

The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:

Document Title: A Model of Static and Dynamic Sex Offender Risk Assessment

Author: Robert J. McGrath, Michael P. Lasher, Georgia F. Cumming

Document No.: 236217

Date Received: October 2011

Award Number: 2008-DD-BX-0013

This report has not been published by the U.S. Department of Justice. To provide better customer service, NCJRS has made this Federally-funded grant final report available electronically in addition to traditional paper copies.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

A MODEL OF SEX OFFENDER RISK ASSESSMENT

A Model of Static and Dynamic Sex Offender Risk Assessment

Grant Award Number: 2008-DD-BX-0013

National Institute of Justice

Robert J. McGrath, Michael P. Lasher, and Georgia F. Cumming

Vermont Department of Corrections

Abstract

The purpose of the present study was to test models of combining static and dynamic risk measures that might predict sexual recidivism among adult male sex offenders better than any one type of measure alone. Study participants were 759 adult male sex offenders under correctional supervision in Vermont who were enrolled in community sex offender treatment between 2001 and 2007. These offenders were assessed once using static measures (Static-99R, Static-2002R and VASOR) based on participants' history at the date of placement in the community. A 22-item dynamic risk measure (SOTNPS) was used multiple times to assess participants, shortly after their entry into community treatment and approximately every six months thereafter. Analyses of SOTNPS scores resulted in the development of a new 16-item dynamic risk measure, the Sex Offender Treatment Intervention and Progress Scale (SOTIPS). At fixed one- and three-year follow-up periods from participants' initial, second, and third dynamic risk assessments, the SOTIPS and Static-99R, the static risk measure selected for further analysis in the present study, each independently showed moderate ability to rank order risk for sexual, violent, and any criminal recidivism and return to prison. A logistic regression model that combined SOTIPS and Static-99R consistently predicted recidivism and outperformed either instrument alone when both instruments had similar predictive power. Participants who demonstrated treatment progress, as reflected by reductions in SOTIPS scores, showed lower rates of recidivism than those who did not.

Keywords: Sex Offender Treatment Intervention and Progress Scale (SOTIPS), dynamic risk assessment, recidivism, treatment change, Static-99R

Table of Contents

Abstract	ii
Table of Contents	iii
Executive Summary	1
I. Introduction	14
A. Statement of the Problem	14
B. Literature Review	15
C. Statement of Hypotheses	21
II. Methods	21
A. Setting	21
B. Program	22
C. Participants	23
D. Measures	24
E. Outcome Measures	27
F. Procedure and Data Analyses	28
III. Results	33
IV. Conclusions	39
A. Discussion of Findings	39
B. Implications for Policy and Practice	46
C. Implications for Further Research	48
V. References	50
VI. Authors' Note	59
VII. Dissemination of Research findings	60

VIII. Tables	61
Table 1. Descriptive and Interrater Reliability Statistics for Static Risk Assessment Instruments	61
Table 2. AUC Values for Static Risk Assessment Instruments	62
Table 3. Predictive Accuracy of SOTIPS for Sexual Recidivism	63
Table 4. Predictive Accuracy of SOTIPS for Violent Recidivism	64
Table 5. Predictive Accuracy of SOTIPS for Any Recidivism	65
Table 6. Predictive Accuracy of SOTIPS for Return to Prison	66
Table 7. Factor Loadings for Exploratory Factor Analysis with Varimax Rotation of SOTIPS Items	67
Table 8. Interrater Reliability for Pooled Samples	68
Table 9. Repeated Measures ANOVAs of SOTIPS Items and Total Scores for Sexual Non-recidivists and Recidivists	69
Table 10. Sexual Recidivism by SOTIPS Risk Category and Score Time Period	70
Table 11. Violent Recidivism by SOTIPS Risk Category and Score Time Period	71
Table 12. Any Criminal Recidivism by SOTIPS risk category and Score Time Period	72
Table 13. Return to Prison Rates by SOTIPS Risk Category and Score Time Period	73
Table 14. Predictive Accuracy of Static-99R, SOTIPS and Combined Scores	74
Table 15. GEE Model Effects and Parameter Estimates for Sexual Recidivism	75
Table 16. GEE Model Effects and Parameter Estimates for Violent Recidivism	76
Table 17. GEE Model Effects and Parameter Estimates for Any Recidivism	77

Table 18. GEE Model Effects and Parameter Estimates for Return to Prison	78
Table 19. Observed and Predicted Sexual Recidivism Rates for Combined SOTIPS and Static-99 Risk Levels	79
Table 20. Observed and Predicted Violent Recidivism Rates for Combined SOTIPS and Static-99 Risk Levels	80
Table 21. Observed and Predicted Any Recidivism Rates for Combined SOTIPS and Static-99 Risk Levels	81
Table 22. Observed and Predicted Return to Prison Recidivism Rates of Combined SOTIPS and Static-99 Risk Levels	82
Table 23. GEE Predicted Sexual Recidivism Rates for Combined SOTIPS and Static-99R Risk Levels Across Times 1, 2, and 3	83
Table 24. GEE Predicted Violent Recidivism Rates for Combined SOTIPS and Static-99R Risk Levels Across Times 1, 2, and 3	84
Table 25. GEE Predicted Any Recidivism Rates for Combined SOTIPS and Static-99R Risk Levels Across Times 1, 2, and 3	85
Table 26. GEE Predicted Return to Prison Rates for Combined SOTIPS and Static-99R Risk Levels Across Times 1, 2, and 3	86
Table 27. Labels for Combined SOTIPS and Static-99R Risk Categories	87
Table 28. Recidivism Rates for Combined Static-99R and SOTIPS GEE risk levels	88
IX. Figures	89
Figure 1. Sexual Recidivism Rates at 3 Years by Combined Static-99R and SOTIPS GEE Risk Levels	89
Figure 2. Interval Censored Survival Curves for Sexual, Violent, and Any Recidivism and Return to Prison Recidivism	90

Executive Summary

Sexual offenses have serious negative consequences to victims, their families and communities. Over the last two decades, these crimes have received particularly heightened public and legislative attention. Legislation targeting sex offenders has recently spanned the areas of sentencing, registration, community notification, residency restrictions, civil confinement, electronic monitoring, supervision, and treatment.

Legislative and other interventions designed to protect the public from sex offenders, however, often have been implemented without regard to important differences among individuals who commit these crimes. Not all sex offenders are the same. Research evidence on effective correctional practice (Andrews & Bonta, 2010) stresses the value of moving beyond “one size fits all” approaches.

One of the most important ways in which sex offenders differ from one another concerns risk to reoffend. This is a critical issue because services that are matched to the risk level of offenders are more effective than those that are not (Hanson, Bourgon, Helmus, & Hodgson, 2009; Lovins, Lowenkamp, & Latessa, 2009). Higher-risk sex offenders should receive higher intensity services, and low-risk offenders should receive minimal or no services.

Because risk assessments can have a profound impact on community safety, offenders’ liberty and wise allocation of public resources, their accuracy is critically important. Although the accuracy of risk assessment methods has improved dramatically over the last two decades, considerable room for improvement still exists.

Advances in risk assessment can be marked in four major phases (Bonta & Wormith, 2008). First-generation risk-assessment approaches rest on unstructured professional judgment, which is not as accurate as structured methods (Grove, Zald, Lebow, Snitz, & Nelson, 2000;

Hanson & Morton-Bourgon, 2009). In the sex offender field, second-generation risk-assessment instruments began emerging in the late 1990's. These structured measures are composed primarily of static risk factors, that is, unchangeable aspects of an individual's history. More recently, researchers have developed third-generation instruments that combine static and relatively inclusive collections of dynamic risk factors. Because dynamic risk factors are potentially changeable, risk-needs instruments offer direction to providers about what offender problems to target in order to reduce risk to reoffend. Risk-needs instruments that fully integrate assessments with ongoing case planning are called fourth-generation risk assessments.

A relatively small number of integrated sex offender risk-needs models now exist in which dynamic risk factors inform rehabilitation and case-management efforts and have added incremental predictive validity to static risk factors. These include the integrated suite of instruments developed by Hanson and his associates, the Static-99, Stable-2007 and Acute-2007 (Hanson, Harris, Scott, & Helmus, 2007), Structured Risk Assessment (SRA) model (Thornton, 2002) and Violence Risk Scale-Sex Offender Version (VRS-SO; Olver, Wong, Nicholaichuk, & Gordon, 2007). In particular, studies using the VRS-SO have shown an association between positive treatment progress and reductions in recidivism (Beggs & Grace, 2011; Olver & Wong, 2011).

A limitation of studies examining integrated sex offender risk-need instruments is that they typically have used assessment paradigms that employ either just a single dynamic assessment (Thornton, 2002) or two assessments, one pre- and one post-service (Beggs & Grace, 2010, 2011; Olver & Wong, 2011). In the case of the one sex-offender study that assessed dynamic risk factors over more than two time periods, changes in dynamic risk scores were not associated with changes in recidivism (Hanson et al., 2007). Assessing dynamic risk at regular

intervals during the course of services may be important in helping providers adjust the intensity and duration of interventions to match individuals' changing risk and needs.

The present study used a repeated measure paradigm over three assessment periods with a sample of 759 adult male sex offenders enrolled in community treatment. Predictive accuracy for sexual, violent, and any criminal reoffending and return to prison was examined for initial static risk-assessment scores, and for dynamic risk-assessment scores at 1, 7, and 13 months after beginning treatment for follow-up periods of one and three years. Analyses were limited to testing for first-time recidivism for each of the four types of recidivism studied.

Hypotheses. Three hypotheses were tested. First, it was expected that one or more static risk measures (Static-99R, Static-2002R, and VASOR) would predict sexual recidivism with moderate accuracy in the sample. Second, a dynamic risk measure (SOTNPS) or a subset of risk factors contained in this measure would also predict sexual recidivism with moderate accuracy and be sensitive to the changes in dynamic risk over time. Third, a combined static and dynamic risk measure would predict sexual recidivism more accurately than either measure alone.

Method. The sample was drawn from adult male sex offenders enrolled in Vermont's statewide network of community sex offender treatment programs (Vermont Treatment Program for Sexual Abusers; VTPSA). In 2001, the VTPSA began requiring contracted community treatment providers to complete risk and needs assessments on all program admissions. The initial assessment included two static measures, the Static-99 (Hanson & Thornton, 2000) and Vermont Assessment of Sex Offender Risk (VASOR; McGrath & Hoke, 2001) and one dynamic needs measure, the Sex Offender Treatment Needs and Progress Scale (SOTNPS; McGrath & Cumming, 2001, 2003). Per contract, trained providers were directed to re-administer the SOTNPS needs assessment to clients in January and July of each year. The goal was for

treatment providers to use assessment results for treatment planning, provide copies to probation and parole officers to inform supervision, and forward copies to VTPSA program directors for quality assurance purposes.

The 759 adult male sex-offender participants in the study were under correctional supervision in Vermont and met four criteria. First, they were convicted of at least one sexual offense against an identifiable child or non-consenting adult victim (Category “A” sexual offense as defined in the Static-99 coding manual; Harris, Phoenix, Hanson, & Thornton, 2003). Second, participants entered a VTPSA sponsored community sex offender treatment program between 2001 and 2007. Third, the time period from when they began VTPSA-sponsored treatment to the end date of the study follow-up period on December 31, 2010 was at least three years. Fourth, their treatment provider evaluated them using the SOTNPS at least once during the study period and submitted the results to the researchers. Approximately 357 sex offenders placed on community supervision in Vermont between 2001 and 2007 did not meet study criteria because they did not attend treatment, or if they did, therapists did not submit assessments.

At the time of community placement, the men were 18 to 75 years old, with an average age of 34.2. Most (86.8%) were serving a sentence for their first sexual offense. Consistent with Vermont’s lack of racial diversity, 96.4% of participants were White. Most participants were employed full time (57.7%).

Using definitions established by the Association for the Treatment of Sexual Abusers (Gordon et al., 1998), the sample was composed of 137 rapists (18.1%), 59 non-contact sex offenders (7.8%), 111 incest offenders (14.6%), and 452 child molesters (59.6%).

Measures. Following an examination of the predictive validity of the three static risk measures used in the study (Static-99R, Static-2002R, and VASOR) and 22 dynamic risk factors

that comprised the SOTNPS, two measures were selected for use in developing an integrated assessment model. These were the Static-99R and SOTIPS, a new dynamic risk measure that was constructed during the course of the study. Descriptions of these two measures are as follows.

Sex Offender Treatment Intervention and Progress Scale (SOTIPS). The SOTIPS, a revised version of the 22-item Sex Offender Treatment Need and Progress Scale (SOTNPS; McGrath & Cumming, 2001, 2003), was used as the dynamic measure for model testing in the present study. The SOTIPS is a provider-administered rating scale and was constructed during the course of this study by extracting 16 items from the SOTNPS that were empirically related to sexual recidivism. Item definitions and scoring instructions for the retained 16 SOTIPS items remained unchanged. Scoring instructions direct providers to score clients at intake and thereafter every six months on a 4-point scale; *minimal to no need for improvement, some need for improvement, considerable need for improvement* and *very considerable need for improvement*. Total scores range from 0 to 48 points and are organized into three risk/need groups: low (0 to 10), moderate (11 to 20) and high (21 to 48).

Static-99R. The Static-99R is a 10-item actuarial instrument designed to assess the recidivism risk of adult males known to have committed at least one sexual offense (Helmus, Thornton, Hanson, & Babchishin, 2011) and was selected for model-testing in this study. In the present sample, its predictive accuracy was about the same as the Static-2002R and VASOR, but it has the advantage of having greater acceptance among clinicians and researchers in the field. Total scores range from -3 to 12 points and are organized into four risk groups; low (-3 to 1), moderate-low (2 to 3), moderate-high (4 to 5) and high (6 to 12) and predict sexual recidivism with moderate accuracy (Hanson & Morton-Bourgon, 2009).

Outcome Measures. Recidivism data was obtained and analyzed for each study participant for all new charges for sexual, violent (sexual or non-sexual violence), and any criminal offenses, and returns to prison. The definition of a new sexual offense included a charge for a violation of supervision conditions if the incident could have been charged as a criminal sexual offense. Charges were counted based on criminal record checks in the states where each participant was known to have resided during the study period. The Vermont DOC computer database was used to identify violation of supervision charges and returns to prison.

Procedure and Data Analyses. A trained and experienced master's level research assistant used DOC case files and databases to code demographic and offense characteristic information on each participant. The research assistant scored participants on three static risk instruments (Static-99R, Static-2002-R, and VASOR) based on their status as of the date of community placement, and when scores already existed, ensured their accuracy. A second rater, the first or third author of this report, independently scored the three static risk instruments on approximately every tenth case to assess interrater reliability. The research assistant also collated SOTNPS score sheets that treatment providers had previously completed during participants' attendance in treatment.

Recidivism was coded as a binary variable (yes or no recidivism, coded 1 or 0). Analyses were limited to testing for first time recidivism for sexual, violent, and any criminal recidivism and return to prison.

To test the first hypothesis, analyses examined the interrater reliability of the three static risk instruments used in the study. The area under the curve of the receiver operating characteristic (AUC) statistic was used to examine the ability of the scales to rank order

participants in terms of risk to reoffend at 3- and 5-year fixed follow-up periods from the dates that participants were placed in the community.

To test the second hypothesis, we used AUCs to examine the predictive accuracy of the 22 individual SOTNPS items and total scores of different combinations of these items for six assessment waves. The six assessment waves were 1- and 3-year follow-up periods following participants' SOTNPS scores at three assessment times, referred to as Times 1, 2, and 3. Time 1 SOTNPS scores were those that providers completed during the first 3 months ($n = 606$; $M = 1.0$ months; $SD = 1.1$) that a participant had been in treatment, Time 2 scores between 4 and 9 months ($n = 665$; $M = 6.7$; $SD = 1.6$) after a participant had been in treatment, and Time 3 scores between 10 and 15 months ($n = 620$; $M = 12.7$; $SD = 1.7$) in treatment.

Based on analyses of AUCs for the 22 SOTNPS items, the scale was revised. The revised scale consisted of 16 SOTNPS items and was renamed the Sex Offender Treatment Intervention and Progress Scale (SOTIPS). Analyses on the new scale examined its factor structure and interrater reliability, compared scores of sexual recidivists to non-recidivists, and set cut-off score ranges for risk-band groups.

To test the third hypothesis, analyses examined the predictive accuracy of the SOTIPS and Static-99R, individually and in combination, for the six assessment waves using the AUC statistic. For these analyses, we used categorized scores for both instruments (SOTIPS scores of low, moderate and high; Static-99R scores of low, moderate-low, moderate-high and high) to reduce the number of between-instrument interactions and increase the likelihood of yielding statistically and clinically useful results.

Using categorized SOTIPS and Static-99R scores, the two measures were combined into a total score for use in logistic regression analyses to test models for predicting reoffending. As

the nature of the SOTIPS is to provide repeated information on the same subject over time, we used the generalized estimating equations (GEE) approach developed by Liang and Zeger (1986) as a means of conducting repeated measures logistic regression analyses for the combined Static-99R+SOTIPS risk groups across the six assessment waves. The strength of GEE when examining repeated measures data is that it accounts for the correlations within subjects when conducting regression analyses and provides more simplified information for practical use and comparison over time. To create an overall predictive model, GEE analyses were conducted using a binary logistic regression model. Finally, we categorized GEE reoffense probabilities into four risk bands (low, moderate-low, moderate-high, and high) and entered them into a second set of GEE regression analyses using an interval-censored survival model (Finkelstein, 1986) to conduct repeated measures survival analyses.

Results. The overall recidivism rates for the entire sample of 759 participants at fixed 3-year follow-up periods from the dates of placement in the community were as follows: sexual, 4.6% (35); any violent, 8.6% (65); and any criminal recidivism, 23.1% (175); and return to prison, 40.6% (308). Subtracting days participants were in prison during the 3-year fixed follow-up period, time-at-risk in the community was an average of 33.6 months.

Participants' mean static risk instruments risk scores were all in the moderate-low risk range, and interrater reliabilities, assessed on approximately every tenth case with intraclass correlation coefficients (ICC), were very good. Participants' mean SOTIPS scores at Times 1, 2, and 3 were all in the moderate range. Interrater reliability of SOTIPS total, factor, and individual scores was acceptable and was based on two independent ratings of 320 pooled cases scored by pairs of experienced treatment providers and supervision officers.

Factor loadings for the 16 SOTIPS items following exploratory principle components analyses with Varimax rotation identified three factors; Sexual Deviance, Criminality, and Social Stability and Supports. Total variance accounted for by these three components was 58.7%.

Multiple ANOVAs found significant between-subjects effects when comparing sexual recidivists and non-recidivists at Times 1, 2, and 3. Recidivists showed no significant changes in SOTIPS scores across time, whereas non-recidivists showed significant reductions in total SOTIPS scores.

Analyses also examined the predictive validity of adult-victim-only (20.4%) and child-victim-only (75.8%) offenders' scores separately across the three time periods. Scores for child-only offenders predicted all four types of recidivism at both 1- and 3-year follow-up periods based on significant AUCs. Adult-only offenders' scores were significant predictors of returns to prison, but were not significant predictors of other types of recidivism, with the exception of Time 2 scores, which predicted violent and any recidivism.

Although a few individual AUC analyses for SOTIPS and Static-99R risk categories were not statistically significant, combined SOTIPS and Static-99R scores proved significant across the six assessment waves for all types of recidivism. Combination models were tested by comparing the best goodness-of-fit Deviance χ^2 . The combination of the SOTIPS and Static-99R fit the observed data better when combining the logistic coefficients of the two measures' main effects in a multiple logistic regression model than when modeling a simple logistic equation based on a Static-99R \times SOTIPS interaction.

As with the individual logistic regression analyses, GEE model testing was conducted to determine which combination of multiple SOTIPS scores with the Static-99R yielded a stronger model. Both models provided predicted reoffense probabilities that adequately fit the observed

data based on goodness-of-fit Deviance χ^2 , but the combination of multiple SOTIPS scores with the Static-99R proved best by combining the coefficients in a multiple logistic regression analysis.

Four risk band categories for combined Static-99R and SOTIPS scores were assigned logically based on the predicted reoffense rates generated by the binomial logistic GEE analyses. The 3-year sexual recidivism rate for the approximately 45% of offenders categorized in the low risk band was 1.2%; for the 27% in the moderate-low risk band was 3.3%; for the 19% in the moderate-high risk was 8.7%; and for the 9% in the high risk band was 12.7%. The SOTIPS+Static-99R combination risk categories showed significant improvements in predictive power over Static-99R risk categories alone for both the one- and three-year follow-up periods. Net reclassification improvements (NRI; Pencina et al., 2007) for one-year follow-up predictions showed improvements between 20% and 43% and for three-year follow-ups between 10% and 30%.

Using these four risk bands, we calculated survival curves based on interval-censored survival GEE analyses for sexual, violent, and any recidivism and return to prison. Likelihood Ratio tests and Mantel-Cox Log Rank tests were conducted to test for significant differences between survival rates and survival curves, respectively, within each of the four recidivism types. The Likelihood Ratio tests showed significant differences between survival rates for the four risk groups among all four recidivism types, but the only significant difference between survival curves, that is, the proportional hazard rate among the four risk groups, was for sexual reoffending.

