THE MINNESOTA SEX OFFENDER SCREENING TOOL-3.1 (MNSOST-3.1): AN UPDATE TO THE MnSOST-3

Research Summary

Since the 1990s, the Minnesota Sex Offender Screening Tool-Revised (MnSOST-R) has been one of the most widely-used sex offender risk assessment tools. Recently, the MnSOST-R was updated through the development of the MnSOST-3. Multiple logistic regression was used to create the MnSOST-3 and relied on bootstrap resampling to not only select items to be included in the instrument, but also to internally validate the model. The MnSOST-3, which is scored in a Microsoft Excel application, contains 11 predictors, two of which are interaction terms. In January 2012, the Minnesota Department of Corrections (MnDOC) began using the MnSOST-3 in place of the MnSOST-R. But after examining the first several hundred cases scored with the MnSOST-3, several issues arose concerning the behavior of the two interaction terms that prompted the development of the MnSOST-3.1, which is strictly a main effects model. In presenting the findings resulting from the creation of the MnSOST-3.1, the study shows that the predictive discrimination and calibration of the MnSOST-3.1 is similar to that of the MnSOST-3. Yet, because the MnSOST-3.1 excludes the two interaction terms, it produces risk assessment output that is somewhat simpler and easier to interpret.

INTRODUCTION

More than 20 years ago, Epperson and colleagues began work on developing the Minnesota Sex Offender Screening Tool (MnSOST) (Epperson, Kaul, Huot, Goldman, and Alexander, 2003). In 1996, they initiated efforts to revise the MnSOST, eventually resulting in the Minnesota Sex Offender Screening Tool-Revised (MnSOST-R). Most recently, the MnSOST-R was updated through the development of the MnSOST-3 (Duwe and Freske, 2012).

The MnSOST-3 is different in a number of ways from its MnSOST predecessors. The MnSOST-R, for example, was developed on a sample of 256 sex offenders released from Minnesota prisons during the late 1980s and early 1990s. Using sex offense rearrest within six years as the outcome measure, Epperson et al. (2003) employed a modified Nuffield (1982) weighting scheme by first cross-tabulating potential individual items with recidivism rates and then comparing those rates with the baseline rate. Weights were assigned to items based on the magnitude of difference between the recidivism rates for individual items and the baseline rate. Individual items were retained in the MnSOST-R if: a) the assigned value was different from 0, b) the item was consistent with existing theory and/or practice, c) the association with sexual recidivism was $p < .10$, and d) the items significantly improved the prediction of sexual reoffending in a hierarchical logistic regression model at the $p \leq .20$ level (Epperson et al., 2003). Altogether, the MnSOST-R contains 16 items and is scored in a pencil-and-paper format, with scores ranging from a low of -12 to a high of 31.

In developing the MnSOST-3, the sample consisted of 2,535 sex offenders released from Minnesota prisons. The 2,535 offenders were drawn from two separate samples: the MnSOST-R cross-validation sample and a contemporary sample of released sex offenders.

The MnSOST-R cross-validation sample contained 220 offenders released from Minnesota prisons during the early 1990s, whereas the contemporary sample included 2,315 sex offenders released from Minnesota prisons between 2003 and 2006. Relying on sex offense reconviction within four years as the outcome measure, multiple logistic regression was used to create the instrument. Moreover, bootstrap resampling was employed to not only select items to be included in the instrument, but also to internally validate the model. The MnSOST-3 contains 11 predictors—nine main effects and two interaction effects. Of the nine main effects, only three were items derived from the MnSOST-R (public place, completion of chemical dependency and sex offender treatment, and age at release).

The MnSOST-3, which is scored in a Microsoft Excel application, provides several measures of sexual recidivism risk. The MnSOST-3 value an offender receives represents his predicted probability of sexual recidivism within four years, which varies from a low of 0 percent to a high of 100 percent. To provide a range in which the true risk of sexual recidivism likely falls, 95 percent confidence intervals (CIs) were calculated around MnSOST-3 estimates. While the MnSOST-3 value and the accompanying 95 percent CIs offer measures of absolute sexual recidivism risk, percentile ranking was also included to provide a measure of relative risk.

To illustrate, an offender with a MnSOST-3 value (i.e., predicted probability) of 10 percent would fall into the 92nd percentile. Moreover, this offender would have a lower CI of 5 percent and an upper CI of 16 percent. Therefore, the MnSOST-3 output for this offender suggests that his likelihood of reconviction for a new sex offense within four years is 10 percent. The CI's, meanwhile, imply a 95 percent likelihood that his true likelihood for a new sex crime reconviction falls between 5 and 16 percent. And the percentile ranking indicates

2

that only 8 percent of the Minnesota sex offenders had a MnSOST-3 value higher than 10 percent.

After development of the MnSOST-3 was completed, the Minnesota Department of Corrections (MnDOC) began using it in place of the MnSOST-R in early January 2012. By the end of January 2012, the MnDOC's Risk Assessment and Community Notification (RACN) Unit had scored more than 200 cases on the MnSOST-3. Upon reviewing these cases, several potential issues were identified with the MnSOST-3, particularly involving the two interaction terms in the model.

First, in the MnSOST-3, both the effects of violations of orders for protection (VOFP) and recent disorderly conduct convictions on sexual recidivism risk vary according to the age of the offender at the time of release. Whereas VOFPs increase the risk for younger offenders, they decrease the risk for older offenders. Conversely, recent disorderly conduct convictions increase the risk for older offenders, while they decrease the risk for younger offenders. The interaction effects may appear counter-intuitive because VOFP and disorderly conduct convictions decrease an offender's risk in some instances. Moreover, given that the VOFP and disorderly conduct convictions are relatively new findings in the sex offender literature, prior research offers little guidance.

Second, although VOFPs and disorderly conduct convictions will reduce the MnSOST-3 score for certain offenders, it does not necessarily mean that these offenders would pose less of a risk for violent recidivism. Additional data are currently being analyzed related to the ability to assess risk for different types of recidivism, including non-sexual violent, non-sexual, felony, and first-time sexual offending. Preliminary findings from these analyses suggest that VOFPs and disorderly conduct convictions increase the risk of non-

sexual violent recidivism. This finding suggests the possibility that the risk for sexual recidivism may drop because the risk for other types of recidivism (non-sexual violence) increases. For example, the increased likelihood for non-sexual violent recidivism may shorten the at-risk period for sexual recidivism, especially if that recidivism leads to reincarceration (resulting in a more limited opportunity to sexually reoffend).

Finally, it was observed that approximately 20 percent of the offenders had a prior VOFP among the cases scored by the RACN Unit, which is more than three times the percentage observed (6 percent) in the MnSOST-3 development sample. Given the lack of certainty in explaining the VOFP-age interaction, the higher percentage of offenders with VOFP convictions raised concerns regarding this item's interaction with age at release. Therefore, in an effort to produce an instrument that is simpler and easier to interpret, the model's performance was assessed without the two interaction terms. As shown later, removing the two interaction terms did not have a detrimental effect on predictive validity. As a result, in February 2012 the MnDOC began using the main effects-only model—the MnSOST-3.1.

PRESENT STUDY

In this study, work on the MnSOST-3 is updated by presenting findings from the development of the MnSOST-3.1. The development of the MnSOST-3.1 is, of course, similar to that of the MnSOST-3, and the discussion of the MnSOST-3.1's development will likely be familiar to those who have reviewed work on the MnSOST-3 (Duwe and Freske, 2012). Where relevant, however, this report will identify the ways in which the MnSOST-3.1 differs from the MnSOST-3. In addition to presenting the MnSOST-3.1 results, this study also

includes those from the MnSOST-3 for purposes of comparison. The concluding remarks will focus mainly on practical use of the instrument.

