PREDICTING RECIDIVISM USING THE OFFENDER SCREENING TOOL

by

Cassandra Valerio

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF PSYCHOLOGY

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

WITH A MAJOR IN CLINICAL PSYCHOLOGY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2020

2

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by: Cassandra Valerio
titled: Predicting Recidivism Using the Offender Screening Tool

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

judith becker — Date: Feb 28, 2020
judith becker

Mary-Frances O'Connor — Date: Feb 18, 2020
Mary-Frances O'Connor

Connie J Beck — Date: 3/5/2020
Connie J. Beck

Date: Feb 28, 2020
Erika Lawrence

Date: Feb 28, 2020
Heidi Hamann

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

We hereby certify that we have read this dissertation prepared under our direction and recommend that it be accepted as fulfilling the dissertation requirement.

judith becker — Date: Feb 28, 2020
judith becker
psychology

Mary-Frances O'Connor — Date: Feb 18, 2020
Mary-Frances O'Connor
Psychology

Acknowledgments

This project was only possible because of the assistance and support of the Pima County Adult Probation Department, who provided access to their offices and records to facilitate data collection. Thank you especially to Barbara Johnson for her support of this project and for coordinating the data collection efforts within the probation department. Thank you also to Dr. Serena Gorgueiro for her assistance in developing the idea for this project.

I was also fortunate to have the knowledge and insight of my committee members to help me along the way on this research project. Thank you to Drs. Judith Becker, Connie Beck, Mary Frances O’Connor, Erika Lawrence, and Heidi Hamann for your guidance and feedback throughout this process.

Table of Contents

Abstract

An estimated 60% of sex offenders serve a term of probation or parole, and the level of risk posed by a particular offender has important implications for the amount of supervision required by probation. Thus, it is especially crucial that probation offices can accurately assess this risk. Probation officers in a growing number of jurisdictions now conduct risk assessments, help create case plans addressing criminogenic needs, match offenders to treatment services, and use motivational techniques to encourage behavior change. The purpose of the current project was to evaluate the Offender Screening Tool (OST), a risk assessment measure used by the probation department in a southwest county in Arizona to determine risk of reoffense and assign a level of supervision to sex offenders on probation. OST scores and recidivism data was collected from 174 adult male sex offenders supervised by the probation department. Participants were also scored on the Static-99R to compare the OST's performance with a well-established risk measure. The OST's accuracy in predicting sexual, violent nonsexual, and nonsexual nonviolent recidivism was measured. Results supported use of the OST for predicting nonsexual, nonviolent recidivism, while the Static-99 performed superiorly in predicting sexual reoffending. Implications for the use of the OST are discussed.

Introduction

Sexual violence represents a major public health concern in the United States. Nearly one in five adult women and one in 59 adult men report experiencing rape at some point in their lives. Additionally, 12.5% of women and 5.8% of men have experienced sexual coercion, and 27.3% of women and 10.8% of men have experienced unwanted sexual contact (Centers for Disease Control [CDC], 2014). Among children, an estimated 1 in 5 girls and 1 in 20 boys are victims of child sexual abuse (National Center for Victims of Crime, 2012). Sexual violence may result in significant psychological consequences for victims, including depression, anxiety, sexual dysfunctions, substance abuse, sleep-wake disorders, and trauma-related disorders (Martin, Macy, & Young, 2011; Långström et al., 2013). Additionally, research indicates that victims of childhood sexual abuse experience poorer long-term health outcomes than non-victims, including general health, pain, reproductive health, gastrointestinal health, cardiopulmonary symptoms, and obesity (Irish, Kobayashi, & Delahanty, 2009).

In order to reduce the high costs associated with sexual violence and assault, society has responded to sexual violence in a number of ways, including specific treatments for survivors of sexual abuse, treatment for adult and adolescent sexual offenders, as well as the development of risk assessment measures to determine offenders' risk to recidivate. The assessment of risk related to reoffending is a critical component of sex offender treatment and management. Risk assessment provides information that may inform a variety of decisions, including level of law enforcement supervision (e.g. restrictions placed on an individual as conditions of probation or parole), length of sentence, and type and intensity of treatment services (Association for the Treatment of Sexual Abusers [ATSA], 2012). Public perception regarding these offenders, their treatment, and the policies needed to maintain public safety have generally been incongruent with research. For example, public perception of the risk posed by these offenders has contributed to the development of laws that require registration, tracking, and community notification of sex offenders in order to protect the public. The Sex Offender Registration and Notification Act (SORNA), enacted in 2006, established a national sex offender registration system that facilitates the monitoring and tracking of sex offenders following their release into the community (Department of Justice [DOJ], 2018). SORNA provides information including name, current location, and prior sexual offenses to law enforcement, as well as publicly accessible websites with information about convicted sex offenders in all 50 states, the District of Columbia, and some United States territories. Many individual states also have additional registration and notification laws (DOJ, 2018). There is little empirical research regarding the effectiveness of these laws in reducing offending or reoffending (Zgoba & Levenson, 2012), and the laws that currently exist have been criticized for being overly inclusive and using limited law enforcement resources to target all sex offenders, when in reality many of these offenders are at a relatively low risk to reoffend (Vess, Day, Powell & Graffam, 2014). Additionally, offenders' noncompliance with these laws has not been shown to predict recidivism (Zgoba & Levenson, 2012).

An area where the intersection of sex offender risk and supervision is particularly important is probation. Nearly 60% of sex offenders serve a term of probation or parole (Meloy, 2005), and the level of risk posed by a particular offender has important implications for the amount of supervision required by probation. Thus, it is especially crucial that probation offices are able to accurately assess this risk. Over the past several years, many probation offices have implemented more evidence-based practices in their work with offenders, and the role of probation officers has changed dramatically (Lewis, Lewis, & Garby, 2013). While probation officers’ primary role traditionally has been the monitoring of offenders’ compliance with court orders, many are now being utilized as an important part of the behavioral change process. Probation officers in a growing number of jurisdictions now conduct risk assessments, help create case plans addressing criminogenic needs, match offenders to treatment services, and use motivational techniques to encourage behavior change (Lewis, Lewis, & Garby, 2013). Recognizing the expanding role that many probation departments have, the Bureau of Justice Assistance (BJA, 2011) released a report outlining steps that probation departments can take to help reduce recidivism among the offenders they supervise. One of the steps detailed in the report involves improving the screening and assessment methods used by probation, and subsequently using the results of assessments to tailor supervision plans. The report notes that while many probation departments have assessment instruments that were developed over time within the department, few of these measures have been tested empirically (BJA, 2011).

The present study will evaluate the utility of a risk and needs assessment instrument for adult sexual offenders used by the adult probation department in a southwest county in Arizona. The results of this research could contribute to a more accurate assessment of both risk of recidivism and more clearly define the treatment needs of these offenders.

In order to predict recidivism and develop effective risk assessment measures, evaluators conducting risk assessments should be aware of empirically supported risk factors. A summary of research findings regarding predictors of sexual recidivism is described below.

Predictors of Recidivism

While research has found that the base rate for sexual recidivism is relatively low, often in the 10-15% range (Helmus, Hanson, Thornton, Babchishin, & Harris, 2012), recidivism rates vary widely across different settings and samples. A number of individual studies and several meta-analyses have identified risk factors for recidivism (e.g. Hanson & Morton-Borgon, 2005). In order to be considered a risk factor, evidence must exist that there is a meaningful relationship between the characteristic or trait of concern and risk of reoffense (Eher, Matthes, Schilling, Haubner-MacLean, & Rettenberger, 2012). Risk factors can be broadly classified as either static or dynamic. Static risk factors are historical and unchanging factors, such as prior convictions for sexual offenses; dynamic risk factors are those that are considered amenable to change and may be the focus of intervention (Hanson & Morton-Bourgon, 2005). Dynamic factors may be further divided into stable and acute dynamic factors. Stable dynamic factors are potentially changeable but often persist over time, such as impulsivity, while acute dynamic factors may change in a short period of time, such as intoxication or other acute stressor (Hanson, Harris, Scott, & Helmus, 2007). In a comprehensive meta-analysis of recidivism among sexual offenders, Hanson and Morton-Bourgon (2005) identified a number of characteristics that predicted sexual, violent, and general recidivism among adult and adolescent sexual offenders. For both adults and adolescents, deviant sexual interests (defined as enduring attraction to illegal or unusual sexual acts) and antisocial orientation (including characteristics such as antisocial personality, a history of rule violations, and antisocial traits such as impulsivity and substance use) had the strongest association with sexual recidivism. Antisocial orientation was also the strongest predictor of violent and general recidivism. Dynamic factors that could potentially be targeted in sex offender treatment, such as sexual preoccupation and general self-regulation deficits, were also identified. Notably, the study found that several variables commonly addressed in sexual offender treatment programs were not predictive of recidivism, indicating that focusing treatment time on these variables may not represent the most effective way to reduce sexual offending. These variables included denial of the sexual offense, low victim empathy, psychological distress, and stated motivation for treatment.