Discussion. The purpose of the present study was to combine static and dynamic risk measures into an overall assessment model that might predict sexual recidivism among adult

male sex offenders better than either type of measure alone. To this end, we tested three hypotheses in succession and found support for each. First, the static risk measures (Static-99R, Static-2002R and VASOR) all predicted sexual recidivism similarly and with moderate accuracy. The Static-99R, owing to its popularity among practitioners and researchers, was selected for subsequent model testing in the study. Second, a new dynamic risk measure, the Sex Offender Treatment Intervention and Progress Scale (SOTIPS), was developed as part of the present research. It also predicted sexual recidivism with moderate accuracy and was sensitive to changes in dynamic risk over time. Third, the combination of the Static-99R and SOTIPS outperformed either instrument alone when both instruments had similar predictive power.

These results are consistent with previous sex offender studies in which dynamic risk factors have added incremental predictive validity to static risk factor schemes (Beggs & Grace, 2010; Hanson et al., 2007; Knight & Thornton, 2007; Olver et al, 2007; Thornton, 2002) and studies in which an association has been found between positive treatment progress and reductions in recidivism (Beggs & Grace, 2011; Olver & Wong, 2011). The fact that the present study used a repeated measures model and found main effects across three SOTIPS score periods is particularly noteworthy in light of previous research in the field which typically has employed less frequent assessments.

Employing repeated measures paradigms yields large amounts of data, which can make it complicated to organize findings in a simple and useful manner. Our solution to this problem was to use generalized estimating equations (GEE; Liang & Zeger, 1986) to show the combinations of the Static-99R and SOTIPS risk levels across three time periods in one overall risk table. For most applied purposes, such as allocating supervision and treatment services, GEE

tables that categorize offenders into broad relative risk and need groups (e.g., low, moderate-low, moderate-high and high) are valid.

Two of the three broad criminogenic risk factors extracted during factor analyses, namely, Sexual Deviance and Criminality, have consistently predicted sexual recidivism in other studies as well (Hanson & Morton-Bourgon, 2005; Mann, Hanson, & Thornton, 2010). The third factor, which we labeled Social Stability and Supports, contains items that also have been found to be markers for an antisocial orientation.

A caution in interpreting the present study results concerns the low recidivism base rates in the sample. Sexual recidivism rates ranged from a high of 5.3% for the Time 1 three-year follow-up period to a low of 1.6% for the Time 3 one-year follow-up period. On the one hand, it is noteworthy that significant effects were detected given such low base rates. On the other hand, the findings would have been much more robust had the sample size and base rates been larger. Although the SOTIPS predicted sexual and other types of recidivism among the child-victim-only offenders who made up three-quarters of the sample, it did not predict sexual reoffending among the one-fifth of participants who were adult-victim-only offenders.

Studies focused on long follow-up times (i.e., 5 to 20 years) result in higher base rates, but long time frames are not particularly informative to service providers who must decide how to allocate supervision and treatment services over much shorter time periods. This perspective is why we examined relatively short follow-up periods in the present study. Even using follow-up time frames of one year, differences in predicted sexual recidivism rates between offenders in some risk levels were of practical significance.

A challenge in accurately assessing dynamic risk factors is that they are typically more subjective and difficult to score than static risk factors. Although treatment providers and

probation and parole officers in the current study were asked to score offenders on the SOTIPS independently, in practice, scoring cases jointly will lead to better assessments. Different service providers often each have different information about offenders and average measure correlations are certainly higher than single measure ones.

Furthering a collaborative model, we have encouraged clinicians and supervision officers to involve offenders in scoring the SOTIPS. Treatment and supervision targets should not be a secret. Collaborative approaches in which service providers and offenders discuss and set treatment goals together are more successful than non-collaborative ones (Miller & Rollnick, 2002; Shingler & Mann, 2006). As well, we argue that periodic reassessments can help recalibrate supervision and treatment plans, potentially leading to the delivery of more effective services.

Conclusion. The present study contributes to a growing body of research supporting empirically based risk and need assessment schemes for sexual offenders. Integrated risk and need instruments may help providers and jurisdictions better allocate supervision and treatment resources in a rational manner than by using one type of measure alone. As this is the first study to examine the SOTIPS, future replication studies will be important to evaluate its usefulness in other settings and in combinations with other static risk-assessment instruments.

Introduction

Statement of Problem

Of all crimes, sex offenses arguably have received the most public and legislative attention in recent years. Sexual offenses have serious negative consequences to victims, their families and communities and tax overextended criminal justice systems. In the United States, legislators have passed more laws concerning sex offenders during recent legislative sessions than during any other period in our nation's history (Center for Sex Offender Management, 2008; Vandervort-Clark, 2009). Legislation targeting sex offenders has spanned the areas of sentencing, registration, community notification, residency restrictions, civil confinement, electronic monitoring, supervision, and treatment.

Unfortunately, legislative and other interventions designed to protect the public from sex offenders often have been implemented without regard to important differences among individuals who commit these crimes. Not all sex offenders are the same. Research evidence on effective correctional practice (Andrews & Bonta, 2010) stresses the value of moving beyond "one size fits all" approaches.

Interventions that take important differences among sex offenders into account by following three principles of effective correctional intervention, those of risk, need and responsivity (RNR; Andrews & Bonta, 2010), have demonstrated significant reductions in recidivism among sex offenders (Hanson, Bourgon, Helmus, & Hodgson, 2009), as well as among violent (Dowden & Andrews, 2000) and general (Andrews & Bonta, 2010) offenders. The risk principle states that interventions are more effective when they focus on offenders who are most likely to reoffend, that is, those at moderate or higher risk. The need principle states that interventions should target offenders' changeable problems that are closely linked to criminal

behavior. These are commonly called criminogenic needs or dynamic risk factors, examples of which are pro-offending attitudes and offense-related sexual interests. The responsivity principle states that interventions should be matched to an offender's learning style, abilities and other characteristics such as gender, motivation, personality characteristics and cultural background.

Given that a key element of implementing effective correctional practice principles involves assessing sex offenders' risk to reoffend and allocating services differentially based on risk, assessments of that risk are of central importance. Risk assessments inform decisions on sentencing, community registration and notification, treatment, supervision, release from detention, and discharge from services. Consequently, the accuracy of assessments can have a profound impact on community safety, offenders' liberty, and wise allocation of public resources.

Literature Review

Risk assessments commonly focus on whether an individual's risk to sexually reoffend is lower than, similar to, or higher than that of the "average" sex offender (Cumming & McGrath, 2005). These are assessments of "relative risk." They compare an individual's risk to reoffend to other sex offenders and yield categorizations such as low, moderate or high. In a relatively limited number of circumstances, such as civil commitment evaluations, the assessment question concerns whether an individual meets an "absolute risk" threshold, such as "more likely than not" to sexual recidivate.

Regardless of the assessment approach, the recidivism rate of sex offenders is much lower than that commonly perceived by the public (Levenson, Brannon, Fortney, & Baker, 2007). Low rates of detected reoffending (low base rates), however, make it challenging to accurately identify those offenders who are most likely to recidivate (Craig, Browne, & Beech,

2008). The accuracy of risk predictions is greatest when the base rate of reoffending is about 50%, and it diminishes as the base rate diminishes. Low-frequency events are simply more difficult to predict than higher frequency events (Craig et al., 2008; Quinsey, Harris, Rice, & Cormier, 1998).

The base rates of sexual reoffending are found to be considerably lower than 50%. In a United States Department of Justice study of 9,691 adult male sex offenders released from state prisons in 1994, only 5.3% were rearrested for a sex offense within three years of release (Langdon, Schmitt, & Durose, 2003). A meta-analysis of 82 studies involving 29,450 sex offenders found a sexual recidivism rate of only 13.7% over an average follow-up period of about six years (Hanson & Morton-Bourgon, 2005). A more recent analysis of 24 studies involving 8,390 adult male sex offenders found that the sexual recidivism rate at five-year follow-up was 11.1% and at ten years was 16.6% (Helmus, Thornton, Hanson, & Babchishin, 2011).

Nevertheless, the accuracy of risk-assessment methods has improved dramatically over the last two decades, though room for improvement still exists. Advances in risk assessment can be marked in four major phases (Bonta & Wormith, 2008). First-generation risk-assessment approaches rest on unstructured professional judgment. Criticisms of this approach are that it is subjective, inconsistent, biased and not as reliable or accurate as structured, empirically based risk-assessment methods (Dawes, Faust, & Meehl, 1989; Grove, Zald, Lebow, Snitz, & Nelson, 2000; Hanson & Morton-Bourgon, 2009; Mossman, 1994).

In the sex offender field, second-generation risk-assessment instruments began emerging in the late 1990s. These structured actuarial measures are composed primarily of static risk factors, that is, unchangeable aspects of an individual's history, such as prior criminal

convictions and selection of certain types of victims. Static actuarial instruments include the Rapid Risk Assessment for Sexual Offence Recidivism (RRASOR; Hanson, 1997), Risk Matrix 2000 Sexual (RM2000/S; Thornton et al., 2003), Static-99 (Hanson & Thornton, 2000) and Static-2002 (Hanson, Helmus, & Thornton, 2009). Some second-generation actuarial risk instruments also include a small number of dynamic risk factors, sometimes called criminogenic needs. These are changeable risk factors such as pro-offending attitudes and offense-related sexual interests. Examples of instruments that include mostly static but also some dynamic risk factors are the Minnesota Sex Offender Screening Tool-Revised (MnSOST-R; Epperson et al., 1998), Sexual Violence Risk-20 (SVR-20; Boer, Hart, Kropp, & Webster, 1997) and Vermont Assessment of Sex Offender Risk (VASOR; McGrath & Hoke, 2001). The MnSOST-R, SVR-20 and VASOR are arguably second-generation instruments because they do not include a sufficient number of dynamic risk factors to provide professionals much guidance for delivering rehabilitation services.

The predictive accuracy of most second-generation risk-assessment instruments, based on the area under the curve of the receiver operating characteristic (AUC) statistic, is in the moderate range. Moderate predictive accuracy is defined here as an AUC between .70 and .79. We note, however, that there is no clear consensus in the field about terminology for categorizing AUCs (see Craig et al., 2008; Quinsey et al., 1998; Sjöstedt & Grann, 2002). The AUC represents the probability that a randomly selected recidivist will have a higher score on a risk measure than will a randomly selected nonrecidivist. AUC values range from 0 to 1, with 0 representing worse than chance prediction, .5 representing chance-level prediction, and 1 representing perfect prediction. The AUC is a recommended index of predictive accuracy for relatively low base-rate phenomena such as sexual reoffending (Rice & Harris, 1995).

Developing third-generation instruments has increasingly been the focus of contemporary research efforts. Third-generation instruments are those that combine static and relatively inclusive collections of dynamic risk-predictors in a single risk-needs instrument or “set” of instruments. Adding multiple dynamic factors in the risk-assessment equation leads to more comprehensive evaluations and has the potential to add incrementally to the long-term predictive accuracy of static instruments. Because, by definition, dynamic risk factors are potentially changeable, risk-needs instruments also offer direction to providers about how to identify and target offenders’ criminogenic needs in order to reduce their risk to reoffend. Risk-needs instruments that fully integrate assessments with ongoing case planning are called fourth-generation risk-assessment tools (Bonta & Wormith, 2008).

A series of meta-analyses by Hanson and his colleagues (Hanson & Bussiere, 1998; Hanson & Morton-Bourgon, 2004, 2005; Mann, Hanson, & Thornton, 2010) have identified a large number of dynamic risk factors strongly tied to sexual recidivism. Approaches to assessing these dynamic risk factors among sex offenders have included offender self-report psychometric batteries (Allan, Grace, Rutherford, & Hudson, 2007; Beech, Friendship, Erikson, & Hanson, 2002), clinician-administered scales (Anderson, Gibeau, & D’Amora, 1995; Looman, Abracen, Serin, & Marquis, 2005) and collaborative goal-attainment scaling (Barrett, Wilson, & Long, 2003; Green, 1988; Hogue, 1994).

Researchers have also developed a small number of integrated sex offender risk-needs models. The Violence Risk Scale-Sex Offender Version (VRS-SO; Olver, Wong, Nicholaichuk, & Gordon, 2007) consists of 7 static and 17 dynamic risk items. In a sample of 321 incarcerated mixed-type sex offenders followed up for an average of 10 years post-release, the scale predicted sexual and nonsexual violent recidivism with moderate accuracy (Olver et al., 2007). The total

dynamic scale score made a significant incremental contribution to predicting sexual recidivism after controlling for static risk. More recently, Olver and Wong (2011), using the Static-99 as a static risk measure and the VRS-SO dynamic scale as the dynamic measure, found that high-risk/low-change offenders recidivated at significantly higher rates than high-risk offenders who demonstrated greater treatment change. Studies on the VRS-SO in New Zealand replicated some of these findings in a sample of 218 relatively low risk child molesters who participated in a prison treatment program. In these studies, the VRS-SO showed good predictive validity (Beggs & Grace, 2010) and an association between treatment gain and reduced recidivism at an average 12-year follow-up (Beggs & Grace, 2011).

The Structured Risk Assessment model (SRA; Thornton, 2002) integrates static risk measures with those of four dynamic risk domains: sexual interests, distorted attitudes, socio-affective functioning, and self-management. In a sample of 117 adult male sex offenders who served prison sentences in the United Kingdom, Static-99 and SRA dynamic scores (absent unavailable sexual interest domain scores) predicted sexual reconviction better than either the static or dynamic measure alone at a mean time-at-risk of about three years (Thornton, 2002). More recently, the SRA model has predicted sexual recidivism with moderate accuracy in a sample of 513 high-risk civilly-committed adult male sex offenders at 5- and 10-year follow-up periods (Knight & Thornton, 2007) and among a sample of 119 child molesters drawn from one prison and several community sites in the United Kingdom (Craig, Thornton, Beech, & Brown, 2007).

Hanson and his associates have developed and studied an integrated set of static and dynamic instruments (Static-99, Stable-2007, and Acute-2007) in a landmark prospective study involving over 900 sex offenders from across 16 North American sites (Hanson et al., 2007). The

Static-99 is composed of 10 static risk factors; the Stable-2007 contains 13 relatively enduring but changeable dynamic risk factors (e.g., poor problem solving); and the Acute-2007 contains 7 rapidly changeable risk factors (e.g., access to victims) (Hanson et al., 2007; Hanson & Thornton, 2000). Combined initial scores on these measures were more predictive of sexual recidivism at a median 41-month follow-up period than static scores alone (Hanson et al., 2007). Change scores computed for subsequent administrations of the dynamic measures, however, were not related to recidivism. More recently, in a German-speaking sample of 263 sex offenders, the Stable-2007 added incremental predictive accuracy to the Static-99 for violent and general recidivism but not for sexual recidivism (Eher, Matthes, Schilling, Haubner-MacLean, & Rettenberger, 2011).

A limitation of existing sex offender risk-need instruments is that they typically have used assessment paradigms that employ either a single dynamic assessment (Thornton, 2002) or one pre- and one post-service assessment (Beggs & Grace, 2010, 2011; Olver & Wong, 2011). In the case of the one sex offender study that assessed dynamic risk factors over more than two time periods, changes in dynamic risk scores were not associated with changes in recidivism (Hanson et al., 2007). Assessing dynamic risk at regular intervals during the course of services may be important in helping providers adjust the intensity and duration of interventions to match individuals' changing risk and needs.

In the present study, a new dynamic risk instrument was developed and evaluated, the Sex Offender Treatment Intervention and Progress Scale (SOTIPS). It was designed to evaluate individuals at regular intervals and can be scored by treatment providers and probation and parole officers. It is used in conjunction with a static risk measure, the Static-99R. Among a sample of 759 adult male sex offenders enrolled in community treatment, we examined the

ability of the SOTIPS and the Static-99R to rank-order participants in terms of risk to reoffend for sexual, violent and any criminal reoffending and return to prison. Static risk factors were scored retrospectively. Dynamic risk scores were prospective, since treatment providers had previously scored participants on several dynamic risk items approximately every six months during the course of community treatment between 2001 and 2007.

Although the primary focus of the study was on sexual recidivism, data also was collected and analyzed for violent recidivism (sexual and violent recidivism combined), any criminal recidivism, and return to prison. Analyses were limited to testing for first-time recidivism for each of the four types of recidivism studied.

Statement of Hypotheses

The primary purpose of the present study was to test models of combining static and dynamic risk measures that might predict sexual recidivism better than any one measure alone and might be sensitive to participant change during the course of treatment. To this end, we tested in successive order three hypotheses. First, it was expected that one or more static risk measures (Static-99R, Static-2002R and VASOR) would predict sexual recidivism with moderate accuracy. Second, the SOTNPS or a combination of dynamic risk factors selected from this instrument would also predict sexual recidivism with moderate accuracy and be sensitive to changes in dynamic risk. Third, a combined static and dynamic risk measure would predict sexual recidivism more accurately than either measure alone.

Method

Setting

The sample was drawn from adult male sex offenders enrolled in community treatment in Vermont. Vermont is a state of small cities, towns and rural areas with a population of 625,741

(United States Census Bureau, 2010). The Vermont Treatment Program for Sexual Abusers (VTPSA) is the state's integrated network of three prison and 13 outpatient programs operated by the Vermont Department of Corrections (DOC). There are no state or county jails in Vermont. The DOC contracts with private agencies and treatment providers to deliver treatment services. The program utilizes a primarily cognitive-behavioral group treatment model and is designed to have treatment providers and supervision officers work in collaborative teams (McGrath, Cumming, Livingston, & Hoke, 2003; McGrath, Hoke, & Vojtisek, 1998).

Program

In 2001, the VTPSA began requiring contracted community treatment providers to complete risk and needs assessments on all program admissions. The initial assessment included two static risk measures, the Static-99 (Hanson & Thornton, 2000) and Vermont Assessment of Sex Offender Risk (VASOR; McGrath & Hoke, 2001) and one dynamic risk measure, the Sex Offender Treatment Needs and Progress Scale (SOTNPS; McGrath & Cumming, 2001, 2003). Per contract, providers were directed to re-administer the SOTNPS needs assessment to clients in January and July of each year. Treatment staff and probation and parole officers completed one-day trainings and follow-up supervision on how to administer, score and interpret these measures. The trainings included a review of the scoring manual, case exemplars for each item, and practice scoring cases with feedback. The goal was for treatment providers to use assessment results for treatment planning, provide copies to probation and parole officers to inform supervision services, and forward copies to VTPSA program directors for quality assurance and research purposes.

Participants

Participants were 759 adult male sex offenders under correctional supervision in Vermont who met four criteria. First, they had committed the requisite types of sexual offenses in order to be scored on each of the static risk instruments used in the study (Static-99R, Static-2002R and VASOR). This meant that participants were convicted of at least one sexual offense against an identifiable child or non-consenting adult victim (Category “A” sexual offense as defined in the Static-99 coding manual; Harris et al., 2003). Using this definition, individuals whose sexual crimes were limited to offenses such as prostitution, statutory rape, or child pornography possession were excluded from the study. Second, participants entered a VTPSA sponsored community sex offender treatment program between 2001 and 2007. Third, the time period from when they began VTPSA-sponsored treatment to the end date of the study follow-up period on December 31, 2010 was at least three years. Fourth, their treatment provider evaluated them using the SOTNPS at least once during the study period and submitted the results to the researchers. Based on analyses of DOC databases, an estimated additional 357 sex offenders were placed on community supervision in Vermont between 2001 and 2007 but did not meet study criteria because they did not attend treatment, or if they did, therapists did not submit assessments. Further information was not available on these non-participants.

At the time of community placement, the men were 18 to 75 years old, with an average age of 34.2 ($SD = 12.8$). Most (86.8%) were serving a sentence for their first sexual offense. Consistent with Vermont’s lack of racial diversity, 96.4% of participants were White. Most participants were employed full time (57.7%). A small number of participants (2.4%) met eligibility criteria to receive state services for a developmental disability. One-fifth (21.5%) of participants had received sex offender treatment while in prison.

Most participants (80.2%) were on probation, and approximately two-fifths (40.7%) of these participants had been sentenced to some period of incarceration prior to being released on probation. Less than one-fifth of all participants (17.1%) were on community-furlough release from prison and only a few (2.6%) were on parole.

Most participants (93.8%) committed sexual offenses that did not result in a physical victim injury requiring medical attention. About three-quarters (75.8%) of participants had only child victims (age 15 or younger), one-fifth (20.4%) had only adult victims, and a small percentage (3.8%) was known to have sexually offended against both children and adults.

Using definitions established by the Association for the Treatment of Sexual Abusers (Gordon et al., 1998), the sample was composed of 137 (18.1%) rapists (a contact sexual offense against an individual age 16 or older), 59 (7.8%) non-contact sex offenders (convicted for such offenses as exhibitionism or voyeurism), 111 (14.6%) incest offenders (who committed an offense against the offender's own biological offspring or child who has been parented by the offender for at least two years; offenses against an offender's nieces, nephews, cousins, and grand-children were considered child molestations) and 452 (59.6%) child molesters (who committed a non-incest hands-on sexual offense against a child age 15 years or younger). Of the 452 child molesters, 388 of the total sample (51.1%) molested girls only and 64 (8.4%) molested at least one boy.

Measures

During the data collection phase of the study, participant scores were collected on three static risk-assessment measures (Static-99R, Static-2002R and VASOR) and one dynamic risk-assessment measure (SOTNPS). Following initial data analyses, the Static-99R was selected as the static measure for model testing. The dynamic measure used for model testing was a new

instrument, the SOTIPS, which was constructed during the course of this study and is a revised version of the SOTNPS. Descriptions of all of these instruments are as follows.

