

ASSESSING THE RISK AND NEEDS OF SUPERVISED SEXUAL OFFENDERS

A Prospective Study Using STABLE-2007, Static-99R, and Static-2002R

R. KARL HANSON

Public Safety Canada

LESLIE-MAAIKE HELMUS

Public Safety Canada

ANDREW J. R. HARRIS

Forensic Assessment Group

Effective intervention with offenders requires accurate identification of their risk-relevant propensities. In this prospective study, 139 Canadian community supervision officers were trained to assess the risk factors and criminogenic needs of adult male sexual offenders using structured risk tools. Recidivism outcomes were recorded for 768 offenders (average age of 41 years, approximately half had child victims, 14% Aboriginal) during an average 7-year follow-up period. All forms of recidivism (sexual, violent, any) were predicted by sex crime specific risk tools based on static, historical factors (Static-99R; Static-2002R) and by tools designed to assess psychologically meaningful risk factors of sexual offenders (STABLE-2000; STABLE-2007). Professional overrides of the Static-99 scores did not improve predictive accuracy. STABLE-2007 scores added incrementally over STATIC scores for all recidivism outcomes, but only for complete cases, suggesting meaningful variation in the extent to which community supervision officers can assess psychologically meaningful risk factors for sexual offenders.

Keywords: risk assessment; Static-99R; Static-2002R; STABLE-2007; sexual offenders; recidivism

Managing the risk presented by sexual offenders is a preoccupation of the criminal justice system. Even a single instance of sexual recidivism by a known offender can invoke considerable public scrutiny. Most sexual offenders in Canada and the United States will serve part of their sentence under community supervision. The effectiveness of community supervision

AUTHORS' NOTE: *The views expressed are those of the authors and not necessarily those of Public Safety Canada. We would like to thank the community supervision officers who took the time to contribute data to this project, Marie Coligado and Jessica Woodley for double-running analyses, and the following individuals for assistance with data coding and management: Terri-Lynne Scott, Jennifer Cooney, Erik Gaudreault, Shannon Hodgson, Cathrine Pettersen, Shelley Price, Chelsea Sheahan, and Kimberly Smallshaw. Correspondence concerning this article should be addressed to R. Karl Hanson, Research Division, Community Safety and Countering Crime Branch, Public Safety Canada, 340 Laurier Avenue West, Ottawa, Ontario, Canada K1A 0P8; e-mail: karl.hanson@ps-sp.gc.ca.*

CRIMINAL JUSTICE AND BEHAVIOR, 2015, Vol. 42, No. 12, December 2015, 1205–1224.

DOI: 10.1177/0093854815602094

© 2015 International Association for Correctional and Forensic Psychology

for sexual offenders requires the accurate identification of risk-relevant propensities. Sexual offenders vary in their risk for recidivism, as well as in the reasons for that risk. For example, community supervision officers would want to set different conditions and treatment plans for an otherwise prosocial man who has molested boys as a sports coach compared with a substance abusing pimp with a history of vindictive rape.

During the past 20 years, there has been considerable research identifying the risk-relevant propensities of sexual offenders (Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005; Mann, Hanson, & Thornton, 2010). These factors, also called *criminogenic needs* (Andrews & Bonta, 2010), fall within the broad categories of general criminality (e.g., negative peer associations, low self-control, substance abuse) and sex crime specific criminality (e.g., deviant sexual interests, sexual preoccupations, emotional congruence with children). For general criminality, there are a number of validated risk tools that are commonly used by community supervision officers, such as the Level of Service Assessments (Andrews, Bonta, & Wormith, 2010) and the Statistical Information on Recidivism–Revised (SIR-R1; Nafekh & Motiuk, 2002), which also predict general recidivism among sexual offenders (Hanson & Morton-Bourgon, 2009; Wormith, Hogg, & Guzzo, 2012).

There is much less research concerning structured risk scales for the assessment of sex crime specific needs. The most commonly used risk scales for sexual offenders have focused on static, historical factors (e.g., Static-99; McGrath, Cumming, Burchard, Zeoli, & Ellerby, 2010). Although static factors, such as the number of prior offenses, can be efficient risk markers, it is not obvious how they can be used to infer risk-relevant propensities (Bonta, 1996). Consequently, there is a need for empirically validated risk tools for sexual offenders that help community supervision officers identify the source of the risk, and to track changes over time.