Several static factors are also associated with increased risk, such as those included in common risk assessment measures (Hanson & Thornton, 1999). Static factors associated with increased risk include younger age, prior sexual offenses, offenses against unrelated or stranger victims, offenses against male victims, and the offender's relationships history (where lack of long-term relationship history indicates higher risk). These factors are historical and therefore not subject to change through intervention.

While research has identified several empirically supported risk factors for recidivism, the methodology for determining how these factors collectively contribute to risk has changed over time. The prediction of behavior is an important tool for psychologists in general, and an especially crucial task of those involved in the treatment and management of offenders.

Actuarial and Clinical Prediction

Early predictions of risk primarily relied on unstructured clinical judgment, or an evaluator's unstructured assessment of an offender's risk level based on their own knowledge of the offender or experience A number of authors have questioned the validity of using unstructured clinical judgment to predict risk (e.g. Meehl, 1954; Quinsey, Harris, Rice, & Cormier, 2006). Dawes, Faust, and Meehl (1989) specified a number of problems with using pure clinical judgment that could be eliminated by use of actuarial methods. For example, actuarial methods always reach the same decision for a given data set, while human judges may have random fluctuations in judgment. Additionally, the authors noted that human judges may make predictions based on illusory correlations. In other words, humans tend to have difficulty distinguishing valid from invalid predictive variables and are susceptible to developing false beliefs about associations between variables. Actuarial methods, on the other hand, ensure that variables contribute to a given conclusion based on their actual predictive power and thus eliminate illusory correlations and random fluctuations in judgment. Further, clinical judgment may result in a “self-fulfilling prophecy,” where a clinician who has predicted a particular outcome may then act in ways that influence or make that outcome more likely. Meehl (1954) consistently found that statistical (or actuarial) prediction, or prediction that specifies well-defined decision-making rules in order to make a prediction, was at least as accurate, and in many cases more accurate, than human judgment alone. A more recent meta-analysis found similar results, with actuarial predictions outperforming clinical predictions by a margin of eight to one in the prediction of medical and psychological diagnoses and prognoses (Grove, Zald, Lebow, Snitz, & Nelson, 2000). Regarding the prediction of violent behavior specifically, the use of clinical judgment alone may result in inconsistent classification of risk across practitioners, as well as a tendency to overclassify the risk of offenders. Because of the low base rate of sexual offending, the risk of a false positive prediction is high (Witt, DelRusso, Oppenheim, & Ferguson, 1996).

An additional option for risk assessment is the use of structured clinical judgment, in which risk-relevant items to be considered are specified, but the ultimate evaluation of risk is determined by the evaluator’s professional judgment (Hanson & Morton-Bourgon, 2009). Previous research has generally found that actuarial measures of risk demonstrate better prediction of recidivism among sexual offenders (e.g. Barbaree, Seto, Langton, & Peacock, 2001; Hanson & Morton-Bourgon, 2009). Hanson and Morton-Bourgon’s (2009) meta-analysis of risk assessment measures found that empirically derived actuarial measures predicted sexual, violent, or any recidivism more accurately than unstructured professional judgment, while the accuracy of structured professional judgment was intermediate between the other two types of measures. The authors noted the limited number of studies investigating structured professional judgment, however, and thus the limited conclusions that could be drawn from the results.

Research on clinical and statistical prediction has provided insight to psychologists who assess individuals' risk of recidivism. The field of risk assessment has evolved in three generation identified by Bonta (1996) that mirror the methods of prediction outlined above.

Evolution of Risk Assessment

Bonta (1996) identified three generations of risk assessment. The first generation risk assessment measures were based on unstructured clinical judgment. Clinicians followed no standard set of questions to ask an offender or determine from collateral information, but rather clinicians asked any questions they deemed necessary to determine risk. The second generation of risk assessment relies on a standardized set of empirically derived factors to determine risk of reoffending. This method addresses the individual bias and arbitrary nature of classification that may occur when using unstructured clinical judgment alone. Second generation risk assessment generally includes only static factors to determine risk, and while these factors are theoretically related to recidivism, reliance only on static factors creates additional limitations. By definition, historical factors are unchanging, and therefore reassessment of an offender will produce the same level of risk, regardless of any intervention efforts that may reduce risk of recidivism. Additionally, second generation risk assessment does not take into consideration responsivity factors that may impact an offender's response to treatment. Rather, these measures may be most useful to determine the appropriate level of supervision for a particular offender (Ferguson, 2002). An additional limitation includes resistance from clinicians, who may believe their professional judgment to be more accurate than a standardized measure of risk. Third generation risk assessment builds on the strengths of second-generation assessment by including some static factors in the prediction of risk, but also includes dynamic factors. The inclusion of dynamic factors allows for reassessment of offenders, which may be useful in determining if an intervention is working for an offender and if their risk level has consequently been reduced (Bonta, 1996). Additionally, by viewing risk and needs as related concepts and including them in the same assessment, an assessment is strengthened in its utility in that it can assess level of risk or appropriate level of supervision, but can also determine what services are needed by the offender (Ferguson, 2002). A recent meta-analysis found that dynamic risk assessment instruments significantly predict recidivism among male sexual offenders and add modest predictive value over static risk assessments (van den Berg et al., 2018).

Many of the most commonly used risk assessment measures use an actuarial approach to determining risk, congruent with findings by Hanson and Morton Bourgon (2009) that these measures predict recidivism more accurately than structured or unstructured professional judgment. A smaller number of measures currently in use rely on structured professional judgment, as described below.

Existing Measures of Risk Assessment

A number of risk assessment instruments are currently in use by evaluators to predict risk of sexual and non-sexual recidivism. Some (e.g. the Static-99) have a large body of research supporting their use, while other measures are less well established. Several commonly used measures with at least some research support are outlined in this section.

Static-99/R and Static-2002/R

The most widely used risk assessment measure is the Static-99 (Hanson & Thornton, 1999), which is a ten-item actuarial measure for use with adult male sexual offenders. The measure's ten items include research-based, static risk factors for sexual recidivism, including the offender's age, prior convictions for sexual offenses, and victim characteristics such as relationship to the offender and victim sex. Risk factors are scored as present or absent, and an individual's score on each item is totaled to place the individual in a risk category: low, moderate-low, moderate-high, or high (Harris, Phenix, Thornton, & Hanson, 2003). While it is the most widely used measure, the Static-99 may not provide a comprehensive assessment of risk given that it measures only static factors. Dynamic factors that research indicates may motivate sexual offending, such as deviant sexual interests, are not assessed on the Static-99 (Hanson, 2006), though McGrath, Cumming, and Burchard (2003) found that about half of treatment programs surveyed used this measure. Many studies have investigated the predictive validity of the Static-99. Scores on the Static-99 have been found to be correlated with sexual and violent recidivism (e.g. Hanson & Thornton, 1999) and to significantly predict sexual recidivism in offenders following treatment (Beech, Beckett, & Fisher, 2000). Barbaree and colleagues (2001) found that the Static-99 demonstrated concurrent validity with several other risk assessment measures. It also significantly predicted sexual, serious, and general recidivism. The Static-99 has also shown utility in predicting sexual and violent recidivism in culturally diverse offenders, such as Sjöstedt and Långström's (2000) study of Swedish sexual offenders.

The Static-2002 (Hanson & Thornton, 2003) was designed to improve upon the Static-99. It is a brief actuarial measure that assesses many of the same predictors as the Static-99, the Static-2002 made several changes to standardize the measure's coding rules and to increase predictive accuracy. Research has supported the use of the revised version. Hanson, Helmus, and

Thornton (2010) found that the Static-2002 showed moderate ability in predicting sexual, violent, and general recidivism (respective areas under the receiver operating characteristic curve [AUCs] of .68, .71, and .70). Further, this study found that the Static-2002 was more accurate than the Static-99. Further, Martens, Rettenberger, and Eher (2015) found that the German adaptation of the Static-2002 was found to have large effect sizes for predicting sexual, violent, and general recidivism (AUCs .78, .75, and .75, respectively). It should be noted that this study found that the measure had incremental predictive validity beyond the Static-99 only for general recidivism. In 2009, the weighting of the age variable in both the Static-99 and Static-2002 was revised in order to reflect the growing number of older (50+) sexual offenders in the justice system, as risk tends to decline with older age (Helmus, Babchishin, & Hanson, 2009). These revised criteria resulted in the Static-99R and Static-2002R.

