Minnesota Sex Offender Screening Tool–Revised (MnSOST-R) Technical Paper: Development, Validation, and Recommended Risk Level Cut Scores

Douglas L. Epperson, Ph.D.
Iowa State University

James D. Kaul, Ph.D.
Minnesota Department of Corrections

Stephen Huot, M. Eq.
Minnesota Department of Human Services

Robin Goldman, M.A.
Minnesota Department of Corrections

Will Alexander, M.S.
Minnesota Department of Corrections

December 2003

Abstract

This paper describes the development, reliability, and validity of the Minnesota Sex Offender Screening Tool – Revised (MnSOST-R), as well as recommended risk levels and cut scores. Variables from multiple dimensions, both static and dynamic, were reviewed for inclusion in the MnSOST-R. Final items were selected and scored empirically based on clearly defined criteria. The resulting 16 items that comprise the MnSOST-R maximize the positive predictive power of the tool, and perform significantly better than previous versions of the MnSOST. This newest version correlate achieves impressive hit rates with rapists and extrafamilial sex offenders, the population for which the instrument was developed. Very high true positive rates were achieved depending on the selected cut score.

Like other researchers, we found the pattern for intrafamilial sex offenders on potential items and the relationship of that pattern to sexual recidivism to be dramatically different from that for other sex offenders, with the exception of intrafamilial sex offenders who essentially raped their related victim(s).

The MnSOST-R is currently being used by the Minnesota Department of Corrections as a screening tool for referral for commitment under the state's Sexual Psychopathic Personality and Sexually Dangerous Person laws, and as part of the state's Community Notification Act.

Minnesota Sex Offender Screening Tool–Revised (MnSOST-R):

Development, Validation, and Recommended Risk Level Cut Scores

Determinations of risk occur throughout the criminal justice system, despite decades of concerns by legal scholars and social scientists about the accuracy of the process. From decisions to waive a juvenile to adult court, from bail to sentencing, and from custody level to conditions of release, decisions about risk level must be made. For example, Witt, DelRusso, Oppenheim, and Ferguson (1996) reported that the Bail Reform Act of 1984 permits the federal courts to deny an individual bail based solely upon the accused's potential for causing future harm, and that the U.S. Supreme Court has rejected challenges to this act based on problems in quantifying future dangerousness. These authors also noted that the federal courts have upheld the Federal Special Offender Statute, observing that the courts found that the "... likelihood of future criminality and the potential danger to society are determinations implicit in sentencing decisions". Finally, they reported that the claim that the prediction of risk of future behavior is unconstitutionally vague has been rejected by every court of appeals that has considered the question. Similarly, Janus and Meehl (1997) concluded that while there are statutory and evidentiary standards limiting prediction testimony:

> "...it seems well established that there is no constitutional impediment to using predictions of dangerousness in legal proceedings up to and including those that result in a loss of liberty or death. As a legal matter, prediction is not, in all of its forms and for all purposes, so inaccurate as to violate the due process clause." (p. 36)

Sex offenders, however, present some unique challenges in the legal system. In addition to the risk assessment that occurs at every stage of the criminal justice system with every offender, there has been a recent proliferation of sex offender registration and community

notification laws, and a number of states have recently enacted sex offender commitment statutes (Schulhofer, 1996). The policy justification for both sex offender community notification and commitment statutes is the protection of the public from the "most dangerous" sex offenders (Witt, et. al., 1996; Janus, 1996).

Even though some courts have struck down sex offender commitment laws as unconstitutional, there has been a resurgence of commitments as a number of states have enacted new legislation or increased the use of existing commitment statutes (Alexander, 1995; Janus and Meehl, 1997; Schulhofer , 1996). One of the constitutionally required elements for such commitments is the prediction of dangerousness, and the probability of future acts of sexual violence greatly influences how the risk of error is to be distributed between offenders and the public (Janus and Meehl, 1997). If the probability of dangerousness is over-predicted (false positives), many offenders are unnecessarily deprived of their liberty and placed in treatment that is both expensive and prolonged. If the probability of dangerousness is predicted too conservatively (false negatives), dangerous sex offenders are released without appropriate supervision and may commit new sexual offenses.

Although a number of states already had sex offender registration and community notification laws, federal legislation in 1996 ("Megan's Law") required the fifty states to implement registration and notification laws within two years or face the loss of federal funds. The bill, which took two days to pass through Congress, provided no guidelines as to how community notification was to work (Rudin, 1996). The trend to date has been for notification statutes to incorporate a three-tier system of determining the level of risk presented by the offender (low, medium or high risk) rather than to identify the specific types of offenders that are to be subject to

notification (Finn, 1997). In Minnesota, for example, the Community Notification Act of 1996 called for a risk assessment scale to be implemented by January 1, 1997, and identified a series of factors to be used in assessing risk and assigning risk levels. As in other states, those sex offenders determined to present the greatest risk receive the highest level of community notification.

While the accurate assessment of a sex offender's risk of recidivism has become increasingly important to many areas of the criminal justice system, there has been little effort until recently to develop an empirical base for guiding the assessment of dangerousness in this population. Despite many years of hand wringing by mental health experts, the practice of sex offender assessments has, by necessity, continued to be largely clinical or subjective in nature until relatively recently.

Actuarial Prediction of Sex Offense Recidivism

Meehl's general thesis over 40 years of writing, confirmed in many different contexts, is that clinical/subjective judgement is at best as good as, but often worse than, actuarial methods (Grove and Meehl, 1996; Janus and Meehl, 1997). Indeed, a large body of literature over the last several decades has consistently demonstrated the general superiority of actuarial prediction over clinical prediction in virtually every decision-making situation for which the issue has been studied (Harris, Rice, and Quinsey, 1993; Jones, 1998). In the particular case of predicting violence, it is well documented that mental health professionals possess no special expertise in the prediction of violence, and that reliance on clinical judgements alone results in numerous inaccurate predictions of violent recidivism (Rice and Harris, 1995). Nonetheless, as reported by Monahan (1996), the courts have upheld the constitutionality of laws that relied on clinical

prediction, even when confronted with empirical evidence of the low accuracy of such predictions. The focus has begun to shift, however, from whether violence could be predicted to how violence prediction could be improved. According to Monahan (1996), the solution to improved violence prediction is the same as for the improvement of clinical predictions in general -- the use of actuarial methods.

There is now agreement that it is possible to predict general criminal recidivism with moderate accuracy by using objective risk scales that not only specify the risk factors to be considered, but also assign relative weights to each of these factors (Hanson, 1997, 1998; Jones, 1998; Quinsey, Harris, Rice, & Cormier). However, while these statistical prediction instruments have been reasonably successful in the prediction of general criminal recidivism, such instruments have generally been less successful at predicting violent recidivism (Harris, Rice, and Quinsey, 1993). One exception to this has been the Hare Psychopathy Checklist - Revised (Hare, 1996). While not designed to predict recidivism, it has proven to be a robust predictor of violence and violent recidivism, and it correlates highly with other actuarial risk scales.