The Sexual Violence Risk-20 (Boer, Hart, Kropp, & Webster, 1997) was one of the earliest of the current generation of structured risk scales for sexual offenders. Based on a structured professional judgment (SPJ) approach, evaluators rate 20 factors relevant to case management, and then provide an overall assessment (low/moderate/high) concerning the extent to which the offenders have life problems worthy of clinical intervention (Hart & Boer, 2010). This instrument is primarily aimed at forensic mental health practitioners, including items such as major mental illness, negative attitude toward intervention, and psychopathy (Hare, 2003).

Thornton's Structured Risk Assessment (SRA) framework is another plausible approach to assessing the criminogenic needs of sexual offenders (Thornton, 2002). Originally an SPJ measure (Thornton, 2002), this framework has been organized into a structured scale (Structured Risk Assessment–Forensic Version [SRA-FV]; Thornton & Knight, 2015). The SRA framework, however, may not be ideally suited to the context of community supervision because SRA scores are largely determined by life history and are relatively insensitive to changes in current functioning.

The Violence Risk Scale–Sexual Offender version (VRS-SO; Olver, Wong, Nicholaichuk, & Gordon, 2007) combines 7 static and 17 dynamic risk factors into empirically derived recidivism rate estimates. The dynamic factors are grouped into three broad categories of sexual deviance, criminality, and treatment responsiveness. It is the only measure for sex offenders that provides empirically derived adjustments to initial risk ratings based on how the offender progresses (or not) in treatment (Olver, Beggs Christofferson, Grace, & Wong, 2014). Given the VRS-SO's focus on treatment, it is unclear how this scale should be applied to offenders who have not recently been in treatment.

One risk tool that was specifically designed for community supervision of sexual offenders is the Sex Offender Treatment Intervention and Progress Scale (SOTIPS; McGrath, Lasher, & Cumming, 2012). It contains 16 items grouped into three categories: (a) sexual deviance, (b) criminality, and (c) social stability and support. Initial research has been promising (McGrath et al., 2012), and an extended SOTIPS replication study is currently underway (Michael Miner, personal communication, February 28, 2014).

The current study focuses on the STABLE-2007 (Hanson, Harris, Scott, & Helmus, 2007), which is the most commonly used measure of sexual offender needs in Canada and the United States (Kelley, Barahal, Thornton, & Olson, 2015; McGrath et al., 2010). It contains 13 items related to sexual self-regulation (e.g., sexualized coping, deviant sexual interests), general self-regulation (e.g., poor cognitive problem solving), social relationships (e.g., lack of concern for others), and cooperation with supervision. Currently, there are translations of the scoring manual (Fernandez, Harris, Hanson, & Sparks, 2014) available in French, German, Dutch, Swedish, and Estonian. STABLE-2007 was the product of a program of research informed by quantitative reviews (Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2004, 2005), by interviews with community supervision officers (Hanson & Harris, 2000, 2001), and other risk scales (Beech, 1998; Thornton, 2002). Hanson and Harris created STABLE-2000 (in 2000), tested it in the Dynamic Supervision Project (DSP; see Hanson & Harris, 2013; Harris & Hanson, 2003, 2010), and used the results of that study to create STABLE-2007 (Hanson et al., 2007).

Independent validation studies of STABLE-2007 have been conducted by Eher, Matthes, Schilling, Haubner-MacLean, and Rettenberger (2012); Eher et al. (2013); Looman and Abracen (2012); and Sowden (2013). STABLE-2007 significantly predicted sexual recidivism in all studies except Sowden (2013) and was significantly incremental to Static-99/R in Eher et al. (2013) and Looman and Abracen (2012; but not Eher et al., 2012, or Sowden, 2013). In all four studies, STABLE-2007 was scored by research psychologists based on files of offenders serving custodial sentences.