Violence Risk Appraisal Guide

Another commonly used risk assessment measure is the Violence Risk Appraisal Guide (VRAG; Harris, Rice, & Quinsey, 1993). VRAG scores are obtained by weighting the measure's 12 items: nine items related to the offender or the offense (e.g. age, victim gender, marital status, offense history, history of alcohol problems); whether the offender meets Diagnostic and Statistical Manual of Mental Disorders, 3rd edition (DSM-III; American Psychiatric Association [APA], 1980) criteria for schizophrenia or any personality disorder; and the offender's score on the Psychopathy Checklist Revised (PCL-R; Hare, 1991). The PCL-R is a particularly important part of this measure and contributes more than any other item to the total score. The VRAG has been shown to predict violent recidivism among male offenders, including sexual offenses that involve victim contact, though this does not include non-contact sexual offenses (Barbaree et al., 2001). Rice and Harris's (1997) study of nearly 300 sex offenders found that in the average 10 year follow-up period, VRAG scores correlated .44 with violent recidivism in general and .17 with sexual recidivism specifically. Additionally, Hastings, Krishnan, Tangney, and Stuewig (2011) investigated the predictive validity of the VRAG for both male and female offenders. They found that the measure predicted institutional misconduct and recidivism in the first year post-release for males but did not replicate these results for females.

While the VRAG is designed to assess the risk of violence generally, evaluators may also use the Sex Offender Risk Appraisal Guide (SORAG; Quinsey, Harris, Rice, & Cormier, 1998), a modification of the VRAG developed to assess risk of violent recidivism, including both contact and non-contact sexual offenses, among male sex offenders. The SORAG contains the same items as the VRAG with a few additions specifically related to sexual offending, including sexual offense history and phallometrically assessed deviant sexual interest (Quinsey et al., 1998). This measure has also shown utility in predicting violent recidivism, including sexual recidivism. Rice and Harris (1999) found that the VRAG and SORAG were highly correlated with one another, which is not unexpected given the large overlap in items. Additionally, both measures significantly predicted violent and sexual recidivism.

Risk Matrix 2000

The Risk Matrix 2000 (RM2000; Thornton et al., 2003) is a measure developed in the United Kingdom to assess risk of sexual and violent nonsexual recidivism among adult males convicted of sex offenses (Kingston, Yates, Firestone, Babchishin, & Bradford, 2008). The RM2000 uses a stepwise approach to assessing risk, such that risk for sexual aggression is determined by first considering historical factors such as number of sexual offenses and offender age, followed by assessment of aggravating factors including having male victims or stranger victims and noncontact offense history. The presence of two or three of these aggravating factors increases risk by one level, while presence of all four increases risk by two levels. Risk levels for the sexual subscale (RM: Sexual) fall into categories of low, medium, high, and very high risk (Kingston et al., 2008). The RM: Violent subscale assesses risk for violent recidivism using three items: offender age, violent convictions, and prior convictions for burglary. Both subscales (RM: Sexual and RM: Violent) are combined to produce an overall risk level for sexual or nonsexual violence (Kingston et al., 2008).

Thornton and colleagues (2003) validated the RM2000 on samples in the United Kingdom. The Sexual subscale significantly predicted sexual recidivism in two of the three cross-validation samples, while the Violent subscale significantly predicted violent recidivism in all three samples. The RM2000 has also shown utility in assessing risk in North American samples. Using an average follow-up period of 12 years, Kingston and colleagues (2008) found that the measure was significantly correlated with both the Static-99 and SORAG. Although overall predictive accuracy was somewhat lower than Thornton and colleagues' (2003) validation studies, the authors found medium to large effect sizes for the RM2000's predictive accuracy. Both subscales and the combined scale predicted both sexual and nonsexual violent recidivism. Both the Static-99 and SORAG were slightly more accurate in predicting sexual recidivism, and the SORAG was more accurate in predicting violent and general recidivism. Barnett, Wakeling, and Howard (2010) found overall moderate predictive accuracy over a two to four year follow-up period. Importantly, this study found that the RM2000 was promising in predicting risk for diverse subgroups of sex offenders, including those with internet sexual offenses (e.g. possession of child pornography).

STABLE-2007/ACUTE-2007

While the measures described above consider static risk factors only, there are also a number of dynamic risk factors associated with recidivism. Risk assessment measures that take into account these potentially changeable factors may have particular utility in treatment settings where dynamic factors may be targeted for intervention among sex offenders. The STABLE-2007/ACUTE-2007 (Hanson et al., 2007) combines static and dynamic risk factors to aid in decision-making about treatment and supervision levels. The authors recommend obtaining an offender's score on the Static-99 at intake, obtaining the score on the STABLE-2007 every six months, and obtaining the ACUTE-2007 score "at each offender contact" in order to determine risk of reoffense during the next month. Examples of stable dynamic risk factors assessed on the STABLE-2007 include poor problems solving skills, impulsivity, sex drive preoccupation, negative emotionality, and significant social influences, among others. Items are scored as 0 if the factor is not a problem, a 1 if the factor presents some concern, and a 2 if the factor is a definite concern. Acute dynamic risk factors the authors recommend assessing frequently with the ACUTE-2007 include victim access, hostility, rejection of supervision, emotional collapse, and substance abuse, among others. These factors are scored using the same coding scheme as the STABLE-2007 but the evaluator has the additional option to code an item as "IN-intervene now."

Several studies have investigated the utility of the STABLE-2007/ACUTE-2007 in predicting recidivism. Eher, Matthes, Schilling, Haubner-MacLean, and Rettenberger (2012) found that the STABLE-2007 was associated with general, violent, and sexual recidivism. Additionally, they found that STABLE-2007 added incrementally to the predictive accuracy of the Static-99 for general and violent recidivism, and added incrementally to the SORAG for predicting sexual recidivism. Similarly, Hanson, Helmus, and Harris (2015) recently found that psychologically meaningful risk factors assessed by the STABLE-2007 predicted sexual, violent, and general recidivism. Further, for cases that were complete, STABLE-2007 scores added incrementally to predictive accuracy over using Static-99 scores.

Historical Clinical Risk Management-20

In addition to the actuarial risk measures described previously, a number of evaluators rely on measures based on a structured professional judgment model, such as the Historical Clinical Risk Management-20 (HCR-20; Douglas, Hart, Webster, & Belfrage, 2013). Though not specific to predicting sexual violence, the third version (HCR-20V3) assesses 20 factors associated with risk of general violence and their relevance to the individual being assessed, including a history of violent behavior, history of antisocial behavior, substance abuse, unstable relationships, and history of problems with treatment and supervision response (Douglas et al., 2014). Risk factor are rated as not present, possibly or partially present, or present, and examples are provided to assist the evaluator in rating the presence of a particular factor. In addition to the presence of each factor, evaluators also rate the relevance of each factor to the evaluee, including the factor’s contribution to past violence, its likely influence on future violence, its impact on the evaluee’s capacity to use nonviolent problem-solving strategies, and the necessity to intervene with each particular variable in order to reduce risk (Strub, Douglas, & Nicholls, 2014).

Recent research investigating the HCR-20V3 has indicated the measures utility in estimating violence risk in several populations. Strub, Douglas, and Nicholls (2014) tested the validity of the risk factor presence and relevance ratings, as well as summary risk ratings in a sample of offenders and civil psychiatric patients released into the community. At the 4-6 week and 6-8 month follow-up times, they found evidence for the validity of both risk factor presence and relevance ratings, with effect sizes averaging in the moderate range, though it should be noted that analyses did not indicate incremental validity of the relevance ratings compared to presence ratings. Effect sizes for summary risk ratings ranged from moderate to large. The authors noted that the findings were not moderated by either subsample or gender. Additionally, research has indicated that Version 3 of the HCR-20 is more systematic and detailed in its approach to assessing risk compared to earlier versions (Eidhammer, Selmer, Bjøkly, 2013).

Sexual Violence Risk-20

The Sexual Violence Risk 20 (SVR-20) was developed as a counterpart to the HCR-20 to assess sexual violence risk. This measure contains 20 items divided into three clusters: psychosocial adjustment, sexual offenses, and future plans. Scores on these items may be summed to assess risk (i.e. mechanical approach), or an evaluator's structured professional judgment may be made based on an individual's score on the items. Risk level is rated as high, medium, or low based on the clinician's weighting of items. Hanson and Morton-Bourgen's (2009) meta-analysis found that the mechanical approach to scoring the measure produced a mean AUC of .68 for sexual recidivism and a mean AUC of .61 for any violent (including sexual) recidivism. Studies investigating the structured professional judgment approach to scoring the SVR-20 are limited. The same meta-analysis found that of the studies available, results were widely varied, with some indicating excellent predictive accuracy and others finding nonsignificant results.

In sum, evaluators have several empirically supported risk assessment instruments to choose from depending on the goal of the evaluation and the characteristics of the particular offender being evaluated. Hanson and Morton-Bourgon (2009) argue that the ideal instrument would have the largest relationship to recidivism, the smallest confidence interval, and the least variability across samples. The authors state that "[b]y these criteria, no single measure has yet to establish itself as clearly more accurate than other, similar measures" (p. 10). In line with this view, a large number of risk assessment instruments have been developed by individual programs or jurisdictions to evaluate risk of recidivism, which is the case in the county where the current study will be conducted.

Present Study

The current study evaluated a risk assessment tool developed by a county in Arizona. The goals of developing this measure included providing a broad assessment of offender risks and needs, assessing static and dynamic risk factors predictive of criminal behavior, providing information that could be used to determine an offender's risk of recidivism and to guide case management decisions, and being acceptable to staff (Stinson, 2002).