Another major advance in actuarial risk assessment has developed out of a series of studies of violence and mentally disordered offenders in Canada (Monahan, 1996). Specifically, the Violence Risk Appraisal Guide (VRAG), an actuarial tool for predicting violent criminal recidivism (Harris, et. al, 1993), emerged from this researdh. The VRAG consists of twelve items empirically scored using Nuffield's (1982) method. The items include the Psychopathy Checklist Score (PCL-R); separation from parents before age 16; victim injury for index offense; DSM-III criteria for schizophrenia; never married; elementary school maladjustment; female victim - index offense; failure on prior conditional release; property offense history, age at index offense; alcohol

abuse history; and DSM-III criteria for personality disorder. The total score of these items correlated .44 with violent recidivism, and the PCL-R score proved to be the single best predictor. While the VRAG was developed to assess violent recidivism in mentally disordered offenders, subsequent research indicated that the instrument may also predict sexual recidivism, but to a lesser degree (Hanson, 1997, 1998).

Hanson (1997) reported that, although such risk scales work quite well at predicting general recidivism and non-sexual violent recidivism among sex offenders, the scales are not as effective in predicting sexual recidivism. Indeed, Hanson and Bussiere's (1996) meta-analysis suggested that sexual recidivism could probably best be predicted by a different set of factors than those predicting general or non-sexual recidivism. Using an initial pool of items selected from this meta-analysis, Hanson (1997) developed the Rapid Risk Assessment of Sex Offender Recidivism (RRASOR), a brief risk scale designed to be used to screen sex offenders into relative risk levels. The four items on the RRASOR are: number of prior sexual offenses, age at release, any victims who were strangers, and any male victims. On average, the RRASOR correlated .27 with sexual recidivism, which was significantly higher than the average found by Hanson and Bussiere (1996) for clinical assessments ("an unimpressive r = .10"; p. 19).

The Minnesota Sex Offender Screening Tool

Development of the MnSOST began in response to a 1991 Minnesota Department of Corrections special report calling for a more formal and uniform process to identify predatory and violent sex offenders. Few prediction models existed at that time, and many of those were dependent on clinical and/or phallometric data. A major limitation of such models is the fact that many sex offenders are not offered or refuse treatment and/or phallometric assessment, making

clinical and phallometric data unavailable. From the outset, then, our goal was to create a reliable and valid predictive instrument that could be easily scored by correctional case managers using only information routinely available to them from correctional records.

Aware of the difficulties in predicting dangerousness and recidivism in general, and with sex offenders in particular, the MnSOST Task Force launched an extensive review of the general literature on sex offenders (e.g., Abel, Mittleman, Becker, Rathner & Rouleau, 1988; Marshall, Jones, Ward, Johnson & Barbaree, 1991;Working Group, Sex Offender Treatment Review, 1990), probation and parole risk assessment instruments (e.g., National Council on Crime and Delinquency, 1990; Hare, 1991), and previous risk prediction inventories designed specifically for sex offenders (e.g., Bemus, 1988; Doke, 1989). Based on this review, the Task Force developed an initial inventory of 14 items and assigned a priori weights to these items based on clinical judgment.

Results of two preliminary studies provided evidence that the MnSOST had acceptable reliability and modest predictive validity. These two studies also resulted in changes to some of the items and the addition of several others, yielding the 21-item MnSOST-Research Edition. The reliability and predictive validity of this instrument was assessed across a sample of 256 sex offenders released from the Minnesota Department of Corrections, and those results were presented at the 1995 ATSA convention (Epperson, Kaul, and Huot, 1995). In summary, the results of that study showed that the MnSOST, in any of the forms assessed, could increase the accuracy of predictions substantially above chance levels, particularly in regards to predicting sexual recidivism. Sex offenders arrested for another sex offense upon release scored significantly higher on the MnSOST than did those who were arrested for a non-sex offense or

who had no arrests over a five- year follow-up. This version of the MnSOST was more successful in predicting recidivism for rapists and extrafamilial child molesters than for intrafamilial child molesters.

Based on these results, a major revision of the MnSOST was initiated in 1996 that focused on two elements. One major element in the revision process was to look at the more recent research on prediction of sexual recidivism (e.g., Monahan, 1994; Quinsey, Rice, and Harris, 1995; Rice and Harris, 1995; Quinsey, Lalumiere, Rice, and Harris, 1995) in an attempt to identify additional potential variables for the MnSOST. The Task Force also examined other newer inventories, such as the Psychopathy Checklist - Screening Version (Hart, Cox, and Hare, 1995) and the Level of Service Inventory - Revised (Andrews and Bonta, 1995).

The second element of this revision was a change to empirical methods for item selection and scoring. The previous MnSOST produced a total score that was used in an actuarial manner, but the scoring of individual items was clinically based. In contrast, the MnSOST-Revised (MnSOST-R) utilized empirical methods for item selection and scoring. Given the strong support for the general superiority of systematically derived empirical risk assessments over intuitive or even trained clinical predictions, as summarized above, it was assumed that the predictive validity of the MnSOST-R would be significantly improved by using empirically based, rather than clinically based, item selection and scoring.

The purpose of the present study was to evaluate the validity of the MnSOST-R by assessing its ability to predict sexual recidivism in a sample of nearly 400 offenders who had been released from a Minnesota Correctional Facility at least six years prior to the date of the study, thus allowing for a significant time at risk to reoffend.

Method

Samples

Development sample. The development sample for the MnSOST-R was comprised of 256 sex offenders incarcerated in the state of Minnesota for felony sex offenses. The sample included sex offenders with a history of extrafamilial child molestation or rape of related or unrelated victims. For the purposes of our research, offenses involving vaginal or anal penetration of a victim age 13 or younger were classified as rape. Forcible penetration was required for offenses against older victims to be classified as rape. Thus, the only group of felony sex offenders excluded from the sample were those with an exclusive history of intrafamilial offenses that would not be classified as rape using our criteria. This group was excluded because preliminary analyses indicated that this group's pattern of responses on potential predictor variables was substantially different than that for other types of offenders. Consistent with this, case managers and therapist reported that this population seemed qualitatively different to them as well. Over time, it also became clear that this population presented fewer concerns regarding three potential decisions that risk assessments might inform: level of supervision upon release, level of community notification, or possible referral for commitment.

The development sample included a random sub-sample of sex offenders released in 1988 (N = 107), a random sub-sample of sex offenders released in 1990 (N = 108), and a sub-sample of offenders readmitted to the Minnesota Department of Corrections during the time the sample was being put together, regardless of release year (N = 41). Most of the people in the last sub-sample were sexual recidivists (N = 29). This group was deliberately over-sampled to

provide more stability to any observed relationships between sexual recidivism and potential predictor variables.