MnSOST-3 AND 3.1 SAMPLE

As noted above, 2,535 sex offenders were examined who were drawn from two separate samples: the MnSOST-R cross-validation sample and a contemporary sample of released sex offenders. Included among the 2,535 sex offenders were 99 offenders whose only sex offense conviction(s) occurred as a juvenile, 53 "intrafamilial fondlers" (a group of incest-only offenders whose only sex offenses consisted exclusively of non-penetration sexual contact for whom the MnSOST-R has had limited predictive accuracy), and 12 offenders whose only sex-related offense(s) involved possession of child pornography. These groups of offenders were included in the development sample because they have at least one prior sex or sex-related offense, which triggers the need to assess their risk for sexual recidivism, as evidenced by the fact that MnSOST-R assessments were administered to these offenders while they were in prison.

The MnSOST-R cross-validation sample contains 220 offenders released from Minnesota prisons during the early 1990s, whereas the contemporary sample includes 2,315 sex offenders released from Minnesota prisons between 2003 and 2006. During this four-year period, there were 134 sex offenders who were released from prison but were not at risk to reoffend because they were civilly committed. Due to the absence of an at-risk period, the 134 civilly committed offenders were excluded. Yet, as discussed later, assessments were completed from both versions of the MnSOST-3 on these offenders to further assess the validity of the instrument.

The contemporary and MnSOST-R cross-validation samples were used to develop both versions of the MnSOST-3 for a few reasons. First, due to the recent decline in sexual recidivism rates and to concerns raised about the inflated baseline rate for the MnSOST-R development sample (Vrieze & Grove, 2008; Wollert, 2002), it was necessary to select a group of sex offenders who had recently been released from prison. Second, as noted above, the MnSOST-R development and cross-validation samples contained sex offenders released from Minnesota prisons during the late 1980s and early 1990s. Since that time, however, sex offenders released from Minnesota prisons have been more likely to be civilly committed, subjected to broad community notification, intensively supervised, have their supervised release revoked for a technical violation, and incarcerated for longer periods of time. The growing use of these external constraints has likely been responsible, at least to some extent, for the declining sexual recidivism rates observed in Minnesota since the early 1990s (Minnesota Department of Corrections, 2007). To ensure that both versions of the MnSOST-3 predict sexual recidivism risk without constraints as accurately as the MnSOST-R, it was also necessary to include the MnSOST-R cross-validation sample. Although data were available on the sample used to develop the MnSOST-R, this sample was not used to develop either version of the MnSOST-3 because it oversampled for recidivists. As shown later, however, the MnSOST-R development sample was used to help cross-check the predictive validity of the MnSOST-3.1.

All 2,535 sex offenders in this study were scored at least once on the MnSOST-R. In some instances, offenders received more than one MnSOST-R assessment during the same sentence. For the offenders in the contemporary sample who had more than one MnSOST-R assessment during their confinement, the most recent score prior to their release date was

selected. Minnesota prisoners receive a MnSOST-R assessment if they have at least one sex offense in their history for which documentation is available. Of the 2,535 offenders, 67 percent were incarcerated for a sex offense while the remaining 33 percent had a non-sexual index offense.

MnSOST-3 AND 3.1 ITEMS

To recalibrate the weights assigned to the 16 items on the MnSOST-R, binary measures were created for the dichotomous items (e.g., under any form of supervision, sex offense committed in a public place, force used, multiple acts, offended against a 13-15-year-old victim). For example, on the MnSOST-R, offenders who have committed a sex offense in a public location are given a value of "2", whereas those who have not committed a sex offense in public are assigned a value of "0". The scoring of these items were modified in the multiple logistic regression analyses by giving them values of either "0" or "1". The categorical measures on the MnSOST-R (e.g., length of sexual offending history, different age groups, stranger victims, adolescent antisocial behavior, pattern of recent alcohol or drug abuse, employment history, chemical dependency treatment, and sex offender treatment) were transformed into dichotomous dummy variables. For example, on the MnSOST-R, offenders whose history of sexual offending is less than one year receive a value of "-1", offenders with a history between one and six years are given a value of "3", while those with a history in excess of six years are assigned a value of "0". For the multiple logistic regression analyses, the following three variables were created for length of sexual offending history: Less than one year (Yes = 1; No = 0), One to six years (Yes = 1; No = 0), and More than six years (Yes = 1; No = 0). Less than one year was the reference in the statistical

analysis. Finally, three of the MnSOST-R items—number of sex offenses, discipline convictions, and age at release—were transformed into continuous variables.

To identify whether there are, in addition to the 16 MnSOST-R items, other factors predictive of sexual recidivism, all of the data collected by the MnDOC and maintained in the Correctional Operations Management System (COMS) were gathered on the 2,535 offenders. The data included information relating to demographics, prior criminal history (e.g., total number of convictions, age at first conviction, type of offense, etc.), educational level (e.g., presence or absence of high school degree or general equivalency diploma at admission and release from prison), institutional misconduct (e.g., whether the offender received any disciplinary sanctions, the total number of disciplinary convictions, the type of institutional misconduct, etc.), gang membership (i.e., security threat group), involvement in institutional programming (e.g., anger management classes, critical thinking courses, etc.), prison visitation (e.g., whether offenders were visited in prison, the number of times they were visited in prison, number of prison visits divided by length of stay, etc.), length of stay in prison during the most recent incarceration period prior to release, total prison time served during the current sentence, type of offense (e.g., sex offense, assault, robbery, failure to register as a predatory offender, etc.), type of prison admission (e.g., new court commitment, probation violator, and supervised release violator), whether they were released to supervision and, if so, what type of supervision (e.g., regular supervision and intensive supervised release) (a full list of the variables used can be obtained from the authors upon request). To facilitate valid and reliable scoring of both versions of the MnSOST-3, this study focused on identifying items that significantly predicted sexual recidivism, were consistent

with existing theory and/or research, and were relatively objective measures that are consistently available in COMS.

A difference between MnSOST-3 and MnSOST-3.1 items is that, in order to avoid inflated risk estimates resulting from extreme values among the five continuous items on the instrument, the values for some items on the MnSOST-3.1 are capped based on the highest value observed in the MnSOST-3/MnSOST-3.1 development sample. Therefore, the largest value an offender can receive for the sentences with male victims item is four. Predatory offense sentences are capped at 25, felony offense sentences as 20, VOFP convictions at 5, and recent disorderly conduct convictions at 2.

MEASURING SEXUAL RECIDIVISM

Sex offense reconviction data were collected on the 2,535 sex offenders through the end of December 31, 2010. For the offenders in the contemporary sample released toward the end of 2006, four years was the maximum follow-up period. Because logistic regression assumes that each offender has the same amount of time in which to reoffend, the follow-up period was limited to four years for all 2,535 offenders in this study.

Sexual recidivism was defined as a reconviction for a new sex crime within four years of release. In operationalizing sex crimes, only hands-on sex offenses were included. In doing so, non-contact, sex-related offenses such as possession of child pornography or indecent exposure were excluded. Reconviction was used as the recidivism measure because it reduces the likelihood of including false positives (i.e., cases that are not truly instances of sexual recidivism). Although rearrest is arguably a more sensitive measure of recidivism and, thus, increases the chances of identifying more true positives (i.e., actual sex reoffenses), it also increases the odds of including more false positives. In addition, information on the

date(s) when the reoffense occurred was seldom available in the rearrest data but was consistently present in the conviction data. Offense date information was necessary to exclude cases of "pseudo recidivism," as there were a handful of offenders who returned to prison for a "new" sex offense that had been committed prior to the beginning of their previous prison term, e.g., an offender who was incarcerated from 2002-2005 is reconvicted in 2008 for an offense committed in 1998. In these instances, the reconviction was not considered to be a recidivism event.