Offender Screening Tool (OST)

As part of an initiative to implement evidence-based practices, in 1998 the Maricopa County Adult Probation Department (MCAPD) developed the Offender Screening Tool (OST). Although the literature reviewed above supports the use of several existing risk assessment measures for sex offenders, the MCAPD elected to create a new measure for two major reasons. First, given the number of assessments conducted by the MCAPD each year, the cost to purchase an existing measure would be high. Second, there was concern that changes to the method for conducting an assessment would encounter resistance from staff. Development of a new measure allowed for input from probation department staff to facilitate successful implementation of the measure (Ferguson, 2002).

The OST is administered at the presentence phase and information used to score the OST is collected as part of a presentence interview of the offender and from collateral records

(Ferguson, 2002). The OST contains 44 total items that encompass both static factors (n=14) and dynamic factors (n=30) in 10 domains: vocational/financial (5 items), education (3 items), family and social relationships (8 items), residence and neighborhood (2 items), alcohol use (3 items), drug use (3 items), mental health (2 items), physical health (2 items), attitude (7 items), and criminal behavior (9 items) (Stinson, 2002). The OST focuses more intensely on those factors that are the strongest predictors of recidivism. For example, the location where a person lives can affect criminality; however, pro-criminal attitudes have a stronger relationship with reoffending. Thus, the residence and neighborhood category has two items, while the attitude category has seven items. This allows the stronger predictor to be weighted more heavily and make a greater contribution to an offender's overall score (Ferguson, 2002).

Each individual item is scored as 0 (risk factor is not present) or 1 (risk factor is present). Scores on each item are then added to form a total score. Cut-off scores for males are as follows: low risk (0-5), medium-low risk (6-10), medium high risk (11-17), and high risk (18 or more). These risk categories generated from the OST score also determine the level of probation supervision required for each offender. Those who fall in the low risk category require minimum supervision, medium-low and medium-high risk require medium supervision, and high risk requires maximum probation supervision. Thus, an individual's score on the OST is directly related to restrictions put on an individual's freedom during their term of probation.

Use of the OST was implemented in probation departments statewide in 2005 (Arizona Supreme Court, 2005) and the existing research supports its use with non-sexual offenders. A study conducted by independent researchers at the University of Cincinnati examined the predictive validity of the OST in fifteen counties throughout Arizona with a sample of over 3000 offenders. The study found that individuals' risk level as measured by the OST was positively associated with recidivism during the follow-up period, such that individuals who were classified as low risk had significantly fewer arrests than those who were classified as high risk (Lowenkamp, Latessa, & Betchel, 2008). Similarly, total OST score was significantly correlated with number of arrests during the follow-up period. Offenders with a higher overall OST score (and thus higher risk) had more arrests than those with lower scores (Lowenkamp et al., 2008). While these promising findings support use of the OST with nonsexual offenders, it has not been validated for use with sexual offenders and its accuracy in determining level of risk in this population is not yet known. Further, the normative data available for the OST is based on a sample of non-sexual offenders on probation in Arizona. Given the importance of assessing risk and needs for sexual offenders, determining whether the OST is appropriate for use with this population is of critical importance.

In the Arizona county where the present study was conducted, the adult probation department uses the OST to determine level of risk and the necessary level of supervision for both sexual and non-sexual offenders, though has not yet been evaluated whether this method accurately predicts the likelihood of sexual offenders on probation to reoffend. The current study aimed to determine whether the OST demonstrates utility in predicting sexual, violent, and non-violent non-sexual recidivism among sexual offenders.

Method

Specific Aims

The aims of the current study included: 1) determining whether the OST predicted sexual recidivism among sexual offenders, 2) determining whether the OST predicted violent nonsexual recidivism among sexual offenders, 3) determining whether the OST predicted nonviolent nonsexual recidivism among sexual offenders, and 4) determining whether the OST predicted each type of recidivism as well as currently available measures.

Measures

OST. As described above, the OST is a 44-item measure that encompasses both static (n=14) and dynamic (n=30) factors believed to be associated with risk of reoffending. Individual items on the OST fall into the following domains: vocational/financial (5 items), education (3 items), family and social relationships (8 items), residence and neighborhood (2 items), alcohol use (3 items), drug use (3 items), mental health (2 items), physical health (2 items), attitude (7 items), and criminal behavior (9 items). Risk factors are rated dichotomously as present (score of 1) or not present (score of 0). Scores on individual items are summed to create risk categories of low risk (0-5), medium-low risk (6-10), medium high risk (11-17), and high risk (18 or more) for male offenders. Participants’ scores on the OST that were obtained at the pre-trial stage were collected.

Static-99R. Information from the participants’ records were used to give each participant a score on the Static-99R. Offenders’ scores on this commonly used and well-established risk assessment measure were compared to scores obtained by the OST to compare the two measures’ predictive accuracy. The Static-99 is a 10-item measure comprised of static risk factors.

Procedures

The current study utilized the records from the Adult Probation Department located in a southwestern Arizona county. Participants included male sexual offenders who have been supervised by the probation department since the implementation of the OST in 2005. Data was collected from the probation department’s records of these offenders. Basic requirements for inclusion in the sample consisted of an OST score and information on recidivism (i.e. whether the offender was convicted of or pled guilty to any new offense during their term of probation). The two major reasons potential participants were excluded were because: 1) they were under the age of 18 at the time of their offense or the time the OST was administered, given that the validity of the OST with juvenile offenders has not been established; 2) the OST was administered and the offender subsequently served a prison sentence, as the OST score may no longer be valid following a period of incarceration (offenders who were administered the OST after release from prison were not excluded). Offenders' risk level (low, medium-low, medium-high, or high) obtained by the OST was recorded, as well as their total OST score. Demographic information including age, ethnicity, and mental health diagnoses (if any) were also obtained. Additionally, offense characteristics were recorded for both the index sexual offense for which the offender was under probation supervision, as well as any subsequent sex offenses in the cases where recidivism occurred. Offense characteristics collected included 1) whether the offense was a contact or non-contact offense, 2) victim relationship to perpetrator (familial or non-familial), 3) victim age (child or adult), and 4) victim sex (male or female).

Recidivism data was also collected for all participants used in the sample. For this study, recidivism was defined as a new conviction or plea bargain during the period of probation. Three types of recidivism were considered: 1) sexual recidivism, which included any new sexual offense, 2) violent recidivism, which included any violent but non-sexual offense, and 3) non-sexual non-violent recidivism, which included any additional offenses that were neither sexual nor violent in nature. In cases where nonsexual recidivism occurred, the type of crime and victim characteristics (if a victim was involved) were recorded. During data analysis, it was determined that too few cases of nonsexual violent recidivism occurred to conduct any meaningful statistical analyses. Therefore, the analyses conducted focused on nonviolent nonsexual recidivism, sexual recidivism, and a combined variable consisting of sexual and violent recidivism (described below). Data was collected at the probation department and no identifying information was removed from the site. IRB approval, as well as approval from the probation department, were obtained prior to data collection.

Results

Preliminary and Descriptive Analysis

Demographics. The sample used in the analysis consisted of 174 male offenders who had committed at least one sexual offense. Table 1 lists the reported race of participants. Slightly over half of the sample was white (56%, n=98), while a relatively large minority (28%, n=49) were Hispanic. Average age at the time of the index sexual offense (i.e. the offense for which they were on probation) was just under 37 (range 18-87). Among those who committed any type of reoffense, average age at reoffense was just over 35 (range 19-65).

Mental health. The majority of the sample (54.9%) had no documented mental health diagnosis, though it is possible that some participants did not disclose previous diagnoses and/or records were not available to probation officers. Of those who had a diagnosed mental health condition, the most common diagnoses were mood disorders (26.3%) and PTSD (4.6%).

Criminal history and index sex offense characteristics. Offenders in the sample had been charged with between one and five sexual offenses (mean = 1.1) and between zero and 64 nonsexual offenses (mean = 4.4). Table 2 lists the characteristics of the index sex offenses as well as reoffenses for the portion of the sample who committed another sex offense while on probation. The index sexual offense for a majority of the sample was a contact sexual offense (73.1%). The majority of victims were children (74.9%) and female (78.9%).

Recidivism. Recidivism rates in the sample were generally low. 46.3% of the sample (n = 81) were convicted or pled guilty to any new offense during the follow-up time. The majority of recidivism consisted of nonsexual nonviolent offenses (82.7%, n = 67). 13.6% of recidivism consisted of a new sexual offense (n = 11). This represented 6.3% of the sample overall. Table 2 lists characteristics of sexual reoffenses. Interestingly, recidivism data indicated that nearly three-quarters (72.7%) of sexual reoffenses were noncontact offenses, though females (58.3%) and children (66.7%) remained the primary victims. Violent nonsexual recidivism made up a very small portion of reoffending. Only 3.7% of new offenses (n = 3) fell into this category. Tables 4 and 5 show recidivism rates for each risk category of the OST and the Static-99R.