Sex Offender Treatment Needs and Progress Scale (SOTNPS). The SOTNPS is a 22-item provider-administered dynamic risk scale designed to aid clinicians and probation and parole officers in identifying and monitoring the supervision and treatment needs of adult male sex offenders (McGrath & Cumming, 2001, 2003). Items were conceptually derived on the basis of professional consensus, literature review, and theory, and have defined scoring rules. It was designed to score clients at intake and thereafter every six months on a four-point scale; *minimal to no need for improvement, some need for improvement, considerable need for improvement and very considerable need for improvement*. A scoring manual details coding rules for each item using the four-point scale (McGrath & Cumming, 2003). Scores are supposed to reflect the offender's functioning during the previous six months and relative treatment and supervision need on each factor. Total scores are intended to provide an estimation of an individual's overall level of supervision and treatment need and risk for sexual recidivism. The SOTNPS has undergone minimal empirical study (McGrath, Cumming, & Livingston, 2005) but has been used in about one-fifth of community sex offender programs in the United States (McGrath, Cumming, Burchard, Zeoli, & Ellerby, 2010). The 22 SOTNPS items are shown in Table 3 and are labeled as "SOTIPS items" ($k = 16$) and "SOTNPS deleted items" ($k = 6$).

Sex Offender Treatment Intervention and Progress Scale Sex (SOTIPS). The SOTIPS is a revised version of the SOTNPS and was the primary dynamic measure used in the present study. Based on analyses conducted in the present study, 16 SOTNPS items were extracted from the SOTNPS and it was renamed the SOTIPS. Item definitions and scoring instructions for the retained 16 items remain unchanged. Total scores range from 0 to 48 points and, based on

analyses in the present study, are organized into three risk/need groups: low (0 to 10), moderate (11 to 20) and high (21 to 48).

Static-99R. The Static-99R is a 10-item actuarial instrument designed to assess the recidivism risk of adult males known to have committed at least one sexual offense (Helmus, Thornton, Hanson, & Babchishin, 2011; Harris, Phenix, Hanson, & Thornton, 2003). Items are identical to the Static-99 (Hanson & Thornton, 2000; Harris et al., 2003), with the exception of updated age weights (Helmus et al., 2011). The ten items pertain to sexual and nonsexual offense history, victim characteristics, and offender demographics. Total scores range from -3 to 12 points and are organized into four risk groups; low (-3 to 1), moderate-low (2 to 3), moderate-high (4 to 5) and high (6 to 12). A recent meta-analysis of 63 studies found a moderate relationship between Static-99 and sexual recidivism (Hanson & Morton-Bourgon, 2009). The authors of the Static-99/R now recommend that evaluators use the revised version of the scale (Helmus et al., 2011).

Static-2002R. The Static-2002R is a 14-item actuarial instrument designed to assess the recidivism risk of adult males known to have committed at least one sexual offense. Items are identical to the Static-2002 (Hanson, Helmus, & Thornton, 2009; Phenix, Doren, Helmus Hanson, & Thornton, 2009) with the exception of slight readjustments to the weighting of age categories (Helmus et al., 2010). The original version was designed to improve the Static-99 by simplifying scoring instructions, adding more items and grouping them into meaningful subscales. Subscales are age, persistence of sexual offending, deviant sexual interests, relationship to victims and general criminality. Total scores range from -2 to 14 points and are organized into five risk groups; low (-2 to 2), moderate-low (3 to 4), moderate (5 to 6), moderate-high (7 to 8) and high (9 to 12). A recent meta-analysis of 8 studies found a moderate

relationship between Static-2002R and sexual recidivism (Hanson & Morton-Bourgon, 2009). The authors of the Static-2002/R now recommend that evaluators use the revised version of the scale (Helmus et al., 2011).

Vermont Assessment of Sex Offender Risk (VASOR). The VASOR (McGrath & Hoke, 2001) is a risk scale designed to aid probation and parole officers in making placement and supervision decisions about convicted adult male sexual offenders. A 13-item reoffense risk scale is composed of many of the unchangeable risk factors found on the Static-99R and Static-2002R, as well as a small number of changeable risk factors, namely, alcohol and drug use, residence and employment stability and treatment cooperation. Total risk scale scores range from 0 to 125 and are organized into four risk groups: low (0 to 25), moderate-low (26 to 50), moderate-high (51 to 70) and high (71 to 125) (McGrath, Lasher & Hoke, 2009). Two studies have found a moderate relationship between the VASOR risk scale and sexual recidivism (Langton, Barbaree, Harkins, Seto, & Peacock, 2002; McGrath, Hoke, Livingston, & Cumming, 2001). A six-item violence scale concerns violence history and offense severity and is not meant to be associated with recidivism risk.

Outcome Measures

Recidivism data was obtained and coded for each study participant for all new charges for sexual, violent (sexual and nonsexual violence), and any criminal reoffenses and return to prison. The definition of a new sexual offense included a charge for a violation of supervision conditions if the incident could have been charged as a criminal sexual offense. Charges were counted based on criminal record checks in the states where each participant was known to have resided during the study period. The Vermont DOC computer database was used to identify violation of supervision charges and returns to prison.

Procedure and Data Analyses

A trained and experienced master's level research assistant used DOC case files and databases to collect demographic, offense characteristic, and dynamic risk factor score (SOTNPS) information on each participant. The research assistant also scored participants on the Static-99R, Static-2002R and VASOR based on their status as of date of community placement, and when scores already existed, ensured their accuracy. A second rater, the first or third author of this paper, independently scored these three static risk instruments on approximately every tenth case to assess interrater reliability.

The primary focus of the study was on examining the predictive accuracy of the static and dynamic risk measures used in the study, individually and in combination, for sexual, violent, and any recidivism and return to prison. Coding manuals for the Static-99R, Static-2002R and VASOR instruct coders to score offenders as of the date that they were placed in the community after conviction for a sexual offense. Predictive accuracy of these risk instruments is then typically examined by following offenders from that date to some specified time period. In order to help others compare the reoffense base rates and performance of the static risk instruments in the present study with these in other similar studies, one set of analyses followed this approach using fixed 3-year and 5-year follow-up periods

In contrast, all other predictive accuracy analyses using the static and dynamic risk measures in the study followed participants from the dates that treatment providers scored them on the dynamic measure for six assessment waves. The six assessment waves were 1- and 3-year follow-up periods following participants' SOTNPS scores at three assessment times, referred to as Times 1, 2 and 3. Time 1 SOTIPS scores were completed during the first 3 months ($n = 606$; $M = 1.0$ months; $SD = 1.1$) that a participant was in treatment, Time 2 scores between 4 and 9

months ($n = 665$; $M = 6.7$; $SD = 1.6$) after a participant had been in treatment, and Time 3 scores between 10 and 15 months ($n = 620$; $M = 12.7$; $SD = 1.7$) in treatment. Analyses of Times 4 to 10 SOTNPS and SOTIPS scores did not show predictive accuracy and are not reported.

Recidivism was coded as a binary variable (yes or no recidivism, coded 1 or 0). Analyses were limited to testing for first-time recidivism for sexual, violent, and any reoffending, and return to prison. Consequently, participants who recidivated following a SOTNPS Time 1, 2 or 3 score were removed from the sample for further analyses concerning that type of recidivism. Sexual recidivists were removed from the sample as of the date of their first new sexual offense charge.

Incidents of recidivism occurring before the beginning each of the six assessment wave periods (i.e., before a participant was scored on the SOTNPS) was not considered in AUC analyses for those waves. During the time frame between the dates participants were placed in the community and began treatment ($M = 88$ days; $Mdn = 54$ days; $SD = 143$ days) and between beginning treatment and their first SOTNPS score ($M = 31$ days; $Mdn = 14$ days; $SD = 49$ days), 7 (0.9%) individuals were charged with a new non-sexual offense and 39 (5.1%) were reincarcerated for these new offenses or for technical violations.

Assessment wave sample sizes varied due to participants recidivating, dropping out of treatment, not having enough time in the community to be included in the 3-year assessment waves and having records that were missing data. Most missing data concerned providers not completing SOTNPS score sheets in a timely manner. For example, almost all of the 153 (20.2%) missing Time 1 SOTNPS score sheets were missing because providers did not score these participants until they entered the Time 2 assessment waves.

To test the first hypothesis, analyses examined the interrater reliability of the three static risk instruments used in the study and the area under the curve of the receiver operating characteristic (AUC) statistic to examine the ability of the scales to rank order participants in terms of risk to reoffend at 3- and 5-year fixed follow-up periods from the dates that participants were placed in the community. These analyses used uncategorized risk scores.

To test the second hypothesis, we used AUCs to examine the predictive accuracy of the 22 individual SOTNPS items and total scores of different combinations of these items for the six assessment waves. Based on analyses of these findings and consideration of various clinical and practical issues discussed in the Results section, we revised the scale. The revised scale consisted of 16 items from the SOTNPS and was renamed the Sex Offender Treatment Intervention and Progress Scale (SOTIPS). Analyses of the SOTIPS examined its factor structure and interrater reliability, compared scores of sexual recidivists to non-recidivists, first using multiple ANOVAs to test for significant between-subjects differences for Times 1, 2 and 3 and then using Repeated Measures ANOVAs (RMANOVA) to test for within-subjects differences across time for sexual recidivists and sexual non-recidivists, and set cut-off score ranges for risk band groups.

To test the third hypothesis, analyses examined the predictive accuracy of the SOTIPS and Static-99R individually and in combination for the six assessment waves using the AUC statistic. The Static-99R was selected for these analyses because in the present sample its predictive accuracy was about the same as the Static-2002R and VASOR, but it has the advantage of having greater acceptance among clinicians and researchers in the field. For these analyses, we used categorized scores for both instruments (SOTIPS scores of low, moderate and high; Static-99R scores of low, moderate-low, moderate-high and high) to reduce the number of

between-instrument interactions and increase the likelihood of yielding statistically and clinically useful results.

Using the categorized SOTIPS and Static-99R scores, the two measures were combined into a total score for use in logistic regression analyses to test models for predicting reoffending. Logistic regression can predict probabilities in conditions which have not been observed, can account for non-linear data distributions, and uses distributions in which the probability of reoffense must be greater than 0% but less than 100%. Best possible models for examining all four types of recidivism during each of the three time periods were developed by testing for the best goodness-of-fit Deviance χ^2 between the cumulative logistic coefficients of the main effects in a multiple logistic regression model and the Static-99R×SOTIPS interaction in a simple logistic regression model.

To calculate reoffense probabilities from logistic coefficients, we followed Sofroniou and Hutcheson's (2002) recommended formula

$$p = \frac{e^{\text{logit}(p)}}{1 + e^{\text{logit}(p)}}$$

where $\text{logit}(p) = \alpha + \beta_1 + \beta_2$, α is the intercept of the logistic equation and β_1 and β_2 are the logistic coefficients of the Static-99R and SOTIPS categories. The 95% confidence interval for p was obtained by calculating the asymptotic standard error (ASE) and using the equation

$$95\% \text{ CI of } p = \frac{e^{\text{logit}(p) \pm 1.96(\text{ASE})}}{1 + e^{\text{logit}(p) \pm 1.96(\text{ASE})}}$$

where $\text{ASE} = \sqrt{(\text{Var}(\alpha) + \text{Var}(\beta_1) + \text{Var}(\beta_2) + 2(\text{Covar } \alpha, \beta_1, \beta_2))}$ and 1.96 is the large sample approximation of t for a two-tailed 95% confidence interval.

As the purpose of the SOTIPS is to provide repeated information on the same subject over time, we used the generalized estimating equations (GEE) approach developed by Liang

and Zeger (1986) as a means of conducting repeated measures logistic regression analyses for combined Static-99R+SOTIPS risk groups across Times 1, 2, and 3 for each of the four types of recidivism studied. The strength of GEE when examining repeated measures data is that it accounts for the correlations within subjects when conducting regression analyses and provides more simplified information for practical use and comparison over time.

To create an overall predictive model, GEE analyses were conducted using a binary logistic regression model. As these GEE analyses were an extension of the individual logistic regression analyses, model comparisons again tested for the best combination model between the cumulative logistic coefficients in a multiple logistic regression model and the Static-99R×SOTIPS interaction in a simple logistic regression model. We accomplished this by testing for the lowest possible Quasi Likelihood Under Independence Model Criterion (QIC) or Corrected Quasi Likelihood Under Independence Model Criterion (QICC), an adaptation of Akaike Information Criteria (AIC) for repeated measures in GEE, and testing for acceptable goodness-of-fit Deviance χ^2 . Sofroniou and Hutchesson's (2002) recommendations were used for calculating reoffense probabilities and confidence intervals for logistic coefficients generated by GEE analyses.

Finally, we categorized GEE reoffense probabilities into four risk bands (low, low-moderate, moderate-high, and high). We then entered them into a second set of GEE regression analyses using an interval-censored survival model (Finkelstein, 1986) to conduct repeated measures survival analyses. The parameter estimates generated by these regression analyses were then used to plot survival curves for the four types of recidivism. Likelihood Ratio tests were used to test for overall significant differences between survival rates between risk groups and Mantel-Cox Log Rank (Collett, 2003; Mantel, 1966) tests were then conducted to test for overall

significant differences between survival curves for each of the four types of recidivism studied. The final categorized combined model scores were compared to Static-99R categorized scores using Net Reclassification Improvement (NRI; Pencina, D'Agostino, D'Agostino, & Vasan, 2007). The NRI statistic indicates the degree of improvement in an AUC when comparing two different predictors with similar classification schemes.

All statistical analyses were conducted with SPSS (version 17.0), except for the recidivism probability and confidence interval calculations for logistic regression and binary logistic GEE analyses (Sofroniou & Hutcheson, 2002) and the Mantel-Cox Log Rank (Collett, 2003, p. 43) tests used for the interval censored survival GEE analyses, which were calculated using Microsoft Excel 2003 macros written for this study.

Results

Static Risk Assessment Instrument Analyses. As shown in Table 1, participants' mean risk scores for the static risk instruments were all in the moderate-low range. Interrater reliabilities of these measures, assessed on approximately every tenth case with intraclass correlation coefficients (ICC), ranged from acceptable to very good (.81 to .89).

Table 2 shows recidivism rates and AUCs for each of four types of recidivism at fixed 3- and 5-year follow-up periods beginning on the dates participants were first placed in the community. Of particular note, sexual recidivism rates were relative low, 4.6% at three years and 5.6% at five years. Table 2 also shows that all of the static risk instruments predicted sexual recidivism with moderate accuracy at the 3-year follow-up with AUCs ranging between .71 and .74, and slightly lower at five-year follow-up with AUCs ranging between .64 and .69, all at a significance level of at least $p < .01$. Table 2 also shows that the Static-99R had AUCs similar to

the other static risk instruments. Its predecessor, the Static-99, is by far the most commonly used static risk instrument in North America (McGrath et al., 2010).

Dynamic Risk Assessment Instrument Analyses. Tables 3-6 show AUC values for the dynamic risk variables considered in the present study for sexual (Table 3), violent (Table 4) and any recidivism (Table 5) and return to prison (Table 6). Based on analyses of these findings, we reduced the original 22-item SOTNPS to 16 items and renamed it the SOTIPS. Items retained were those that showed some statistically significant relationships to sexual recidivism (see Table 3). Items discarded were those that showed minimal or no statistically significant relationships to sexual recidivism. When we deemed it was a “close call” as to whether to retain an item, we also considered the item’s AUCs for violent and other types of recidivism, its clinical utility to practitioners, its contribution to the factor structure of the scale and its impact on AUCs for total scores.

Table 7 shows the factor loadings for the 16 SOTIPS items following exploratory principle components analyses with Varimax rotation for scores at Time 2. Three factors were identified; Sexual Deviance, Criminality, and Social Stability and Supports. Analyses were also conducted on Times 1 and 3 scores, but only the results for Time 2 are shown. Just as scores from Time 1 showed some reduced statistical power on AUC analyses when compared to values from Times 2 and 3, factor loadings from Time 1 showed some abnormalities when compared to Times 2 and 3, which loaded similarly within a three-factor solution. This was based on extracting components with eigenvalues greater than 1. Total variance accounted for by these three components was only slightly greater at Time 2 (58.66%) than at Time 3 (57.82%). Tables 3-6 also include the AUCs for the three factors identified in these analyses for each of the four types of recidivism studied.

Pooled interrater reliability data for each SOTIPS item, three factor scores and total score is presented in Table 8. They are based on two independent ratings from pairs of 17 experienced treatment providers and 24 probation and parole officers who volunteered to participate in an earlier study with this sample (McGrath, et al., 2005) and as part of the present study. Because these pairing were random, a one-way, random-effects ANOVA model intraclass correlation coefficient (ICC) was used (Shrout & Fleiss, 1979). As is shown in Table 8, the total SOTIPS score single measure ICC was .77 and the average measure was .87. For each of the three SOTIPS factors (see Factor Analyses section), scores also showed acceptable single and average measure ICC's; for Sexual Deviance it was .68 and .81 respectively, Criminality was .76 and .86, and Social Stability and Supports was .69 and .82. Individual item ICC's were typically lower and had a wide range for both single measure ICC's (.39-.71) and average measure ICC's (.56-.83). All ICCs were significant at $p < .001$. The 95% confidence intervals for all item, factor and total scores overlapped between the two score periods.

The SOTIPS also showed acceptable internal consistency. Cronbach's alpha for the pooled total score was .87 and the Gutman split-half reliability was .87. The standard error of measurement (SEM) using the pooled total score ICC₁ and ICC₂ were 3.45 and 2.59 respectively, both at the 68% confidence level.

Multiple ANOVAs found significant between-subjects effects when comparing sexual recidivists and non-recidivists at Time 1 ($F(48, 552) = 1.68, p < .01$), Time 2 ($F(49, 584) = 3.00, p < .001$) and Time 3 ($F(49, 499) = 1.90, p < .001$). In Table 9, multiple repeated measure analyses of variance (RMANOVA) compared item and total SOTIPS scores for within-subjects effects for 18 sexual recidivists and 434 non-recidivists. Recidivists showed no significant changes in scores across time, whereas non-recidivists showed significant reductions in total and

item scores except for Criminal Behavior and Cooperation with Supervision. As only about five percent of all participants were charged with new sexual offenses, five percent of non-recidivists were randomly selected by computer and the RMANOVAs were reconducted with 17 non-recidivists to adjust for inflation of Type I error. This resulted in generally lower F scores and less statistical power, but all items with significant differences over time continued to show significant reductions with the additional exceptions of Emotion Management, Employment, Residence, and Social Influences.

As shown in Tables 10-13, SOTIPS scores were categorized in three risk levels (low, moderate and high). This was achieved by maximizing predictive accuracy based on AUCs and minimizing overlap in the 95% confidence intervals of each level for sexual (Table 10), violent (Table 11), and any (Table 12) recidivism, and return to prison (Table 13).

Analyses also examined the predictive validity of adult-victim only (20.4%) and child-victim only (75.8%) offenders' scores separately across the three time periods. Scores for child-victim only offenders predicted all four types of recidivism at both 1- and 3-year follow-up periods based on significant AUCs. Adult-victim only offenders' scores were significant predictors of returns to prison, but were not significant predictors of other types of recidivism (with the exception of Time 2 scores, which predicted serious and any recidivism as well).

Combined Static and Dynamic Risk Assessment Instrument Analyses. Table 14 shows the AUC values for the combination of the SOTIPS and Static-99R. Although a few individual AUC analyses for SOTIPS and Static-99R categories were not statistically significant, the combination scores proved significant across all time periods, types of recidivism, and follow-up periods. Combination models were tested by comparing the best goodness-of-fit Deviance χ^2 , and the combination of the SOTIPS and Static-99R fit the observed data better when combining

the logistic coefficients of the two measures' main effects in a multiple logistic regression model than modeling a simple logistic equation based on a Static-99R×SOTIPS interaction.

Examination of Table 14 also shows a pattern of the static measure (Static-99R) predicting better than the dynamic measure (SOTIPS) at Time 1, both predicting about the same at Time 2 and the dynamic measure predicting better than the static measure at Time 3.

As with the individual logistic regression analyses, GEE model testing was conducted to determine which combination of multiple SOTIPS scores with the Static-99R yielded a stronger model. Both models provided predicted reoffense probabilities which adequately fit the observed data based on goodness-of-fit Deviance χ^2 , but the combination of multiple SOTIPS scores with the Static-99R proved best by combining the coefficients in a multiple logistic regression analysis, as this model resulted in the lower QIC or QICC for the multiple conditions. An additional analysis of variance indicated that the SOTIPS accounted for 11% of the variance in this model and the Static-99R accounted for 6%. Tables 15-18 show the specific model effect and parameter estimate information of the binary logistic GEE analyses for sexual (Table 15), violent (Table 16), and any recidivism (Table 17), and return to prison (Table 18).

Tables 19-22 show the observed and predicted recidivism rates based on the combination of the Static-99R and SOTIPS for each individual logistic regression analysis and the predicted recidivism rates generated by the GEE analyses for sexual (Table 19), violent (Table 20), and any (Table 21) recidivism, and return to prison (Table 22). Issues of poor nomological validity within some logistic regression analyses were evident (e.g., a moderate-high risk offenders' predicted reoffense rate that were lower than the low-moderate offenders' predicted reoffense rate in one analysis), but these were resolved in the predictive models developed by the GEE analyses. For simpler comparisons, multiaxial Tables 23-26 show reoffense probabilities based

on the repeated measures combination of the SOTIPS and Static-99R for sexual (Table 23), violent (Table 24), and any (Table 25) recidivism and return to prison (Table 26).