Reconviction data were obtained from both the Minnesota Bureau of Criminal Apprehension (BCA) and the Federal Bureau of Investigation (FBI). Whereas the BCA data include only convictions that occur in Minnesota, the FBI criminal history data contain information on convictions that took place outside Minnesota. As with any recidivism study, official criminal history data will likely underestimate the actual extent to which the sex offenders examined here recidivated.

The recidivism data revealed that 102 (4.0 percent) of the 2,535 offenders had been reconvicted of a new sex offense within four years of their release from prison. The four-year sexual reconviction rate was 12.3 percent in the MnSOST-R validation sample and 3.3 percent in the contemporary sample.

DEVELOPING THE MNSOST-3 AND 3.1

Existing research has identified three types of validity important for predictive regression modeling: apparent, internal and external (Harrell, Lee, & Marks, 1996). Apparent validity refers to performance on the sample used to develop the prediction model. In examining the performance of the model on the population underlying the sample, internal validity is concerned with whether the model can be reproduced. External validity,

meanwhile, focuses on the generalizability of the model by looking at how well it performs on a related, but slightly different, population. Applied to the present study, apparent validity addresses the performance of the MnSOST-3 on the sample used to develop it. While internal validity tells us how well the MnSOST-3 would likely perform on other samples of Minnesota sex offenders, external validity would assess MnSOST-3 performance on non-Minnesota sex offender populations. Apparent and internal validity are the focus of this study.

To assess apparent validity, statistics such as ROC curves may be estimated on the development sample to determine the predictive accuracy of the model. As for internal validity, three main methods have been developed to determine the reproducibility of a prediction model. The split-population, or data splitting, method has been the most popular approach in the development of sexual recidivism risk assessment tools. With this method, a portion (e.g. one-half or two-thirds) of the sample is used to develop the prediction model. The developed model is then applied to the remaining portion in order to test the internal validity of the model. Despite its popularity, this approach wastes data (Harrell et al., 1996).

Cross-validation, or k-fold validation, is more efficient than the split-population approach because it involves repeated data splitting. Research has demonstrated, however, that bootstrap resampling is the most efficient internal validation technique (Steyerberg, Harrell, Borsboom, Eijkemans, Vergouwe, & Habbema, 2001; Steyerberg, Bleeker, Moll, Grobbee, & Moons, 2003). Developed by Efron (1979), bootstrap resampling involves pulling many smaller samples from the overall sample in order to generate estimates of error. In doing so, it makes full use of the data set for developing and validating models while also providing error estimates that have relatively low variability and minimal bias (Harrell, 2001;

Steyerberg et al., 2001). As discussed shortly, bootstrap resampling was used to not only refine our selection of items for the MnSOST-3/MnSOST-3.1, but also to calculate estimates of optimism due to overfitting for both versions of the MnSOST-3.

SELECTION OF PREDICTORS

Stepwise variable selection procedures are frequently used in the development of prediction models. Although there are a variety of stepwise methods available, the two main approaches are forward selection and backward selection. Under forward selection, a variable does not enter the model unless it is statistically significant at a predetermined level (e.g., alpha = 0.05). With backward selection, a variable is removed from the model if its level of statistical significance exceeds the established alpha level. Stepwise routines have been criticized on a number of grounds, especially for producing biased regression coefficients (Tibshirani, 1996) and for capitalizing on chance features of the data (Judd, McClelland, & Ryan, 2008). Still, because backward selection is generally preferable to forward selection (Harrell et al., 1996), it is the approach that was used here.

Multiple logistic regression analyses were conducted on the offenders in the development sample to identify significant predictors of sexual recidivism. In addition to including the 16 items from the MnSOST-R, a host of variables derived from COMS data were examined. Using an alpha of .10, more than 100 potential predictors were analyzed. Following Efron and Gong (1983), predictors were added one at a time until no further single addition achieved significance level a = .10. Among the 10 significant predictors, there were 45 possible two-way interaction effects that were tested. Using an alpha of .05, six interaction effects were found that were statistically significant.

In an effort to develop a more parsimonious prediction model, bootstrap resampling was used to refine the selection of predictors included in the MnSOST-3/MnSOST-3.1. More specifically, only the predictors that were consistently significant in the bootstrap samples were retained. Although the bootstrap variable selection method has been discussed in the literature (Efron & Gong, 1983), there is no widely accepted "rule of thumb" threshold for retaining or removing predictors. Zhao (1998) recommended using at least a 40 percent cutoff (i.e., predictors are retained in at least 40 percent of the bootstrap samples), whereas Cooke and colleagues (2009) used a 60 percent threshold. Here, a relatively high threshold (70 percent) was used to determine whether predictors should be included in the model.

Table 1. Multiple Logistic Regression Model for MnSOST-3.1

Predictors	MnSOST-3.1		MnSOST-3	
	<u>B</u>	<u>p</u>	<u>B</u>	<u>p</u>
Predatory Offenses	0.289	.001	0.292	.001
Male Victims	0.905	.000	0.874	.000
Public Place	0.653	.009	0.747	.003
Felony Offenses	0.126	.000	0.129	.000
VOFP/Stalking/Harassment	0.436	.009	3.271	.001
Disorderly Conduct (Last 3 Years)	0.936	.000	-1.742	.057
Complete SO/CD Treatment	-1.491	.017	-1.557	.013
Age at Release (Years)	-0.044	.000	-0.044	.001
Unsupervised Release	1.726	.000	1.783	.000
VOFP X Age			-0.099	.010
Recent Disorderly Conduct X Age			0.074	.008
Constant	-3.254	.000	-3.247	.000
N	2,535		2,535	
Log-likelihood	699.107		686.075	
Nagelkerke R^2	0.209		0.225	

After estimating 1,000 bootstrap samples from the 16-predictor model, there were five predictors (one main effect and four interaction terms) that were statistically significant at the .05 level in less than 70 percent of the samples. After removing these five predictors,

another 1,000 bootstrap samples were estimated. The results showed that 11 predictors (nine main effects and two interaction effects) were statistically significant at the .05 level in at least 70 percent of the bootstrap samples. For the MnSOST-3.1, two interaction terms were removed, which left nine main effects in the model. The results for the main effects model are presented in Table 1.

Discussion of Multiple Logistic Regression Results

Of the nine main effects in the model, three are items derived from the MnSOST-R (public place, completion of chemical dependency and sex offender treatment, and age at release). Although the predatory offense sentences item is somewhat similar to the number of sex/sex-related convictions item on the MnSOST-R, it is arguably a much broader measure of sexual offending history. Moreover, even among the three items derived directly from the MnSOST-R, it is worth noting that they are measured differently for both versions of the MnSOST-3. For example, public place is a dichotomous measure (as opposed to a categorical item), completion of both chemical dependency and sex offender treatment is a dichotomous measure that merges these two categorical items on the MnSOST-R, while age at release is a continuous, rather than a dichotomous, measure. Although a visual inspection of the residuals did not reveal any signs of nonlinearity for either age at release or number of predatory offenses, this study tested for nonlinearity by estimating a model with a logarithmic transformation of both predictors. Neither coefficient, however, was statistically significant at the .10 level, which suggests that recidivism or, more specifically, the logit of the recidivism measure used here is linearly related to age at release and number of predatory offenses.

The results presented in Table 1 are generally consistent with existing research. The risk of sexual recidivism was significantly less for offenders who completed both chemical dependency and sex offender treatment in prison, a finding that dovetails with prior research on offenders from Minnesota (Duwe, 2010; Duwe and Goldman, 2009) and in general (Lösel & Schmucker, 2005; Mitchell, Wilson, & MacKenzie, 2007). Similar to prior research on sex offenders (Hanson & Morton-Bourgon, 2004) and, more narrowly, those from Minnesota (Epperson et al., 2003), the risk was significantly greater for younger sex offenders and those with more prior predatory offenses, more predatory offenses that involved male victims, and a history of committing a sex-related offense in a public location.

