sacks & Pearson, 2003). A search of the mental health literature identified over 50 instruments to screen for the existence of mental disorders in general, and numerous other screening instruments for specific psychiatric disorders. These instruments are of varying quality and validated on diverse populations in a wide variety of settings, with psychometric properties that can vary according to the cut-off scores applied to a particular population or screening criteria. None of the identified instruments included additional items that would provide information of value to DOCs, such as potential for disruptive behavior, violence, and suicide in a prison setting. In view of these limitations, the need to validate or develop an appropriate COD screening device in the context of the prison setting is clear. The instrument developed may well incorporate items from existing instruments; it is possible that existing standardized substance abuse and mental health screening instruments will be selected in total, combined and consolidated with a small number of additional scales comprising component items that incorporate the particular concerns of criminal justice providers. These decisions will be determined through a process of literature review, stakeholder input and expert panel review.

Caveats (limitations) of Screening Instruments

The role of the screening instrument is to identify co-occurring mental and substance use disorders of sufficient severity to warrant specialized services. The screening instrument is *not* designed to be a comprehensive diagnostic interview, but only to indicate a level of substance use and mental disorder that warrants referral for further assessment

and an expanded mental status workup. It remains the function of the program intake protocol to conduct a more complete clinical evaluation, including diagnostic assessment, for purposes of treatment planning.

Research Design and Methods

Approach

The investigators identified, selected and reviewed screening instruments in the public domain that predicted the presence of substance use (alcohol or drugs) and mental disorders, and that match criteria for inclusion (listed in Table 1). The mental disorder screening devices were to be combined with those for substance disorder to create a screening instrument specifically for COD. A small number of additional scales were also examined to address specific issues of concern to the stakeholders, such as risk for violence, disruptive behavior, and suicide as well as any mental health, substance abuse or behavioral items not covered by the selected scales. A panel of experts reviewed the screening battery for feasibility and comprehensiveness. The panel suggested that the Texas Christian University Drug Screen (TCUDS; Broome, Knight, Joe, & Simpson, 1996; Simpson, 1995; Simpson, Joe, Rowan-Szal, & Greener, 1997) be accepted as the substance disorder screening instrument and that the study test only mental disorder screening instruments. This was suggested on the basis of studies already available on the validity of the TCUDS as a substance disorder screening instrument for offender populations. The panel approved three mental disorder instruments to include in the screening battery, The Global Appraisal of Individual Needs (GSS, version 1.0; Dennis, Chan, & Funk, 2006), the Mental Health Screening Form (MHSF; Carroll & McGinley, 2001), and the Modified MINI Screen (MMS; Sheehan, Lecrubier, Sheehan et al., 1998).

The investigators tested the screening battery, made up of the GSS, MHSF, and MMS, against the Structured Clinical Interview for DSM-IV (SCID; First, Spitzer, Gibbon, et al., 2002) in a sample of 280 inmates participating in prison-based substance abuse treatment programs at four CJDATS research centers.

Step 1— Instrument Selection and Modification

A literature search was conducted to identify the best available mental health and substance abuse screening instruments consistent with the initial criteria - good psychometric properties, in the public domain, requiring no more than 20 minutes to complete, and suitable for administration by the subject or by correctional/substance treatment staffs who do not have specific mental health training.

Selection Criteria. In order to fit selection criteria, instruments needed to sample items measuring Axis I and II disorders relevant to criminal justice populations, including psychotic, mood, anxiety, and Post-Traumatic Stress Disorder (PTSD), on Axis I and Axis II personality disorders consistent with a COD classification. The investigators tried to locate instruments with high (.80 or higher) test-retest reliability, and with known validity for diagnosing substance use and mental disorders when compared to other instruments of known validity, such as the *Structured Clinical Interview for DSM-IV* (SCID). Instruments were considered in agreement on a diagnosis when both instruments either identified or did not identify a disorder. The search for instruments was separated into two broad categories, *Alcohol and Drug Use Disorders* and *Mental Disorders*.