Follow up time: Offenders included in the sample had been sentenced to periods of probation ranging from three years to lifetime probation. For this study, follow up time was measured as beginning at the start of each participant's period of probation and ending either a) when recidivism occurred or b) the last day of data collection, which was used as an artificial end date. Follow up time ranged from 37 days to 15.1 years (mean= 3.9 years, SD = 3.1 years).

Current probation status. Table 3 lists the current probation status of offenders included in the sample. Just under half of the sample (46.5%) successfully completed their term of probation and were no longer under probation supervision at the time of data collection (one offender had been unsuccessfully terminated from probation). 43.7% had returned to jail or prison and were also no longer under supervision. Only 4% of the sample were still on probation at the time of data collection.

ROC Analysis

Receiver operating characteristic (ROC) analysis was used to test the predictive accuracy of the OST and the Static-99R for different types of reoffending. The ROC plots the number of offenders who were correctly identified as recidivists against those who were incorrectly identified as recidivists. The resulting area under the curve (AUC) statistic therefore is a measure of both tools' accuracy in classifying individuals into distinct categories of recidivists and non-recidivists. An AUC of .5 indicates that a measure predicts at chance level, while higher scores indicate greater predicative power. Rice and Harris (2005) compared AUC values with Cohen's *d* and suggest that an AUC value of .64 represents a moderate effect size, while .71 represents a large effect size. These cutoffs were used to evaluate the AUC values obtained for the OST and Static-99R.

Tables 6 and 7 show the AUC statistics for the total score of both the OST and the Static-99R for sexual, nonsexual and nonviolent, any violent (i.e. sexual and nonsexual violent combined), and any (i.e. all types combined) reoffending. An AUC statistic for violent nonsexual reoffending was not calculated because of the very small number of offenders in this category ($n = 3$). The AUC for the OST for detected sexual recidivism was .54, indicating that the OST predicted sexual recidivism at close to chance level. Similarly, the AUC for the OST for combined sexual and violent recidivism was .55. The OST demonstrated much greater predictive power for nonsexual nonviolent reoffending. The AUC for this measure of recidivism was .72 ($p < .001$, see Figure 1). This suggests the OST has strong predictive accuracy for nonsexual nonviolent reoffending. The AUC for all types of reoffending combined was also .72 ($p < .001$, see Figure 2).

AUC values were also calculated for the Static-99R to compare the predictive power of this measure to the OST. The AUC for the Static-99R for sexual reoffending was .74 (p = .007), indicating a large effect size in predicting sexual recidivism. Combined sexual and/or violent reoffending produced a similarly large AUC of .71 (p = .008). The AUC for nonsexual nonviolent reoffending was .69 (p < .001). All types of reoffending combined produced an AUC of .75 (p < .001). These results indicate that the Static-99R demonstrated strong predictive accuracy for any type of recidivism, sexual recidivism, and combined sexual and/or violent recidivism. It had modest accuracy for nonsexual nonviolent recidivism. The Static-99R performed more accurately than the OST in predicting each type of recidivism except for the nonsexual nonviolent type.

Cox Regression Analysis

Cox regression survival analyses were conducted on the OST and Static-99R, as well as the risk categories for each measure. Survival analysis was chosen as an analytic strategy because it allows a comparison of different risk groups while taking into account their detected recidivism rate and the length of time that passed before reoffense once they were at risk (i.e. once they were released into the community under probation supervision). Participants were followed until they either committed another offense or their period of probation ended, or until the end of the follow-up period (i.e. the end of data collection)

First, Cox regression analyses were performed using the OST total score alone as the predictor variable and using each type of recidivism as an outcome variable. OST total score was a significant predictor for any recidivism (all types combined) (β = .125, p < .001, hazard ratio = 1.133), combined sexual and violent recidivism (β = .086, p = .04, hazard ratio = 1.09), and nonsexual nonviolent recidivism (β = .112, p < .001, hazard ratio = 1.118). Total OST score was not a significant predictor of sexual recidivism.

Second, Cox regression models were produced using the OST risk category (low, medium-low, medium-high, and high) as the predictor variable and time to the various types of recidivism as the outcome. This analysis indicated that OST risk level did significantly predict the hazard rate for any recidivism and nonviolent nonsexual recidivism. In both models, the rate of detected reoffending varied significantly between each risk group. Higher risk offenders reoffended more quickly and at a higher rate than lower risk offenders. Figures 3 and 4 show the survival curve for any recidivism and nonsexual nonviolent recidivism, respectively. OST risk category did not significantly predict the hazard rates for sexual or combined sexual and violent recidivism.

Similarly, Cox regression models were produced using Static-99R total score and risk category as predictor variables. Static-99R total score was a significant predictor of any recidivism (all types combined) (β = .299, p < .001, hazard ratio = 1.35), sexual recidivism (β = .49, p = .001, hazard ratio = 1.63), combined sexual and violent recidivism (β = .449, p = .<.001, hazard ratio = 1.57), and nonsexual nonviolent recidivism (β = .236, p < .001, hazard ratio = 1.27). Static-99R categories also significantly predicted the hazard rate for any recidivism and nonsexual nonviolent recidivism. The Static-99R risk categories did not significantly predict the hazard rate for sexual or combined sexual and violent recidivism. However, the low number of offenders who reoffended sexually or violently meant that few offenders fell into each of the Static-99R's risk categories, making statistical analyses less robust.

Further analyses were conducted to determine the incremental validity of the OST and the Static-99R. Given the Static-99R's superior performance predicting most types of recidivism, Cox regression models were produced to determine whether OST scores could enhance the predictive accuracy of Static-99R scores in predicting sexual recidivism, combined sexual and violent recidivism, or any recidivism. The first and second models found that adding the OST to the Static-99R did not increase accuracy in predicting sexual reoffending (OST: $\beta = -.018$, $p = .706$, hazard ratio = .982) or combined sexual and violent reoffending (OST: $\beta = .011$, $p = .80$, hazard ratio = 1.011). The third model showed that the OST did increase predictive accuracy for any recidivism when added to the Static-99R (OST: $\beta = .091$, $p < .001$, hazard ratio = 1.095).

Because the OST demonstrated greater predictive power in predicting nonsexual nonviolent recidivism, an additional Cox regression model was produced to determine whether the Static-99R increased predictive accuracy when added to the OST in predicting this type of reoffending. The model showed that the Static-99R did not improve upon the performance of the OST for prediction of nonsexual nonviolent reoffending (Static 99: $\beta = .102$, $p = .114$, hazard ratio = 1.107).

Discussion

The aim of the current study was to test the predictive validity of a risk assessment measure for adult male sexual offenders used by the adult probation department in a southwestern Arizona county. While the OST is used to determine the level of risk and the level of supervision for individuals on probation, its validity for use with sexual offenders has not previously been tested. Given the significance of the OST results in terms of both public safety and offenders' lives, determining its ability to predict reoffense for this population is of critical importance.

For the purposes of analysis in the present study, three types of reoffense were considered: 1) nonsexual nonviolent reoffense, 2) sexual reoffense, and 3) a combined category of sexual and violent nonsexual reoffense. Statistical analyses were not conducted for violent nonsexual reoffending alone due to the very small number of cases in this category.

ROC analysis, a measure of relative risk, of the OST produced an AUC statistic of .72 for all types of reoffense combined. This robust statistic was due primarily to the OST's predictive power for nonsexual nonviolent reoffending (AUC=.72). However, the OST performed near chance level for prediction of sexual and combined sexual and violent reoffending. These results indicate that while the OST may be useful for predicting events such a probation noncompliance or other general, nonviolent reoffending, it predicts other types of reoffense at no better than chance levels.

For comparison with the OST, AUC statistics were also produced for the Static-99R, given its widespread use and well-established validity for predicting recidivism among adult sexual offenders. All types of reoffense combined produced an AUC statistic of .75. Consistent with previous research, the Static-99R produced an AUC statistic of .74 for sexual reoffending, indicating a large effect size for predicting sexual recidivism. Combined sexual and violent reoffending produced a similarly large AUC of .71. The AUC for nonsexual nonviolent reoffending was .69 ($p < .001$), representing a moderate effect size. Notably, the OST performed better than the Static-99R for this specific subtype of recidivism.

Cox regression analysis was also used to evaluate the predictive accuracy of both the OST and the Static-99R. Because offenders included in the study began and ended their terms of probation at different times over a period of more than ten years, it was important to include a statistical analysis that could account for the time an individual was at risk to reoffend. For the purposes of this analysis, offenders were considered at risk from the time their period of probation began until they either reoffended or completed probation (in cases where no reoffense occurred). For individuals whose term of probation had not yet ended, the last day of data collection was used as an artificial end date. Follow up time ranged from 37 days to 15.1 years (mean= 3.9 years, SD = 3.1 years).