Table 27 shows four recommended risk band categories for combined Static-99R and SOTIPS scores assigned logically based on the predicted reoffense rates generated by the binomial logistic GEE analyses; Table 28 shows recidivism rates for each of these bands for sexual, violent, and any recidivism, and return to prison. These risk bands categorize approximately 45% of offenders as low risk, 27% moderate-low risk, 19% moderate-high risk and 9% high risk. Net Reclassification Improvements (Pencina et al., 2007) show that AUCs for the SOTIPS+Static-99R final model were significantly greater than AUCs for the Static-99R alone for 1-year follow-ups of sexual recidivism (NRI = .423, $p < .001$), violent recidivism (NRI = .299, $p < .01$), any recidivism (NRI = .202, $p < .001$), and return to prison (NRI = .426, $p < .001$), as well as for 3-year follow-ups of sexual recidivism (NRI = .279, $p < .001$), violent recidivism (NRI = .162, $p < .01$), any recidivism (NRI = .104, $p < .05$), and return to prison (NRI = .301, $p < .001$).

Figure 1 shows graphically an example of changes in reoffense rates within Static-99R risk groups as a function of changes in SOTIPS dynamic risk for sexual reoffense rates at 3-years. As can be seen, risk band levels for combined Static-99R and SOTIPS scores were the same as the Static-99R risk levels when SOTIPS scores showed a moderate treatment need (i.e., moderate risk level). For SOTIPS scores reflecting a high treatment need, the combined Static-99R and SOTIPS scores went up one risk band level from the Static-99R risk levels and down one risk band level for SOTIPS scores reflecting a low treatment need.

Using these four risk bands, we calculated survival curves based on interval-censored survival GEE analyses for sexual, violent, and any recidivism, and return to prison (see Figure

2). Likelihood Ratio tests and Mantel-Cox Log Rank tests were conducted to test for significant differences between survival rates and survival curves, respectively, within each of the four recidivism types. While the Likelihood Ratio tests showed significant differences between survival rates for the four risk groups among all four recidivism types (sexual reoffending $\chi^2(4, n = 1,784) = 1,921.68, p < .001$; violent reoffending $\chi^2(4, n = 1,772) = 342.31, p < .001$; any reoffending $\chi^2(4, n = 1,717) = 292.21, p < .001$; and return to prison $\chi^2(4, n = 1,636) = 187.59, p < .001$), the only significant difference between survival curves, that is, the proportional hazard rate among the four risk groups, was for sexual reoffending ($\chi^2(3, n = 1,784) = 24.26, p < .001$). Based on the Mantel-Cox Log Rank tests, we cannot confirm there were significant differences between survival curves for violent reoffense ($\chi^2(3, n = 1,772) = 9 \times 10^{-15}$), any reoffense ($\chi^2(3, n = 1,717) = 3 \times 10^{-14}$), and return to prison ($\chi^2(3, n = 1,636) = 2 \times 10^{-14}$).

Conclusions

Discussion of Findings

The purpose of the present study was to combine static and dynamic risk measures in an overall assessment model that might predict sexual recidivism among adult male sex offenders better than either type of measure alone. To this end, we tested three hypotheses in succession and found support for each. First, the static risk measures (Static-99R, Static-2002R and VASOR) all predicted 3-year sexual recidivism similarly and with moderate accuracy. The Static-99R, owing to its popularity among practitioners and researchers, was selected for subsequent model testing in the study. Second, a new dynamic risk measure, the Sex Offender Treatment Intervention and Progress Scale (SOTIPS), was developed as part of the present research. Under most conditions, it also predicted sexual recidivism with moderate accuracy and was sensitive to changes in dynamic risk over time. Third, the combination of the Static-99R and

SOTIPS outperformed either instrument alone when both instruments had similar predictive power.

These results are consistent with previous sex offender studies in which dynamic risk factors have added incremental predictive validity to static risk factor schemes (Beggs & Grace, 2010; Hanson et al., 2007; Knight & Thornton, 2007; Olver et al, 2007; Thornton, 2002), as well as studies in which an association was found between positive treatment progress and reductions in recidivism (Beggs & Grace, 2011; Olver & Wong, 2011). The fact that the present study used a repeated measures model and found main effects across three SOTIPS score periods, however, is particularly noteworthy in light of previous research in the field. Other studies of sex offenders typically have employed less frequent assessments, such as a single dynamic assessment (Thornton, 2002) or a paradigm of one pre- and one post-treatment assessment (Beggs & Grace, 2011; Olver & Wong, 2011). In the case of the one other repeated-measures sex offender study of which we are aware, changes in scores on dynamic risk factors were not associated with changes in recidivism (Hanson et al., 2007).

Employing repeated measures paradigms yields large amounts of data and this can make it complicated to organize findings in a simple and useful manner. Our solution to this problem was to use generalized estimating equations (GEE; Liang & Zeger, 1986) to show the combinations of the Static-99R and SOTIPS risk levels across three time periods in one overall risk table (see Table 28). For most applied purposes, such as allocating supervision and treatment services, GEE tables that categorize offenders into broad relative risk and need groups (e.g., low, moderate-low, moderate-high, and high) are valid and sufficient. For purposes requiring absolute predicted recidivism rates, tables based on non-GEE grouped risk scores would be required.

It is noteworthy that the SOTIPS+Static-99R combination risk categories showed significant improvements in predictive power over Static-99R risk categories alone for both the one- and three-year follow-up periods. Net reclassification improvements (NRI; Pencina et al., 2007) for one-year follow-up predictions showed improvements between 20% and 43% and for three-year follow-ups between 10% and 30%. The greater improvements for shorter term predictions are not surprising given that static risk-assessment tools like the Static-99R are typically evaluated using follow-up periods of five or more years (Hanson & Morton-Bourgon, 2009) and the current analyses of the SOTIPS showed sensitivities to short-term changes in most participants' dynamic risk.

Although combining the Static-99R and SOTIPS added predictive value to each other, independently, they performed similarly if judged by the fact that their AUC 95% confidence intervals in every comparison overlapped considerably. This is not surprising given that recent meta-analyses have found that the predictive efficiency of most validated sexual and violence risk-assessment instruments are all quite similar (Hanson & Morton-Bourgon, 2009; Yang, Wong, & Coid, 2010). Nevertheless, comparing Static-99R and SOTIPS AUCs showed some interesting trends (see Table 14). These were that the Static-99R seemed to predict sexual recidivism better than the SOTIPS at Time 1, both predicted about the same at Time 2 and the SOTIPS predicted better than the Static-99R at Time 3. These and other similar findings (Olver & Wong, 2011; Studer & Reddon, 1998) point out the limitations of relying solely on static measures to predict risk. As well, the salience of static and dynamic risk factors over time can change in important ways. In the present case, providers' SOTIPS assessments may have been more accurate at Time 2 versus Time 1, because they knew the offenders better. On the other

hand, offenders' static risk factors simply may have become less important and predictive of reoffending the more they successfully addressed their criminogenic needs.

It is likely that the failure to find significant effects for Time 4 SOTIPS scores and beyond was due to decreasing base rates of sexual recidivism and attrition within the sample, but other possible explanations exist. Participants in the present study may have reached a treatment gain ceiling in about 18 months and further treatment conferred no additional benefit. It is also possible that the SOTIPS was insensitive to participant change after a certain benefit threshold was met. Regardless, although considerable data about treatment dosage delivered in sex offender programs exists, little is known about what is an optimal dosage (McGrath et al., 2010). In the United States, the median treatment dose for community programs for adult sex offenders is about 200 hours over 24 months and for prison programs about 350 hours over 24 months. Canadian programs typically provide less than half the treatment dose of United States programs (McGrath et al., 2010). Treatment dose is an important clinical and resource issue requiring further research.

Two of the three broad criminogenic risk factors extracted during factor analyses, namely Sexual Deviance and Criminality, have consistently predicted sexual recidivism in other studies as well (Hanson & Morton-Bourgon, 2004, 2005; Mann et al., 2010). The third factor, labeled Social Stability and Supports, contains items that also have been found to be markers for an antisocial orientation. These factors are unemployment, negative social influences, and poor problem solving (Hanson & Morton-Bourgon, 2004). The Sexual Deviance factor predicted sexual recidivism about as well as the other two factors across SOTIPS score periods and follow-up times, but as might be expected, the Criminality and Social Stability and Supports factors had slightly higher AUCs for violent and any recidivism and return to prison.

The current study has several limitations. Participants in the study were from a small rural state that, compared to other jurisdictions in the United States, has a relatively low crime rate and a small non-white population. Although the static risk instruments used in the present study predicted similarly as in studies involving more diverse populations (Hanson & Morton-Bourgon, 2009) and the dynamic risk factors that comprise the SOTIPS have been found to predict sexual reoffending in numerous other studies (Mann et al., 2010), whether the present study results will generalize in other jurisdictions needs to be examined.

Ideally, participants' treatment needs would have been evaluated immediately upon placement in the community, but there was an average lag time of a few months before providers scored participants on the SOTIPS. Some participants reoffended during this lag time, and a small but unknown number who would have met eligibility criteria for the study were incarcerated before being referred to treatment and evaluated. As well, the characteristics of offenders who would have met eligibility for the study but failed to participate in treatment or who lacked SOTIPS scores is not known. Other risk-assessment paradigms have failed to adequately study some sub-groups of sex offenders, such as statutory rapists and child-pornography-only offenders (see "Category B" sex offender in Harris et al., 2003; Phenix et al., 2009) and the present study has this limitation as well.

Another caution in interpreting the present study results concerns the low recidivism base rates in the sample. Sexual recidivism rates ranged from a high of 5.3% for the Time 1 three-year follow-up period to a low of 1.6% for the Time 3 one-year follow-up period. On the one hand, it is noteworthy that significant effects were detected given such low base rates. On the other hand, the findings would have been much more robust had the sample size and base rates been larger. Although SOTIPS predicted sexual and other types of recidivism among the child only offenders

who made up three-quarters of the sample, it did not predict sexual reoffending among the one-fifth of participants who were adult only offenders.

Likely also owing in part to the low base rates and low sample cell sizes for some of the SOTIPS+Static-99R risk level combinations, the 95% confidence interval for predicted recidivism rates varied widely and many were quite large. Replication studies are needed, ideally with large sample sizes. Although structured actuarial risk assessment clearly has a better predictive track record than unaided clinical judgment, it still leaves unaccounted for a considerable amount of predictive variance.

Barbaree (1997) has noted that low base rates are an ongoing challenge in sex offender research and this has perhaps become even more evident in recent years as sexual abuse and assault rates have continued to decline (Finkelhor, Jones, & Shattuck, 2010; Mishra & Lalumière, 2009). In the present study, participants were under supervision and enrolled in treatment, which may have reduced the recidivism base rate.

Studies focused on long follow-up times (e.g., 5 to 20 years) result in higher base rates (e.g., Harris & Hanson, 2004; Knight & Thornton, 2007), but long time frames are not particularly informative to service providers who must decide how to allocate supervision and treatment services over much shorter time periods. This administrative reality is the reason we examined relatively short follow-up periods in the present study. Even using a follow-up time frame of one year, differences in predicted sexual recidivism rates among offenders at various risk levels were arguably of practical significance (see Table 23). Of course, studies with longer follow-up times will be needed to examine whether SOTIPS scores over the short term predict reoffending over the long term.

A challenge to accurately assessing dynamic risk factors is that they are typically more subjective and difficult to score than static risk factors. For example, the degree to which an individual endorses antisocial attitudes is a more challenging assessment matter than whether an individual has a criminal record of a prior sexual offense. Certainly, SOTIPS interrater reliability coefficients were lower than for the static risk measures used in the study. Because dynamic risk measures are more subjective, provider scoring bias is also likely more of a problem than with static risk measures. Although treatment providers and probation and parole officers in the current study were asked to score offenders on the SOTIPS independently, in practice, scoring cases jointly may lead to better assessments. Different service providers often have different information about offenders they supervise and treat, and average-measure correlations are certainly higher than single measure ones.

Furthering a collaborative model, we also have encouraged clinicians and supervision officers to involve offenders in scoring their SOTIPS. Treatment and supervision targets should not be a secret. Collaborative approaches in which service providers and offenders discuss and set treatment goals together are more successful than authoritarian approaches (Miller & Rollnick, 2002; Shingler & Mann, 2006). Periodic reassessments can help recalibrate supervision and treatment plans, a step that may lead to the delivery of more effective services.

The identification of dynamic factors that were not predictive in the present study may be as informative as some of those that were. Of the original 22 dynamic risk factors examined in this study, several were weakly or not at all associated with sexual reoffending; however, as the definition of recidivism widened from sexual to violent to any recidivism and to return to prison, stronger predictive association were found. With respect to individual risk factors, the lack of an association between major mental illness and any type of recidivism is consistent with research

showing that this population has the same markers for violence as their healthy counterparts do and that psychiatric symptoms play a minimal role in predicting violent behavior (Skeem, Manchak, & Peterson, 2011). To our surprise, substance abuse also had little relationship to sexual recidivism but, as might be expected (Bonta, Law, & Hanson, 1998), was strongly correlated with any criminal recidivism and returns to prison.

Consistent with other findings in recent meta-analyses (Hanson & Morton-Bourgon, 2005; Mann, Hanson, & Thornton, 2010), admission of sexual offending behavior was not a strong predictor of sexual recidivism. It is important to note, though, that most providers during the time of the study did not accept categorical deniers into treatment; therefore this subgroup of offenders was poorly represented in the sample.

Implications for Policy and Practice

The present study adds to a growing body of research supporting empirically based risk prediction for offenders (Hanson & Morton-Bourgon, 2009; Skeem & Monahan, 2011; Yang, Wong, & Coid, 2010). Of course, no risk prediction scheme will be entirely accurate, and the measures described in this report are far from perfect. Nevertheless, the principal finding that a dynamic risk measure can add incremental predictive value to a static risk-assessment scheme and measure treatment change that is associated with variations in reoffense rates have several practical policy and practice implications.

If validated in replication studies, the current model may help jurisdictions better employ the risk principle (Andrews & Bonta, 2010; Hanson, Bourgon et al., 2009) for classifying offenders and allocating resources in an informed and rational manner. Broadly, this practical application involves providing more intensive treatment, supervision, and management services to higher risk sex offenders and less intensive services to lower risk individuals. Decisions anchored to validated risk

instruments are typically more efficient and uniform than those that do not involve a structured, empirically-based process (Cumming & McGrath, 2005). Ultimately, if services are matched appropriately to offenders' risk level, the likelihood of services reducing reoffense rates is increased (Hanson, Bourgon et al., 2009; Lovins, Lowenkamp, & Latessa, 2009). Implementing the risk principle also has the effect of directing scarce correctional treatment and supervision resources toward people and programs where they are likely to have the greatest impact.

The facts that the SOTIPS is a relatively simple and straightforward instrument and that treatment providers and probation and parole officers can score it reliably increase its utility. As well, in our clinical experience, SOTIPS items appear to be easily understood by and are face valid to clients. Consequently, the SOTIPS has potential for use in therapist/client collaborative treatment planning activities (Shingler & Mann, 2006).

Finally, consistent with the findings of Olver and Wong (2011), the present results lend support to the notion that even sex offenders who are designated high risk on static actuarial risk instruments may in fact be able to substantially reduce their risk to sexually recidivate as a result of treatment intervention. Assessment instruments that can reliably measure such treatment changes have important policy implications for the treatment and management of high-risk sex offenders, many of whom are civilly committed after their prison terms are completed in the United States at costs that often exceed \$100,00 per individual per year (LaFond, 2005).

Finally, professionals conducting sex offender risk assessments are advised to continue use those instruments that have the greatest empirical support (e.g., Static-99 and Static99R). Currently, these established instruments are composed primarily of static risk factors. SOTIPS and other similar dynamic risk measures (Stable-2007, SRA; SVR-SO) should be used cautiously in combination with established static risk instruments. The literature on how to

combine static and dynamic risk factors in an overall sex offender risk-assessment scheme is still quite young.

Implications for Further Research

First and foremost, the current results will need to be replicated by other researchers in other jurisdictions. The present study was designed to assist service providers in making management decisions over the short term and the usefulness of the SOTIPS in predicting sexual recidivism over the long term has yet to be studied. Additionally, the ability of the SOTIPS to predict sexual recidivism in subgroups other than child-victim only sex offenders needs to be examined further. Although the Static-99R is a commonly used static sex-offender risk instrument, some jurisdictions use other static instruments; how those instruments can be combined with the SOTIPS to improve predictive accuracy is another area for future research.

The present study examined the SOTIPS on a sample of sex offenders under community correctional supervision and enrolled in treatment. Future research should examine its usefulness with sex offenders in incarcerated settings as well.

Perhaps the most important sub-group of sex offenders on whom further SOTIPS research should focus is made up of those at highest risk to sexually recidivate. Whereas sex offenders in the present study who scored as low risk on the Static-99R had 3-year sexual recidivism rates that were relatively low (between approximately 1% and 5%) regardless of their SOTIPS risk score risk level, the sexual recidivism rates of those who scored as high risk on the Static-99R ranged widely depending on their SOTIPS risk level and were all relatively high (between approximately 10% and 30%). The potential opportunity for identifying which high-risk offenders are truly high risk and whether they can successfully reduce their risk is a critical research question with important implications for practice, policy and community safety. As the

sample size of high-risk offenders in the present study was quite small, study of the SOTIPS with a larger group of high-risk sex offenders is warranted.

References

- Allan, M., Grace, R. C., Rutherford, B., & Hudson, S. M. (2007). Psychometric assessment of dynamic risk factors for child molesters. *Sexual Abuse: A Journal of Research and Treatment, 19*, 347–367.
- Anderson, R. D., Gibeau, D., & D'Amora, D. A. (1995). The Sex Offender Treatment Rating Scale: Initial reliability data. *Sexual Abuse: A Journal of Research and Treatment, 7*, 221-227.
- Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). Cincinnati: Anderson.
- Barbaree, H. E. (1997). Evaluating treatment efficacy with sexual offenders: The insensitivity of recidivism studies to treatment effects. *Sexual Abuse: A Journal of Research and Treatment, 9*, 111-128.
- Barrett, M., Wilson, R. J., & Long, C. (2003). Measuring motivation to change in sexual offenders from institutional intake to community treatment. *Sexual Abuse: A Journal of Research and Treatment, 15*, 269-283.
- Beech, A., Friendship, C., Erikson, M., & Hanson, R. K. (2002). The relationship between static and dynamic risk factors and reconviction in a sample of U.K. child abusers. *Sexual Abuse: A Journal of Research and Treatment, 14*, 155–167.
- Beggs, S. M., & Grace, R. C. (2010). Assessment of dynamic risk factors: An independent validation study of the Violence Risk Scale: Sexual Offender Version. *Sexual Abuse: A Journal of Research and Treatment, 22*, 234–251.
- Beggs, S. M., & Grace, R. C. (2011). Treatment gain for sexual offenders against children predicts reduced recidivism: A comparison validity study. *Journal of Consulting and*

- Clinical Psychology*, 79, 182-192.
- Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). *Manual for the Sexual Violence Risk – 20: Professional Guidelines for Assessing Risk of Sexual Violence*. Vancouver, B.C.: The Institute Against Family Violence.
- Bonta, J., Law, M., & Hanson, R. K. (1998). The prediction of criminal and violent recidivism among mentally disordered offenders: A meta-analysis. *Psychological Bulletin*, 123, 123-142.
- Bonta, J., & Wormith, S. J. (2008). Risk and need assessment. In G. McIvor & P. Raynor (Eds.), *Developments in social work with offenders*. (pp. 131-152). London: Jessica Kingsley Publishers.
- Center for Sex Offender Management (2008). *Legislative trends in sex offender management*. Silver Spring, MD: Author.
- Collett, D. (2003). *Modeling survival data in medical research* (2nd ed.). Boca Raton, FL: CRC Press.
- Craig, L. A., Browne, K. D., & Beech, A. R. (2008). *Assessing risk in sex offenders: A practitioner's guide*. West Sussex, England: John Wiley and Sons.
- Craig, L. A., Thornton, D., Beech, A., & Browne, K. D. (2007). The relationship of statistical and psychological risk markers to sexual reconviction in child molesters. *Criminal Justice and Behavior*, 34, 314-349.
- Cumming, G., & McGrath, R. J. (2005). *Supervision of the sex offender*. Brandon, VT: Safer Society.
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243, 1668-1674.

- Dowden, C., & Andrews, D. A. (2000). Effective correctional treatment and violent reoffending: A meta-analysis. *Canadian Journal of Criminology*, *42*, 449-476.
- Eher, R., Matthes, A., Schilling, F., Haubner-MacLean, T., & Rettenberger, M. (2011). Dynamic risk assessment in sexual offenders using the Stable-2000 and the Stable-2007: An investigation of predictive and incremental validity. *Sexual Abuse: A Journal of Research and Treatment*, *15*, 269-283. Advanced online publication.
- Epperson, D. L., Kaul, J. D., Huot, S. J., Hesselton, D., Alexander, W., & Goldman, R. (1998). *Minnesota Sex Offender Screening Tool – Revised (MnSOST-R): Development, performance, and recommended risk level cut scores*. Minneapolis, MN: Minnesota Department of Corrections.
- Finkelhor, D., Jones, L., & Shattuck, A. (2010). *Updated trends in child maltreatment - 2008*. Durham, NH: Crimes Against Children Research Center.
- Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics*, *42*, 845-854.
- Gordon, A., Harris, A., Murphy, W., Seto, M., Hanson, R. K., Marques, J. K., & Quinsey, V. L. (1998, July). *Collaborative database of sex offender treatment outcome*. Beaverton, OR: Association for the Treatment of Sexual Abusers.
- Green, R. (1988). Comprehensive treatment planning for sex offenders. In B. K. Schwartz (Ed.), *A practitioner's guide to treating the male sex offender* (pp. 71-84). Washington, DC: U.S. Department of Justice.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, *12*, 19-30.