Alcohol and Drug Use Disorders. Alcohol and drug use disorders screening instruments were previously evaluated by Peters et al. (2000) for use in prison settings. Seven instruments were selected on the basis of common usage and good psychometric support,

and evaluated for their psychometric properties in prison settings; the *Alcohol Dependence Scale* (ADS); the *Addiction Severity Index-Drug Use* (ASI-D); the *Drug Abuse Screening Instrument* (DAST-20); the *Michigan Alcoholism Screening Test* short version (SMAST); the *Substance Abuse Subtle Screening Inventory-2* (SASSI-2); the *Simple Screening Instrument* (SSI); and the TCUDS. Section E–Substance Use Disorders module- of the SCID, provided the criteria for alcohol and drug abuse problems. Two instruments, the ADS for alcohol disorders and the TCUDS for drug use disorders, best met the stated criteria; namely, overall accuracy (a combination of sensitivity and specificity), brevity, can be found in the public domain, and capable of administration by subject or by non- [mental health] professional staff. These instruments were proposed for use in study.

Mental Disorder Screening Instruments. In addition to a general review of the relevant literature, the search included four principal sources. First were two articles, one (Peters & Bartoli, 1997) that reviews screening and assessment in criminal justice, and that included mental health instruments such as the *Beck Depression Inventory* (BDI) the *Brief Symptom Inventory* (BSI), and/or the *Symptom Checklist 90 – Revised* (SCL-90-R), and one (Peters & Hills, 1993) that recommended the *Referral Decision Scale* (RDS). The second primary source was the TIP *Substance Abuse Treatment For Persons With Co-Occurring Disorders* (CSAT, Sacks & Reis, 2005), which provided an extended listing of 16 recommended instruments, including the *Global Appraisal of Individual Needs* (GAIN), and the *Mini-International Neuropsychiatric Interview* (MINI). Third were the scales and items from the *National Co-Morbidity Survey* (Kessler et al., 1994; 2003). The fourth source comprised other suggested instruments being used in community based substance abuse treatment programs, such as the *Psychiatric Diagnostic Screening Questionnaire* (Zimmerman et al., 2004) and the *MHSF*.

Instruments reported in the research literature that appeared relevant were examined further if they met appropriate psychometric standards, were in the public domain, or could be used with permission of the author(s), including a willingness to make the instrument, once it is found to have good psychometrics and to be clinically useful, available to service providers. Table 1 shows the initial breakdown of how instruments were evaluated; stakeholder criteria were added to the table following completion of Step 2.

Table 1

Criteria	Instrument 1	Instrument 2	Instrument 3
Purpose			
Screening			
Assessment			
Clinical utility: Diagnosis			
Clinical utility: Placement			
Clinical utility: Treatment Planning			
Clinical Utility: Outcome			
Severity Measure			
Reporting			

Methodological Considerations			
Diagnostic categories included			
Criminal justice norms available			
Reliability on re-administration			
Internal consistency (with alpha)			
Validity			
Sensitivity			
Selectivity			
Overall accuracy			
Effects of demographics or background on validity			
Setting Applicability			
Cut-off score specified			
Administration			
Tech. support available and free			
Tech. support available at minimal cost			
Computer admin.			
Computer scoring			
Interpretive			
Time taken for admin.			
Complexity of scoring			
Required skill level			
Intensity of required training			
Self-administered			
Reading level (for self-administered)			
Public domain/cost			

Step 2— Development of Criteria for Selecting a Mental Health Screening Instrument— Stakeholder Input

Composition. The stakeholders group consisted of 12 members representing the research community, the criminal justice system (e.g., director of DOC, wardens, substance abuse treatment program directors), and treatment providers. To broaden the relevance of the final instrument to systems issues, stakeholders were drawn from both the prison and the community, including community mental health and substance abuse treatment providers, along with criminal justice professionals.