The number of felony sentences a sex offender had significantly increased the odds of reoffending sexually. Also, the risk of sexual recidivism was significantly greater for offenders with convictions for VOFP, stalking, or harassment. In addition to measuring impulsivity, this measure may tap into rule noncompliance and intimacy deficits, which have been found to be salient predictors in previous research (Hanson & Morton-Bourgon, 2004). The results showed that offenders with disorderly conduct convictions in the three years preceding their commitment to prison had a significantly elevated risk of recidivism.

Offenders who were released to no supervision (because their sentence had expired) were significantly more likely to reoffend sexually than those who were released to some form of post-prison supervision. Offenders were typically released to no supervision if they had multiple stays in prison as a release violator or had accumulated substantial extended incarceration disciplinary time stemming from institutional misconduct or failure to complete a sex offender treatment directive. The finding regarding the absence of post-release supervision is consistent with recent research on offenders in general, which has shown that

prisoners who "max out" are significantly more likely to reoffend (Ostermann, 2009; Schlager & Robbins, 2008). Moreover, in their validation study of the MnSOST-R and Static-99, Boccaccini and colleagues (2009) found that the risk of sexual recidivism was significantly greater for sex offenders who were discharged (i.e., released to no supervision).

ASSESSING PREDICTIVE ACCURACY

The validity, or accuracy, of a prediction model is often assessed by examining its predictive discrimination and calibration (Harrell et al., 1996; Steyerberg, 2009). With the MnSOST-3.1, predictive discrimination looks at how well it separates recidivists from non-recidivists. Calibration, on the other hand, examines the extent to which there is agreement between the predicted probabilities of recidivism and the actual rates of reoffending. In light of the recent decline in sexual recidivism, one of the concerns raised about tools such as the MnSOST-R is that, despite having good predictive discrimination, it overestimates the risk of sexual recidivism (Wollert, 2006). With a well-calibrated model, however, the predicted probabilities closely correspond with the observed recidivism rates. In the ensuing sections, predictive discrimination of the MnSOST-3.1 is examined before moving on to an assessment of its calibration with actual rates of sexual recidivism.

Predictive Discrimination

The apparent predictive discrimination for the MnSOST-3.1 was first analyzed by estimating a ROC curve for the predicted probabilities derived from the main effects model shown in Table 1. The apparent AUC value for the main effects model is .818, which is slightly lower than the apparent AUC obtained for the MnSOST-3 interaction model (.821). To determine the extent to which this value overestimates predictive discrimination due to

overfitting, an optimism value was estimated based on the method described by Efron and Tibshirani (1993).

First, as shown above, an upwardly biased (i.e., overly optimistic) AUC estimate of apparent predictive discrimination was obtained based on the full sex offender sample (N = 2,535) examined here. Second, a bootstrap sample was drawn from the full offender sample and then obtained maximum likelihood estimates of beta weights based on the bootstrap sample. Third, an AUC value was calculated for that bootstrap sample. Fourth, the beta weights developed from the bootstrap model were applied to the full offender sample and AUC values for these results were obtained. Fifth, optimism estimates were generated by calculating the differences in AUC values obtained during the third and fourth steps. Sixth, steps two through five were repeated 200 times, and the differences obtained at each iteration were recorded. Seventh, the average of the 200 differences generated during step six was used as the "bootstrap estimate" of optimism for each model. Finally, an optimism-corrected AUC estimate was calculated by subtracting the optimism average obtained during the seventh step from the apparent AUC value produced during the first step.

The optimism value for the main effects model (MnSOST-3.1) was .022, which is slightly lower than that observed for the MnSOST-3 interaction model (.025). As a result, the optimism-corrected AUC values are .796 for both the interaction (MnSOST-3) and main effects (MnSOST-3.1) models. The results suggest that removing the interaction terms produced a more stable model. The optimism-corrected AUC value of .796 for the MnSOST-3.1 provides an unbiased estimate that adjusts for overfitting. It may also represent an upper-level estimate as to what may be expected in validation studies on non-Minnesota sex offenders.

Table 2. MnSOST-3.1 and MnSOST-3 Predictive Discrimination Across Samples

Sample	AUC	Lower Bound	Upper Bound	N
MnSOST-3.1				
Full Sample	.818	.776	.861	2,535
Contemporary Sample	.824	.777	.871	2,315
MnSOST-R Cross-Validation	.789	.697	.882	220
MnSOST-R Development	.749	.682	.816	243
Rapist	.726	.634	.817	140
Molester	.781	.686	.877	103
MnSOST-3				
Full Sample	.821	.777	.865	2,535
Contemporary Sample	.824	.772	.875	2,315
MnSOST-R Cross-Validation	.792	.700	.884	220
MnSOST-R Development	.752	.686	.819	243
Rapist	.733	.642	.824	140
Molester	.781	.686	.876	103

In examining the predictive discrimination of the MnSOST-3.1, it is worth comparing its performance not only among several different samples but also with the MnSOST-R. For the offenders released from prison between 2003 and 2006 (contemporary sample), the AUC was .824 compared to .550 for the MnSOST-R. For the cross-validation sample, the MnSOST-3.1 had an AUC value of 0.789 in comparison to .758 for the MnSOST-R. As noted by Epperson et al. (2003), the MnSOST-R development sample contained 256 sex offenders released from prison during the late 1980s and early 1990s. Yet, because the data needed to fully score the MnSOST-3.1 were unavailable for 13 offenders in the MnSOST-R development sample, the analyses were limited to the remaining 243 offenders. The AUC value for the MnSOST-R was .758 compared to .749 for the MnSOST-3.1, a difference that was not statistically significant ($p = 0.817$) using the DeLong, DeLong, and Clarke-Pearson (1988) method for ROC curve comparison. The AUC values for the MnSOST-R

development and cross-validation samples are not the same as those reported by Epperson and colleagues (2003) due to the different definition of sexual recidivism used here; i.e., Epperson et al. (2003) defined sexual recidivism as a new sex offense rearrest within six years. Overall, the findings suggest that while the MnSOST-3.1 has higher predictive accuracy for offenders recently released from prison who are subject to significant external constraints, it does not perform significantly worse than the MnSOST-R for offenders released from prison more than 20 years ago who were exposed to relatively few external constraints.

In Table 2, a closer look is taken at the predictive discrimination of the MnSOST-3.1 on the MnSOST-R development sample. Epperson and colleagues (2003) distinguished the offenders in the MnSOST-R development sample on the basis of whether they were rapists or molesters. Of the 243 offenders from the MnSOST-R development sample who were examined in this study, 140 had been classified as rapists and the other 103 as molesters. The AUC values for the MnSOST-3.1 were lower for rapists (0.733) but higher for child molesters (0.781).

In Table 3, additional performance measures are presented for the MnSOST-3.1. The top one percent of offenders had a MnSOST-3.1 value of 40 percent or higher. In other words, only 25 of the 2,535 offenders (i.e., the top one percent) had a predicted probability of sexual recidivism (within four years) of 40 percent or higher. The top five percent had a value of 14 percent or higher, the top 10 percent had a value of 8 percent or higher, and the top 15 percent had a value of 5.8 percent or higher. Among the 262 offenders with a MnSOST-3.1 value of 8 percent or higher (the top 10 percent), there were 53 who were recidivists, which amounts to a reconviction rate of 20 percent. Considering the sexual

recidivism rate was 4 percent for the sample, the reconviction rate for the top 10 percent is

five times greater than the overall rate. For every true positive (i.e., recidivist) identified at

the 8 percent cut point, there were nearly four false positives (non-recidivists). Because there

were a total of 102 recidivists, the 53 recidivists with MnSOST-3.1 values of 8 percent or

higher accounted for 52 percent (capture rate) of the total recidivists.