Although there is general consensus concerning the risk-relevant propensities of sexual offenders, and evidence that these propensities can be assessed reliably by trained professionals using structured risk tools, we know little about the qualifications and training required to implement these measures with adequate fidelity. Given that certain constructs included in STABLE-2007 require considerable psychological inference (e.g., sexualized coping), it is an open question whether they can be reliably assessed by community supervision officers as part of their routine duties. In the current study, we examined the ability of community supervision officers to assess the criminogenic needs of adult male sexual offenders using STABLE data from the DSP (Hanson et al., 2007). Officers were trained to assess a range of static and dynamic risk factors, and then their risk scores were used by the officers to guide their case management. Once trained, the officers were asked to score consecutive cases, resulting in a relatively complete sample of adjudicated sexual offenders that included all types (child victims, adult victims, non-contact, intrafamilial) in the proportions naturally found in each jurisdiction. In summary, this study examined the applied use of these risk tools with routine, unselected sexual offenders using a prospective design.

METHOD

The original DSP study included approximately 1,000 sexual offenders from Canada and two U.S. states (Iowa and Alaska), who started supervision between 2001 and 2005. Only

the Canadian samples were included in the current study because of the difficulty obtaining reliable recidivism information from the U.S. states. As well, four Canadian female offenders were deleted. These women (age range from 23 to 62) had all sexually offended against children (age range of 6-14) who were family members or acquaintances. Three of the four women had been hospitalized overnight for a psychiatric condition.

All offenders were adult males starting a period of community supervision (probation or parole) following a recent conviction for a sexual offense. Offenders were excluded if they had been in the community for 6 months prior to the first assessment. For the vast majority of offenders, their most recent conviction was a sexually motivated offense; however, certain cases were included whose most recent offense was not sexually motivated provided that they began supervision for a sexually motivated offense within the past 2 years. In all cases, the supervising officers considered the offenders' primary problem to be sexual offending and were supervising them as such.

Data on offenders were obtained from 139 supervision officers from 14 different jurisdictions: all Canadian provinces and territories, as well as the Atlantic region of the Correctional Service of Canada (CSC). Descriptive data on the supervision officers were not collected. Based on current practices at the time, the sample of officers would have included men, women, diverse levels of expertise, sex crime specialists, and officers with general caseloads. The use of the risk tools was mandated by the jurisdictions, but the submission of data was voluntary. Consequently, the sample does not include all adjudicated sex offenders in these jurisdictions. Of 887 Canadian male offenders registered in this project, 119 were removed because no relevant assessment information was provided. Of the 768 included, 764 had Static-99R scores, 710 had Static-2002R scores, 616 had at least one STABLE-2000 assessment, and 615 had at least one STABLE-2007 assessment.

On average, the offenders were 41 years old ($n = 768$, $SD = 13.5$, range: 18-84 years) and serving their first sentence for a sexual offense conviction (69% [490/710] had no prior sentencing occasions for a sex offense). Half the sample had child victims (22% [162] were incest offenders, and 27% [206] had any unrelated child victim), 35% (267) had adult victims, 10% (76) had only non-contact offenses, and 6% (44) were mixed offenders ($n = 738$ for offense type classification). Most of the sexual offenses (77%, 588/763) involved physical contact but not overt physical injury. Eleven percent of the offenders (82/746) had ever been hospitalized overnight for a psychiatric condition, and 5% (37/746) had been diagnosed as developmentally delayed (DD; low intellectual functioning). Two out of three offenders (67.5%) had ever lived with a lover (married or common law) for 2 years (516/764). Three out of 10 offenders (29.5%, 224/760) had been separated from their biological parents prior to age 16. Fourteen percent of the offenders self-identified as being of Aboriginal heritage (106/760).

MEASURES

Stable-2000

The STABLE-2000 (Hanson et al., 2007) is a mechanical risk tool assessing dynamic risk factors among adult male sex offenders. It has 16 items organized into six subsections (significant social influences, intimacy deficits, sexual self-regulation, attitudes, general self-regulation, and cooperation with supervision). Each item is assessed using a 3-point rating system, where 0 refers to *no problem*, 1 refers to *some concern/slight problem*, and 2

refers to *present/definite concern*. Total scores on the STABLE-2000 are calculated by summing the highest score on each subsection, resulting in total scores ranging from 0 to 12, where scores of 0 to 4 indicate low risk, 5 to 8 indicate moderate risk, and scores of 9 or higher indicate high risk. Although officers submitted multiple STABLE assessments, only the first assessment was used.