Results of the Cox regression analysis revealed that the OST total score was a significant predictor of nonsexual nonviolent recidivism and combined sexual and violent recidivism, as well as all types of recidivism combined. Total score was not a significant predictor of sexual recidivism. Additional Cox regression analyses found that OST risk category (i.e. low, medium-low, medium-high, and high) was also a significant predictor of nonsexual nonviolent recidivism as well as all types combined. In both models, offenders classified as high risk by the OST offended sooner and at a higher rate than those in lower risk categories.

Consistent with previous research, the Static-99R total score significantly predicted each type of recidivism. The Static-99R’s risk categories were also significant predictors of any recidivism and nonsexual nonviolent recidivism.

Lastly, Cox regression models were produced to determine whether adding the results of the OST to the Static-99R added any predictive power. For sexual reoffense and combined sexual and violent reoffense, the addition of the OST did not increase predictive power. However, predictive accuracy for nonsexual nonviolent reoffense was increased by adding the OST to the Static-99R.

Strengths and Limitations of the Current Study

The current study had several strengths. First, the study has good external validity. Data was collected in a community setting with few restrictions on the individuals selected for inclusion in the sample. It included a diverse sample in terms of demographics and offense characteristics, representative of the population whose risk management is increasingly assessed and monitored by probation officers (Lewis, Lewis, & Garby, 2013).

Second, while the development of the OST was guided by researchers and includes empirically supported risk factors for recidivism, input was also obtained from probation officers who work directly with and supervise individuals on probation. Research literature frequently identifies the divide between researchers and practitioners as a barrier to better client care and outcomes (e.g. Kazdin, 2008; Goldfried, 2019). Including those who work directly with offenders in the development of the OST represents an attempt to bridge this gap.

An additional strength of the current study was the addition of the Static-99R to the analysis. This measure has well established reliability and validity and is widely used to assess recidivism risk among sexual offenders. Obtaining participants' score on this measure allowed for direct comparison of the OST's performance with another measure using the same sample. Comparing the two measures revealed that while the Static-99R had greater predictive accuracy for most types of recidivism, the OST performed better for nonviolent nonsexual reoffense, which was the most common type of reoffense observed over the follow-up period. Indeed, research indicates that the majority of sexual offenders are more likely to reoffend nonsexually, and only a small percentage commit a future sex offense (Hanson & Morton-Bourgon, 2005).

Lastly, the follow-up time was a strength of this study. Although follow-up time varied between participants, the length of some offenders' probation terms allowed a follow-up time of more than ten years.

The current study should also be considered in the context of its limitations. First, reoffense rates reported in this study include only detected recidivism. It is possible that some reoffenses went undetected, and thus the reoffense rates reported here may be underestimates. Related to this, offenders included in the sample were only followed for the length of their probation term. It is possible that some offenders committed another offense after their term of probation ended. While follow-up time was relatively long for some offenders in the sample, this is nonetheless a limitation of the data available through the probation department.

Additionally, sample size was a limitation of this study. While the overall sample size of 174 was reasonable, the reoffense rate was relatively low. There was an insufficient number of violent nonsexual reoffenses to conduct statistical analyses on this category. This is a notable limitation because previous research indicates that factors that predict general recidivism are similar to those that predict violent nonsexual recidivism, while predictors of sexual recidivism appear to be distinct (Hanson & Morton-Bourgen, 2005). Thus, the OST may have demonstrated some predictive accuracy for this type of reoffense if further analysis had been possible. Further, the sample was not large enough to conduct separate analyses on subtypes of offenders. The index sexual offense for most participants was a contact sexual offense against a female child. A larger sample would be needed to determine the OST’s accuracy with subtypes of offenders such as rapists, internet sexual offenders, and noncontact offenders.

Future Directions

The current study supports continued use of the OST for assessing the risk for nonsexual, nonviolent reoffending. Its superior performance in predicting this type of recidivism compared to another well-validated measure is notable, especially given the prevalence of this type of reoffense among offenders in the sample. As noted above, it is possible that a larger sample size could determine the OST's predictive accuracy for violent nonsexual reoffense, as well as elucidate whether the OST has more or less utility with different subtypes of offenders. Additionally, the OST's utility with female sexual offenders has not yet been assessed.

While these results were promising, the current study did not support the use of the OST for assessing risk of sexual recidivism. This is perhaps not surprising given that the OST's risk domains do not include some of the risk factors most strongly associated with sexual reoffense, such as deviant sexual interests and sexual preoccupation (Hanson & Morton-Bourgen, 2005). Although the majority of sexual offenders do not commit another sex offense, identifying those most at risk to do so is critically important for public safety. This study found that the Static-99R had a high degree of predictive power in regard to sexual recidivism and a combined category of sexual and violent nonsexual recidivism. Thus, the probation department whose data was analyzed as part of this study should consider adding a second measure to their risk assessment protocol in order to best identify those offenders in need of the highest level of supervision. The Static-99R may be an ideal choice given that it is available free of charge and demonstrated utility on the current sample of sexual offenders.

Conclusion

Accurately predicting risk to reoffend is a critically important task for those who work with sexual offenders. Increasingly, probation officers are taking on a greater role in the assessment of risk and needs among the offenders they supervise (Lewis, Lewis, & Garby, 2013). The current study examined the predictive utility of a risk assessment measure presently used by the adult probation department in a southwest county in Arizona. Risk level predicted by the Offender Screening Tool is used to assign the level of supervision an offender will have while on probation, so assessing the OST's predictive accuracy is important for community safety and to ensure that offenders are given neither unnecessary nor insufficient supervision.

The results of the current study supported the OST's use for predicting offenders who are likely to commit nonsexual nonviolent reoffenses, such as probation violations, drug offenses, or property offenses. This is significant given that most instances of reoffense in the sample fell into this category. However, the OST performed at chance level in its prediction of sexual reoffending, while the Static-99R by comparison showed strong predictive power for sexual recidivism. With a larger sample size, the OST's utility in predicting violent nonsexual reoffense, as well as predicting reoffense among different subtypes of sexual offenders, could be established. The current results suggest that a risk assessment protocol consisting of the OST and an additional measure with greater validity for predicting sexual recidivism may be the most promising way both to promote community safety and to ensure appropriate and fair policies towards sexual offenders.

Tables

Table 1

Participant Race

Race	**Percent**
Caucasian, non-Hispanic	56.0
African-American	8.5
Hispanic	28.0
Native American	3
Other	4.5

Table 1. Participant Race

Table 2

Sex Offense Characteristics

	Index Offense (%)	Re-offense (%)
Contact Sex Offense	73.1	27.3
Non-contact Sex Offense	25.7	72.7
Adult Victim	16.0	33.3
Child Victim	74.9	66.7
Male Victim	8.0	41.7
Female Victim	78.9	58.3
Relative Victim	22.3	8.3
Stranger Victim	24.0	66.7
Acquaintance Victim	43.4	25.0

Table 2. Sex Offense Characteristics

Table 3

Participant Current Probation Status

Status	Percent
Successfully completed probation	46.5
Still on probation	4.0
Jail or prison	43.7
Deceased	4.0
Other	1.7

Table 3. Participant Current Probation Status

OST Risk Category	Total n(%)	Any Recidivism n(%)	Sexual Recidivism n(%)	Any Violent Recidivism n(%)	Nonsexual Nonviolent Recidivism n(%)
Low	24 (13.8)	5 (20.8)	1 (4.2)	1 (4.2)	4 (16.7)
Medium-Low	57 (52.8)	21 (36.8)	5 (8.8)	7 (12.3)	14 (24.6)
Medium-High	61 (35.1)	32 (52.5)	3 (4.9)	3 (4.9)	29 (47.5)
High	32 (18.4)	23 (71.9)	2 (6.3)	3 (9.4)	20 (62.5)

Table 4. Recidivism by OST Risk Category

Static-99R Risk Category	Total n(%)	Any Recidivism n(%)	Sexual Recidivism n(%)	Any Violent Recidivism n(%)	Nonsexual Nonviolent Recidivism n(%)
Very Low	4 (2.3)	0 (0)	0 (0)	0 (0)	0 (0)
Below Average	35 (20.1)	6 (17.1)	0 (0)	0 (0)	6 (17.1)
Average	85 (48.9)	37 (43.5)	5 (5.9)	7 (8.2)	30 (35.3)
Above Average	34 (19.5)	25 (73.5)	2 (5.9)	2 (5.9)	23 (67.6)
Well Above Average	16 (9.2)	13 (81.2)	4 (25)	5 (31.3)	8 (50)

Table 5. Recidivism by Static-99R Risk Category

Recidivism Type	AUC	Significance	95% Confidence Interval
Any Recidivism	.721	p < .001	.646 - .796
Sexual Recidivism	.539	p = .667	.364 - .713
Any Violent Recidivism	.545	p = .578	.388 - .701
Nonsexual Nonviolent Recidivism	.718	p < .001	.64 - .797