- Hanson, R. K. (1997). *The development of a brief actuarial risk scale for sexual offense recidivism* (User Rep. 97-04). Ottawa: Department of the Solicitor General of Canada.
- Hanson, R. K., Bourgon, G., Helmus, L., & Hodgson, S. (2009). The principles of effective correctional treatment also apply to sexual offenders. *Criminal Justice and Behavior, 36*, 865-891.
- Hanson, R. K., & Bussière, 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., Harris, A. J. R., Scott, T. L., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project*. Ottawa: Public Safety Canada.
- Hanson, R. K., Helmus, L., & Thornton, D. (2009). Predicting recidivism amongst sexual offenders: A multi-site study of Static-2002. *Law and Human Behavior, 34*, 198-211.
- Hanson, R. K., & Morton-Bourgon, K. (2004). *Predictors of sexual recidivism: An updated meta-analysis* (Corrections User Report No. 2004-02). Ottawa: Public Safety and Emergency Preparedness Canada.
- Hanson, R. K., & Morton-Bourgon, K. E. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology, 73*, 1154–1163.
- Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment, 21*, 1–21.

- 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.
- Harris, A. J. R., & Hanson, R. K. (2004). *Sex offender recidivism: A simple question* (User Report 2004-03). Ottawa: Public Safety and Emergency Preparedness Canada.
- Harris, A., Phenix, A., Hanson, R. K., & Thornton, D. (2003). *Static-99 coding rules: Revised 2003*. Ottawa: Department of the Solicitor General of Canada.
- Helmus, L., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2011). *Improving the predictive accuracy of Static-99 and Static-2002 with older sex offenders: Revised age weights. Sexual Abuse: A Journal of Research and Treatment*. Advance online publication.
- Hogue, T. E. (1994). Goal attainment scaling: A measure of clinical impact and risk assessment. *Issues in Criminological & Legal Psychology, 21*, 96-102.
- Knight, R. A., & Thornton, D. (2007, March). *Evaluating and improving risk assessment schemes for sexual recidivism: A long-term follow-up of convicted sexual offenders* (Document No.: 217618). Washington, DC: U.S. Department of Justice.
- LaFond, J. Q. (2005). *Preventing sexual violence: How society should cope with sex offenders*. Washington, DC: American Psychological Association.
- Langton, C. M., Barbaree, H. E., Harkins, L., Seto, M. C., & Peacock, E. J. (2002, October). *Evaluating the predictive validity of seven risk assessment instruments for sexual offenders*. Paper presented at the 21st Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Montreal, Quebec, Canada.
- Levenson, J. S., Brannon, T. N., Fortney, T., & Baker, J. N. (2007). Public perceptions about sex offenders and community protection policies. *Analyses of Social Issues and Public Policy, 7*, 1-25.

- Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.
- Looman, J., Abracen, J., Serin, R., & Marquis, P. (2005). Psychopathy, treatment change and recidivism in high-risk, high-need sexual offenders. *Journal of Interpersonal Violence*, 20, 549–568.
- Lovins, R., Lowenkamp, C. T., & Latessa, E. J. (2009). Applying the risk principle to sex offenders: Can treatment make some sex offenders worse. *The Prison Journal*, 89, 344-357.
- Mann, R. E., Hanson, R. K., & Thornton, D. (2010). Assessing risk for sexual recidivism: Some proposals on the nature of psychologically meaningful risk factors. *Sexual Abuse: A Journal of Research and Treatment*, 22, 191-217.
- Mantel, N. (1966). Evaluation of survival data and two rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, 50, 163-170.
- McGrath, R. J., & Cumming, G. F. (2001, 2003). *The Sex Offender Treatment Needs and Progress Scale – Version 2001, 2003*. Middlebury, VT: Author.
- McGrath, R. J., Cumming, G. F., & Livingston, J. A. (2005). *Predictive validity of the Sex Offender Treatment Needs and Progress Scale*. Paper presented at the annual conference of the Association for the Treatment of Sexual Abusers, Salt Lake City.
- McGrath, R. J., Cumming, G. F., Burchard, B. L., Zeoli, S., & Ellerby, L. (2010). *Current practices and emerging trends in sexual abuser management: The Safer Society 2009 North American survey*. Brandon, VT: Safer Society Press.

- McGrath, R. J., Cumming, G., Livingston, J. A., & Hoke, S. E. (2003). Outcome of a treatment program for adult sex offenders: From prison to community. *Journal of Interpersonal Violence, 18*, 3-17.
- McGrath, R. J., & Hoke, S. E. (2001). *Vermont Assessment of Sex Offender Risk Manual*. Middlebury, VT: Author.
- McGrath, R. J., Hoke, S. E., Livingston, J. A., & Cumming, G., (2001, November). *The Vermont Assessment of Sex Offender Risk (VASOR): An initial reliability and validity study*. Paper presented at the 20th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, San Antonio, TX.
- McGrath, R. J., Hoke, S. E., & Vojtisek, J. E. (1998). Cognitive-behavioral treatment of sex offenders: A treatment comparison and long-term follow-up study. *Criminal Justice and Behavior, 25*, 203-225.
- McGrath, R. J., Lasher, M. P., & Hoke, S. E. (2009). *Vermont Assessment of Sex Offender Risk Manual Supplement – 2009 Revised Norms*. Middlebury, VT: Author.
- Miller, W. R., & Rollnick, S. (2002). *Motivational interviewing: Preparing for change* (2nd Edition). New York: Guilford.
- Mishra, S. & Lalumière, M. L. (2009). Is the crime drop of the 1990s in Canada and the USA associated with a general decline in risky and health-related behavior? *Social Science and Medicine, 68*, 39-48.
- Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62*, 783-792.

- Olver, M. E., Wong, S. C. P., Nicholaichuk, T. P., & Gordon, A. E. (2007). The validity and reliability of the Violence Risk Scale–Sexual Offender version: Assessing sex offender risk and evaluating therapeutic change. *Psychological Assessment, 19*, 318-329.
- Olver, M. E., & Wong, S. C. P. (2011). A comparison of static and dynamic assessment of sexual offender risk and need in a treatment context. *Criminal Justice and Behavior, 38*, 113-126.
- Pencina, M. J., D'Agostino, R. B., Sr., D'Agostino, R. B., Jr., & Vasan, R. S. (2007). Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Statistics in Medicine, 27*, 157-172.
- Phenix, A., Doren, D., Helmus, L., Hanson, R. K., & Thornton, D. (2009). *Coding rules for Static-2002*. Ottawa, Ontario: Public Safety Canada.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C.A. (1998). *Violent offenders: Appraising and managing risk* (2nd Ed.). 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.
- Shingler, J., & Mann, R. E. (2006). Collaboration in clinical work with sexual offenders: Treatment and risk assessment. In W. L Marshall et al. (Eds.), *Sexual offender treatment: Controversial issues* (pp. 225-239). West Sussex, England: Wiley.
- Shrout, P. E., & Fleiss, J. L. (1979). Interclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428.

- Skeem, J., Manchak, S., & Peterson, J. (2011). Correctional policy for offenders with mental illness: Creating a new paradigm for recidivism reduction. *Law and Human Behavior, 35*, 110-126.
- Skeem, J. L., & Monahan, J. (2011). Current directions in violence risk assessment. *Current Directions in Psychological Science, 20*, 38-42.
- Sofroniou, N., & Hutcheson, G. D. (2002). Confidence intervals for the predictions of logistic regression in the presence and absence of a variance-covariance matrix. *Understanding Statistics, 1*, 3-18.
- Sjöstedt, G., & Grann, M. (2002). Risk assessment: What is being predicted by actuarial prediction instruments? *International Journal of Forensic Mental Health, 1*, 179-183.
- Studer, L. H., & Reddon, J. R. (1998). Treatment may change risk prediction for sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 10*, 175-181.
- Thornton, D. (2002). Constructing and testing a framework for dynamic risk assessment. *Sexual Abuse: A Journal of Research and Treatment, 14*, 139-153.
- Vandervort-Clark, A. (2009). *Legislating sex offender management: Trends in state legislation 2007 and 2008*. Presentation to the 28th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Dallas, TX.
- United States Census Bureau (2010). *United States Census*. [Online] Available: <http://www.census.gov>.
- Yang, M., Wong, S. C. P., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin, 136*, 740-767.

Authors' Note

This research was supported by National Institute of Justice grant 2008-DD-BX-0013 to the Vermont Department of Corrections. The views expressed in the article do not necessarily represent those of the Vermont Department of Corrections. We thank Tammy Smith, Kristy Hetzel and David Farre for their assistance with data collection and entry, Brian Gilbar for data coding, and Karl Hanson, Naomi Freeman, Jeffery Sandler and Timothy Stickle for their helpful comments at various stages on this research.

Dissemination of Research Findings

Forthcoming Conference Presentations

McGrath, R. J., Lasher, M. P., & Cumming, G. F. (November, 2011). *The Sex Offender Treatment Intervention and Progress Scale (SOTIPS): Revisions and research updates*. Association for the Treatment of Sexual Abusers 30th Annual Research and Treatment conference, Toronto, Canada.

Manuscript in Preparation

McGrath, R. J., Lasher, M. P., & Cumming, G. F. (2011). *The Sex Offender Treatment Intervention and Progress Scale (SOTIPS): Psychometric properties and incremental validity with the Static-99R*. Manuscript in preparation.

Table 1. Descriptive and Interrater Reliability Statistics for Static Risk Assessment Instruments

	<i>N</i>	<i>M</i>	<i>SD</i>	Range	ICC	<i>n</i> scored by two raters
Static-99R	759	2.48	2.09	-3 to 9	.89	109
Static-2002R	759	3.55	2.13	-2 to 11	.89	109
VASOR	759	25.16	15.76	0 to 70	.81	109

Note. ICC = Interclass correlation coefficient. VASOR = Vermont Assessment of Sex Offender Risk.

Table 2. AUC Values for Static Risk Assessment Instruments

Three-year fixed follow-up									
Type of and percent recidivism (recidivists/total n)									
	Sexual 4.6% (35/759)		Violent 8.6% (65/759)		Any 23.1% (175/759)		Return to prison 40.6% (308/759)		
	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI	
Static-99R	.74***	.65-.83	.70***	.64-.76	.67***	.64-.73	.67***	.63-.71	
Static-2002R	.71***	.62-.81	.66***	.60-.73	.67***	.62-.71	.68***	.64-.72	
VASOR	.73***	.63-.82	.64***	.56-.71	.62***	.57-.66	.69***	.65-.73	
Five-year fixed follow-up									
	Sexual 5.6% (33/593)		Violent 12.8% (76/593)		Any 32.7% (194/593)		Return to prison 48.2% (286/593)		
	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI	
Static-99R	.66**	.56-.76	.68***	.62-.74	.71***	.67-.75	.66***	.61-.70	
Static-2002R	.64**	.53-.75	.65***	.59-.72	.68***	.64-.73	.66***	.61-.70	
VASOR	.69***	.58-.80	.63***	.56-.69	.62***	.57-.67	.67***	.62-.71	

Note. AUC = area under the curve of the receiver operating characteristic. CI = confidence interval. VASOR = Vermont Assessment of Sex Offender Risk.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 3. Predictive Accuracy of SOTIPS for Sexual Recidivism

	One-year follow-up						Three-year follow-up					
	% recidivists (recidivists/total n)						% recidivists (recidivists/total n)					
	Time 1		Time 2		Time 3		Time 1		Time 2		Time 3	
	2.3% (14/606)		1.8% (12/665)		1.5% (9/620)		5.3% (32/601)		4.1% (26/634)		3.4% (18/549)	
	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI
SOTIPS total score	.66*	.51-.81	.85***	.77-.93	.78**	.60-.97	.63*	.52-.73	.74***	.63-.84	.75***	.63-.87
SOTIPS factors												
Sexual deviance	.60	.45-.76	.79***	.66-.92	.78**	.61-.95	.61*	.51-.71	.71***	.61-.82	.74***	.63-.85
Criminality	.65*	.52-.79	.82***	.74-.89	.77**	.61-.93	.57	.47-.68	.71***	.61-.81	.72**	.60-.84
Social stability and support	.63	.48-.78	.82***	.74-.89	.74*	.57-.91	.60*	.50-.71	.73***	.63-.83	.72**	.61-.84
SOTIPS items												
Offense responsibility	.57	.41-.73	.76**	.62-.89	.65	.46-.84	.54	.42-.65	.65*	.53-.76	.57	.42-.72
Sexual interests	.55	.40-.70	.67*	.51-.83	.74*	.58-.90	.60	.49-.70	.70***	.59-.80	.72**	.60-.83
Sexual attitudes	.66*	.52-.80	.79***	.68-.90	.69	.48-.91	.62*	.51-.73	.68**	.56-.80	.68*	.54-.82
Sexual behavior	.63	.47-.79	.70*	.54-.86	.72*	.52-.92	.61*	.50-.71	.63*	.52-.75	.75***	.63-.88
Sexual risk management	.57	.42-.72	.69*	.53-.85	.69*	.50-.88	.60	.49-.70	.66**	.54-.77	.67*	.54-.80
Criminal attitudes	.68*	.54-.82	.77**	.66-.87	.80**	.64-.96	.60	.49-.71	.68**	.57-.79	.72**	.59-.84
Criminal behavior	.56	.41-.71	.75**	.63-.88	.72*	.51-.89	.54	.44-.65	.61	.50-.73	.64*	.50-.78
Emotion management	.63	.47-.79	.74**	.61-.87	.81**	.68-.93	.60	.49-.70	.63*	.52-.75	.73***	.63-.85
Problem solving	.68*	.55-.81	.79***	.67-.91	.74*	.56-.92	.63*	.53-.73	.71***	.61-.81	.72**	.60-.84
Impulsivity	.69*	.56-.82	.67*	.50-.85	.58	.30-.80	.60*	.50-.71	.62*	.50-.73	.63	.48-.78
Stage of change	.51	.37-.65	.65	.48-.82	.72*	.52-.91	.51	.41-.61	.62*	.50-.73	.69**	.56-.81
Treatment cooperation	.60	.44-.77	.65	.49-.81	.79**	.64-.94	.54	.43-.65	.65**	.53-.77	.72**	.59-.85
Supervision cooperation	.61	.46-.75	.76**	.63-.89	.71*	.52-.91	.51	.40-.61	.67**	.56-.78	.67*	.53-.81
Employment	.53	.39-.67	.67*	.53-.82	.53	.30-.75	.53	.42-.63	.66**	.54-.77	.57	.43-.73
Residence	.60	.44-.75	.64	.49-.79	.52	.33-.72	.59	.48-.70	.63*	.52-.75	.59	.45-.74
Social influences	.59	.46-.72	.75**	.63-.87	.71*	.54-.87	.55	.45-.65	.68**	.57-.78	.64*	.51-.76
SOTNPS deleted items												
Offense behavior admission	.55	.38-.73	.68*	.52-.85	.62	.42-.83	.50	.39-.62	.54	.41-.67	.60	.45-.75
Substance abuse	.49	.34-.64	.70*	.56-.85	.57	.37-.76	.48	.38-.59	.58	.46-.69	.51	.38-.65
Mental health stability	.61	.45-.77	.72*	.57-.86	.53	.32-.72	.69	.46-.67	.61	.49-.72	.53	.40-.66
Finances	.51	.36-.66	.62	.52-.73	.45	.24-.66	.53	.43-.63	.65**	.56-.75	.54	.39-.69
Adult love relationship	.57	.44-.69	.59	.46-.71	.71*	.55-.89	.55	.46-.65	.58	.48-.67	.61	.47-.74
Social involvement	.58	.47-.70	.55	.39-.71	.44	.26-.62	.51	.42-.61	.46	.34-.58	.56	.42-.69

Note. Time 1 = SOTIPS score at 0 to 3 months after client started treatment; Time 2 = 4 to 9 months; Time 3 = 10 to 15 months. AUC = area under the curve of the receiver operating characteristic. CI = confidence interval. SOTIPS = Sex Offender Treatment Intervention and Progress Scale. SOTNPS = Sex Offender Treatment Needs and Progress Scale.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 4. Predictive Accuracy of SOTIPS for Violent Recidivism

	One-year follow-up						Three-year follow-up					
	% recidivists (recidivists/total n)						% recidivists (recidivists/total n)					
	Time 1		Time 2		Time 3		Time 1		Time 2		Time 3	
	3.8% (23/604)		3.8% (25/661)		2.6% (17/611)		9.5% (57/599)		8.2% (52/631)		7.7% (42/542)	
	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI
SOTIPS total score	.66*	.54-.77	.74***	.64-.83	.71**	.58-.84	.63**	.55-.70	.69***	.60-.77	.67***	.58-.76
SOTIPS factors												
Sexual deviance	.58	.46-.69	.66**	.54-.78	.62*	.52-.71	.57	.49-.65	.63**	.55-.71	.61**	.53-.69
Criminality	.67**	.56-.77	.74***	.65-.83	.69***	.61-.77	.59*	.52-.67	.69***	.62-.77	.69***	.62-.75
Social stability and support	.63*	.52-.75	.73***	.64-.83	.66***	.58-.75	.64***	.57-.72	.69***	.61-.77	.66***	.69-.74
SOTIPS items												
Offense responsibility	.56	.44-.68	.70***	.60-.80	.67*	.54-.80	.53	.45-.61	.60*	.52-.68	.58	.49-.68
Sexual interests	.49	.37-.61	.58	.47-.70	.56	.42-.70	.55	.47-.63	.61*	.53-.69	.56	.47-.65
Sexual attitudes	.64*	.53-.75	.66**	.55-.78	.62	.47-.76	.60*	.52-.68	.62**	.54-.71	.60*	.51-.68
Sexual behavior	.55	.42-.67	.60	.48-.72	.56	.41-.71	.55	.47-.63	.60*	.51-.68	.57	.48-.67
Sexual risk management	.58	.47-.70	.60	.47-.73	.58	.44-.72	.57	.49-.64	.63**	.54-.71	.59*	.51-.68
Criminal attitudes	.69**	.59-.78	.70***	.61-.80	.75***	.64-.87	.61**	.54-.69	.67***	.59-.75	.67***	.59-.75
Criminal behavior	.59	.46-.72	.69**	.58-.80	.62	.48-.77	.54	.46-.63	.61**	.53-.70	.56	.46-.66
Emotion management	.60	.49-.72	.65**	.54-.77	.70**	.58-.82	.60*	.53-.68	.59*	.51-.68	.64**	.56-.73
Problem solving	.67**	.56-.78	.70***	.59-.82	.65*	.51-.79	.64***	.56-.71	.67***	.59-.75	.67***	.59-.62
Impulsivity	.67**	.56-.78	.65*	.53-.77	.61	.46-.76	.61**	.53-.69	.64***	.56-.72	.68***	.60-.76
Stage of change	.53	.42-.64	.59	.47-.71	.66*	.52-.80	.51	.43-.58	.58	.50-.66	.61*	.52-.71
Treatment cooperation	.62	.49-.74	.67**	.56-.78	.73***	.61-.86	.56	.48-.64	.64***	.56-.73	.68***	.60-.77
Supervision cooperation	.61	.49-.73	.68**	.57-.79	.63	.48-.77	.54	.46-.62	.64***	.56-.72	.60*	.50-.69
Employment	.51	.39-.63	.64*	.52-.76	.56	.42-.71	.55	.47-.63	.63**	.55-.72	.60	.47-.65
Residence	.61	.48-.74	.62*	.51-.72	.57	.43-.72	.61**	.53-.69	.63**	.55-.71	.58	.49-.67
Social influences	.62*	.52-.73	.70***	.61-.79	.66*	.54-.78	.59*	.52-.67	.65***	.58-.73	.59*	.51-.68
SOTNPS deleted items												
Offense behavior admission	.55	.42-.68	.65*	.53-.77	.64*	.49-.79	.50	.42-.58	.58	.49-.67	.59*	.49-.69
Substance abuse	.58	.47-.70	.65*	.53-.77	.55	.41-.70	.55	.47-.64	.56	.48-.65	.53	.43-.62
Mental health stability	.55	.43-.67	.62*	.50-.74	.61	.47-.75	.56	.48-.64	.56	.48-.64	.55	.45-.64
Finances	.53	.41-.65	.60	.51-.69	.56	.41-.70	.56	.49-.64	.65***	.57-.72	.59	.50-.68
Adult love relationship	.56	.46-.66	.60	.50-.68	.61	.50-.74	.54	.47-.62	.54	.47-.61	.56	.51-.68
Social involvement	.54	.44-.65	.52	.41-.63	.52	.40-.65	.52	.44-.59	.48	.39-.57	.56	.48-.62

Note. Time 1 = SOTIPS score at 0 to 3 months after client started treatment; Time 2 = 4 to 9 months; Time 3 = 10 to 15 months. AUC = area under the curve of the receiver operating characteristic. CI = confidence interval. SOTIPS = Sex Offender Treatment Intervention and Progress Scale. SOTNPS = Sex Offender Treatment Needs and Progress Scale.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 5. Predictive Accuracy of SOTIPS for Any Recidivism