Stable-2007

The STABLE-2007 (Fernandez et al., 2014; Hanson et al., 2007) is an empirical actuarial risk tool assessing dynamic risk factors among adult male sex offenders. The scale was developed by revising the STABLE-2000 scale based on preliminary results from this dataset (Hanson et al., 2007). Specifically, the three attitude items were removed and then the items for lovers/relationship stability and deviant sexual interest were revised to more clearly incorporate past behavior. Last, the emotional identification with children item was restricted to apply only to offenders with at least one victim less than 14 years old. Consequently, the STABLE-2007 has 13 items organized into five subsections (significant social influences, intimacy deficits, sexual self-regulation, general self-regulation, and cooperation with supervision).

Total scores on the STABLE-2007 are calculated by summing all item scores. Given that the emotional identification with children item is scored only for offenders with child victims, this means that total scores can range from 0 to 26 for offenders with child victims and 0 to 24 for other offender types. Based on the total score, offenders can be assigned to one of three risk categories: low risk (0-3), moderate risk (4-11), and high risk (12+). For these analyses, STABLE-2007 scores were calculated only if there was no more than one item with missing information. Officers submitted scores on an earlier version of the scale (STABLE-2000), and STABLE-2007 was calculated by dropping and modifying the previously noted items (see Hanson et al., 2007).

Static-99R

Static-99R (Hanson & Thornton, 2000; Helmus, Thornton, Hanson, & Babchishin, 2012) is an empirically derived actuarial risk assessment tool designed to predict sexual recidivism in adult male sex offenders (see also www.static99.org). The Static-99R items are identical to those in the Static-99 with the exception of updated age weights. The scale has 10 items assessing criminal history, victim characteristics, age, and relationship history. Static-99R has been found to have moderate predictive accuracy for sexual recidivism (area under the receiver operating characteristic curve [AUC] = .69; Helmus, Hanson, Thornton, Babchishin, & Harris, 2012). In this study, the supervising officers completed the original Static-99 scale. Static-99R was later computed by using the offender's date of birth to calculate the updated age item.

Static-2002R

Static-2002R (Hanson & Thornton, 2003; Helmus, Thornton, et al., 2012) is also an empirical actuarial risk assessment tool for adult male sex offenders (see also www.static99.org). The scale has 14 items grouped into five main subscales: age at release (1 item), persistence of sex offending (3 items), sexual deviance (3 items), relationship to victims (2

items), and general criminality (5 items). The Static-2002R items are identical to the original Static-2002 with the exception of updated age weights. Static-2002R has moderate predictive accuracy ($AUC = .69$), similar to that of Static-99R (Babchishin, Hanson, & Helmus, 2012). Static-2002 was scored in 2005 by two research assistants based on victim information provided by supervising officers in their Static-99 assessment and from official criminal records (Helmus & Hanson, 2007).

Static-99 Override

After completing a file review, recording demographic and offense history, and scoring the Static-99, officers were asked the following: Are there any exceptional circumstances that support an override of the Static risk assessment (e.g., historical offender being low risk because last known offense was 30 years ago, and there is no evidence of subsequent offending or a first time offender is high risk due to a large number of victims)? The response options were (a) there are no exceptional circumstances, (b) there are exceptional circumstances that *increase* risk, and (c) there are exceptional circumstances that *decrease* risk. For the purpose of statistical analyses, these response options were coded as 0, +1, and -1, respectively.

PROCEDURE

Assessments were completed as part of the routine supervision practices of the officers (Hanson et al., 2007). Supervising officers attended a 2-day training session, although in rare cases, officers were trained by apprenticing with other local officers.