The development sample included sex offenders from each Minnesota correctional facility that housed sex offenders. Sex offenders in the sample ranged in age from 17 to 70 years old (mean = 32.42), and the ethnic mix of the sample was 66% white, 23% African-American, 5% Hispanic, 4% Native American, 2% other minority groups. On average, sex offenders in the sample had 0.82 prior sex convictions, with a range from 0 to 8.

Cross-validation sample. The independent cross-validation sample consisted of an exhaustive sample of comparable sex offenders released in 1992 for whom complete data were available (N = 220). Sex offenders in this independent sample ranged in age from 19 to 70 years old, with a mean age of 35.40 years old. Also similar to the development sample, the ethnic mix for this sample was 71% white, 18% African-American, 5% Hispanic, 5% Native American, and 2% other minority groups.

Minnesota reliability sample. Eleven case files from the Minnesota Department of Corrections, which were current files at the time of the Minnesota reliability study, were selected through stratified (on expected score) random sampling. Eight case managers, 2 research assistants, and 3 supervisors with the Minnesota Department of Corrections scored each of the case files on the final version of the MnSOST-R. The 12 staff members varied dramatically in prior experience with the MnSOST-R, ranging from no experience to substantial experience. One of the staff members was not only new to the MnSOST-R but also brand new to the Minnesota Department of Corrections. This individual was actually "shadowing" another staff member selected to participate in the study. Although the individual was permitted to score the

cases, his scores were not included in the reliability analyses because he was not at all representative of the typical evaluator because his unfamiliarity with the structure and contents of DOC files and the associated inability to locate relevant information in the file. A second staff person failed to score all of the cases, so the data from that person were also excluded from analyses.

Florida reliability sample. A reliability study was conducted in Florida with the assistance of Mary Ann Bradley in the Florida Department of Children and Families (Bradley & Epperson, 2001). Participants in this study were twenty-seven Florida mental health professionals. The majority of participants (N = 24) held doctoral degrees in psychology. All participants reported a high degree of familiarity with Florida State DOC files, but prior experience scoring the MnSOST-R was highly variable. Six of the participants had never scored the MnSOST-R or scored it only once, and the median number of prior uses of the MnSOST-R was 13.5.

Ten Florida cases were selected using stratified (expected score) random sampling procedures to ensure that a range of scores would be represented in the sample, and all 10 cases were scored by each of the 27 participants.

Data Collection and Coding Procedures

Prison base files were compiled for each sex offender in the each of the samples described above. The files were created to include only information that would have been available prior to the offender's release. The files generally included the following documents: initial criminal complaints, pre-sentence investigation reports, initial psychological assessments, sex offender assessments, major discipline reports, treatment summaries (if applicable), and case manager summaries (initial, annual, and discharge). The files contained no information on revocations, arrests, or convictions subsequent to the release date.

Development study. Files were screened to ensure that all information subsequent to the offender's targeted release data had been removed. In addition, the names of offenders were deleted from all reports in the case files used in the development sample to minimize the likelihood that a case manager or researcher reviewing the file would be able to identify the offender.

The screened files were randomly assigned to 40 case managers for "blind" reviews on 17 variables identified in previous research with the original Minnesota Sex-Offender Screening Tool (Epperson, Kaul, & Huot, 1995). These variables included number of sex convictions, number of felony convictions, age at first conviction for a sex-related offense, use of a weapon in a sex-related offense, total number of victims, age(s) of victims, use of force, length of sex offending history, felony committed after a previous release, history of drug and alcohol abuse, history of sex offender treatment prior to incarceration, number of significant relationships, employment history, discipline history while incarcerated, chemical dependency treatment while incarcerated, and sex offender treatment while incarcerated. In those few instances where the case managers were randomly assigned a case that they recognized even with the name removed, the file was reassigned to a different case manager.

Each case file in the development sample was also reviewed by a researcher, who extracted data on a number of research dimensions. These dimensions included a range of demographic variables, adolescent maladjustment or antisocial behavior, family relationships and functioning, personal victimization history, employment history, substance abuse history, detailed criminal history, detailed treatment history, supervision failures, victim information (e.g., gender, age, relationship), and detailed offense information (e.g., acts preceding offense, location of offense, acts to achieve compliance, acts included in the actual offense, use and types of

weapons, and acts following the offense). A Psychopathy Checklist – Screening Version (PCL-SV) (Hart, Cox, & Hare, 1995) was also completed based on a file review by researchers who had been appropriately trained by the authors of this instrument.

Minnesota reliability study. As noted earlier, 11 current case files at the time of the reliability study were selected through stratified (on expected score) random sampling. Following a 2-hour training session on scoring the MnSOST-R, participating Minnesota Department of Corrections staff member scored all 11 cases the same day. In many respects, this was the worst possible conditions for reliability because of the very limited training (typical training is one to two days in length) and high degree of stress and fatigue resulting from the pressure to complete the 11 cases in about 6 hours. As a consequence, we considered the resulting estimate of reliability to be a minimal estimate.

Florida reliability study. The 27 Florida participants attended a 1.5 day training seminar on the MnSOST-R. Over the subsequent 3 months, each participant scored each of the 10 cases that were selected through the stratified (on expected score) random sampling procedures. Because the training on the MnSOST-R was more typical for this study, we expected it to produce higher indices of reliability.

Cross-validation study. Case files for the cross-validation study were developed in the same way and contained the same documents as in the development sample. For this confirmatory study, a MnSOST-R was scored for each case based on a file review by trained undergraduate research assistants at Iowa State University.

Criterion Variable

The criterion variable in risk assessment studies must be specified both in terms of the time at risk and the criterion event. Sexual recidivism was defined as arrest for a "hands-on" sex offense within 6 years of release for the development sample. Consistent with Quinsey and his colleagues (c.f., Quinsey, Harris, Rice, & Cormier, 1998, p. 119), we used arrest rather than conviction as the criterion event. One reason for this decision was to partially compensate for the under-estimation of sexual recidivism when using any official criminal justice record that results from the under-reporting of sex offenses to police. Similarly, this approach compensated for the tendency for sex offenses to be plea-bargained down to non-sex offenses even when there is clear evidence that a sex offense occurred. Finally, the arrest date is always more proximal to the actual offense date than is the conviction date, and the difference is often quite large due court delays often as long as a year or more.

The follow-up period of six years after release was selected because re-offense rates had clearly declined by the sixth year and because the size of the development sample began to decline substantially beyond the sixth year. Thus, offenders who were charged for a "hands-on" sex offense within 6 years of release were classified as sexual recidivists and offenders who were not charged with a "hands-on" sex offense within 6 years of release were classified as sexual non-recidivists.