Table 3. MnSOST-3/3.1 Performance Metrics

MnSOST-3/3.1 Values (Percentile)	N	Sex Crime Reconvictions	Reconviction Rate	Risk Ratio	True-False Positive Ratio	Capture Rate
MnSOST-3.1						
>= 40% (Top 1%)	21	8	38.1 %	9.5	1.6	7.8 %
>= 25% (Top 2%)	51	20	40.0 %	9.9	1.5	19.6 %
>= 14% (Top 5%)	124	33	26.7 %	6.6	2.8	32.4 %
>= 8.0% (Top 10%)	262	53	20.2 %	5.0	3.9	52.0 %
>= 5.8% (Top 15%)	381	61	16.0 %	4.0	5.2	59.8 %
>= 2.3% & < 5.8% (16-45%)	760	25	3.3 %	0.8	29.4	24.5 %
< 2.3% (Bottom 55 %)	1,394	16	1.2 %	0.3	86.1	15.7 %
MnSOST-3						
>= 40% (Top 1%)	25	11	44.0 %	10.9	1.3	10.8 %
>= 25% (Top 2%)	51	22	44.0 %	10.7	1.3	21.6 %
>= 13.5% (Top 5%)	127	35	27.6 %	6.9	2.6	33.5 %
>= 8.0% (Top 10%)	251	55	21.9 %	5.4	3.6	53.9 %
>= 5.5% (Top 15%)	380	65	17.1 %	4.3	4.8	63.7 %
>= 2.5% & < 5.5% (16-40%)	637	17	2.7 %	0.7	36.5	9.7 %
< 2.5% (Bottom 59 %)	1,518	20	1.3 %	0.3	74.9	11.4 %

As noted earlier, this study did not examine 134 sex offenders released from

Minnesota prisons between 2003 and 2006 because they were civilly committed. Still, to

further test the validity of the MnSOST-3.1, MnSOST-3.1 values for these offenders were

generated. The average MnSOST-3.1 value for the 134 civilly committed offenders was 10.1

percent, which is 2.5 times higher than the overall average. One of the criteria for civil

commitment or sexually violent predator (SVP) decisions is the determination that the

offender is either "substantially likely" or "more likely than not" to reoffend sexually, which

roughly translates into a probability of 51 percent or higher. Only four of the offenders (three percent), however, had a MnSOST-3.1 value greater than 50 percent, and only nine (seven percent) had an upper 95 percent confidence interval (CI) that exceeded 50 percent. Moreover, as noted above, less than one percent of the 2,535 offenders had a MnSOST-3.1 value that exceeded 50 percent, which is substantially lower than the rate (seven percent) at which Minnesota sex offenders have been civilly committed over the last few decades. These findings should not be considered too surprising, however, given that a recent report on Minnesota's civil commitment program found that county of commitment, which is unrelated to sexual recidivism risk, was a significant factor in determining whether sex offenders were civilly committed (civil commitment decisions are finalized at the county level in Minnesota) (Minnesota Office of the Legislative Auditor, 2011).

Calibration

In Table 4, data are presented on the distribution of MnSOST-3.1 values and the corresponding 95 percent CIs. Although the predicted probabilities from a logistic regression model can vary from 0-100 percent, the MnSOST-3.1 values for the 2,535 offenders ranged from a low of 0 percent to a high of 98 percent. Only 0.5 percent of the sample, or 13 offenders, had a MnSOST-3.1 value of 50 percent or higher, whereas a little more than one percent (N = 33) had an upper CI at or above 50 percent. Two percent of the sample had a MnSOST-3.1 value of 25 percent or higher, while nearly eight percent had a value of 10 percent or higher. Nearly half of the sample (46 percent) had a MnSOST-3.1 value below 2 percent, while roughly one-fifth (19 percent) had a value below 1 percent. Overall, 77 percent had a value below four percent, which was the sexual recidivism rate observed among the 2,535 sex offenders.

Table 4. Distribution of MnSOST-3/3.1 Values and 95 Percent Confidence Intervals

MnSOST-3/3.1 Value	Lower 95% CI	Upper 95% CI	N ≥ to MnSOST-3.1 Value	% of Sample (N = 2,535)
MnSOST-3.1				
50%	29%	77%	13	0.5%
40%	21%	62%	21	0.8%
33%	17%	50%	33	1.3%
30%	16%	48%	37	1.5%
25%	14%	39%	50	2.0%
20%	11%	34%	70	2.8%
15%	8%	26%	111	4.4%
10%	5%	16%	192	7.6%
5%	3%	8%	458	18.1%
4%	3%	7%	578	22.8%
3%	2%	5%	837	33.0%
2%	1%	3%	1,374	54.2%
1%	<1%	2%	2,062	81.3%
MnSOST-3				
50%	22%	77%	19	0.7%
40%	21%	66%	25	1.0%
34%	19%	50%	34	1.3%
30%	17%	47%	38	1.4%
25%	15%	41%	50	2.0%
20%	11%	34%	77	3.0%
15%	8%	27%	116	4.6%
10%	6%	18%	184	7.3%
5%	3%	8%	431	17.0%
4%	2%	7%	566	22.3%
3%	2%	5%	807	31.8%
2%	1%	3%	1,374	54.2%
1%	<1%	2%	2,014	79.4%

Calibration of the MnSOST-3.1 was first assessed by estimating a Hosmer-Lemeshow test in which MnSOST-3.1 values were regressed on sexual recidivism. The test was statistically significant at the .05 level, which suggests the MnSOST-3.1 is not well calibrated with the observed rates of sexual reoffending in the sample. Yet, because the Hosmer-Lemeshow test is sensitive to sample size, a statistically significant test does not

necessarily mean the MnSOST-3.1 model is not well calibrated due to the large sample size

used here.

Table 5. Calibration between Actual Recidivism Rates and MnSOST-3/3.1 Values

MnSOST-3/3.1	Actual Rate	Avg. MnSOST-3.1 Value	N
MnSOST-3.1			
60% or higher	55.6%	73.5%	9
40-59%	25.0%	47.8%	12
30-39%	31.3%	35.7%	16
20-29%	36.4%	24.1%	33
15-19%	19.5%	17.6%	41
10-14%	9.9%	12.1%	81
5.0-9.9%	8.3%	7.0%	266
4.0-4.9%	6.7%	4.4%	120
3.0-3.9%	3.2%	3.5%	252
2.0-2.9%	2.6%	2.4%	544
1.0-1.9%	1.0%	1.5%	688
Less than 1%	0.4%	0.6%	473
Total	4.0%	4.0%	2,535
MnSOST-3			
60% or higher	57.1%	72.3%	14
40-59%	27.3%	48.2%	11
30-39%	33.3%	35.7%	12
20-29%	30.0%	23.8%	40
15-19%	15.4%	17.4%	39
10-14%	17.6%	11.9%	68
5.0-9.9%	8.9%	7.0%	247
4.0-4.9%	3.0%	4.5%	135
3.0-3.9%	2.9%	3.5%	241
2.0-2.9%	2.3%	2.4%	567
1.0-1.9%	1.2%	1.5%	640
Less than 1%	0.6%	0.6%	521
Total	4.0%	4.0%	2,535

To better understand the results from the Hosmer-Lemeshow test, data are presented

in Table 5 that compare average MnSOST-3.1 values and observed sexual recidivism rates

among the offenders in the sample according to 13 discrete categories of MnSOST-3.1

values. Given that the average MnSOST-3.1 values are higher than observed sexual

recidivism rates for offenders with a score of 40 percent or higher, the results suggest the