	One-year follow-up						Three-year follow-up					
	% recidivists (recidivists/total n)						% recidivists (recidivists/total n)					
	Time 1 (64/602)		Time 2 (62/636)		Time 3 (55/577)		Time 1 (148/597)		Time 2 (130/608)		Time 3 (105/512)	
	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI
SOTIPS total score	.60**	.53-.68	.71***	.64-.77	.67***	.60-.74	.62***	.56-.67	.67***	.61-.72	.64***	.59-.70
SOTIPS factors												
Sexual deviance	.56	.48-.63	.62**	.55-.69	.58*	.52-.64	.55*	.50-.61	.61***	.55-.66	.73***	.67-.79
Criminality	.61**	.54-.69	.76***	.70-.82	.66***	.60-.72	.62***	.57-.67	.70***	.65-.75	.73***	.66-.79
Social stability and support	.58*	.50-.66	.65***	.59-.72	.64***	.58-.70	.61***	.56-.66	.64***	.59-.69	.74***	.69-.80
SOTIPS items												
Offense responsibility	.56	.49-.63	.61**	.54-.69	.62**	.54-.70	.56*	.50-.61	.58**	.53-.64	.58*	.52-.64
Sexual interests	.50	.42-.57	.55	.48-.62	.55	.46-.63	.48	.43-.54	.55	.50-.61	.52	.46-.58
Sexual attitudes	.58*	.51-.66	.63**	.55-.70	.60*	.52-.67	.57*	.51-.62	.60***	.54-.66	.57*	.51-.63
Sexual behavior	.51	.44-.59	.57	.50-.65	.56	.48-.64	.50	.45-.56	.57*	.51-.62	.54	.48-.60
Sexual risk management	.54	.46-.61	.58	.50-.65	.55	.47-.63	.55	.50-.61	.60***	.54-.65	.56*	.50-.62
Criminal attitudes	.61**	.53-.68	.73***	.67-.80	.67***	.60-.74	.62***	.56-.67	.68***	.63-.73	.66***	.60-.71
Criminal behavior	.57	.49-.65	.73***	.66-.80	.62**	.54-.70	.56*	.50-.62	.64***	.58-.70	.57*	.51-.63
Emotion management	.53	.45-.60	.60**	.52-.68	.66***	.58-.73	.56*	.51-.61	.57*	.52-.63	.61***	.55-.67
Problem solving	.58*	.51-.63	.65***	.57-.72	.62**	.55-.70	.61***	.56-.66	.65***	.60-.70	.65***	.59-.71
Impulsivity	.60**	.53-.67	.65***	.57-.72	.62**	.54-.69	.61***	.55-.66	.63***	.57-.68	.63***	.57-.69
Stage of change	.56	.48-.63	.60*	.52-.67	.60*	.52-.68	.55	.50-.61	.57**	.52-.63	.59**	.53-.65
Treatment cooperation	.57	.49-.65	.65***	.58-.73	.64***	.56-.72	.59***	.54-.64	.62***	.57-.68	.62***	.56-.68
Supervision cooperation	.60**	.53-.68	.69***	.61-.76	.58*	.50-.66	.58**	.53-.64	.63***	.58-.69	.56	.49-.62
Employment	.53	.46-.61	.56	.48-.63	.58*	.50-.66	.54	.49-.60	.56*	.50-.62	.58**	.51-.64
Residence	.59*	.51-.66	.61**	.54-.67	.60*	.52-.69	.58*	.53-.63	.60***	.54-.65	.56	.50-.62
Social influences	.57	.49-.64	.62**	.56-.69	.61**	.53-.68	.57*	.52-.63	.60***	.55-.65	.59**	.53-.65
SOTNPS deleted items												
Offense behavior admission	.55	.48-.63	.57	.49-.65	.56	.48-.64	.53	.48-.59	.54	.49-.60	.55	.49-.62
Substance abuse	.63***	.56-.71	.66***	.58-.74	.60*	.52-.69	.60***	.54-.65	.60***	.55-.66	.58**	.52-.65
Mental health stability	.50	.43-.58	.53	.45-.60	.56	.48-.64	.49	.44-.55	.51	.47-.57	.51	.46-.57
Finances	.57	.49-.64	.59*	.53-.65	.62**	.52-.68	.57*	.52-.62	.61***	.56-.66	.60**	.54-.66
Adult love relationship	.53	.46-.60	.54	.47-.61	.55	.47-.62	.52	.47-.58	.52	.47-.57	.52	.46-.58
Social involvement	.54	.46-.62	.48	.41-.55	.47	.40-.54	.52	.46-.57	.49	.42-.53	.50	.44-.56

Note. Time 1 = SOTIPS score at 0 to 3 months after client started treatment; Time 2 = 4 to 9 months; Time 3 = 10 to 15 months. AUC = area under the curve of the receiver operating characteristic. CI = confidence interval. SOTIPS = Sex Offender Treatment Intervention and Progress Scale. SOTNPS = Sex Offender Treatment Needs and Progress Scale.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 6. Predictive Accuracy of SOTIPS for Return to Prison

	One-year follow-up						Three-year follow-up					
	% recidivists (recidivists/total n)						% recidivists (recidivists/total n)					
	Time 1 (159/597)		Time 2 (124/596)		Time 3 (80/510)		Time 1 (262/592)		Time 2 (196/577)		Time 3 (135/465)	
	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI
SOTIPS total score	.74***	.69-.78	.77***	.73-.82	.77***	.72-.83	.72***	.68-.76	.76***	.71-.80	.74***	.69-.79
SOTIPS factors												
Sexual deviance	.66***	.61-.71	.70***	.65-.75	.73***	.67-.79	.64***	.59-.68	.68***	.63-.72	.66***	.61-.71
Criminality	.74***	.70-.79	.75***	.70-.80	.73***	.66-.79	.71***	.67-.75	.74***	.70-.79	.71***	.66-.77
Social stability and support	.72***	.68-.77	.75***	.70-.80	.74***	.69-.80	.72***	.68-.76	.74***	.70-.78	.74***	.69-.78
SOTIPS items												
Offense responsibility	.60***	.55-.65	.64***	.58-.70	.64***	.57-.70	.60***	.55-.64	.60***	.55-.65	.58**	.52-.64
Sexual interests	.59**	.54-.64	.63***	.58-.69	.63***	.57-.70	.59***	.54-.64	.61***	.57-.66	.59**	.54-.65
Sexual attitudes	.64***	.58-.69	.66***	.60-.71	.66***	.59-.72	.61***	.57-.66	.64***	.59-.69	.60**	.54-.65
Sexual behavior	.58**	.53-.64	.60***	.54-.65	.63***	.56-.70	.56*	.51-.61	.60***	.55-.65	.58**	.53-.64
Sexual risk management	.61***	.56-.66	.65***	.59-.70	.65***	.59-.72	.60***	.55-.64	.63***	.59-.68	.63***	.57-.68
Criminal attitudes	.70***	.65-.75	.72***	.66-.77	.69***	.62-.75	.69***	.65-.74	.70***	.66-.75	.68***	.62-.73
Criminal behavior	.68***	.63-.73	.71***	.66-.77	.67***	.60-.74	.65***	.60-.69	.67***	.62-.72	.63***	.57-.67
Emotion management	.60***	.55-.66	.63***	.57-.68	.64***	.56-.71	.58***	.54-.63	.62***	.57-.67	.64***	.58-.69
Problem solving	.65***	.60-.70	.65***	.60-.71	.69***	.63-.76	.64***	.59-.68	.67***	.62-.71	.70***	.65-.75
Impulsivity	.66***	.61-.71	.67***	.61-.72	.64***	.58-.71	.67***	.62-.71	.67***	.62-.72	.66***	.60-.71
Stage of change	.64***	.59-.69	.64***	.58-.69	.69***	.63-.75	.62***	.57-.66	.63***	.59-.68	.66***	.60-.71
Treatment cooperation	.69***	.64-.74	.68***	.63-.73	.67***	.60-.73	.66***	.61-.70	.66***	.61-.70	.65***	.59-.71
Supervision cooperation	.70***	.65-.75	.68***	.63-.74	.66***	.59-.73	.64***	.60-.69	.67***	.62-.72	.65***	.59-.71
Employment	.66***	.61-.71	.68***	.62-.73	.63***	.56-.70	.65***	.60-.69	.66***	.62-.71	.60***	.54-.66
Residence	.64***	.59-.69	.68***	.62-.73	.64***	.57-.71	.63***	.59-.68	.66***	.61-.71	.62***	.56-.68
Social influences	.69***	.64-.74	.69***	.64-.74	.69***	.63-.76	.69***	.64-.73	.66***	.62-.71	.68***	.62-.73
SOTNPS deleted items												
Offense behavior admission	.58**	.52-.63	.61***	.55-.66	.61**	.54-.68	.56*	.52-.61	.58**	.53-.63	.55	.49-.61
Substance abuse	.59***	.54-.64	.62***	.56-.68	.61**	.54-.68	.61***	.57-.66	.62***	.57-.67	.61***	.55-.67
Mental health stability	.55	.50-.60	.57*	.51-.63	.57	.50-.64	.55*	.50-.59	.57**	.52-.62	.57*	.51-.63
Finances	.70***	.65-.75	.66***	.61-.71	.65***	.58-.72	.69***	.64-.73	.68***	.63-.72	.64***	.58-.69
Adult love relationship	.63***	.58-.68	.62***	.57-.67	.63***	.56-.69	.61***	.57-.66	.60***	.55-.64	.57*	.51-.62
Social involvement	.64***	.59-.69	.62***	.57-.68	.64***	.58-.71	.62***	.58-.67	.62***	.57-.66	.59**	.53-.64

Note. Time 1 = SOTIPS score at 0 to 3 months after client started treatment; Time 2 = 4 to 9 months; Time 3 = 10 to 15 months. AUC = area under the curve of the receiver operating characteristic. CI = confidence interval. SOTIPS = Sex Offender Treatment Intervention and Progress Scale. SOTNPS = Sex Offender Treatment Needs and Progress Scale.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 7. Factor Loadings for Exploratory Factor Analysis with Varimax Rotation of SOTIPS Items

	Sexual deviance	Criminality	Social stability and supports
Sexual attitudes	.789	.270	.220
Sexual interests	.761	.030	.290
Offense responsibility	.698	.250	.020
Sexual risk management	.630	.380	.170
Sexual behavior	.621	.130	.250
Stage of change	.502	.490	.111
Criminal behavior	.110	.820	.240
Supervision cooperation	.230	.782	.200
Criminal attitudes	.320	.727	.270
Impulsivity	.190	.551	.500
Treatment cooperation	.470	.540	.270
Residence stability	.080	.080	.765
Employment stability	.190	.180	.682
Problem solving	.240	.400	.590
Emotion management	.260	.260	.549
Social influences	.240	.390	.471

Note. Highest factor loading for each item is in boldface.

Table 8. Interrater Reliability for Pooled Samples

	ICC ₁	95% CI	ICC ₂	95% CI	F (df = 319, 320)
Total Scores	.77	.73-.81	.87	.84-.90	7.93
Factor Scores					
Sexual deviancy	.68	.61-.73	.81	.76-.85	5.19
Criminality	.76	.71-.80	.86	.83-.89	7.34
Social stability and supports	.69	.63-.75	.82	.78-.86	5.53
Individual Items					
Acceptance of responsibility	.52	.43-.59	.68	.61-.75	3.15
Sexual interests	.47	.38-.55	.64	.55-.71	2.77
Sexual attitudes	.39	.29-.48	.56	.45-.65	2.28
Sexual behavior	.51	.43-.59	.68	.60-.74	3.12
Sexual risk management	.43	.33-.51	.60	.50-.68	2.49
Criminal attitudes	.43	.34-.52	.60	.50-.68	2.51
Criminal behavior	.63	.55-.69	.77	.71-.82	4.34
Emotion management	.40	.30-.49	.57	.47-.66	2.33
Problem solving	.52	.43-.59	.68	.60-.74	3.13
Impulsivity	.51	.42-.58	.67	.59-.74	3.05
Stage of change	.54	.46-.62	.71	.63-.76	3.39
Treatment cooperation	.63	.56-.69	.77	.72-.82	4.43
Supervision cooperation	.71	.66-.76	.83	.79-.87	5.98
Employment	.68	.62-.74	.81	.76-.85	5.25
Residence	.55	.47-.62	.71	.63-.76	3.41
Social Influences	.43	.33-.51	.60	.50-.68	2.50

Note. ICC = interclass correlation coefficient for a single rating (ICC₁) and for average ratings (ICC₂). For all ICCs, $p < .001$.

Table 9. Repeated Measures ANOVAs of SOTIPS Items and Total Scores for Sexual Non-recidivists and Recidivists at Three-Year Follow-up

	Sexual non-recidivists					Sexual recidivists			
	Time 1 n = 567	Time 2 n = 634	Time 3 n = 598	F ^a n = 434	Adjusted F ^b n = 17	Time 1 n = 39	Time 2 n = 31	Time 3 n = 22	F ^a n = 18
	M (SD)	M (SD)	M (SD)			M (SD)	M (SD)	M (SD)	
Offense responsibility	1.11 (0.90)	0.76 (0.79)	0.51 (0.69)	126.23***	12.35***	1.28 (1.03)	1.52 (2.00)	0.91 (0.97)	2.31
Sexual interests	0.96 (0.79)	0.79 (0.73)	0.67 (0.69)	47.61***	5.42**	1.31 (0.89)	1.39 (0.80)	1.23 (0.61)	2.87
Sexual attitudes	1.06 (0.80)	0.83 (0.73)	0.69 (0.71)	62.73***	9.66**	1.46 (0.88)	1.42 (0.92)	1.32 (0.89)	0.65
Sexual behavior	0.44 (0.69)	0.33 (0.59)	0.28 (0.57)	9.83***	5.85**	0.79 (0.98)	0.68 (0.87)	0.95 (0.84)	0.64
Risk management	1.23 (0.93)	0.87 (0.86)	0.69 (0.77)	75.42***	3.96*	1.54 (0.97)	1.29 (1.04)	1.14 (0.83)	2.13
Criminal attitudes	0.96 (0.90)	0.81 (0.85)	0.72 (0.83)	7.90***	3.51*	1.36 (1.04)	1.45 (1.00)	1.41 (1.10)	0.20
Criminal behavior	0.55 (0.80)	0.47 (0.75)	0.43 (0.71)	0.68	4.55*	0.72 (0.83)	0.81 (0.87)	0.86 (0.99)	1.24
Emotion management	1.08 (0.73)	0.91 (0.71)	0.87 (0.72)	16.38***	2.92	1.36 (0.81)	1.23 (0.85)	1.45 (0.80)	1.34
Problem solving	1.04 (0.77)	0.93 (0.74)	0.82 (0.74)	9.51***	7.51**	1.41 (0.85)	1.52 (0.77)	1.45 (0.86)	0.69
Impulsivity	0.86 (0.78)	0.75 (0.76)	0.63 (0.70)	17.94***	3.43*	1.26 (1.07)	1.16 (0.90)	1.09 (0.97)	0.00
Stage of change	1.46 (0.74)	1.10 (0.68)	0.91 (0.75)	99.36***	10.70***	1.51 (0.72)	1.48 (0.81)	1.50 (0.80)	0.72
Treatment cooperation	0.74 (0.84)	0.67 (0.78)	0.58 (0.75)	5.81**	7.59**	1.00 (1.00)	1.26 (1.06)	1.36 (0.90)	1.38
Supervision cooperation	0.60 (0.83)	0.50 (0.81)	0.43 (0.75)	2.51	1.00	0.69 (0.83)	1.03 (0.98)	0.95 (0.95)	2.68
Employment	0.95 (1.05)	0.80 (0.90)	0.72 (0.86)	5.79**	0.56	1.03 (1.01)	1.42 (1.06)	1.14 (1.13)	1.05
Residence	0.73 (0.84)	0.60 (0.75)	0.52 (0.72)	11.25***	0.58	1.10 (1.00)	1.06 (0.89)	1.00 (1.07)	0.14
Social influences	1.11 (0.83)	0.95 (0.76)	0.86 (0.78)	19.66***	2.86	1.28 (0.76)	1.58 (0.81)	1.23 (0.75)	2.13
Total Score	14.86 (8.43)	12.06 (7.96)	10.33 (7.56)	82.56***	16.77***	19.10 (9.10)	20.29 (10.33)	19.00 (10.15)	0.27

Note. Time 1 = SOTIPS score at 0 to 3 months after client started treatment; Time 2 = 4 to 9 months; Time 3 = 10 to 15 months.

AUC = area under the curve of the receiver operating characteristic. CI = confidence interval.

^a F-test ($df = 2, 902$)

^b Adjusted F-test ($df = 2, 902$) based on 5% of non-sexual recidivists randomly selected to correct for inflation of the F statistic.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 10. Sexual Recidivism by SOTIPS Risk Category and Score Time

		Time 1					
		One-year follow-up (AUC = .60; 95% CI = .45 - .75)			Three-year follow-up (AUC = .61*; 95% CI = .50 - .71)		
Risk category	Risk score	recidivists/ total <i>n</i>	%	recidivism 95% CI	recidivists/ total <i>n</i>	%	recidivism 95% CI
Low	0-10	3/210	1.4	0.0 – 3.1	8/208	3.9	1.2 – 6.5
Moderate	11-20	5/232	2.2	0.3 – 4.0	9/230	3.9	1.4 – 6.4
High	21-48	6/164	3.7	0.8 – 6.6	15/163	9.2	4.7 – 13.7
Overall	0-48	14/606	2.3	1.1 – 3.5	32/601	5.3	3.5 – 7.1
		Time 2					
		One-year follow-up (AUC = .81***; 95% CI = .73 - .90)			Three-year follow-up (AUC = .70***; 95% CI = .59 - .80)		
Low	0-10	0/317	0.0	0.0 – 0.0	5/302	1.7	0.2 – 3.1
Moderate	11-20	5/232	2.1	0.3 – 4.0	10/230	4.3	1.7 – 7.0
High	21-48	7/111	6.3	1.7 – 10.9	11/102	10.8	4.7 – 16.9
Overall	0-48	12/665	1.8	0.8 – 2.8	26/634	4.1	2.6 – 5.6
		Time 3					
		One-year follow-up (AUC = .77**; 95% CI = .59 - .95)			Three-year follow-up (AUC = .72**; 95% CI = .58 - .85)		
Low	0-10	2/354	0.6	0.0 – 1.4	5/316	1.6	0.2 – 3.0
Moderate	11-20	1/185	0.5	0.0 – 1.6	4/162	2.5	0.1 – 4.9
High	21-48	6/81	7.4	1.6 – 13.2	9/71	12.7	4.8 – 20.6
Overall	0-48	9/620	1.5	0.6 – 2.4	18/549	3.3	1.9 – 4.7

Note. Time 1 = SOTIPS score at 0 to 3 months after client started treatment; Time 2 = 4 to 9 months; Time 3 = 10 to 15 months.

AUC = area under the curve of the receiver operating characteristic. CI = confidence interval.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 11. Violent Recidivism by SOTIPS Risk Category and Score Time

		Time 1					
		One-year follow-up (AUC = .62*; 95% CI = .50 - .74)			Three-year follow-up (AUC = .60*; 95% CI = .53 - .68)		
Risk category	Risk score	recidivists/ total <i>n</i>	%	recidivism 95% CI	recidivists/ total <i>n</i>	%	recidivism 95% CI
Low	0-10	5/210	2.4	0.3 – 4.5	13/208	6.3	2.9 – 9.6
Moderate	11-20	7/232	3.0	0.8 – 5.2	20/230	8.7	5.0 – 12.4
High	21-48	11/162	6.8	2.8 – 10.7	24/161	14.9	9.4 – 20.5
Overall	0-48	23/604	3.8	2.3 – 5.3	57/599	9.5	7.1 – 11.8
		Time 2					
		One-year follow-up (AUC = .70***; 95% CI = .60 - .80)			Three-year follow-up (AUC = .64***; 95% CI = .56 - .72)		
Risk category	Risk score	recidivists/ total <i>n</i>	%	recidivism 95% CI	recidivists/ total <i>n</i>	%	recidivism 95% CI
Low	0-10	4/317	1.3	0.0 – 2.5	14/302	4.6	2.3 – 7.0
Moderate	11-20	11/236	4.7	2.0 – 7.4	21/229	9.2	5.4 – 12.9
High	21-48	10/108	9.3	3.7 – 14.8	17/100	17.0	9.5 – 24.5
Overall	0-48	25/661	3.8	2.3 – 5.2	52/631	8.2	6.0 – 10.4
		Time 3					
		One-year follow-up (AUC = .69**; 95% CI = .55 - .83)			Three-year follow-up (AUC = .65**; 95% CI = .56 - .74)		
Risk category	Risk score	recidivists/ total <i>n</i>	%	recidivism 95% CI	recidivists/ total <i>n</i>	%	recidivism 95% CI
Low	0-10	5/353	1.4	0.2 – 2.7	15/315	4.8	2.4 – 7.1
Moderate	11-20	5/181	2.8	0.4 – 5.2	14/160	8.8	4.3 – 13.2
High	21-48	7/77	9.1	2.5 – 15.7	13/67	19.4	9.7 – 29.1
Overall	0-48	17/611	2.8	1.6 – 4.3	42/542	7.7	5.5 – 10.0

Note. Time 1 = SOTIPS score at 0 to 3 months after client started treatment; Time 2 = 4 to 9 months; Time 3 = 10 to 15 months.