An obvious problem with this criterion for sexual non-recidivism is that an offender is classified as a sexual non-recidivist even if he was not at risk for the full six years due to incarceration for a non-sexual offense. Consider the following illustration. Offender A was released in 1990, he remained out of prison for the entire 6-year period following his release, and

he was not charged with any "hands-on" sex offenses. Offender B was also released in 1990, and he, too, was not charged with any new "hands-on" sex offenses within 6 years of release. However, he was incarcerated for a non-sex offense in 1991 and remained in prison on that charge until 1996. Defining sexual non-recidivism as not being charged with a "hands-on" sex offense within 6 years of release results in both offenders being classified as sexual non-recidivists, even though Offender B was not at risk the full 6 years.

Although some consider incarcerations for non-sex offenses as "natural protection" against committing new sex offenses, others are reluctant to label an offender as a sexual non-recidivist unless he was at risk for the prescribed period of time. As a consequence, we constructed a second definition of sexual non-recidivism, which required that the offender be at risk the full 6 years without being charged with a hands-on sex offense to be classified as a sexual non-recidivist. Under the second classification system, offenders who were not charged with a hands-on offense but who were not at risk for the full 6 years would have to be excluded from analyses.

Our cross-validation offense data were sufficiently detailed that we could use both definitions of sexual recidivism. To distinguish the two definitions of sexual recidivism/non-recidivism in the cross-validation analyses, the first definition is labeled release recidivism and the second is labeled six-year risk recidivism.

New criminal charges were checked through computerized searches of the National Crime Index maintained by the Federal Bureau of Investigation and the state criminal data base maintained by the Minnesota Bureau of Criminal Apprehension. In addition, Minnesota

Department of Corrections records of all recidivists were reviewed to determine the nature of the new charge(s).

Results

Development Analyses

Item selection and scoring. The empirical selection and scoring of items for the MnSOST-R involved three steps. First, a modified Nuffield's (1982) procedure was used to score potential items. This intuitive procedure was familiar to the field because it was also used to score items on the Violence Risk Assessment Guide (VRAG) (Harris, Rice, & Quinsey, 1993) and on the Sex Offender Risk Assessment Guide (SORAG) (Quinsey, Rice, & Harris, 1995). The scoring of items involved cross-tabulating the levels of each potential individual item with six-year sexual recidivism rates and comparing those rates to the overall base rate, which was 35% in the development sample. If the sexual recidivism rate associated with an item level was within ±5% of the base rate, that item level was scored 0. Item levels associated with sexual recidivism rates at least 5% greater than the base rate were scored +1 for each 5% increment over the base rate (5% increase = +1, 10% increase +2, etc.). Our modification of this procedure was the additional requirement that, for an item level to be scored greater than one, there had to be a 5% increase over the next lower item level in addition to the required distance from base rate. This is illustrated in the following hypothetical example.

Total Number of Victims	Associated Sexual Recidivism	Difference from Base Rate	Possible Nuffield Score
One	32%	-3%	0
Two	41%	+6%	+1
Three	48%	+13%	+2
Four	50%	+15%	+3

In this hypothetical example, the "three victims" and "four victims" item levels would have been collapsed even though they are on different sides of the "15% difference from the base rate border" and a new associated recidivism rate would have been calculated and scored as follows:

Total Number of Victims	Associated Sexual Recidivism	Difference from Base Rate	Possible Nuffield Score
One	32%	-3%	0
Two	41%	+6%	+1
Three or Four	49%	+14%	+2

Similarly, item levels with sexual recidivism rates 5% lower than the base rate and the rate of the adjacent item level were scored –1 for each 5% decrement (5% decrease = -1, 10% decrease = -2, etc.). In addition, we generally required that an item level contain at least 10% of the sample in order to be scored, and a $\pm$ 4 limit was imposed on each item.

An item was dropped from further review if all levels of that item were scored 0, meaning that the item did not discriminate between recidivists and non-recidivists. If least one level of an item was scored different from zero and the scoring was consistent with existing theory and/or

empirical research, then the item was retained at this step. Consistency with existing theory and/or research was required to help avoid spurious relationships.

In the second step, the resulting score distribution on each retained item was required to be significantly associated with sexual recidivism ($p \leq .10$) as an additional control for capitalization on chance. The p level was slightly relaxed to avoid excluding variables prematurely because our greatest interest was in the relationship of the sum of the items with sexual recidivism and because the planned cross-validation study would provide the ultimate control for capitalization on chance.

The third and final step of the item selection process involved the elimination of overly redundant predictors. Hierarchical logistic regression analyses were used to assess the incremental addition of each item to the prediction of recidivism. For these analyses, variables were entered in the following order: (1) dynamic variables (e.g., discipline history, chemical dependency treatment while incarcerated, sex offender treatment while incarcerated, age at release from prison), (2) criminality/chronicity data (e.g., adolescent antisocial behavior, number of sex offense convictions, number of different age groups offended against, length of sex-offending history), (3) sex offense data (e.g., use or threat of force in any sex offense, 13 to 15 year old victim in any sex offense, any sex offense committed in a public place, any victim who was a stranger, multiple acts in any single event contact, sex offense committed while on supervision), and (4) personal history variables reflecting instability (e.g., unstable employment history, history of substance abuse). The dynamic variables were entered first in the sequential analyses because they are the only variables that can change in a positive direction during incarceration and are, therefore, assessed more closely in time to the offenders scheduled release

date. The order of the other variables reflected the general order of their expected predictiveness based on other research.

The 16 items retained in this last step are listed in Table 1. With the exception of employment history, each variable added incrementally to the prediction of reoffense status at the $p \leq .20$ level (most at the $p \leq .10$ level). Employment history was retained despite its lower level of contribution because it was added last, and the overall predictive ability of the tool was maximized with its inclusion.

Insert Table 1 about here

The scoring of each of the final 16 items is provided in Table 2. It is important to emphasize that item scores of zero reflect average or baseline risk, not the absence of risk. Similarly, negative scores reflect lower than average risk, not the absence of risk. Finally, detailed scoring directions are available from the authors, and these instructions must be followed in scoring the MnSOST-R; the information in Table 2 is not sufficient to adequately score the MnSOST-R.

Insert Table 2 about here

Association of MnSOST-R total score with sexual recidivism. The final 16 items were summed to produce each sex offender's MnSOST-R score. The association of MnSOST-R scores with six-year sexual recidivism status in the development sample was assessed through

correlational analyses and as the area under the receiver operator characteristics (ROC) curve. Area under the ROC curve is the superior index because of its independence from base rates and selection ratios (Quinsey et al., 1998, Rice & Harris,1995). The ROC curve is a plot of the tool's hit rate (sensitivity) against its false alarm rate (1 – specifity). An area under the ROC curve of .50 reflects chance levels of accuracy and an area under the ROC curve of 1.0 reflects perfect prediction. The area under the ROC curve can also be interpreted as the probability that a randomly selected sexual recidivist will have a higher score than a randomly selected sexual non-recidivist (Quinsey et al., 1998; Hanson & Thornton, 2000).