Choosing an instrument. The panel was tasked with choosing instruments for a COD screening test battery; 1) a substance use disorder screen (or one each for alcohol and drug abuse), 2) a mental disorder screen, and 3) a small number of additional scales with a limited number of items that responded to specific stakeholder needs. The investigators provided the panel with the criteria from Table 1 for several instruments that fit the initial criteria (brief, valid, in the public domain, and no special training for administration); stakeholders were asked to respond to these criteria and to provide feedback. The process elicited the types of COD-related information of most value to the stakeholders in a screening instrument, procedures or instruments currently in use, the time available for administration, the type of personnel and maximum training time available, the ease of use and scoring, acceptability of questions to respondents (items that are meaningful to prison inmates and transitional offenders, and do not provoke non-compliance), type of profile generated, and utility estimates (i.e. the trade-offs between the length of the instrument,

the desired accuracy or confidence limits to detect COD, and the cost of a full diagnostic workup. Stakeholders conjointly decided on a screening instrument for alcohol abuse, drug abuse, and mental disorders that best meets stakeholder needs and psychometric standards.

Step 3— Panel of Expert Review and Instrument Revision

A modified Delphi method (Linstone & Turoff, 1975) consisting of three rounds of panel feedback was used to query the panel. During the first round, respondents indicated the importance of items and provided a rationale for items to be dropped, modified, or added. Panel members responded to the comments of other panelists in the second round, and gave their final ratings in the third round. Differences in rating statements were resolved through discussion. The *content validity* or face validity of the additional scales and the feasibility of administering them in prison and transitional settings were evaluated by the panel. The panel also reviewed items on the main scales for substance use and mental health disorders, and suggested the following study revisions:

- (1) They accepted the TCUDS as the substance abuse screening component for a COD Screen;
- (2) Since mental disorder screening instruments, at that point, had not been widely validated in the criminal justice system, they suggested testing three different mental disorder screening instruments to determine the top instrument that could eventually be combined with the TCUDS to form a true COD Screen. The mental disorder screening instruments chosen by the panel were the GSS, MHSF, and the MMS;
- (3) The panel suggested adding specific questions to address BPD and ASPD. The panel left the number and source of these questions up to the investigators to determine.

Step 4— Field Test

Screening Battery. All subjects received the entire battery of screening tests, which consisted of the GSS, MHSF, and the MMS. The order in which the instruments were presented was rotated to control for ordering effects.

Demographic and Background Variables. All subjects were given the CJ-DATS *Intake Interview* to collect demographic and background information.

Validity Criterion Measures. The validity of the instruments was tested relative to the SCID, generally considered to be an excellent measure of mental disorders. The criterion measures for additional issues of concern to stakeholders, such as violence, suicide and disruptive behavior consisted of incident reports in prisoner records.

Criterion Score. Three main criteria were considered in evaluating the instruments. *Specificity* refers to the proportion of individuals who do not have the condition (do not suffer from COD) that is correctly identified. *Sensitivity* refers to the proportion of individuals who have the condition (COD) that is correctly identified. Overall *accuracy* consists of the total percent of people correctly classified and provides a third possible criterion. All of the criteria are scientifically acceptable, and each serves somewhat different purposes.

Cut-off Scores. Cut-off scores were used to classify individuals as having a potential mental disorder. Based on a review of the literature, it is anticipated that at least 50% of the sample will have a co-occurring mental disorder (Sacks, Sacks, De Leon, et al., 1997). Two sets of cut-off scores were investigated; those suggested by previous research with the standardized instruments (if selected and cut off scores were available), and those that were generated empirically to maximize sensitivity and positive predictive values in the sample (optimal cut-off scores).

Pilot test. A pre-test was conducted on a sample of 100 consecutive admissions to prison-based substance abuse treatment programs. The pilot test determined the effectiveness of the procedures for administering the instrument, and the reception subjects gave to the instrument. The pilot test was mainly used to determine cut-off scores so that those scores could be validated with the larger sample.