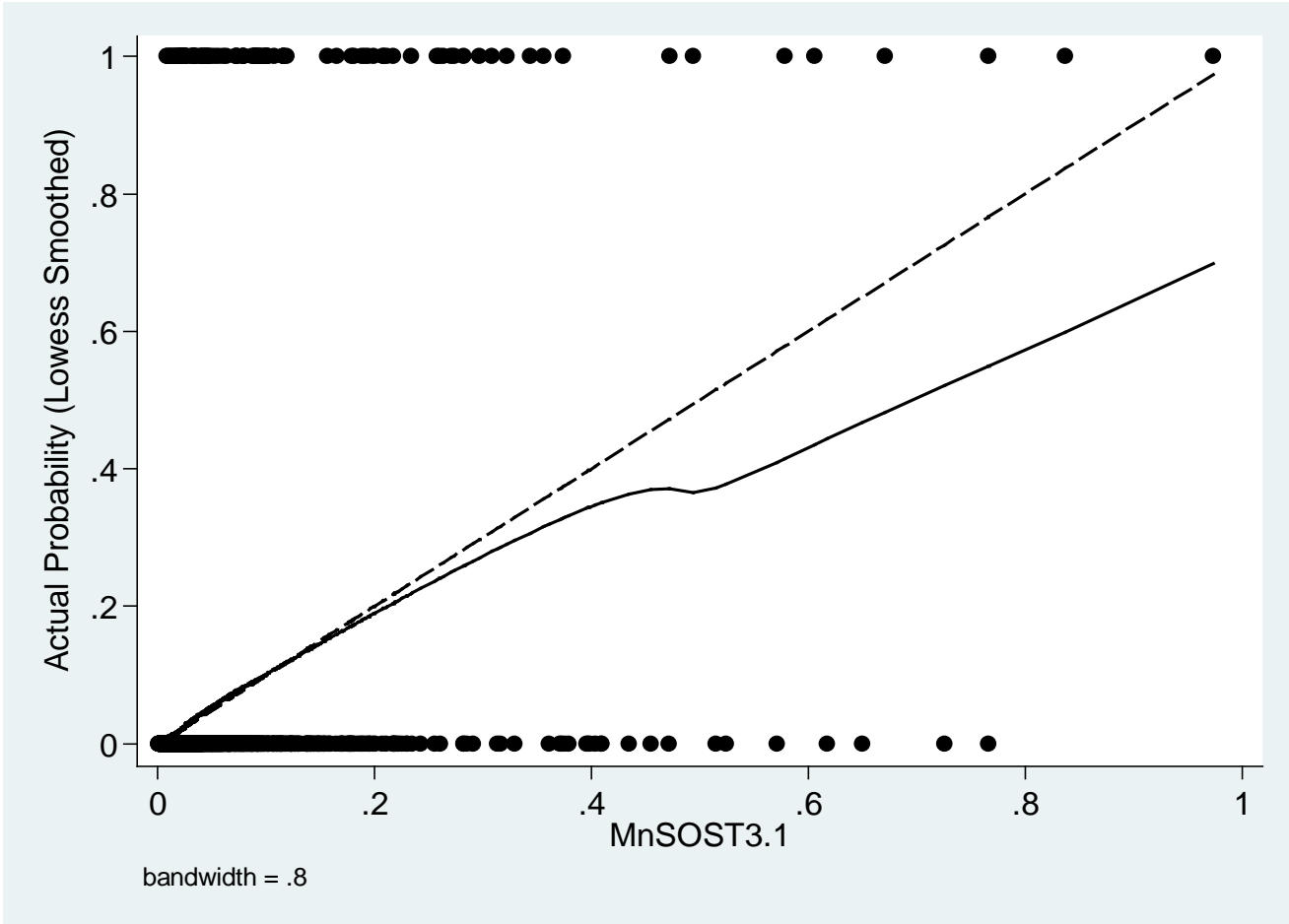MnSOST-3.1 may overestimate risk for the highest-risk offenders.

Figure 1. Lowess Plot for MnSOST-3.1 Values and Observed Sexual Recidivism

The calibration patterns depicted in Table 5 are further illustrated in the Lowess plot, which is shown in Figure 1. Whereas the dotted line represents actual rates of sexual reoffending among the 2,535 sex offenders, the solid line denotes the predicted probabilities (i.e., MnSOST-3.1 values) derived from the logistic regression model. The plot indicates a relatively tight correspondence between actual recidivism rates and predicted probabilities for offenders with MnSOST-3.1 values less than 40 percent, which suggests the MnSOST-

3.1 appears to be well-calibrated with actual sexual recidivism rates for roughly 99 percent of the sample. The two lines begin to diverge, however, at the 40 percent mark. Because the solid line is below the dotted line for MnSOST-3.1 values greater than 40 percent, the plot indicates the MnSOST-3.1 overestimates sexual recidivism risk for the top one percent of offenders in the sample.

Calibration was further assessed by estimating a Brier score, which is a quadratic scoring rule that calculates the squared differences between observed and predicted values. The results indicate a Brier score that was close to zero (0.0339) and a Spiegelhalter's z-score (0.3480) that was not statistically significant (p = 0.3639). Therefore, despite the lack of calibration at the top end of the model, the results suggest sufficient overall calibration between predicted and observed outcomes.

RELIABILITY

In the MnSOST-3 study (Duwe and Freske, 2012), the reliability of scoring the instrument was examined by conducting an inter-rater reliability assessment. Given that the nine main effects are the same in both models, the results are identical to those presented in the MnSOST-3 study except for the total score. For the inter-rater reliability assessment, 20 sex offenders were randomly selected who were released from Minnesota prisons between January 1 and June 30, 2010, on whom a MnSOST-R had been scored. Following a four-hour training session, eight assessors in the MnDOC's RACN Unit each scored the selected cases on the Microsoft Excel application of the MnSOST-3 over a five-day period. An age at release calculator was included on the Excel spreadsheet so as to facilitate the valid and reliable scoring of data for this item. The eight raters in this study had, on average, seven years of experience in scoring sex offender risk assessment instruments. The degree of inter-

rater reliability was analyzed among the eight assessors for these 20 cases by estimating

intraclass correlation coefficients (ICC) using a two-way random effects model.

Table 6. MnSOST-3/3.1 Inter-Rater Reliability Assessment

MnSOST-3/3.1 Items	Consistency			Absolute Agreement		
	ICC	Lower Bound	Upper Bound	ICC	Lower Bound	Upper Bound
MnSOST-3.1 Score	.810	.694	.905	.811	.696	.905
MnSOST-3 Score	.826	.718	.914	.826	.718	.914
Predatory Offense Sentences	.826	.718	.914	.826	.718	.914
Male Victims	.793	.670	.895	.797	.676	.898
Public Place	.928	.875	.966	.929	.876	.966
Felony Offense Sentences	.938	.892	.971	.935	.887	.970
VOFP/Stalking/Harassment	.671	.514	.822	.669	.513	.821
Disorderly Conduct/3 Years	.767	.635	.881	.771	.641	.883
SO/CD Treatment	.796	.674	.897	.798	.677	.898
Release Age (Years)	1.000	1.000	1.000	1.000	.999	1.000
Unsupervised Release	.869	.781	.936	.860	.766	.932

ICC = Intraclass Correlation Coefficient
All coefficients were statistically significant at the .01 level

The results showed that the singular ICC for the eight raters was .810 and .811 for

consistency and absolute agreement of ratings, respectively, for the 20 cases (see Table 6).

These values are slightly lower than the ICC values observed for the MnSOST-3 (.826 for

both consistency and absolute agreement). The item-level data show that ratings were most

consistent for age at release, which may be due in part to the creation of a calculator for this

item. The ratings were least consistent, however, for VOFP/stalking/harassment sentences.

Although most (6) of the items on both versions of the MnSOST-3 are continuous (as

opposed to binary or dichotomous) measures, which presumably increases the margin for

error in scoring items, the items on both instruments are largely objective measures. Overall,

the findings suggest both versions of the MnSOST-3 have an adequate degree of reliability.