AUC = area under the curve of the receiver operating characteristic. CI = confidence interval.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 12. Any Criminal Recidivism by SOTIPS risk category and Score Time

		Time 1					
		One-year follow-up (AUC = .59*; 95% CI = .51 - .66)			Three-year follow-up (AUC = .61***; 95% CI = .56 - .66)		
Risk category	Risk score	recidivists/ total <i>n</i>	%	95% CI	recidivists/ total <i>n</i>	%	95% CI
Low	0-10	16/210	7.6	4.0 – 11.2	34/208	16.3	11.3 – 21.4
Moderate	11-20	23/230	10.0	6.1 – 13.9	57/228	25.0	19.3 – 30.7
High	21-48	25/162	15.4	9.8 – 21.1	57/161	35.4	27.9 – 42.9
Overall	0-48	64/602	10.6	8.1 – 13.1	148/597	24.8	21.3 – 28.3
		Time 2					
		One-year follow-up (AUC = .67***; 95% CI = .60 - .74)			Three-year follow-up (AUC = .62***; 95% CI = .56 - .67)		
Risk category	Risk score	recidivists/ total <i>n</i>	%	95% CI	recidivists/ total <i>n</i>	%	95% CI
Low	0-10	15/310	4.8	2.4 – 7.2	43/296	14.5	10.5 – 18.6
Moderate	11-20	25/224	11.2	11.2 – 15.3	53/217	24.4	18.7 – 30.2
High	21-48	22/102	21.6	13.5 – 29.7	34/95	35.8	26.0 – 45.6
Overall	0-48	62/636	9.7	7.3 – 11.2	130/608	21.4	18.1 – 24.8
		Time 3					
		One-year follow-up (AUC = .63**; 95% CI = .55 - .71)			Three-year follow-up (AUC = .61***; 95% CI = .55 - .67)		
Risk category	Risk score	recidivists/ total <i>n</i>	%	95% CI	recidivists/ total <i>n</i>	%	95% CI
Low	0-10	21/341	6.2	3.6 – 8.7	46/304	15.1	11.1 – 19.2
Moderate	11-20	22/170	12.9	7.8 – 18.0	38/151	25.2	18.2 – 32.2
High	21-48	12/66	18.2	8.6 – 27.7	21/57	36.8	23.9 – 49.8
Overall	0-48	55/577	9.5	7.3 – 11.7	105/512	20.5	17.0 – 24.0

Note. Time 1 = SOTIPS score at 0 to 3 months after client started treatment; Time 2 = 4 to 9 months; Time 3 = 10 to 15 months.

AUC = area under the curve of the receiver operating characteristic. CI = confidence interval.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 13. Return to Prison Rates by SOTIPS Risk Category and Score Time Period

		Time 1					
		One-year follow-up (AUC = .72***; 95% CI = .67 - .76)			Three-year follow-up (AUC = .70***; 95% CI = .65 - .75)		
Risk category	Risk score	recidivists/ total <i>n</i>	%	recidivism 95% CI	recidivists/ total <i>n</i>	%	recidivism 95% CI
Low	0-10	21/208	10.1	6.0 – 14.2	48/206	23.3	17.5 – 29.1
Moderate	11-20	57/230	24.8	19.2 – 30.4	103/228	45.2	38.7 – 51.7
High	21-48	81/159	50.9	43.1 – 58.8	111/158	70.3	63.1 – 77.5
Overall	0-48	159/597	26.6	23.2 – 30.3	262/592	44.3	41.0 – 47.6
		Time 2					
		One-year follow-up (AUC = .74***; 95% CI = .68 - .79)			Three-year follow-up (AUC = .71***; 95% CI = .66 - .75)		
Risk category	Risk score	recidivists/ total <i>n</i>	%	recidivism 95% CI	recidivists/ total <i>n</i>	%	recidivism 95% CI
Low	0-10	27/307	8.8	5.6 – 12.0	55/295	18.6	14.2 – 23.1
Moderate	11-20	51/209	24.4	18.5 – 30.3	90/207	43.5	36.7 – 50.3
High	21-48	46/80	57.5	46.4 – 68.6	54/77	70.1	59.7 – 80.6
Overall	0-48	124/596	20.8	17.7 – 23.9	199/579	34.4	30.5 – 38.2
		Time 3					
		One-year follow-up (AUC = .73***; 95% CI = .67 - .80)			Three-year follow-up (AUC = .68***; 95% CI = .63 - .74)		
Risk category	Risk score	recidivists/ total <i>n</i>	%	recidivism 95% CI	recidivists/ total <i>n</i>	%	recidivism 95% CI
Low	0-10	22/322	6.8	4.1 – 9.6	52/293	17.7	13.4 – 22.2
Moderate	11-20	35/142	24.6	17.5 – 31.8	59/130	45.4	36.7 – 54.1
High	21-48	23/47	48.9	34.1 – 63.8	24/42	57.1	41.5 – 72.8
Overall	0-48	80/511	15.7	12.5 – 18.9	135/465	29.0	24.9 – 33.2

Note. Time 1 = SOTIPS score at 0 to 3 months after client started treatment; Time 2 = 4 to 9 months; Time 3 = 10 to 15 months.

AUC = area under the curve of the receiver operating characteristic. CI = confidence interval.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 14. Predictive Accuracy of Static-99R, SOTIPS and Combined Scores

	One-year follow-up						Three-year follow-up					
	Time 1		Time 2		Time 3		Time 1		Time 2		Time 3	
	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI
Sexual Recidivism												
% recidivism (recidivists/total)	2.3% (14/606)		1.8% (12/665)		1.5% (9/620)		5.3% (32/601)		3.9% (26/634)		2.9% (18/549)	
Static-99R	.80***	.68-.91	.81***	.72-.91	.57	.40-.75	.73***	.64-.83	.72***	.62-.82	.62	.49-.74
SOTIPS	.66*	.51-.81	.85***	.77-.93	.78**	.60-.97	.63*	.52-.73	.74***	.63-.84	.75***	.62-.87
Static-99 and SOTIPS	.74**	.63-.86	.89***	.83-.95	.73*	.57-.89	.70***	.61-.79	.76***	.67-.86	.72**	.61-.82
Violent Recidivism												
% recidivism (recidivists/total)	3.8% (23/604)		3.8% (25/661)		2.7% (17/611)		9.4% (57/599)		7.8% (52/631)		6.8% (42/542)	
Static-99R	.77***	.67-.86	.73***	.64-.82	.58	.46-.69	.72***	.65-.78	.70***	.63-.76	.64**	.57-.71
SOTIPS	.66*	.54-.77	.74***	.64-.83	.71**	.58-.84	.63**	.55-.70	.69***	.60-.77	.67***	.58-.76
Static-99 and SOTIPS	.75***	.66-.83	.78***	.69-.87	.69**	.58-.80	.70***	.67-.77	.71***	.64-.79	.69***	.62-.77
Any Recidivism												
% recidivism (recidivists/total)	10.6% (64/602)		9.3% (62/636)		8.9% (55/577)		24.4% (148/597)		19.5% (130/608)		16.9% (105/512)	
Static-99R	.67***	.61-.74	.71***	.64-.77	.65***	.59-.72	.69***	.65-.74	.69***	.65-.74	.68***	.63-.73
SOTIPS	.60**	.53-.68	.71***	.64-.77	.67***	.60-.74	.62***	.56-.67	.67***	.61-.72	.64***	.59-.70
Static-99 and SOTIPS	.67***	.60-.73	.74***	.68-.80	.68***	.61-.75	.69***	.65-.74	.70***	.65-.75	.70***	.64-.75
Return to Prison												
% recidivism (recidivists/total)	26.2% (159/597)		18.6% (124/596)		12.9% (80/510)		43.2% (262/592)		29.9% (196/577)		21.9% (135/465)	
Static-99R	.66***	.61-.71	.68***	.63-.73	.65***	.55-.67	.67***	.63-.71	.69***	.64-.73	.67***	.62-.72
Total SOTIPS	.74***	.69-.78	.77***	.73-.82	.77***	.72-.83	.72***	.68-.76	.76***	.71-.81	.74***	.69-.79
Static-99 and SOTIPS	.74***	.70-.78	.78***	.73-.82	.77***	.71-.82	.74***	.70-.78	.76***	.72-.80	.74***	.70-.79

Note. Time 1 = SOTIPS score at 0 to 3 months after client started treatment; Time 2 = 4 to 9 months; Time 3 = 10 to 15 months.

AUC = area under the curve of the receiver operating characteristic. CI = confidence interval;

* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 15. GEE Model Effects and Parameter Estimates for Sexual Recidivism

One-year follow-up						
Scale	Model effects			Parameter estimates		
	df	χ^2	Scale level	β	Std. Err.	Wald (df=1)
Intercept	1	270.67***		-1.55	.41	14.28***
Static-99R	3	14.31**	Low	-2.39	.72	11.09***
			Moderate-low	-1.83	.60	9.31**
			Moderate-high	-1.11	.56	3.87*
SOTIPS	2	14.44***	High ^a	0		
			Low	-1.93	.54	12.72***
			Moderate	-1.08	.40	7.27**
			High ^a	0		
Three-year follow-up						
Scale	Model effects			Parameter estimates		
	df	χ^2	Scale Level	β	Std. Err.	Wald (df=1)
Intercept	1	195.54***		-.89	.42	4.53*
Static-99R	3	13.21**	Low	-2.15	.69	9.81**
			Moderate-low	-1.60	.51	9.97**
			Moderate-high	-1.06	.52	4.23*
SOTIPS	2	15.52***	High ^a	0		
			Low	-1.35	.42	10.40**
			Moderate	-1.03	.30	11.98***
			High ^a	0		

Note. GEE = Generalized estimating equations.

^a Parameter is redundant to the intercept.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 16. GEE Model Effects and Parameter Estimates for Violent Recidivism

One-year follow-up						
Scale	Model effects		Scale Level	Parameter estimates		
	df	χ^2		β	Std. Err.	Wald (df=1)
Intercept	1	274.22***		-1.39	.38	13.21***
Static-99R	3	15.28**	Low	-2.35	.64	11.09***
			Moderate-low	-1.30	.45	9.31**
			Moderate-high	-.76	.47	3.87*
			High ^a	0		
SOTIPS	2	13.35**	Low	-1.34	.37	13.01***
			Moderate	-.78	.30	6.67**
			High ^a	0		

Three-year follow-up						
Scale	Model effects		Scale Level	Parameter estimates		
	df	χ^2		β	Std. Err.	Wald (df=1)
Intercept	1	184.44***		-.69	.37	3.38
Static-99R	3	17.44***	Low	-2.44	.59	16.94**
			Moderate-low	-.99	.39	6.36*
			Moderate-high	-.74	.42	3.06
			High ^a	0		
SOTIPS	2	11.81**	Low	-1.35	.42	10.40**
			Moderate	-1.03	.30	11.98***
			High ^a	0		

Note. GEE = Generalized estimating equations.

^a Parameter is redundant to the intercept.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 17. GEE Model Effects and Parameter Estimates for Any Recidivism

One-year follow-up						
Scale	Model effects		Scale Level	Parameter estimates		
	df	χ^2		β	Std. Err.	Wald (df=1)
Intercept	1	264.85***		-1.11	.34	10.84***
Static-99R	3	21.86***	Low	-1.49	.46	10.42**
			Moderate-low	-.61	.36	2.85
			Moderate-high	-.10	.37	.07
			High ^a	0		
SOTIPS	2	15.56***	Low	-.94	.24	13.01***
			Moderate	-.49	.22	6.67**
			High ^a	0		

Three-year follow-up						
Scale	Model effects		Scale Level	Parameter estimates		
	df	χ^2		β	Std. Err.	Wald (df=1)
Intercept	1	93.60***		-0.02	.33	<.01
Static-99R	3	37.07***	Low	-2.11	.43	24.45***
			Moderate-low	-.51	.34	2.31
			Moderate-high	-.18	.36	.26
			High ^a	0		
SOTIPS	2	15.62***	Low	-.81	.21	15.58***
			Moderate	-.47	.18	6.85**
			High ^a	0		

Note. GEE = Generalized estimating equations.

^a Parameter is redundant to the intercept.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 18. GEE Model Effects and Parameter Estimates for Return to Prison

One-year follow-up						
Scale	Model effects		Scale Level	Parameter estimates		
	df	χ^2		β	Std. Err.	Wald (df=1)
Intercept	1	71.38***		1.16	.37	10.06**
Static-99R	3	31.79***	Low	-1.81	.38	21.42***
			Moderate-low	-1.20	.36	11.23
			Moderate-high	-.66	.38	3.08
SOTIPS	2	130.15***	High ^a	0		
			Low	-2.31	.20	129.82***
			Moderate	-1.22	.17	48.95***
			High ^a	0		
Three-year follow-up						
Scale	Model effects		Scale Level	Parameter estimates		
	df	χ^2		β	Std. Err.	Wald (df=1)
Intercept	1	.867		1.72	.40	18.28***
Static-99R	3	42.11***	Low	-1.98	.41	22.71***
			Moderate-low	-.93	.39	5.69*
			Moderate-high	-.49	.41	1.45
SOTIPS	2	91.32***	High ^a	0		
			Low	-1.97	.21	87.41***
			Moderate	-.98	.19	26.60***
			High ^a	0		

Note. GEE = Generalized estimating equations.

^a Parameter is redundant to the intercept.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 19. Observed and Predicted Sexual Recidivism Rates for Combined SOTIPS and Static-99 Risk Levels

Risk levels		One-year follow-up						GEE predictions		
		Percent recidivism observed (recidivists/total n) and logistic regression predictions						Predicted	95% C.I.	
Static-99R	SOTIPS	Time 1		Time 2		Time 3				Overall
		Observed	Predicted	Observed	Predicted	Observed	Predicted			
Low	Low	0 (0/95)	0.52	0 (0/129)	<0.01	0.67 (1/149)	0.60	0.28	0.08	1.02
Low	Moderate	0 (0/53)	0.50	0 (0/50)	<0.01	0 (0/31)	0.61	0.65	0.20	2.08
Low	High	0.03 (1/33)	0.73	0 (0/12)	<0.01	9.09 (1/11)	8.28	1.91	0.52	6.68
Moderate-low	Low	0 (0/75)	0.72	0 (0/124)	<0.01	0 (0/142)	0.63	0.49	0.19	1.26
Moderate-low	Moderate	0 (0/107)	0.70	0.86 (1/116)	1.45	1.03 (1/97)	0.63	1.14	0.47	2.74
Moderate-low	High	0.03 (2/70)	1.02	4.84 (3/62)	3.73	9.76 (4/41)	8.54	3.28	1.17	8.88
Moderate-high	Low	8.57 (3/35)	3.94	0 (0/51)	<0.01	0 (0/52)	0.26	1.01	0.27	3.67
Moderate-high	Moderate	3.57 (2/56)	3.83	3.51 (2/57)	3.28	0 (0/45)	0.26	2.32	0.87	6.01
Moderate-high	High	2.22 (1/45)	5.51	7.69 (2/26)	8.19	5.00 (1/20)	3.73	6.54	3.10	13.25
High	Low	0 (0/5)	11.83	0 (0/13)	<0.01	9.09 (1/11)	0.69	3.01	0.97	8.95
High	Moderate	18.75 (3/16)	11.52	16.67 (2/12)	10.32	0 (0/12)	0.69	6.72	2.66	15.97
High	High	12.50 (2/16)	16.03	22.22 (2/9)	23.23	0 (0/9)	9.34	17.52	8.68	32.18
Total		2.31 (14/606)	2.31	1.80 (12/665)	1.80	1.45 (9/620)	1.45	1.85	1.21	2.81
		Three-year follow-up								
Low	Low	2.15 (2/93)	2.16	0.81 (1/123)	0.71	0.75 (1/134)	1.17	1.23	0.40	3.75
Low	Moderate	0 (0/52)	1.58	2.08 (1/48)	1.53	3.33 (1/30)	1.64	1.68	0.54	5.10
Low	High	6.06 (2/33)	3.53	0 (0/11)	3.60	9.09 (1/11)	8.52	4.57	1.52	12.96
Moderate-low	Low	1.33 (1/75)	3.01	1.67 (2/120)	1.54	0.79 (1/127)	1.99	2.10	1.00	4.37
Moderate-low	Moderate	0.93 (1/107)	2.21	1.75 (2/114)	3.29	3.49 (3/86)	2.77	2.85	1.51	5.32
Moderate-low	High	8.70 (6/69)	4.89	10.34 (6/58)	7.58	16.22 (6/37)	13.75	7.61	3.73	14.92
Moderate-high	Low	11.43 (4/35)	7.49	2.08 (1/48)	2.99	4.26 (2/47)	1.42	3.55	1.40	8.72
Moderate-high	Moderate	7.27 (4/55)	5.56	9.26 (5/54)	6.26	0 (0/36)	1.98	4.80	2.24	10.02
Moderate-high	High	6.67 (3/45)	11.82	8.70 (2/23)	13.86	6.67 (1/15)	10.14	12.40	6.94	21.18
High	Low	20.00 (1/5)	22.20	9.09 (1/11)	7.66	12.50 (1/8)	2.96	9.64	3.55	23.58
High	Moderate	25.00 (4/16)	17.20	14.29 (2/14)	15.25	0 (0/10)	4.11	12.74	6.09	24.75
High	High	25.00 (4/16)	32.11	30.00 (3/10)	30.23	14.29 (1/7)	19.32	29.06	15.26	48.23
Total		5.32 (32/601)	5.32	4.10 (26/634)	4.10	3.28 (18/549)	3.28	4.26	3.00	5.77

Note. GEE = Generalized estimating equations. Time 1 = SOTIPS score at 0 to 3 months after client started treatment; Time 2 = 4 to 9 months; Time 3 = 10 to 15 months.

Table 20. Observed and Predicted Violent Recidivism Rates for Combined SOTIPS and Static-99 Risk Levels

Risk levels		One-year follow-up						GEE predictions		
		Percent recidivism observed (recidivists/total n) and logistic regression predictions								
Static-99R	SOTIPS	Time 1		Time 2		Time 3		Overall	95% C.I.	
		Observed	Predicted	Observed	Predicted	Observed	Predicted	Predicted		
Low	Low	0 (0/95)	0.52	0.78 (1/129)	0.29	1.34 (2/149)	1.10	0.61	0.19	1.93
Low	Moderate	0 (0/53)	0.42	0 (0/50)	0.87	0 (0/31)	2.00	1.06	0.39	2.87
Low	High	3.03 (1/33)	0.87	0 (0/12)	1.59	9.09 (1/11)	6.75	2.30	0.77	6.72
Moderate-low	Low	0 (0/75)	2.16	1.61 (2/124)	1.30	0 (0/141)	1.84	1.73	1.01	2.96
Moderate-low	Moderate	0.93 (1/107)	1.76	2.61 (3/115)	3.77	4.21 (4/95)	3.33	2.99	1.73	5.12
Moderate-low	High	7.25 (5/69)	3.61	8.33 (5/60)	6.75	15.38 (6/39)	10.89	6.32	3.22	12.04
Moderate-high	Low	11.43 (4/35)	5.90	1.96 (1/51)	2.68	3.85 (2/52)	1.22	2.94	1.28	6.61
Moderate-high	Moderate	5.36 (3/56)	4.83	8.77 (5/57)	7.60	2.33 (1/43)	2.21	5.02	2.50	9.79
Moderate-high	High	4.55 (2/44)	9.61	12.00 (3/25)	13.19	0 (0/19)	7.44	10.37	6.01	17.31
High	Low	20.00 (1/5)	16.39	0 (0/13)	4.97	9.09 (1/11)	1.22	6.09	2.58	13.70
High	Moderate	18.75 (3/16)	13.69	21.43 (3/14)	13.51	0 (0/12)	2.23	10.17	4.83	20.20
High	High	18.75 (3/16)	24.94	18.18 (2/11)	22.38	0 (0/8)	7.48	19.88	10.48	34.48
Total		3.81 (23/604)	3.81	3.78 (25/661)	3.78	2.78 (17/611)	2.78	3.52	2.55	4.69
		Three-year follow-up								
Low	Low	2.15 (2/93)	1.99	1.63 (2/123)	1.29	1.49 (2/134)	1.83	1.64	0.63	4.25
Low	Moderate	0 (0/52)	1.99	2.08 (1/48)	2.08	3.33 (1/30)	2.75	2.26	0.87	5.75
Low	High	6.06 (2/33)	3.37	0 (0/11)	3.72	9.09 (1/11)	6.58	4.20	1.58	10.68
Moderate-low	Low	4.00 (3/75)	7.50	6.67 (8/120)	5.80	4.76 (6/126)	7.21	6.67	4.28	10.24
Moderate-low	Moderate	7.48 (8/107)	7.52	7.08 (8/113)	9.10	11.76 (10/85)	10.55	8.98	6.02	13.20
Moderate-low	High	16.18 (11/68)	12.25	17.54 (10/57)	15.36	28.57 (10/35)	22.70	15.76	10.16	23.63
Moderate-high	Low	17.14 (6/35)	11.89	6.25 (3/48)	8.00	12.77 (6/47)	6.13	8.44	4.63	14.89
Moderate-high	Moderate	12.73 (7/55)	11.91	16.67 (9/54)	12.37	8.57 (3/35)	9.01	11.30	6.76	18.29
Moderate-high	High	13.64 (6/44)	18.84	13.64 (3/22)	20.38	0 (0/15)	19.78	19.45	12.72	28.58
High	Low	40.00 (2/5)	27.25	9.09 (1/11)	14.60	12.50 (1/8)	7.23	16.14	7.97	29.96
High	Moderate	31.25 (5/16)	27.30	21.43 (3/14)	21.74	0 (0/10)	10.57	21.01	11.81	34.56
High	High	31.25 (5/16)	39.19	40.00 (4/10)	33.50	33.33 (2/6)	22.74	33.51	19.55	51.12
Total		9.52 (57/599)	9.52	8.24 (52/631)	8.24	7.75 (42/542)	7.75	8.52	6.76	10.68

Note. GEE = Generalized estimating equations. Time 1 = SOTIPS score at 0 to 3 months after client started treatment; Time 2 = 4 to 9 months; Time 3 = 10 to 15 months.