Sample. The final instrument was administered to a sample of 280 new admissions to prison-based substance abuse treatment programs. The pilot test used 100 of those cases to determine cut-off scores and the remaining 180 cases were used to validate those cut-off scores. All testing was completed within the first two weeks subsequent to entry into the program. The sample size of 280, generally considered adequate for a validity study of this type (Nunnally & Bernstein, 1994), was distributed across the four research centers. *NDRI Rocky Mountain* in Colorado (N=117), *Lifespan* at Brown University in Rhode Island (N=75), the *Institute for Behavioral Research* at Texas Christian University in Texas (N=60), and the *Integrated Substance Abuse Programs* at UCLA in California (N=28). This allocation represented diverse institutions and states, shared the burden of data collection equitably among the collaborating centers, and ensured inclusion of women and minorities. Identifying those in the prison treatment program who have a co-occurring mental disorder will have the clinically significant consequence of allowing such offenders to receive the special services necessary for the effective treatment of COD.

The study drew its subjects from prison substance abuse treatment programs, and was restricted to new admissions to prevent any influence from the program itself. The rate of occurrence of COD was a factor favoring substance abuse treatment programs as the source of study participants. Evidence suggests that a significant number of the inmates in substance abuse treatment programs have a co-existing mental disorder. For example, data from the *Diagnostic Interview Schedule* (DIS; Horton, Compton & Cottler, 1999; Robins, Helzer, Cottler & Goldring, 1989) employed in the Amity study (Prendergast, Hall, Wexler, Melnick & Cao 2004) diagnosed an Axis I disorder among 36% of the inmates who participated; prescreening inmates for mental disorders may be the reason that relatively few (2.1%) were diagnosed with psychotic disorders. If the general prison population were to be used, a much larger sample would have been needed to accumulate a statistically sufficient number of positive identifications to evaluate the utility of the CODSI.

Restricting the sample to those who are entering substance abuse treatment optimized the heterogeneity of the sample; some reduction in heterogeneity likely occurred, given the potential for the most obviously disturbed individuals to be screened out at entry into prison system. This truncation at the high end of the mental health scale is likely to reduce the predictive validity "correlation" to some extent. Finally, the ability to identify COD offenders among a substance-abusing population is a very useful step leading towards a COD screening tool for the general population, the assumption being that an instrument with proven validity in the more homogeneous population will continue to have validity in the more heterogeneous population.

Test-retest sub-sample. Test/re-test reliability was calculated from a randomly selected sample of 60 subjects tested one week apart.

Compensation to subjects. Participants in the NDRI Rocky Mountain, Rhode Island and UCLA samples were paid the equivalent of \$5 for their participation in both the screening instrument and SCID interviews. Subjects participating in the test-retest received an additional payment of \$5. Where payments to prisoners were prohibited, the offender received a gift equivalent.

Administration and Training. Trained research assistants administered the instruments. Administration began with the interviewer describing the study and the participant's role. Those who agreed to participate were then guided through the consent process, which begins with the interviewer reading the study information sheet aloud to the subject. The interviewer ensured that the subject fully understood his/her role in the study and the details of the consent form. The SCID was administered separately by a specially

trained professional within two weeks of administration of the screening instruments. Each investigative team identified a professional with extensive prior training and experience with the SCID and related diagnostic instruments.

Statistical Analysis of Field-tested Instrument

Scale Reliability. The internal consistency of factor-based scales was assessed by Cronbach's alpha (a coefficient of internal consistency that is widely recognized and used to establish reliability) and factor-analytic based reliability indices (McDonald 1999; Raykov, 1997). This is consistent with procedures for determining the construct validity of an instrument. The internal consistency reliability, measured by Cronbach's alpha, tests whether multiple items designed to measure the same construct intercorrelate with one another. Alpha can also be computed separately for women and men to determine whether the scales have sufficient internal consistency in both of these groups. To supplement coefficient alpha, factor-analytic based reliability coefficients will be considered (McDonald, 1999, Raykov, 1997). These coefficients do not assume (as coefficient alpha does) that the different items making up a scale are tau-equivalent (equal factor loadings); they only assume that the items are congeneric. Some approaches do not even require that the items are congeneric (Raykov & Shrout, 2002). In addition to relaxing some of the assumptions of coefficient alpha, in the case of congeneric measures, these structural equation modeling approaches yield standard errors of reliability coefficients and allow one to form confidence intervals around the point estimate (Raykov, 1998). Test-retest reliability was also tested by calculating the correlations between the tests on a subset of 60 patients.