Correlation coefficients and areas under the ROC curves with 95% confidence intervals are provided in Table 3 for the total development sample and for relevant subsets of the development sample (rapists versus child molesters and non-minorities versus minorities). As indicated there, MnSOST-R scores were significantly associated with sexual recidivism and performed equally well across the various sub-groups.

Insert Table 3 about here

Cross-Validation Analyses

The MnSOST-R was "tailor-made" for the development sample and may have involved some degree of capitalization on chance, so it was important to establish the predictive validity of the MnSOST-R with an independent sample. Because of our more detailed follow-up data for the cross-validation sample, we were able to use two different definitions of sexual non-recidivism, as described earlier in the Criterion Variable subsection of the Method section.

Correlational and ROC analyses were performed with the independent cross validation sample for each definition of sexual non-recidivism, and the resulting correlation coefficients and areas under the ROC curves with 95% confidence intervals are presented in Table 4.

Insert Table 4 about here

Because the sexual recidivism base rate for the development sample (35%) was substantially higher than for cross-validation "release recidivism" sample (20%), and because differences in base rates impact correlation coefficients, the amount of shrinkage is best evaluated by comparing the areas under the ROC curves for these two samples. Overall, the area under the ROC curve decreased only slightly, from .77 to .73.

In examining the performance of the MnSOST-R with the sample subgroups, it is clear that the screening tool performed equivalently for all of the subgroups in the six-year risk recidivism cross-validation sample, which required that sexual non-recidivists be at risk for the full six years without being charged with a new "hands-on" sex offense. With the exception of minorities, the MnSOST-R also performed equally well with the subgroups in release recidivism cross-validation sample. The difference in the screening tool's performance with minorities in the two overlapping cross-validation samples indicated that high-scoring minorities were more likely to be re-incarcerated for non-sex offenses than were non-minorities and were, therefore, not fully at risk for sexual recidivism.

Reliability Analyses

Many reliability studies report inter-rater reliability data for a single pair of raters, raising questions about the generalizability of the findings to raters in general. The approach used in this both reliability studies reported here included multiple raters and multiple cases, permitting greater generalizability of the findings. The reliability index used with such a methodology is the singular intraclass correlation coefficient (ICC) calculated using a two-way random effects model.

In the Minnesota reliability study, the singular ICC for the 10 raters was .80 for consistency of ratings and .76 for absolute agreement of ratings, indicating that the ratings of individual raters were reasonably reliable, particularly give the harsh conditions for the raters. The Florida reliability study better reflected the conditions under which the MnSOST-R is typically scored in real-life situations, and this study yielded higher reliability coefficients: ICC = .87 for relative agreement of ratings and .86 for absolute agreement of ratings.

Risk Levels and Associated Rates of Sexual Recidivism

Based on our review of the data, examination of typical and relevant Department of Corrections' issues, and consideration of the types of decisions that MnSOST-R scores might inform, we recommended presumptive risk levels and associated cut scores to Minnesota Department of Corrections. Minnesota's statutes, and the statutes of many other states, call for three risk levels. Because Minnesota and at least 13 other states currently have a commitment statute, the presumptive action for a subset of the high risk is referral to county attorney for possible commitment. These recommended risk levels have been adopted by the Minnesota Department of Corrections. The presumptive risk levels and associated cut scores are listed in Table 5.

Insert Table 5 about here

It is important to emphasize two points about the risk levels. The first is that we strongly recommend that scores within a single risk level be treated as equivalent. The second is that these are presumptive risk levels or actions. Certain limited special considerations could conceivably result in an assigned risk level or action other than that indicated by the MnSOST-R score.

Ultimately, questions emerge about how high is high risk or how low is low risk and require answers, particularly in a forensic setting. The best way to answer such questions is by reporting the associated sexual recidivism rates in relevant representative samples. The MnSOST-R development sample was not representative because it over-sampled sexual recidivists. The 1988 and 1990 release year cohorts within the development sample were representative, however, because they were developed through random selection. The cross-validation sample was also representative because it was an exhaustive sample of rapists and extrafamilial child molesters released in 1992 for whom complete data were available. The release recidivism rates associated with each risk level for these two samples, individually and in aggregate, are provided in Table 6, along with 95% confidence intervals (Fleiss, 1981). Because the aggregated sample is the largest, representative sample that we have, the sexual recidivism rates for that sample are considered the most reliable.

Insert Table 6 about here

Six-year risk recidivism rates and 95% confidence intervals (Fleiss, 1981) in the cross-validation sample are reported in Table 7.

Insert Table 7 about here

Discussion

The MnSOST-R was developed in response to a departmental mandate to develop a more formal and uniform process to identify high-risk sex offenders at the time of their release from prison. Our goal was to develop a reliable and valid measure of long-term risk of sexual recidivism that was relatively brief and that could be scored by correctional case managers and other non-clinical staff using only the information routinely available for all sex offenders.

The MnSOST-R appears to meet these criteria. The 16-item instrument is relatively brief and it requires only information routinely available in typical base correctional files. In addition, it can be reliably scored with one to two days of training.

Two reliability studies in two different states attest to the reliability of raters using the MnSOST-R. The Minnesota study, which was performed under very substandard conditions (only 2 hours of training and the requirement to score 11 cases in the next 6 hours under "boiler room" conditions) yielded a reliability coefficient of .76. In contrast, the Florida study, which was performed under more standard conditions (1.5 days of training and scoring of 10 cases over an extended period of time in normal office conditions), yielded a reliability coefficient of .86.

Although both coefficients are reasonably high, the benefit of the more standard training conditions is evident.

The validity indices for the MnSOST-R in the cross-validation study (r = .35, ROC-AUC = .73) are comparable to those reported for other similar sex offender actuarial risk assessment tools. For example, Hanson and Thornton (2000) reported a correlation of .33 and an ROC-AUC of .71 for the Static-99, and a correlation of .28 and an ROC-AUC of .68 for the Rapid Risk Assessment of Sex Offender Recidivism (RRASOR). Each of these actuarial risk assessment tools has documented substantial incremental validity over clinical judgment, which had an average correlation with sexual recidivism of .10 in a meta-analysis by Hanson and Bussiere (1998).