Table 6. AUC Statistics for OST Total Score

Recidivism Type	AUC	Significance	95% Confidence Interval
Any Recidivism	.745	p < .001	.673 - .817
Sexual Recidivism	.743	p = .007	.597 - .889
Any Violent Recidivism	.714	p = .008	.579 - .85
Nonsexual Nonviolent Recidivism	.691	p < .001	.612 - .77

Table 7. Area under the curve (AUC) statistics for Static-99R Total Score

Figures

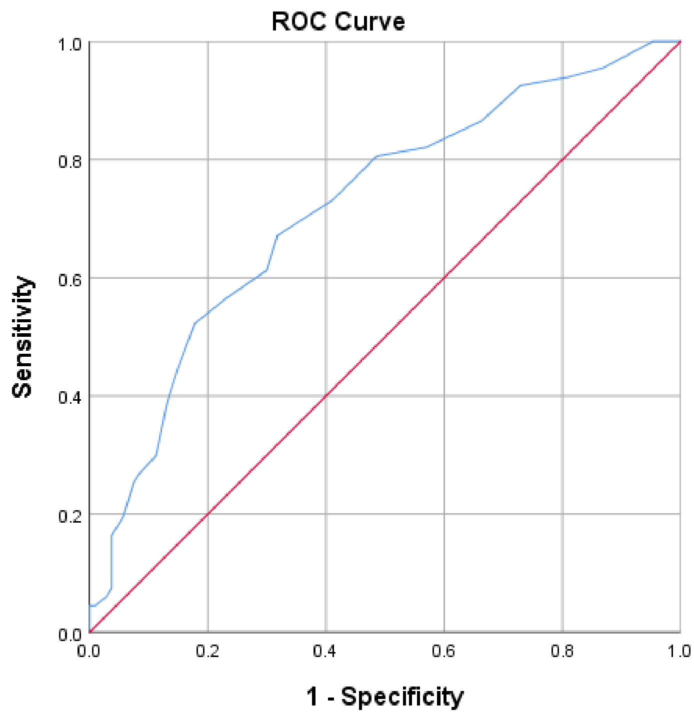

Figure 1. Receiver Operator Characteristic (ROC) Curve: OST, Nonsexual Nonviolent Recidivism

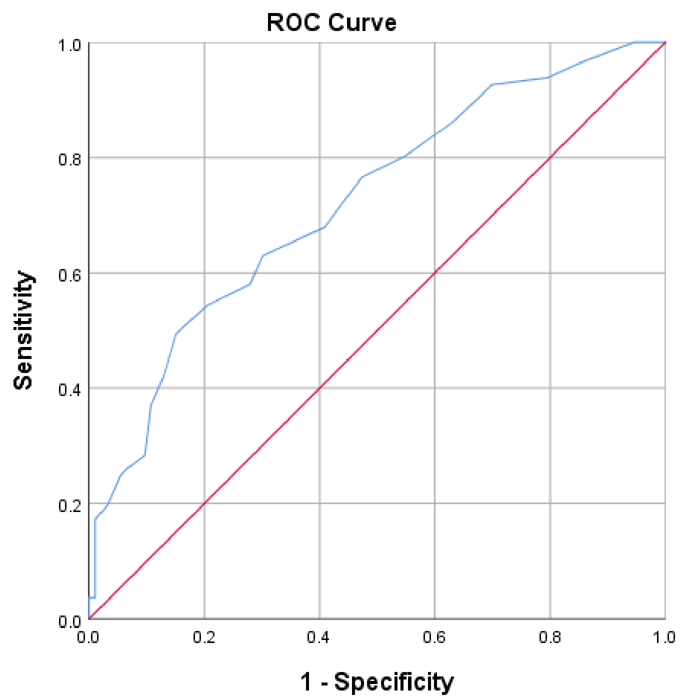

Figure 2. Receiver Operator Characteristic (ROC) Curve: OST, Any Recidivism

Figure 3. Survival curve of any recidivism over time by OST risk category.

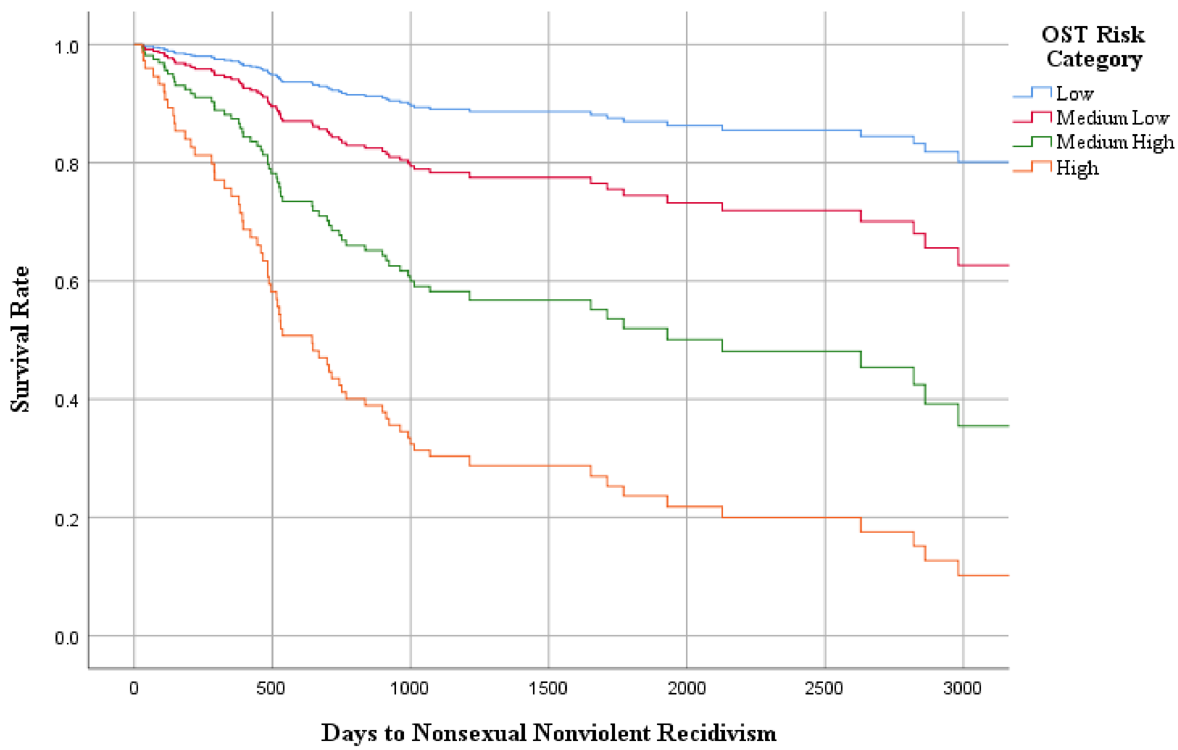

Figure 4. Survival curve of nonsexual nonviolent recidivism over time by OST risk category.

References

American Psychiatric Association. (1980). Diagnostic and statistical manual of mental disorders (3rd ed.). Washington, DC: Author

Arizona Supreme Court (2005). Administrative Order No. 2005-12: Adopting the standardized assessment and reassessment tool and conducting a pilot program for reassessment timeframes for adult intensive probationers. Phoenix, AZ: Author. Retrieved from http://www.azcourts.gov/portals/22/admorder/orders05/2005-12.pdf

Association for the Treatment of Sexual Abusers (2012). *Sex Offender Risk Assessment.* Retrieved from: https://www.atsa.com/pdfs/SexOffenderRiskAssessmentBriefWithBibliography2012.pdf

Barbaree, H. E., Seto, M. C., Langton, C. M., & Peacock, E. J. (2001). Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders. *Criminal Justice and Behavior*, *28*(4), 490-521.

Beech, A., Beckett, R., & Fisher, D. (2000). Outcome data of representative UK sex offender treatment programs: Short-term effectiveness and some preliminary re-conviction data. Unpublished manuscript.

Bonta, J. (1996). Risk-needs assessment and treatment. In A. T. Harland (Ed.), *Choosing correctional options that work: Defining the demand and evaluating the supply* (pp. 18-32). Thousand Oaks, CA: Sage.

Centers for Disease Control (2014). *Understanding Sexual Violence*. Retrieved from http://www.cdc.gov/violenceprevention/pdf/sv-factsheet.pdf

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243*(4899), 1668-1674. doi:10.1126/science.2648573

Department of Justice (2018). Sex Offender Registration and Notification Act (SORNA). Retrieved from: https://www.justice.gov/criminal-ceos/sex-offender-registration-and-notification-act-sorna

Department of Justice Assistance (2011). A Ten Step Guide to Transforming Probation Departments to Reduce Recidivism. Retrieved from: https://www.bja.gov/Publications/CSG_10Step_Guide_Probation.pdf

Douglas, K. S., Hart, S. D., Webster, C. D., & Belfrage, H. (2013). *HCR-20V3: Assessing risk of violence – User guide*. Burnaby, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University

Douglas, K. S., Shaffer, C., Blanchard, A. J. E., Guy, L. S., Reeves, K., & Weir, J. (2014). *HCR-20 violence risk assessment scheme: Overview and annotated bibliography*. HCR-20 Violence Risk Assessment White Paper Series, #1. Burnaby, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.