Exploring Measurement Invariance. Although the sample was not stratified for race/ethnicity, the collection of additional minority cases through the CODSI-MS resulted in a large percentage of minority patients and provided a sufficient number of subjects for additional analysis of African American and Latino offenders. Separate analyses described above were conducted to determine differences in instrument performance among women and race/ethnic groups. To assess more formally the invariance of the factor structure by gender and ethnicity, multigroup confirmatory factor analysis will be considered, in which equality constraints on various parameters across groups are evaluated for consistency with the data (Floyd & Widaman, 1995; Reise et al., 1993; Rensvold & Cheung, 1998; Steenkamp & Baumgartner, 1998). Confirmatory factor analysis was used in the fitting of a second-order factor model (Hoyle, 2000) to determine if a single higher-order factor may account for correlations among the scale scores of the CJ-CODSI. Such an analysis would provide some insight into the feasibility and utility of creating a total composite from CJ-CODSI items that cut across the expected separate facets.

Scale Cut-offs. Two types of cut-off scores were evaluated, cut-off criteria suggested in the literature or in the scale development procedures detailed above, and an additional cut-off score based on *Receiver Operating Characteristics* (ROC) in the current data set. ROC curves (Campbell, 1994; Linnet, 1989; Swets, 1986; Swets, 1988) were used to determine optimal cut-off scores for each scale, and to describe each scale's diagnostic accuracy (Pepe, 2000) with respect to screening the potential for mental disorders. The area under a ROC curve summarizes how well a given cut-off performs in terms of maximizing both the proportion of true negatives (specificity) and the proportion of true positives (sensitivity). False negatives are viewed as more costly errors than false positives; on balance, it would be better to intervene unnecessarily with some clients than to fail to intervene with clients who will not adhere without intervention. Hastie et al. (2001) describe how to weight the importance of each kind of error. Because we are interested in maximizing sensitivity, possibly at the expense of reducing specificity, we will weight false negative errors more heavily than false positive errors, setting criteria for false positive at no more than 40% and false negatives at 90%.

Concurrent Validity. This was examined with bivariate and multivariate statistics summarizing the association between each CODSI scale scores and the criterion measures (e.g., SCID and other stakeholder specified criteria). Correlations were used to determine the extent to which each of the scales and the total CODSI score predict the criterion measures. Because there will be multiple criterion measures, canonical correlation analysis (Thompson, 1991) were considered to summarize the relations between these two domains. ANOVA was used to determine the main effects (e.g. presence of mental illness, propensity for violence and disruption, etc.). Gender and other independent variables (such as education, employment history, crime, primary drug, etc.) that correlate significantly with the SCID and other criterion variables (such as disruptive behavior, etc.) were used as covariates in multiple regression analyses to determine the unique relation between the screening scales and the criterion measures, controlling for these other variables. Logistic regressions tested the effects of the screening instrument scales and other variables against a dichotomous (Yes/No) cut-off score, 90% or greater agreement with the SCID and the other dependent variables. Thus, in addition to the examination of different groups, the analyses used the independent variables from the CJ-DATS *Intake Interview* (e.g., demography, previous crime, substance abuse treatment history) as covariates to partial out the effects of these variables on the relationship between the screening battery and the criterion measures.

Sample Size and Power. MacCallum, Widaman, Zhang, et al., (1999) discuss the issue of sample size in factor analysis in some detail. The proposed sample size ($n=280$) exceeds a variety of minimum sample size guidelines (Cattell, 1978; Gorsuch, 1983; Guilford, 1954; Kline, 1979), and yields an $N:p$ ratio of 10 for up to 30 items, which Gorsuch (1997) suggests may be a larger ratio than the minimum needed. Gorsuch also argues that $n=280$ is an appropriate sample size for item factor analysis even when some factor loadings are expected to be low (.40). MacCallum, Widaman, Zhang, et al. (1999) go beyond these rules of thumb and identify other aspects on which needed sample size depends. Two aspects they highlight are communality and over-determination of factors (factors defined by multiple items). When communalities are high (0.60) and factors are over-determined, sample sizes as low as $n=100$ may be sufficient. Given the narrowness of the constructs of interest, communalities are expected to be relatively high in the proposed research. Even with low communalities and just 3 or 4 indicators per factor, MacCallum et al. (1999) suggest a sample size of $n=300$ is sufficient. On balance, the proposed sample size seems to be adequate for the proposed exploratory factor analyses.