Despite the demonstrated incremental validity of these instruments over clinical judgment, some argue that the accuracy of these instruments is not sufficiently high to inform criminal justice decision-making (e.g., Wollert, 2002). However, overall accuracy, or percent of variance accounted for, is not the relevant issue for many of those decisions (Doren & Epperson, 2001; Hanson & Thornton, 2000). Rather, for most such decisions, the critical issue or index is positive predictive power or negative predictive power. Positive predictive power is the probability that offenders with scores at or above a specified cut point will sexually recidivate, and its inverse is the probability of false positive predictions. Negative predictive power is the probability that offenders scoring below a specified cut score will not sexually recidivate, and its inverse is the probability of false negative predictions.

Positive predictive power is of more concern when the costs associated with false positive predictions are perceived to be greater than those associated with false negative

decisions. An example might be commitment decisions. Because commitment results in a continued loss of personal liberty despite having discharged a prison sentence, the costs associated with false positive predictions are high. In many states, the alternative to commitment is not just to simply release the offender into the community. Rather, the state can exercise a number of options governing the level of supervision and level of community notification that may reduce the threat that this person poses to the community, making the costs associated with false negative predictions arguably lower than the costs associated with false positive rates. As a consequence, in commitment decisions, the focus is on positive predictive power/false positive prediction rates.

One of the real benefits of actuarial risk assessment tools, including the MnSOST-R, is that estimates of these rates are known for each potential cut score, which is almost never the case with clinical prediction. As a consequence, a cut score can be selected that is sufficiently high to meet or exceed the required positive predictive power for a specific decision. In some juridisctions, the risk of sexual recidivism required for commitment is concrete and specific, a greater than 50% risk for example. In such jurisdictions a cut score can be selected that yields a positive predictive power of greater than 50%. In other jurisdictions, the criteria may be less concrete and decision-makers must determine what degree of false positive predictions they are willing or able to tolerate in making such decisions. In those situations, a cut score can also be identified that met the selected criteria, again, because the associated rates are known.

For other decisions, negative predictive power will emerge as central. Determining which sex offenders to release into the community with only standard supervision is an example of decisions in which the costs associated with false negative predictions are perceived to be greater

than those associated with false positive predictions. With this decision, the cost of false negative predictions is likely to be further victimization because the supervision may not be intensive enough to catch the offender in pre-offense behaviors. The cost of false positive predictions to the offender is relatively small because personal liberty will not be denied; rather, the offender will simply have to submit to closer supervision. In addition, there would be a monetary cost to the state for false positive predictions (the added cost of intensive closer supervision), but most would agree that the cost of false negative predictions (further victimization) in this example would still be greater. Consequently, a cut score that was sufficiently low to ensure that there would be high negative predictive power, with the associated low rate of false negative predictions, should be selected in such cases.

The ability to use different cut scores for different decisions, optimized in each case to provide the required positive or negative predictive power for the specific decisions, is a major strength of actuarial risk assessment tools, including the MnSOST-R. Unfortunately, this ability also makes these tools vulnerable to inappropriate or unfair criticism. For example, critics of an actuarial tool can select a high cut score on that tool to maximize positive predictive power at the expense of overall accuracy, but then use overall accuracy to compare the performance of the tool to that of clinical judgment or "betting the base rate". To avoid this unfair comparison of apples to oranges, it is critical that performance be compared on the same criteria used to select the cut score. Furthermore, in criminal justice settings, it is rare that overall accuracy would be the index of greatest interest. As discussed above, different cut scores are selected to meet the requirements of different decisions. In other words, you determine if your decision requires

greater negative predictive accuracy, positive predictive accuracy, or overall predictive accuracy, and then you select a score that maximizes the type of accuracy that you need.

MnSOST-R scores and associated risk levels are used to anchor a range of end-of-confinement decisions, including decisions regarding referral for commitment review, intensiveness of supervision, and level of community notification. Occasionally, a risk level other than that suggested by the MnSOST-R score is assigned to an offender based on the presence of one or more of a finite number of special considerations[1]. The requirement for such special considerations is that (1) they be outside the "actuarial frame" of the MnSOST-R (the variable was not evaluated for inclusion in the MnSOST-R) and (2) their potential relationship with sexual recidivism is self-evident and compelling (it nearly impossible to ignore this information in making decisions). As a consequence, such special considerations are almost always low frequency events and, therefore, do not enter into decisions for the majority of offenders.

The rationale for giving any weight at all to special considerations is that some potentially relevant risk factors occurred so infrequently in our sample that they could not be empirically evaluated for inclusion in the MnSOST-R. Examples of such variables include offenders' direct statements of intent to re-offend upon release and profound physical limitations with an onset during incarceration (e.g., a stroke in prison resulting in quadriplegia). In addition, some variables that we did evaluate (e.g., number of major disciplines for sexual behavior) had a small number of outliers with very high numbers of these events. Although we could evaluate for the presence or absence of a major discipline for sexual behavior, we could not evaluate the predictive power of having a high number of such disciplines because of the small number of offenders meeting that criterion.

If an offender's statement that he will commit more sex offenses upon release is documented in the file, it would have to be weighed heavily in determining his risk level. Similarly, if an offender has repeatedly acted out sexually even with the structure and sanctions of the correctional system, it would difficult to not factor this into a decision about risk level. Finally, if an offender who is exclusively a power rapist by history becomes permanently paralyzed while in prison, it is difficult to justify releasing him as a high-risk offender with all the associated costs of that classification.

Because the occurrence of such special considerations is fairly infrequent, so should deviations from the risk level associated with specific MnSOST-R scores. In general, assigned risk levels in the Minnesota Department of Corrections conform to those suggested by MnSOST-R scores.

References

Abel, G., Mittelman, M., Becker, J.V., Rathner, J., & Rouleau, J. (1988). Predicting child molester's response to treatment. In R.A. Prentky and V.L. Quinsey (Eds.) Human sexual aggression: current perspectives. New York: New York Academy of Science.

Alexander, R. (1995). Employing the mental health system to control sex offenders after penal incarceration. Law & Policy, 17, 111-130.

Andrews, D.A., & Bonta, J.L. (1995). The Level of Service Inventory - Revised. Toronto: Multi-Health Systems, Inc.

Bemus, B. (1988). Assessment of sex offender risk. Perspectives, 16-19.

Doke, J.D., (1989). Risk assessment scale. Paper presented at the annual conference of the Association for the Behavioral Treatment of Sexual Abusers, Seattle, WA.

Doren, D. M., & Epperson, D. L. (2001). Great analysis, but problematic assumptions: A critique of Janus and Meehl (1997). Sexual Abuse: A Journal of Research and Treatment, 13, 45-51

Epperson, D.L., Kaul, J.D., & Huot, S.J. (1995, October). Predicting risk of recidivism for incarcerated sex offenders: Updated development on the Sex Offender Screening Tool (SOST). Paper presented at the 14th annual conference of the Association for the Treatment of Sexual Abusers, New Orleans, LA.