DISCUSSION

In February 2012, the MnDOC began using the MnSOST-3.1 because, compared to

the MnSOSST-3, it is simpler, easier to interpret, and does not have an adverse impact on

predictive accuracy. Yet, due to the same optimism-corrected AUC values, both versions of the MnSOST-3 will still be available to the public on the MnDOC website (http://www.doc.state.mn.us/publications/MnSOST3/default.htm). Use of the MnSOST-3.1, however, will result in several modest changes relating to cut scores and scoring the instrument.

In Minnesota, one of the main purposes for using the MnSOST-3.1 is to assess sexual recidivism risk for community notification. Since the inception of the Community Notification Act in 1997, Minnesota has used a tiered risk management system in which the level of community notification is based on the offender's predicted risk of sexual recidivism. Sex offenders with a high predicted risk of sexual recidivism are given the most extensive level (Level 3) of notification (i.e., community meetings held by law enforcement, publication of the offender's photograph and offense description on the MnDOC's website, etc.), whereas those with lower risk (Levels 1 and 2) are given more limited forms of notification. Because the MnSOST-3.1 is used by End of Confinement Review Committees (ECRC) within the MnDOC to determine risk levels for offenders, it anchors Minnesota's tiered risk management system. Yet, because the ECRC considers additional factors that may either increase or decrease the risk of reoffense (e.g., an offender's stated intention to reoffend following release or a debilitating illness or physical condition that mitigates the risk of reoffense), the ECRC may override the risk level suggested by the MnSOST-3.1. As a result, the risk levels implied by either version of the MnSOST-3 are considered presumptive.

Historically, the MnDOC has given Level 3 assignments to approximately 15 percent of released sex offenders, Level 2 assignments to about 30 percent of offenders, and Level 1

assignments to the remaining 55 percent. Cut scores for presumptive risk levels within

Minnesota are therefore based on this distribution. Under the MnSOST-3, the cut scores for

presumptive risk levels are:

Level 3 = 5.50 percent or higher

Level 2 = 2.30 – 5.49 percent

Level 1 = 2.29 percent or lower

The 5.5 percent MnSOST-3 value was selected as the cut score for a presumptive

Level 3 assignment because this threshold represented the 85th percentile. In other words, sex

offenders with MnSOST-3 values of 5.5 percent or higher constitute the top 15 percent with

respect to predicted likelihood of sexual recidivism in four years. The values of 2.30 and 5.49

percent were selected as the cut scores for the presumptive Level 2 assignment range because

they represent the 55th and 84th percentiles, respectively. And the value of 2.29 percent was

the presumptive Level 1 cut score selected because MnSOST-3 values below 2.30 percent

comprise the bottom 55 percent in terms of sexual recidivism risk.

Under the MnSOST-3.1, the Level 1 cut score remains the same, although there is a

slight modification for the Level 2 and 3 cut scores. The MnSOST-3.1 cut scores for

presumptive risk levels are:

Level 3 = 5.80 percent or higher

Level 2 = 2.30 – 5.79 percent

Level 1 = 2.29 percent or lower

For offenders without VOFP and/or disorderly conduct convictions, scores from the

two versions of the MnSOST-3 will vary only slightly. The major differences in scores for

the two instruments will be observed among offenders who have VOFP and/or disorderly

conduct convictions. Under the MnSOST-3.1, VOFP and disorderly conduct convictions will increase risk regardless of the offender's age at the time of release.

Even though the MnDOC will be using the MnSOST-3.1, efforts will be made to better understand the interaction findings from the MnSOST-3 and further assess the impact these interactions have on assessments of risk. In particular, by estimating the impact of items such as VOFP and disorderly conduct convictions on other types of recidivism besides sexual reoffending, it is anticipated that work on the global risk-assessment tool may help decrease the extent to which the interaction findings appear to be counter-intuitive.

Data will also be collected on offenders scored on the MnSOST-3.1 to determine what their scores would have been on the MnSOST-3. It is expected that analyzing data on a larger number of cases, especially those scored recently, will enable the MnDOC to more fully comprehend the behavior of the interaction terms in the MnSOST-3. And the MnDOC intends to share the results of these analyses with the broader corrections and forensic communities when they become available.

Although the MnSOST-3.1 offers a modest improvement in several ways over the MnSOST-3, the caveats raised about using the MnSOST-3 on non-Minnesota sex offender populations also apply to the MnSOST-3.1. For example, with either version of the MnSOST-3, no attempt was made to specifically develop a widely applicable instrument. As a result, the relatively high predictive accuracy of either version of the MnSOST-3 may not generalize to sex offender populations in other jurisdictions. After all, Minnesota is, in several potentially important ways, different from the rest of the United States. Even though Minnesota is, compared to the other 49 states, generally in the middle of the pack for population size and crime rate, it has the second lowest incarceration rate in the nation

(Guerino, Harrison, and Sabol, 2011). Because Minnesota relies more heavily on local sanctions (e.g., jail and community supervision), prison beds are generally reserved for offenders who have committed very serious offenses and/or have lengthy criminal histories. Further, unless offenders receive extended incarceration disciplinary time, prisoners in Minnesota are typically released after serving two-thirds of their sentence. This may make release at expiration of sentence less common in Minnesota relative to some other jurisdictions.

Use of either version of the MnSOST-3 outside of Minnesota may also be limited by the level of data needed to accurately score the instrument. In particular, given that six of the nine items relate, in some form or another, to criminal history (both sexual and non-sexual), access to complete and accurate criminal history data is imperative. The instrument would therefore have diminished value for agencies that have limited access to these data or in jurisdictions where the criminal history data are less than complete. In addition, although it is anticipated that the items included on the instrument would likely be significant predictors of sexual recidivism for populations of non-Minnesota sex offenders, the weights (i.e., coefficient values) applied to these items are less likely to generalize to other populations.

These limitations notwithstanding, the relatively high optimism-corrected AUC for either version of the MnSOST-3 suggests it still may be among the better risk scales even if there is reduction in its predictive accuracy for other sex offender populations. Nevertheless, determining the extent to which the instrument is generalizable to non-Minnesota sex offender populations ultimately depends on the completion of validation studies. Accordingly, it is recommended that jurisdictions outside Minnesota consider using either version of the MnSOST-3 alongside externally validated risk assessment instruments (e.g.,

Static-99/Static-99R, Static-2002/Static-2002R, SORAG, MnSOST-R, etc.) until results from validation studies are available.

Given that the MnSOST-3.1 sample contains prisoners whose index offenses included both sexual and non-sexual crimes, the instrument can be used to assess post-release sexual recidivism risk for offenders who have at least one documented sex offense in their history regardless of whether their index offense is a sex crime. The sample also included 53 intrafamilial fondlers, 99 offenders whose only sex offense conviction(s) occurred as a juvenile, and 12 child pornography offenders—a group that has expanded in size over the last decade (Wolak, Finkelhor, & Mitchell, 2011). Due to these relatively small numbers, it is recommended that a great deal of caution be exercised in using the instrument on sex offenders who fall into one of these three groups. Again, it is anticipated that external validation studies will help reveal the extent to which the instrument has predictive validity for these groups of offenders.

In an effort to facilitate the completion of validation studies and the use of the instrument in other jurisdictions, this report provides descriptions of how the nine individual items were coded in the Appendix. Moreover, a more detailed coding manual has been prepared and the instrument has been developed so that it can be scored in a Microsoft Excel spreadsheet. Both the coding manual and the Microsoft Excel applications of the MnSOST-3 and 3.1 can be found at: http://www.doc.state.mn.us/publications/MnSOST3/default.htm.

APPENDIX

The following lists the nine items on the MnSOST-3.1 and describes how they were measured. The coding manual for the MnSOST-3.1, which provides a more complete description of these items, can be downloaded here:

http://www.doc.state.mn.us/publications/MnSOST3/default.htm.