Confirmatory factor analysis draws on the work of MacCallum and colleagues (MacCallum, Browne, Sugawara, 1996) to estimate power given the proposed sample size ($n=280$). With alpha of .05 for a test of close fit, models need to have at least 32 degrees of freedom to maintain .80 power or better. A model with 20 CODSI items could have as many as 178 freely estimated parameters without dipping below 32 degrees of freedom. If the overall sample is split to examine gender or ethnic differences, CFA models would need about 82 degrees of freedom to maintain .80 power in tests of close fit. With 20 CODSI items, up to 128 parameters could be freely estimated. For both the overall CFA and for subgroup analyses, we do not expect to estimate the maximum number free parameters allowable to maintain .80 power, so the proposed sample size provides adequate power to detect lack of fit in CFA models.

Logistic regression is one of the primary ways in which concurrent validity will be assessed, determining the relation between criterion measures (e.g., mental health status, risk for violent or disruptive behavior, suicide, etc.) and CODSI scale scores. A logistic regression of a binary response variable (Yes/No on cut-off score) on a continuous, normally distributed variable (CODSI scale) with a sample size of 300 observations achieves 80% power at a 0.05 significance level to detect a change in $\text{Prob}(Y=1)$ from the value of 0.50 at the mean of the CODSI scale to 0.58 when the CODSI scale score is increased to one

standard deviation above the mean. This change corresponds to an odds ratio of 1.382, which is a small effect size (Lipsey & Wilson, 2001).

Subgroup Analyses. For each of the quantitative approaches to scale refinement, differences by key subgroups defined by gender, ethnicity, sexual orientation, and drug use were explored. In some cases, such as gender, a more formal approach to examining differences was taken, using CFA as described above, since subgroup sizes were sufficiently large for these formal tests. For other subgroups that were too small to support these formal tests, an informal approach was taken, simply applying the proposed analyses separately in each subgroup of interest. Any sizeable differences in CJ-CODSI scale reliability, factor structure, and predictive validity were noted and marked for further scrutiny in subsequent research.

References

- Broome, K. M., Knight, K., Joe, G. W., & Simpson, D. D. (1996). Evaluating the drug-abusing probationer: Clinical interview versus self-administered assessment. *Criminal Justice & Behavior*, 23(4), 593-606.
- Campbell, G. (1994). Advances in statistical methodology for the evaluation of diagnostic and laboratory tests. *Statistics in Medicine*, 13, 499-508.
- Carroll, J. F. X. & McGinley, J. J. (2001). A screening form for identifying mental health problems in alcohol/other drug dependent persons. *Alcoholism Treatment Quarterly*, 19(4), 33-47.
- Cattell, R.B. (1978). *The scientific use of factor analysis*. New York: Plenum.
- Center for Substance Abuse Treatment. (2005). *Substance Abuse Treatment for Persons with Co-Occurring Disorders. Treatment Improvement Protocol (TIP) Series, Number 42*. S. Sacks, Chair & R. Reis, Co-Chair, Consensus Panel. DHHS Pub. No. (SMA) 05-3992. Rockville, MD: Substance Abuse and Mental Health Services Administration.
- Dennis, M. L., Chan, Y., & Funk, R. R. (2006). Development and validation of the GAIN Short Screener (GSS) for internalizing, externalizing and substance use disorders and crime/violence problems among adolescents and adults. *The American Journal on Addictions*, 15(S1), 80-91.
- Ditton, P. M. (1999) *Mental health and treatment of inmates and probationers*. Washington, DC: Bureau of Justice Statistics [Web Page]. URL <http://www.ojp.usdoj.gov/bjs/pub/ascii/mhtip.txt>.
- Drake, R.E., McHugo, G.J., Clark, R.E., Teague, G.B., Xie, H., Miles, K. & Ackerson, T.H. (1998) Assertive community treatment for patients with co-occurring severe mental illness and substance use disorder: A clinical trial. *American Journal of Orthopsychiatry* 68(2), 201-215.
- First, M.B., Gibbon, M., Spitzer, R.L. & Williams, J.B.W. (1996) User's Guide for the Structured Clinical Interview for DSM-IV Axis I Disorders - Research Version (SCID-I, Version 2.0, February 1996 FINAL Version) New York: Biometrics Research Department, New York State Psychiatric Institute, 722 W 168th Street, New York 10032.
- Floyd, F.J., & Widaman, K.F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7, 286-299.