Eher, R., Matthes, A., Schilling, F., Haubner-MacLean, T., & Rettenberger, M. (2012). Dynamic risk assessment in sexual offenders using STABLE-2000 and the STABLE-2007: An investigation of predictive and incremental validity. *Sexual Abuse: Journal Of Research And Treatment*, *24*(1), 5-28. doi:10.1177/1079063211403164

Eidhammer, G., Selmer, L.E., & Bjørkly, S. (2013, June). Internal consistency and clinical utility. In S. Bjørkly, S. (Symposium Moderator) Risk assessment and management: Clinical experiences with the HCR-20 and the HCR-V3 in the SAFE pilot project.

Symposium presented at the annual conference of the International Association of Forensic Mental Health Services, Maastricht, Netherlands.

Ferguson, J. L. (2002). Putting the 'what works' research into practice: An organizational perspective. *Criminal Justice And Behavior,29* (4), 472-492. doi:10.1177/0093854802029004007

Goldfried, M. R. (2019). Obtaining consensus in psychotherapy: What holds us back? *American Psychologist*, *74*(4), 484–496.

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*(1), 19-30.

Hanson (2006). Does Static-99 predict recidivism among older sexual offenders? *Sexual Abuse: A Journal of Research and Treatment, 18,* 343-355.

Hanson, R. K., & Thornton, D. (1999). Static-99: Improving actuarial risk assessments for sex offenders (User Report No.1999–02). Ottawa: Department of the Solicitor General of Canada.

Hanson, R. K., & Thornton, D. (2003). Notes on the development of Static-2002. (Corrections Research User Report No. 2003-01). Ottawa: Department of the Solicitor General of 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*(6), 1154–1163.

Hanson, R. K., Harris, A. J. R., Scott, T., & Helmus, L. (2007). Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project (Corrections Research User Rep. No. 2007-05). Ottawa, ON: Public Safety Canada

Hanson, R. K., Helmus, L., & Thornton, D. (2010). Predicting recidivism amongst sexual offenders: A multi-site study of Static-2002. *Law and Human Behavior*, *34*(3), 198-211. doi:10.1007/s10979-009-9180-1

Hanson, R. K., Helmus, L. M., & Harris, A. J. (2015). Assessing the Risk and Needs of Supervised Sexual Offenders A Prospective Study Using STABLE-2007, Static-99R, and Static-2002R. *Criminal Justice and Behavior*,*42*(12), 1205-1224.

Hanson, R.K., & Morton-Bourgon, K. (2004). Predictors of Sexual Recidivism: An Updated Meta-Analysis. Ottawa, Ontario, Canada: Public Safety and Emergency Preparedness Canada.

Harris, A., Phenix, A., Thornton, D., & Hanson, R. K. (2003). *Static 99: Coding rules revised 2003*. Ottawa, ON: Solicitor General Canada.

Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. Criminal Justice and Behavior, 20, 315–335.

Hastings, M. E., Krishnan, S., Tangney, J. P., & Stuewig, J. (2011). Predictive and incremental validity of the Violence Risk Appraisal Guide scores with male and female jail inmates. *Psychological assessment*, *23*(1), 174-183.

Helmus, L., Babchshin, K. M., Hanson, R. K., & Thornton, D. (2009). Static-99R: Revised Age Weights. Retrieved from: http://www.static99.org/pdfdocs/static-99randage20091005.pdf

Helmus, L., Hanson, R. K., Thornton, D., Babchishin, K. M., & Harris, A. J. (2012). Absolute Recidivism Rates Predicted By Static-99R and Static-2002R Sex Offender Risk Assessment Tools Vary Across Samples A Meta-Analysis. *Criminal Justice and Behavior*, *39*(9), 1148-1171.

Irish, L., Kobayashi, I., & Delahanty, D. L. (2010). Long-term physical health consequences of childhood sexual abuse: A meta-analytic review. *Journal of Pediatric Psychology*, *35*(5), 450-461. doi:10.1093/jpepsy/jsp118

Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist*, *63*(3), 146–159.

Kingston, D. A., Yates, P. M., Firestone, P., Babchishin, K., & Bradford, J. M. (2008). Long-term predictive validity of the Risk Matrix 2000: A Comparison with the Static-99 and the Sex Offender Risk Appraisal Guide. *Sexual Abuse: Journal of Research And Treatment*, *20*(4), 466-484. doi:10.1177/1079063208325206

Långström, N., Enebrink, P., Laurén, E., Lindblom, J., Werkö, S., & Hanson, R. K. (2013). Preventing sexual abusers of children from reoffending: Systematic review of medical and psychological interventions. *BMJ: British Medical Journal*, *347*

Lewis, K. R., Lewis, L. S., & Garby, T. M. (2013). Surviving the trenches: The personal impact of the job on probation officers. *American Journal of Criminal Justice*, *38*(1), 67-84.

Lowenkamp, C. T., Latessa, E., & Bechtel, K. (2008). A reliability and validation study of the Offender Screening Tool (OST) and Field Reassessment Offender Screening Tool (FROST) for Arizona. Cincinnati: Center for Criminal Justice Research, University of Cincinnati.

Martens, R., Rettenberger, M., & Eher, R. (2015). The predictive and incremental validity of the German adaptation of the Static-2002 in a sexual offender sample released from the prison system. *Legal and Criminological Psychology*.

Martin, S. L., Macy, R. J., & Young, S. K. (2011). Health and economic consequences of sexual violence. In J. W. White, M. P. Koss, & A. E. Kazdin (Eds.), *Violence against women and children. Vol. 1: Mapping the terrain* (pp. 173–195). Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/12307-008

McGrath, R. J., Cumming, G. F., & Burchard, B. L. (2003). Current practices and trends in sexual abuser management: The Safer Society 2002 Nationwide Survey. Brandon, VT: Safer Society.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.

Meloy, M. L. (2005). The sex offender next door: An analysis of recidivism, risk factors, and deterrence of sex offenders on probation. *Criminal Justice Policy Review*, *16*(2), 211-236.

National Center for Victims of Crime (2012). *Child Sexual Abuse Statistics.* Retrieved from: https://victimsofcrime.org/media/reporting-on-child-sexual-abuse/child-sexual-abuse-statistics.

Nichols, H. R., & Molinder, I. (1984). Multiphasic sex inventory manual. *Tacoma, WA: Author*.

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

Rice, M. E., & Harris, G. T. (1999, May). A multi-site follow-up study of sex offenders: The predictive accuracy of risk prediction instruments. Third Annual Research Day of the University of Toronto Forensic Psychiatry Program, Penetanguishene, Canada

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC, Cohen's d, and r. Law and Human Behavior, 29, 615-620.

Sjöstedt, G., & Långström, N. (2000, November). Actuarial assessment of risk for criminal recidivism among sex offenders released from Swedish prisons 1993-1997. In *Poster presented at the 19th Annual Conference of the Association for the Treatment of Sexual Abusers, San Diego, California*.

Stinson, P. (2002). Development of a risk assessment instrument to be used in bail release decisions in Maricopa County, Arizona. Phoenix, AZ: Maricopa County Pretrial Services Agency. Retrieved from https://www.ncsc.org/~/media/Files/PDF/Education%20and%20Careers/CEDP%20Papers/2002/Risk_Assessment.ashx

Strub, D. S., Douglas, K. S., & Nicholls, T. L. (2014). The validity of Version 3 of the HCR-20 violence risk assessment scheme amongst offenders and civil psychiatric patients. *The International Journal of Forensic Mental Health*, *13*(2), 148-159. doi:10.1080/14999013.2014.911785

Thornton, D., Mann, R., Webster, S., Blud, L., Travers, R., Friendship. C., et al. (2003). Distinguishing and combining risks for sexual and violent recidivism. In R. A. Prentky, E. S. Janus, & M. C. Seto (Eds.), *Sexually coercive behavior: Understanding and management* (pp. 225-235). New York: New York Academy of Sciences

van den Berg, J. W., Smid, W., Schepers, K., Wever, E., van Beek, D., Janssen, E., & Gijs, L. (2018). The predictive properties of dynamic sex offender risk assessment instruments: A meta-analysis. *Psychological Assessment*, *30*(2), 179.

Vess, J., Day, A., Powell, M., & Graffam, J. (2013). International sex offender registration laws: research and evaluation issues based on a review of current scientific literature. *Police Practice and Research*, *14*(3), 205-218.

Witt, P. H., DelRusso, J., Oppenheim, J., & Ferguson, G. (1996). Sex offender risk assessment and the law. *Journal of Psychiatry & Law*, *24*(3), 343-377.

Zgoba, K. M., & Levenson, J. (2012). Failure to register as a predictor of sex offense recidivism: The big bad wolf or a red herring? *Sexual Abuse: Journal of Research and Treatment*, *24*(4), 328-349. doi:10.1177/1079063211421019