Epperson. D.L., Kaul, J.D., Huot, S.J., Hesselton, D., & Alexander, W. (1997, October). Climbing the slippery slope of risk prediction: Updated developments on the Minnesota Sex Offender Screening Tool (MnSOST). Paper presented at the 16th annual conference of the Association for the Treatment of Sexual Abusers, Arlington, VA.

Finn, P. (1997). Sex offender community notification. National Institute of Justice: Research in Action.

Grove, W.M., & Meehl, P.E. (1996). Comparative efficiency of formal (mechanical, algorithmic) and informal (subjective, impressionistic) prediction procedures: The clinical/statistical controversy. Psychology, Public Policy & Law, 2, 293-323.

Hanson, R.K. (1997). The development of a brief actuarial risk scale for sexual offense recidivism. User Report No. 1997-04. Ottawa: Department of the Solicitor General.

Hanson, R.K., & Bussiere, M..T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. Journal of Consulting and Clinical Psychology, 66. 348-362.

Hanson, R.K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. Law and Human Behavior, 24, 119-136.

Hare, R.D. (1991). The Hare Psychopathy Checklist - Revised. Toronto: Multi-Health Systems, Inc.

Hare, R.D. (1996). Psychopathy and the PCL-R: Clinical and Forensic Application. Workshop Presentation: Minnesota Department of Corrections, St. Paul, MN.

Harris, G.T., Rice, M.E., & Quinsey, V.L. (1993). Violent recidivism of mentally disordered offenders. Criminal Justice and Behavior, 20, 315-335.

Hart, S.D., Cox, D.N., & Hare, R.D. The Hare PCL-SV. (1995). Toronto: Multi-Health Systems, Inc.

Janus, E.S. (1996). Preventing sexual violence: Setting principled constitutional boundaries on sex offender commitments. Indiana Law Journal, 72, 157-214.

Janus, E.S., & Meehl, P.E. (1997). Assessing the legal standard for predictions of dangerousness in sex offender commitment proceedings. Psychology, Public Policy, and Law, 3, 33-64.

Jones, P.R. (1998). Risk prediction in criminal justice. Public protection through offender risk reduction: putting research into practice. National Institute of Corrections.

Marshall, W.R., Jones, R., Ward, T., Johnston, P., & Barbaree, H.E. (1991). Treatment outcome with sex offenders. Clinical Psychology Review, 11, 468-485.

Monahan, J. (1996). Violence prediction: The past twenty years and the next twenty years. Criminal Justice and Behavior, 23, 107-120.

Monahan, J., & Steadman, H.J. (1994). Toward a rejuvenation of risk assessment research. In J.Monahan and H. Steadman (Eds.), Violence and mental disorder: developments in risk assessment. Chicago: University of Chicago Press.

Nuffield, J. (1982). Parole decision-making in Canada: Research towards decision guidelines. Ottawa: Solicitor General Canada.

Quinsey, V.L., Lalumiere, M.L., Rice, M.E., & Harris, G.T. (1995). Predicting sexual offenses. In J.C. Campbell (Ed.), Assessing dangerousness. Sage Publications.

Quinsey, V.L., Rice, M.E., & Harris, G.T. (1995). Actuarial prediction of sexual recidivism. Journal of Interpersonal Violence, 10, 85-106.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). Violent offenders: Appraising and managing risk. Washington, D. C.: American Psychological Association.

Rice, M.E., & Harris, G.T. (1995). Violent recidivism: Assessing predictive validity. Journal of Consulting and Clinical Psychology, 63. 737-748.

Rudin, J.B. (1996). Megan's Law: Can it stop sexual predators – and at what cost to constitutional rights? Criminal Justice, 17, 3-10.

Schulhofer, S.J. (1996). Two systems of social protection: Comments on the Civil-criminal distinction, with particular reference to sexually violent predator laws. Journal of Contemporary Legal Issues, 7, 69-96.

Witt, P.H., DelRusso, J., Oppenheim, J., & Ferguson, G. (1996). Sex offender risk assessment and the law. Journal of Psychiatry & Law, 24, 343-377.

Wollert, R.W. (2002). The importance of cross-validation in actuarial test construction: Shrinkage in the risk estimates for the Minnesota Sex Offender Screening Tool – Revised. Journal of Threat Assessment, 2, 87-102.

Working Group: Sex Offender Treatment Review (1990). The management and treatment of sex offenders. Report of the Minister of Supply and Services, Canada (MAS JS42-35/1990E).

Table 1

Bivariate Correlations of the Final Items with Sexual Recidivism and Additive Contributions to the Prediction of Sexual Recidivism Based on Hierarchical Regression in the Development Sample

	Bivariate Correlations with Sexual Recidivism		Hierarchical Logistic Regression with Sexual Recidivism		
Item Description	r	p	df	$X^2_{improvement}$	p
Dynamic Variables					
Discipline history while incarcerated	.11	.070	1	3.27	.071
Sex offender treatment while incarcerated	.11	.089	1	2.29	.130
Chemical dependency treatment history while incarcerated	.21	.001	1	7.66	.006
Age at release from prison	.12	.058	1	2.16	.142
Criminality/Chronicity Variables					
Adolescent antisocial behavior	.15	.015	1	4.98	.026
Number of sex offense convictions	.18	.005	1	6.34	.012
Number of different age groups victimized	.17	.006	1	2.91	.088
Length of sex offending history	.21	.001	1	4.57	.032
Offense-Related Variables					
Use of threat or force to achieve compliance in any sex offense	.13	.046	1	3.56	.059
Any victim 13 to 15 years of age and a 5-year age difference	.12	.055	1	3.02	.082
Any sex offense committed in a public place	.12	.052	1	1.69	.193
Any victim who was a stranger to the offender	.20	.001	1	9.83	.002
Any offense involving multiple sex acts in a single event contact	.16	.010	1	3.57	.059
Any sex offense committed while on supervision	.13	.037	1	5.21	.023

Table 1 (continued)

Bivariate Correlations of the Final Items with Sexual Recidivism and Additive Contributions to the Prediction of Sexual Recidivism Based on Hierarchical Regression in the Development Sample

	Bivariate Correlations with Sexual Recidivism		Hierarchical Logistic Regression with Sexual Recidivism		
Item Description	r	p	df	$X^2_{improvement}$	p
Unstable Life Style Variables					
Substance abuse	.13	.034	1	3.01	.083
Unstable employment history	.16	.012	1	0.70	.402

Table 2

MnSOST-R Item Scores

Item Number	Item Description	Item Score
	Static/Historical Items	
1	Number of sex/sex-related convictions (including current conviction):	
	One	0
	Two or more	+2
2	Length of sexual offending history:	
	Less than one year	–1
	One to six years	+3
	More than six years	0
3	Was the offender under any form of supervision when they committed any sex offense for which they were eventually charged or convicted?	
	No	0
	Yes	+2
4	Was any sex offense (charged or convicted) committed in a public place?	
	No	0
	Yes	+2
5	Was force or the threat of force ever used to achieve compliance in any sex offense (charged or convicted)?	
	No force in any offense	–3
	Force present in at least one offense	0
6	Has any sex offense (charged or convicted) involved multiple acts on a single victim within any single contact event?	
	No	–1
	Yes	+1