Predatory Offenses: this item, which closely corresponds with the offenses that trigger predatory offense registration in Minnesota, measures the number of predatory offense sentences for which an offender has been convicted, including the index offense(s), up to a maximum of 25. Predatory offenses include all criminal sexual conduct crimes (1st-6th degree), murder in the first degree committed while the offender was committing (or attempting to commit) a criminal sexual conduct offense, kidnapping, false imprisonment (if the victim was not the minor dependent of the offender), indecent exposure, soliciting a minor to engage in prostitution, soliciting a minor to engage in sexual conduct, using a minor in a sexual performance, possession of child pornography, and incest.

Male Victims: this item measures the number of predatory offense sentences, as defined above, committed in which a male was the victim or one of the victims, up to a maximum of four sentences.

Public Place: similar to the MnSOST-R, this item measures whether any sexual activity with any sex offense was committed in a public place, which is defined as an area maintained for, or used by, the people or community, or an area open to the scrutiny of others (Epperson et al., 2005). Offenders who have committed a sex offense, charged or convicted, in a public place received a value of "1," whereas those who did not received a value of "0."

Felony Offenses: this item measures the total number of felony-level offense sentences for which an offender has been convicted, including the index offense(s). The value entered for felony sentences is the total number of felony sentences, whether juvenile or adult, index or prior, predatory or non-predatory, up to a maximum of 20.

VOFP/Stalking/Harassment: this item measures the total number of sentences (adult or juvenile; index or prior; petty misdemeanor, misdemeanor, gross misdemeanor, or felony) an offender has for stalking, harassment, or violations of orders for protection, up to a maximum of five sentences.

Recent Disorderly Conduct: this item measures the number of sentences (adult or juvenile; index or prior; petty misdemeanor, misdemeanor, gross misdemeanor, or felony) an offender has for disorderly conduct convictions in the three years preceding his most recent commitment to prison, up to a maximum of two.

Completion of Sex Offender and Chemical Dependency Treatment: this item measures whether offenders have completed both sex offender and chemical dependency treatment while in prison for the index offense. Treatment completions are not included here if they occurred during a prior prison sentences or if the offender has received a new treatment directive. Offenders who complete prison-based sex offender and chemical dependency treatment while incarcerated for the index offense(s) receive a value of "1," whereas those who do not complete both types of treatment are given a value of "0."

Age at Release: the age of the offender in years at the time of release based on the date of birth and release date.

Unsupervised Release: this item measures whether offenders are released to correctional supervision. Offenders who are released to no supervision (i.e., discharged) receive a value

33

of "1", whereas offenders released to some form of correctional supervision receive a value of "0".

REFERENCES

Boccaccini, M.T., Murrie, D.C., Caperton, J.D., & Hawes, S.W. (2009). Field validity of the Static-99 and MnSOST-R among sex offenders evaluated for civil commitment as sexually violent predators. *Psychology, Public Policy and Law*, 15, 278-314.

Cooke, C.R., Shah, C.V., Gallop, R., Bellamy, S., Ancukiewicz, M., Eisner, M.D., Lanken, P.N., Localio, A.R., & Lanken, P.N. (2009). A simple clinical predictive index for objective estimates of mortality in acute lung injury. *Critical Care Medicine* 37(6): 1,913-1,920.

DeLong, E.R., DeLong, D.M., & Clarke-Pearson, D.L. (1988). Comparing the areas under two or more correlated receive operating characteristic curves: A nonparametric approach *Biometrics*, 44, 837-845.

Duwe, G. (2010). Prison-based chemical dependency treatment in Minnesota: An outcome evaluation. *The Journal of Experimental Criminology*, 6, 57-81.

Duwe, G. & Freske, P. (2012). Using logistic regression modeling to predict sex offense recidivism: The Minnesota Sex Offender Screening Tool-3 (MnSOST-3). *Sexual Abuse: A Journal of Research and Treatment* DOI: 10.1177/1079063211429470.

Duwe, G. & Goldman, R. (2009). The impact of prison-based treatment on sex offender recidivism: Evidence from Minnesota. *Sexual Abuse: A Journal of Research and Treatment*, 21, 279-307.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7, 1-26.

Efron, B. & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *American Statistician*, 37, 36-48.

Efron, B. & Tibshirani, R. (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Epperson, D.L., Kaul, J.D., Huot, S., Goldman, R. & Alexander, W. (2003). *Minnesota Sex Offender Screening Tool-Revised (MnSOST-R) Technical Paper: Development, Validation, and Recommended Risk Level Cut Scores*. Retrieved from: http://www.psychology.iastate.edu/~dle/TechUpdatePaper12-03.pdf.

Guerino, P., Harrison, P.M., & Sabol, W.J. (2011). *Prisoners in 2010*. Washington, DC: Bureau of Justice Statistics, U.S. Department of Justice.

Hanson, R.K. & Morton-Bourgon, K. (2004). *Predictors of Sexual Recidivism: An Updated Meta-Analysis* (User Report No. 2004-02). Ottawa, Canada: Department of the Solicitor General of Canada.

Harrell, F.E. (2001). *Regression Modeling Strategies with Application to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer Verlag.

Harrell, F.E., Lee, K.L., & Marks, D. B. (1996). Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, 15, 361-387.

Judd, C.M., McClelland, G.H., & Ryan, C.S. (2008). *Data Analysis: A Model Comparison Approach*. New York: Routledge.

Lösel, F. & Schmucker, M. (2005). The effectiveness of treatment for sexual offenders: A comprehensive meta-analysis. *Journal of Experimental Criminology*, 1, 117-146.

Minnesota Department of Corrections (2007). *Sex Offender Recidivism in Minnesota*. St. Paul, MN.

Minnesota Office of the Legislative Auditor (2011). *Evaluation Report: Civil Commitment of Sex Offenders*. St. Paul, MN.

Mitchell, O., Wilson, D.B., & MacKenzie, D.L. (2007). Does incarceration-based drug treatment reduce recidivism? A meta-analytic synthesis of the literature. *The Journal of Experimental Criminology*, 3, 353-375.

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

Ostermann, M. (2009). An analysis of New Jersey's day reporting center and halfway back programs: Embracing the rehabilitative ideal through evidence based practices. *Journal of Offender Rehabilitation*, 48, 139-153.

Schlager, M. & Robbins, K. (2008). Does parole work? Revisited: Reframing the discussion of the impact of postprison supervision on offender outcome. *The Prison Journal*, 88, 234-251.

Steyerberg, E.W. (2009). *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. New York: Springer.

Steyerberg, E.W., Harrell, F.E., Borsboom, G.J.J.M., Eijkemans, M.J.C., Vergouwe, Y., & Habbema, J.D.F. (2001). Internal validation of predictive models: Efficiency of some

procedures for logistic regression analysis. *Journal of Clinical Epidemiology*, 54, 774-781.

Steyerberg, E.W., Bleeker, S.W., Moll, H.A., Grobbee, D.E., & Moons, K.G. (2003). Internal and external validation of predictive models: A simulation study of bias and precision in small samples. *Journal of Clinical Epidemiology*, 56, 441-447.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, 58, 267-288.

Vrieze, S.I., & Grove, W.M. (2008). Predicting sex offense recidivism. I. Correcting for item overselection and accuracy overestimation in scale development. II. Sampling error-induced attenuation of predictive validity over base rate information. *Law and Human Behavior*, 32, 279-291.

Wolak, J., Finkelhor, D., & Mitchell, K. (2011). Child pornography: Trends in offender and case characteristics. *Sexual Abuse: A Journal of Research and Treatment*, 23, 22-42.

Wollert, R. (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.

Wollert, R. (2006). Low base rates limit expert certainty when current actuarials are used to identify sexually violent predators: An application of Bayes's Theorem. *Psychology, Public Policy, and Law*, 12, 56-85.

Zhao, F. (1998). Bootstrap variable selection and model validation for Cox's proportional hazards regression models—with applications to the identification of factors predictive of overall and post-relapse survival in advance epithelial ovarian cancer. Master's thesis: Queen's University.