- Gorsuch, R.L. (1983). *Factor analysis* (2nd Edition). Hillsdale, NJ: Lawrence Erlbaum.
- Gorsuch, R.L. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment*, 68, 532-560.
- Guilford, J.P. (1954). *Psychometric methods* (2nd Ed.). New York: McGraw-Hill.
- Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer.
- Horton, J., Compton, W.M. & Cottler, L.B. (1999) *Assessing Psychiatric Disorders Among Drug Users: Reliability of the Revised DIS-IV*. In Harris, L. (Ed.) NIDA Research Monograph – Problems of Drug Dependence 1998. Washington, DC.
- Hoyle, R. H. (2000). Confirmatory factor analysis. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 466–497). San Diego, CA: Academic Press.
- Kessler, R.C., Berglund, P., Demler, O., Jun, R., Merikangas, K.R., Rush, A.J., Walters, E.E. & Wang, P.S. (2003) The epidemiology of major depressive disorder: Results from the National Comorbidity Survey Replication (NCS-R). *JAMA: Journal of American Medical Association*, 289(23), 3095-3105.
- Kessler, R.C., McGonagle, K.A., Zhao, S., Nelson, C.B., Hughes, M., Eshleman, S., Wittchen, H.U., & Kendler, K.S. (1994). Lifetime and 12-month prevalence of DSM III-R psychiatric disorders in the United States: Results from the National Comorbidity Study. *Archives of General Psychiatry*, 51, 8-19.
- Kleinsasser L. D., & Michaud, J. T. (2002). Identifying and tracking seriously mentally ill offenders in the Colorado Department of Corrections. *Presentation to the Mental Health in Corrections Consortium*.
- Kline, P. (1979). *Psychometrics and psychology*. London: Academic Press.
- Linnet, K. (1989). Assessing diagnostic tests by a strictly proper scoring rule. *Statistics in Medicine*, 8, 609-618.
- Linstone, H.A. & Turoff, M. (1975). *The Delphi Method: Techniques and Application*. Upper Saddle River, NJ: Addison-Wesley.
- Lipsey, M.W., & Wilson, D.B. (2001). Practical meta-analysis. *Applied Social Research Methods Series* (Vol. 49). Thousand Oaks, CA: SAGE Publications.
- MacCallum, R.C., Browne, M.W., & Sugawara, H.M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1, 130-149.
- MacCallum, R.C., Widaman, K.F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4, 84-99.
- McDonald, R.P. (1999). *Test theory: A unified treatment*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

Office of the Surgeon General, (1999). Report on Mental Health. Publication #017-024-01653-5, Superintendent of Documents, Washington, DC.

Pepe, M.S. (2000). Receiver operating characteristic methodology. *Journal of the American Statistical Association*, 95, 308-311.

Peters, R.H. & Hills, H.A. (1993) Inmates with co-occurring substance abuse and mental health disorders. In Steadman, H.J. & Cocozza, J.J. (Eds.), *Mental Illness in America's Prisons*, pp. 159-212. Seattle, WA: National Coalition for the Mentally Ill in the Criminal Justice System.

Peters R. & Bartoli, M. (1997). *Screening and assessment of co-occurring disorders in the justice system*. Albany, NY: National GAINS Center.