Item Number	Item Description	Item Score
7	Number of different age groups victimized across all sex/sex-related offenses (charged or convicted):	
	[] Age group of victims: (check all that apply)	
	[] Age 6 or younger	
	[] Age 7 to 12 years	
	[] Age 13 to 15 years and the offender is more than five years older than the victim	
	[] Age 16 or older	
	No age group or only one age group checked	0
	Two or more age groups checked	+3
8	Offended against a 13- to 15-year-old victim and the offender was more than five years older than the victim at the time of the offense (charged or convicted):	
	No	0
	Yes	+2
9	Was the victim a stranger in any sex/sex-related offense (charged or convicted)?	
	No victims were strangers	–1
	At least one victim was a stranger	+3
	Uncertain due to missing information	0
10	Is there evidence of adolescent antisocial behavior in the file?	
	No indication	–1
	Some relatively isolated antisocial acts	0
	Persistent, repetitive pattern	+2
11	Pattern of substantial drug or alcohol abuse (12 months prior to arrest for instant offense or revocation):	
	No	–1
	Yes	+1

Item Number	Item Description	Item Score
12	Employment history (12 months prior to arrest for instant offense):	
	Stable employment for one year or longer	–2
	Homemaker, retired, full-time student in good standing, or officially disabled	–2
	Part-time, seasonal, unstable employment	0
	Unemployed or significant history of unemployment	+1
	File contains no information	0
	Dynamic/Institutional Items	
13	Discipline history while incarcerated (does not include discipline for failure to follow treatment directives):	
	No major discipline reports or infractions	0
	One or more major discipline reports	+1
14	Chemical dependency treatment while incarcerated:	
	No treatment recommended / Not enough time / No opportunity	0
	Treatment recommended and successfully completed or in program at time of release	–2
	Treatment recommended but offender refused, quit, or did not pursue	+1
	Treatment recommended but terminated by staff	+4
15	Sex offender treatment history while incarcerated:	
	No treatment recommended / Not enough time / No opportunity	0
	Treatment recommended and successfully completed or in program at time of release	–1
	Treatment recommended but offender refused, quit, or did not pursue	0
	Treatment recommended but terminated	+3
16	Age of offender at time of release:	
	Age 30 or younger	+1
	Age 31 or older	–1

Table 3

Association of MnSOST-R Total Scores with Sexual Recidivism in the Development Sample

Groups	r	ROC-AUC[a]	95% CI[b]
Total Sample	.45	.77	.71-.83
Rapists	.47	.79	.71-.87
Molesters	.41	.75	.65-.84
Minorities	.42	.75	.64-.84
Non-Minorities	.46	.77	.69-.84

[a]Receiver operator characteristics area under the curve

[b]95% confidence interval for the ROC-AUC

Table 4

Association of MnSOST-R Total Scores with Sexual Recidivism in the Cross-Validation Sample

	Release Recidivism Sample[a]			Six-Year Risk Recidivism Sample[b]		
Groups	r	ROC-AUC[c]	95% CI[d]	r	ROC-AUC[c]	95% CI[d]
Total Sample	.35	.73	.65-.82	.46	.77	.69-.86
Rapists	.37	.77	.62-.92	.52	.82	.66-.98
Molesters	.34	.73	.63-.83	.42	.77	.67-.86
Minorities	.22	.64	.49-.80	.45	.74	.59-.90
Non-Minorities	.41	.77	.66-.87	.48	.79	.69-.89

[a]Sexual non-recidivism defined as not being charged with a new "hands-on" sex offense within 6 years of release regardless of how much of the six years the offender was actually at risk (N = 220).

[b]Sexual non-recidivism defined as being at risk for the full 6 years without being charged with a new "hands-on" sex offense (N = 170).

[c]Receiver operator characteristics area under the curve.

[d]95% confidence interval for the ROC-AUC.

Table 5

Presumptive Risk Levels and Associated MnSOST-R Cut Scores

Presumptive Risk Level	MnSOST-R Score
1 (low)	3 and below
2 (moderate)	4 to 7
3 (high)	8 and above
Refer to county attorney	13 and above

Note. [a]The referral group is a subset of the high risk group.

Table 6

MnSOST-R Risk Levels, Associated Release Recidivism Rates, and 95% Confidence Intervals for Relevant Representative Samples

	Representative 1988 & 1990 Cohorts[a]			Cross-Validation Sample[b]			Aggregated Samples[c]		
Risk Level	Proportion of Sample N (%)	Recidivism Rate	95% CI	Proportion of Sample N (%)	Sexual Recidivism Rate	95% CI	Proportion of Sample N (%)	Sexual Recidivism Rate	95% CI
1	126 (59%)	14%	9% - 22%	126 (57%)	10%	5% - 16%	252 (58%)	12%	8% - 17%
2	48 (22%)	31%	19% - 46%	52 (24%)	19%	10% - 33%	100 (23%)	25%	17% - 35%
3	41 (19%)	61%	45% - 75%	42 (19%)	52%	37% - 68%	83 (19%)	57%	45% - 67%
Refer[d]	12 (6%)	92%	60% - 100%	13 (6%)	54%[e]	26% - 80%	25 (6%)	72%	50% - 87%

Note. Sexual non-recidivism is defined as not being charged with a new "hands-on" sex offense within 6 years of release regardless of whether or not the offender was at risk for the full 6 years.

[a] N = 215, recidivism base rate = 27%. These are the representative cohorts from the development sample.

[b] N = 220, recidivism base rate = 20%

[c] N = 435, recidivism base rate = 23%

[d] The referral group is a subset of the high risk group. The n's for the referral groups in the first two separate samples are very small, resulting in very wide confidence intervals.

[e] This rate may have been lowered by the fact that 6 1992 releasees were civilly committed and not at risk to recidiviate

Table 7

MnSOST-R Risk Levels and Associated Six-Year Risk Recidivism Rates and 95% Confidence Intervals for Cross-Validation Sample

Risk Level	Proportion of Sample N (%)	Sexual Recidivism Rate	95% CI
1	102 (60%)	12%	6% - 20%
2	38 (22%)	26%	14% - 43%
3	30 (18%)	73%	54% - 87%
Refer[a]	8 (6%)	88%	47% - 99%

Note. Sample N = 170 and the sexual recidivism base rate is 26%. Non-recidivism is defined as being at risk the full 6 years without being charged with a new "hands-on" sex offense.

[a] The referral group is a subset of the high risk group. The n for the referral group is extremely small, resulting in a very wide confidence interval.