Table 21. Observed and Predicted Any Recidivism Rates for Combined SOTIPS and Static-99 Risk Levels

Risk levels		One-year follow-up						GEE predictions		
		Percent recidivism observed (recidivists/total n) and logistic regression predictions								
		Time 1	Time 2	Time 3	Overall	95% C.I.				
Static-99R	SOTIPS	Observed	Predicted	Observed	Predicted	Observed	Predicted	Predicted		
Low	Low	5.26 (5/95)	3.93	1.57 (2/127)	1.85	3.45 (5/145)	3.11	2.82	1.47	5.34
Low	Moderate	3.77 (2/53)	4.14	6.00 (3/50)	3.60	3.23 (1/31)	5.53	4.35	2.24	8.30
Low	High	3.03 (1/33)	6.29	0 (0/12)	7.16	10.00 (1/10)	7.81	6.92	3.56	13.01
Moderate-low	Low	4.00 (3/75)	7.85	4.92 (6/122)	4.82	5.84 (8/137)	7.60	6.58	4.66	9.21
Moderate-low	Moderate	6.67 (7/105)	8.24	5.66 (6/106)	9.13	13.95 (12/86)	13.06	9.94	7.32	13.37
Moderate-low	High	18.84 (13/69)	12.26	23.64 (13/55)	17.18	23.33 (7/30)	17.86	15.27	10.42	21.81
Moderate-high	Low	20.00 (7/35)	15.71	12.50 (6/48)	11.54	12.24 (6/49)	10.75	10.52	6.96	15.60
Moderate-high	Moderate	17.86 (10/56)	16.43	22.22 (12/54)	20.56	19.51 (8/41)	18.02	15.56	10.85	21.81
Moderate-high	High	18.18 (8/44)	23.41	29.17 (7/24)	34.84	16.67 (3/18)	24.14	23.13	16.47	31.47
High	Low	20.00 (1/5)	17.66	7.69 (1/13)	9.57	20.00 (2/10)	8.13	11.46	6.02	20.74
High	Moderate	25.00 (4/16)	18.45	28.57 (4/14)	17.35	8.33 (1/12)	13.91	16.86	9.20	28.88
High	High	18.75 (3/16)	26.03	18.18 (2/11)	30.25	12.50 (1/8)	18.96	24.88	14.64	39.00
Total		10.63 (64/602)	10.63	9.75 (62/636)	9.75	9.53 (55/577)	9.53	9.98	8.34	11.89
Risk levels		Three-year follow-up						GEE predictions		
		Time 1	Time 2	Time 3	Overall	95% C.I.				
		Static-99R	SOTIPS	Observed	Predicted	Observed	Predicted	Observed	Predicted	Predicted
Low	Low	7.53 (7/93)	6.32	3.31 (4/121)	4.19	3.85 (5/130)	4.64	5.02	2.65	9.31
Low	Moderate	5.77 (3/52)	7.94	10.42 (5/48)	6.11	6.67 (2/30)	6.41	6.91	3.60	12.86
Low	High	12.12 (4/33)	12.09	0 (0/11)	9.09	20.00 (2/10)	10.48	10.63	5.58	19.32
Moderate-low	Low	18.67 (14/75)	21.25	21.19 (25/118)	19.46	21.31 (26/122)	22.19	20.68	15.33	27.30
Moderate-low	Moderate	23.81 (25/105)	25.65	21.15 (22/104)	26.46	28.21 (22/78)	28.65	26.79	20.74	33.85
Moderate-low	High	41.18 (28/68)	35.49	42.31 (22/52)	35.61	46.15 (12/26)	40.70	36.97	27.48	47.59
Moderate-high	Low	31.43 (11/35)	29.67	26.09 (12/46)	25.79	28.89 (13/45)	24.55	26.66	18.78	36.37
Moderate-high	Moderate	40.00 (22/55)	35.03	37.25 (19/51)	34.10	33.33 (11/33)	31.42	33.78	25.21	43.58
Moderate-high	High	38.64 (17/44)	46.24	36.36 (8/22)	44.30	26.67 (4/15)	43.91	44.99	34.90	55.52
High	Low	40.00 (2/5)	35.90	18.18 (2/11)	28.20	28.57 (2/7)	26.46	30.35	17.74	46.83
High	Moderate	43.75 (7/16)	41.72	50.00 (7/14)	36.89	30.00 (3/10)	33.63	37.95	23.40	55.03
High	High	50.00 (8/16)	53.31	40.00 (4/10)	47.34	50.00 (3/6)	46.41	49.50	33.68	65.43
Total		24.79 (148/597)	24.79	21.38 (130/608)	21.38	20.51 (105/512)	20.51	22.30	19.46	25.43

Note. GEE = Generalized estimating equations. Time 1 = SOTIPS score at 0 to 3 months after client started treatment; Time 2 = 4 to 9 months; Time 3 = 10 to 15 months.

Table 22. Observed and Predicted Return to Prison Recidivism Rates of Combined SOTIPS and Static-99R Risk Levels

Risk levels		One -year follow-up						GEE predictions		
		Percent recidivism observed (recidivists/total n) and logistic regression predictions								
Static-99R	SOTIPS	Time 1		Time 2		Time 3		Overall	95% C.I.	
		Observed	Predicted	Observed	Predicted	Observed	Predicted	Predicted		
Low	Low	4.26 (4/94)	6.94	4.76 (6/126)	4.96	1.46 (2/137)	3.00	4.94	3.36	7.20
Low	Moderate	18.87 (10/53)	15.72	7.50 (3/40)	13.13	16.00 (4/25)	9.81	13.44	9.22	19.19
Low	High	39.39 (13/33)	36.80	66.67 (6/9)	38.88	33.33 (2/6)	24.04	34.41	25.21	44.95
Moderate-low	Low	9.33 (7/75)	9.05	6.61 (8/121)	8.41	11.45 (15/131)	9.39	8.66	6.32	11.77
Moderate-low	Moderate	16.82 (18/107)	19.94	20.95 (22/105)	21.01	22.37 (17/66)	26.72	22.10	17.70	27.22
Moderate-low	High	48.48 (32/66)	43.74	57.78 (26/45)	52.82	53.85 (14/26)	51.50	48.93	41.03	56.88
Moderate-high	Low	23.53 (8/34)	18.70	12.77 (6/47)	14.06	6.52 (3/46)	9.53	14.01	9.82	19.61
Moderate-high	Moderate	40.74 (22/54)	36.54	39.22 (20/51)	32.16	32.26 (10/31)	27.05	32.76	25.32	41.18
Moderate-high	High	55.56 (25/45)	64.25	50.00 (9/18)	66.61	50.00 (6/12)	51.91	62.20	52.49	71.03
High	Low	40.00 (2/5)	26.54	53.85 (7/13)	30.55	25.00 (2/8)	15.12	24.03	13.39	39.27
High	Moderate	43.75 (7/16)	47.48	46.15 (6/13)	56.04	40.00 (4/10)	38.53	48.60	32.63	64.85
High	High	73.33 (11/15)	73.84	62.50 (5/8)	84.29	33.33 (1/3)	64.60	76.15	60.91	86.75
Total		26.63 (159/597)	26.63	20.81 (124/596)	20.81	15.66 (80/511)	15.66	21.30	18.88	23.93
		Three-year follow-up								
Low	Low	9.78 (9/92)	13.00	6.56 (8/122)	8.54	7.14 (9/126)	7.95	9.73	6.82	13.72
Low	Moderate	28.85 (15/52)	24.80	17.50 (7/40)	20.39	24.00 (6/25)	20.89	22.54	16.10	30.62
Low	High	51.52 (17/33)	48.93	87.50 (7/8)	42.72	33.33 (2/6)	29.28	43.68	32.56	55.46
Moderate-low	Low	33.33 (25/75)	27.34	19.66 (23/117)	21.51	22.88 (27/118)	23.94	23.53	18.27	29.75
Moderate-low	Moderate	41.12 (44/107)	45.37	45.19 (47/104)	42.91	48.57 (34/70)	49.03	45.37	38.62	52.28
Moderate-low	High	70.77 (46/65)	70.70	68.18 (30/44)	68.64	6.67 (16/24)	60.13	68.87	60.43	76.22
Moderate-high	Low	35.29 (12/34)	38.62	40.00 (18/45)	32.00	33.33 (14/42)	26.39	32.16	24.09	41.47
Moderate-high	Moderate	60.38 (32/53)	58.13	54.00 (27/50)	56.34	50.00 (13/26)	52.29	56.13	46.60	65.24
Moderate-high	High	80.00 (36/45)	80.13	64.71 (11/17)	78.98	40.00 (4/10)	63.21	77.32	68.03	84.52
High	Low	40.00 (2/5)	48.14	54.55 (6/11)	45.47	28.57 (2/7)	37.88	43.74	26.88	62.17
High	Moderate	75.00 (12/16)	67.20	69.23 (9/13)	69.57	66.67 (6/9)	65.09	67.72	50.24	81.35
High	High	80.00 (12/15)	85.61	75.00 (6/8)	86.94	100.00 (2/2)	74.51	84.83	71.75	92.48
Total		44.26 (262/592)	44.26	34.37 (199/579)	34.37	29.03 (135/464)	29.03	36.38	32.91	39.99

Note. GEE = Generalized estimating equations. Time 1 = SOTIPS score at 0 to 3 months after client started treatment; Time 2 = 4 to 9 months; Time 3 = 10 to 15 months.

Table 23. GEE Predicted Sexual Recidivism Rates for Combined SOTIPS and Static-99R Risk Levels Across Times 1, 2, and 3

Static-99R Risk level and score		SOTIPS risk level and score					
		Recidivism rate at one year (n = 25/754)			Recidivism rate at three years (n = 39/749)		
		Low (0 to 10)	Moderate (11 to 20)	High (21 to 48)	Low (0 to 10)	Moderate (11 to 20)	High (21 to 48)
Low	(-3 to 1)	0.28	0.65	1.91	1.23	1.68	4.57
Moderate-low	(2 to 3)	0.49	1.14	3.28	2.10	2.85	7.61
Moderate-high	(4 to 5)	1.01	2.32	6.54	3.55	4.80	12.40
High	(6 to 12)	3.01	6.72	17.52	9.64	12.74	29.06

Note. GEE = Generalized estimating equations. Time 1 = SOTIPS score at 0 to 3 months after client started treatment; Time 2 = 4 to 9 months; Time 3 = 10 to 15 months.

Table 24. GEE Predicted Violent Recidivism Rates for Combined SOTIPS and Static-99R Risk Levels Across Times 1, 2, and 3

Static-99R Risk level and score		SOTIPS risk level and score					
		Recidivism rate at one year (n = 54/749)			Recidivism rate at three years (n = 80/746)		
		Low (0 to 10)	Moderate (11 to 20)	High (21 to 48)	Low (0 to 10)	Moderate (11 to 20)	High (21 to 48)
Low	(-3 to 1)	0.61	1.06	2.30	1.64	2.26	4.20
Moderate-low	(2 to 3)	1.73	2.99	6.32	6.67	8.98	15.76
Moderate-high	(4 to 5)	2.94	5.02	10.37	8.44	11.30	19.45
High	(6 to 12)	6.09	10.17	19.88	16.14	21.01	33.51

Note. GEE = Generalized estimating equations. Time 1 = SOTIPS score at 0 to 3 months after client started treatment; Time 2 = 4 to 9 months; Time 3 = 10 to 15 months.

Table 25. GEE Predicted Any Recidivism Rates for Combined SOTIPS and Static-99R Risk Levels Across Times 1, 2, and 3

		SOTIPS risk level and score					
		Recidivism rate at one year (n = 121/745)			Recidivism rate at three years (n = 213/740)		
Static-99R Risk level and score		Low (0 to 10)	Moderate (11 to 20)	High (21 to 48)	Low (0 to 10)	Moderate (11 to 20)	High (21 to 48)
Low	(-3 to 1)	2.82	4.35	6.92	5.02	6.91	10.63
Moderate-low	(2 to 3)	6.58	9.94	15.27	20.68	26.79	36.97
Moderate-high	(4 to 5)	10.52	15.56	23.13	26.66	33.78	44.99
High	(6 to 12)	11.46	16.86	24.88	30.35	37.95	49.50

Note. GEE = Generalized estimating equations. Time 1 = SOTIPS score at 0 to 3 months after client started treatment; Time 2 = 4 to 9 months; Time 3 = 10 to 15 months.

Table 26. GEE Predicted Return to Prison Rates for Combined SOTIPS and Static-99R Risk Levels Across Times 1, 2, and 3

Static-99R Risk level and score		SOTIPS risk level and score					
		Recidivism rate at one year (n = 256/734)			Recidivism rate at three years (n = 353/729)		
		Low (0 to 10)	Moderate (11 to 20)	High (21 to 48)	Low (0 to 10)	Moderate (11 to 20)	High (21 to 48)
Low	(-3 to 1)	4.94	13.44	34.41	9.73	22.54	43.68
Moderate-low	(2 to 3)	8.66	22.10	48.93	23.53	45.37	68.87
Moderate-high	(4 to 5)	14.01	32.76	62.20	32.16	56.13	77.32
High	(6 to 12)	24.03	48.60	76.15	43.74	67.72	84.83

Note. GEE = Generalized estimating equations. Time 1 = SOTIPS score at 0 to 3 months after client started treatment; Time 2 = 4 to 9 months; Time 3 = 10 to 15 months.

Table 27. Labels for Combined SOTIPS and Static-99R Risk Categories

Static-99R Risk level and score		SOTIPS risk level and score		
		Low (0 to 10)	Moderate (11 to 20)	High (21 to 48)
Low	(-3 to 1)	Low	Low	Moderate-low
Moderate-low	(2 to 3)	Low	Moderate-low	Moderate-high
Moderate-high	(4 to 5)	Moderate-low	Moderate-high	High
High	(6 to 12)	Moderate-high	High	High

Table 28. Recidivism Rates for Combined Static-99R and SOTIPS GEE risk levels

	Sex		Violent		Any		Return to prison	
	1 Year	3 Year	1 Year	3 Year	1 Year	3 Year	1 Year	3 Year
Low	0.12	1.25	0.59	3.13	4.19	11.55	7.36	16.73
Moderate-low	1.36	3.25	3.33	8.98	9.50	23.77	20.52	43.05
Moderate-high	3.89	8.73	7.65	16.21	20.12	38.76	45.15	60.92
High	7.69	12.74	9.70	17.76	20.25	39.47	52.86	72.59
Totals	1.85	4.26	3.52	8.52	9.98	22.30	21.30	36.38

Figure 1. Sexual Recidivism Rates at 3 Years by Combined Static-99R and SOTIPS GEE Risk Levels

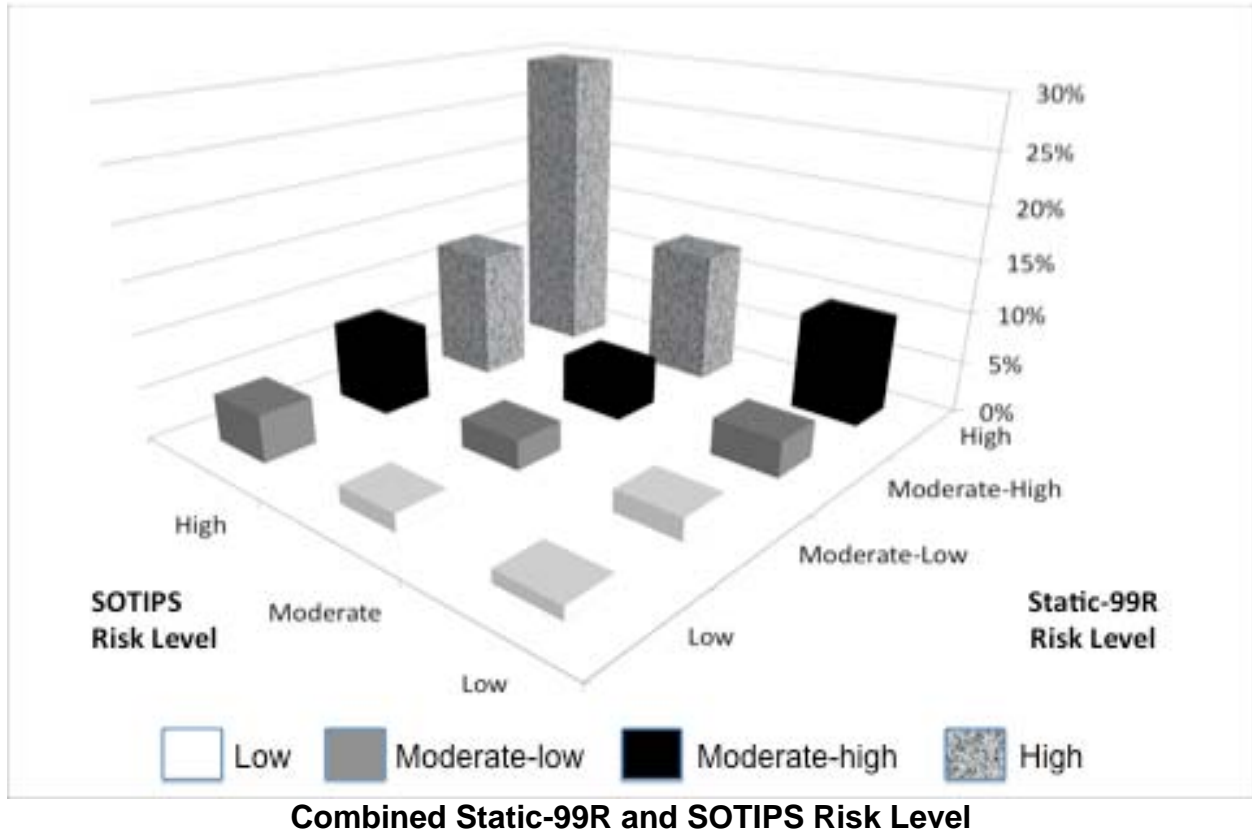
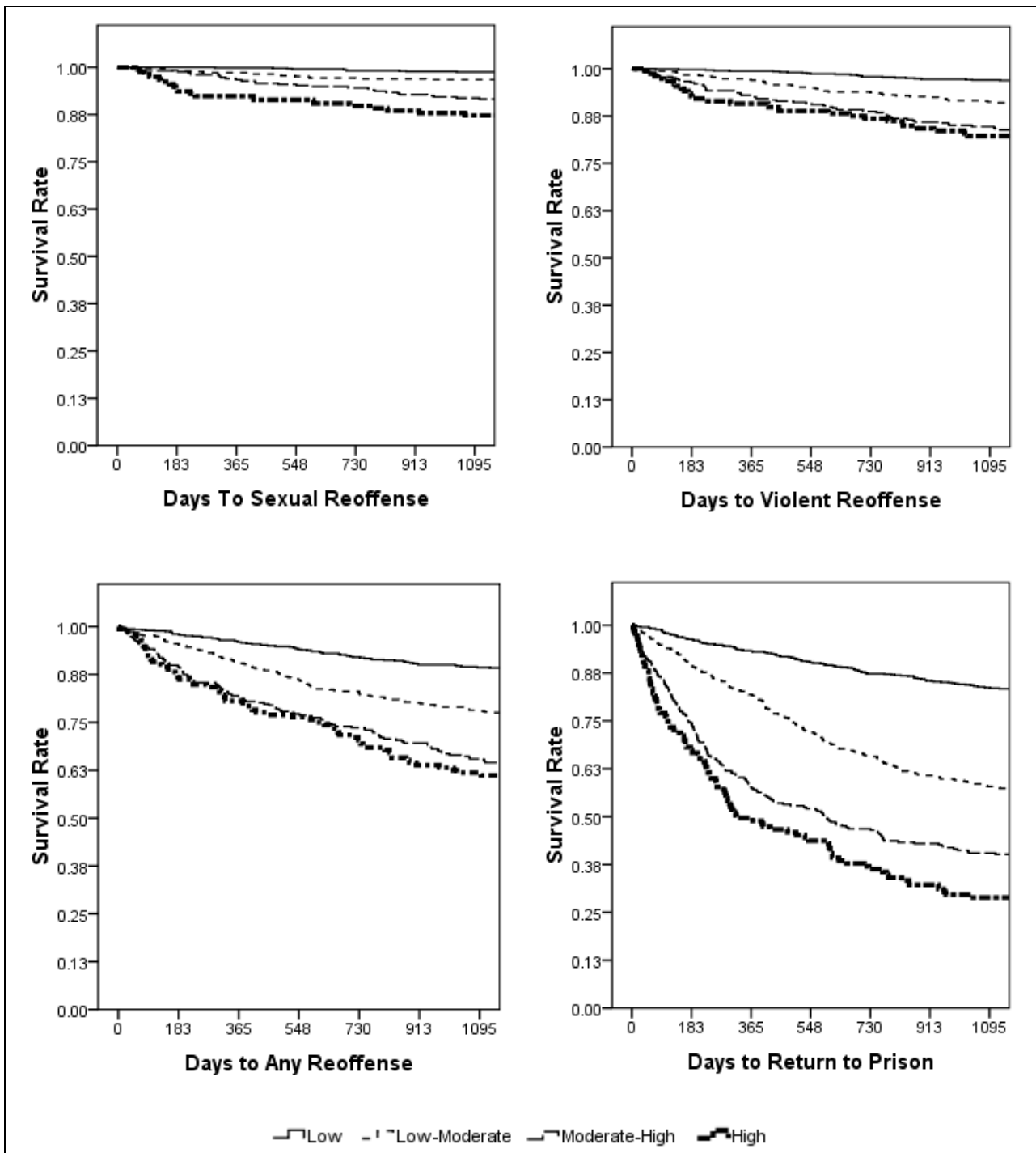


Figure 2. Interval-censored Survival Curves Based on GEE Analyses



Note: Survival rate shown as survival proportion of n during repeated 36-month follow-up; Days to Sexual Reoffense ($n = 749$), Days to Violent Reoffense ($n = 746$), Days to Any Reoffense ($n = 740$) and Days to Return to Prison ($n = 729$).