Peters, R. H., & Hills, H. A. (1997). *Intervention Strategies for Offenders with Co-Occurring Disorders: What Works?* Delmar, NY: National GAINS Center.

Peters, R.H., Greenbaum, P.E., Steinberg, M.L., and Carter, C.R. (2000). Effectiveness of screening instruments in detecting substance use disorders among prisoners. *Journal of Substance Abuse Treatment* 18(4):349–358, 2000.

Prendergast, M.L., Hall, E.A., & Wexler, H.K., Melnick, G., & Cao, Y. (2004). Amity prison-based therapeutic community: Five-year outcomes. *The Prison Journal*, 84(1), 36-60.

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, 21, 173-184.

Raykov, T. (1998). A method of obtaining standard errors and confidence intervals of composite reliability for congeneric measures. *Applied Psychological Measurement*, 22, 369-374.

Raykov, T., & Shrout, P.E. (2002). Reliability of scales with general structure: Point and interval estimation using a structural equation modeling approach. *Structural Equation Modeling*, 9, 195-212.

Regier, D.A., Farmer, M.E., Rae, D.A., Locke, B.Z., Keith, S.J., Judd, L.L. & Goodwin, F.K. (1990) Co-morbidity of mental disorders with alcohol and other drug abuse: Results from the Epidemiologic Catchment Area (ECA) study. *Journal of the American Medical Association*, 264 (19), 2511-2518.

Reise, S.P., Widaman, K.F., & Pugh, R.H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, 114, 552-566.

Rensvold, R.B. & Cheung, G.W. (1998). Testing measurement models for factorial invariance: A systematic approach. *Educational and Psychological Measurement*, 58, 1017-1034.

- Robins, L., Helzer, J., Cottler, L. & Goldring, M.A.T. (1989) *NIMH Diagnostic Interview Schedule, Version III, Revised*. Rockville, MD: US Department of Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration.
- Sacks, S., Sacks, J. Y., De Leon, G., Bernhardt, A. I., & Staines. G. L. (1997). Modified therapeutic community for mentally ill chemical abusers: Background; influences: Program description: Preliminary findings. *Substance Use and Misuse*, 32(9), 1217-1259.
- Sacks, S. & Pearson, F. (2003) Co-Occurring substance use and mental disorders in offenders. *Federal Probation*, 67(2), 32-39.
- Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., and Dunbar, G.C. (1998) The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry* 59(Suppl 20):22–33.
- Simpson, D. D. (1995). Toward the year 2000: Issues in treatment process and services research. *International Journal of the Addictions*, 30(7), 875-879.
- Simpson, D. D., Joe, G. W., Rowan-Szal, G. A., & Greener, J. M. (1997). Drug abuse treatment process components that improve retention. *Journal of Substance Abuse Treatment*, 14(6), 565-572.
- Steadman, H.J., Fabisiak, S., Dvoskin, J., and Holohean, E.J., Jr. A survey of mental disability among state prison inmates. *Hospital and Community Psychiatry* 38(10):1086–1090, 1987.
- Steenkamp, J.E.M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25, 78-90.
- Stein, M.D., Rich, J.D., Maksad, J., Chen, M.H., Hu, P., Sobota, M., Clark, J. (2000). Adherence to antiretroviral therapy among HIV-infected methadone patients: Effect of ongoing illicit drug use. *American Journal of Alcohol Abuse*, 26(2), 195-205.
- Stommel, J. (2005). Personal communication, March 16, 2005.
- Swets, J.A. (1986). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin*, 99, 100-117.
- Swets, J.A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285-1293.
- Teplin, L. (1990). Detecting disorder: The treatment of mental illness among jail detainees. *Journal of Consulting & Clinical Psychology*, 58, 233-236.
- Thompson, B. (1991). A primer on the logic and use of canonical correlation analysis. *Measurement & Evaluation in Counseling & Development*, 24, 80-95.
- Zimmerman, M., Sheeran, T., Chelminski, I., & Young, D. (2004) Screening for psychiatric disorders in outpatients with DSM-IV substance use disorders. *Journal of Substance Abuse Treatment*, 26(3):181-188.