On average, each of the 139 participating officers submitted information on 5.5 cases ($Mdn = 3$). The rate of participation varied ($SD = 6.9$, range: 1-57), with 33 officers reporting on only one case, and 29 officers reporting on only two cases. There was also variability in the rate at which the submitted data for a case were complete (343 out of 768 cases), defined as including the Static-99, the Static-99 override, at least one STABLE-2000, and at least one ACUTE-2000.¹

Rater reliability was assessed by six experts re-scoring 92 cases. Given our selection procedures, the reliability files contained a disproportionate number of complete cases. For the total scores, rater reliability was high: For Static-99, the intraclass correlation coefficient (ICC) was .91 ($n = 88$) and for STABLE-2000, the ICC was .89 ($n = 87$). The ICC for the individual STABLE-2000 items ranged from .66 to .92 ($Mdn = .83$). The only item with unacceptable levels of rater reliability was the Static-99 override rating (ICC = .15, $n = 74$). The above ICCs likely overestimate reliability because the second raters used case files prepared by the officer who originally scored the case. As well, although the second raters were directed to make independent ratings, the original scores were contained in the case files.

RECIDIVISM

All of the risk measures (Static-99R, Static-2002R, STABLE-2000, and STABLE-2007) were designed to assess the likelihood of new sexually motivated offenses for individuals who have already been charged or convicted for a sexually motivated offense against an identifiable victim. Nevertheless, other types of recidivism were considered because the predictors of general recidivism overlap with the predictors of sexual recidivism, and non-sexual recidivism is

more common than sexual recidivism among individuals convicted of sexual offenses. Specifically, we examined five different recidivism outcomes: (a) sexual recidivism—any sexually motivated offense (regardless of the charge or conviction), including Category B sex offenses as defined by the Static-99 coding rules (e.g., possession of child pornography, prostitution; Harris, Phenix, Hanson, & Thornton, 2003); (b) sexual recidivism including breaches (“any sex”)—all sexual recidivism as well as sexual breaches, defined as official sanctions for sexually motivated violations of the conditions of community supervision (e.g., being in the company of children); (c) violent recidivism—all crimes involving direct confrontation with the victim. This category included contact sexual offenses, but excluded non-contact sex offenses and sexually motivated breaches; (d) any criminal recidivism—included all crimes (sexual, violent, or non-violent) but excluded technical offenses; and (e) any recidivism—included all crimes (sexual, violent, non-violent), as well as all technical offenses (e.g., breach of conditional release).

Information concerning new offenses was gathered from provincial and national criminal history records, supervising officers, and local police jurisdictions. National criminal history records maintained by the Royal Canadian Mounted Police (RCMP) were received in August 2005, June 2006, and March 2011. Provincial records were received from British Columbia (January 2006), Manitoba (April 2005), and Ontario (December 2005). The Offender Management System of CSC was checked in May 2005, for the CSC offenders registered in the project. Police jurisdictions were contacted to obtain information on all new sexual and violent recidivism incidents identified from official records (as well as technical or non-violent incidents where a serious sanction was applied). Police contacts also resulted in identification of new recidivism incidents not appearing on official criminal records. Last, newspaper databases were searched for each offender in the summer of 2011 to identify new charges or convictions that may not be on criminal records.

Recidivism was considered to have occurred if the agency reporting the information believed that the offense occurred. For breaches, however, an official record of parole revocation or a new conviction for violation of conditional release was required. Given that criminal history records were the major source of recidivism information, the vast majority of recidivism events were linked to an officially recorded charge or conviction.

For violent and sexual recidivism incidents, information about the date and circumstances of the offense were obtained in all but a few cases. For other recidivism types, the date and offense circumstances were known only if the supervising officer provided this information. Classification of offense types was based on the behavior (if available) as opposed to the official charge. So, for violent and sex offenses, the date recorded in the dataset was generally the date of offense. For non-violent recidivism outcomes, it was often the date of charge or conviction. The follow-up period was calculated from the date of the first assessment information to the date that the last recidivism information was received (or until death or deportation). The offender start dates ranged from February 2001 to October 2005. Eighty-four percent of cases were followed until March 2011 (range: 0.2-10.1 years, $M = 7.4$, $SD = 2.2$, $Mdn = 7.9$).

OVERVIEW OF ANALYSES

All analyses were run independently by two people (the first two authors and two undergraduate students [psychology, criminology]). Discrimination accuracy was measured

TABLE 1: Descriptive Statistics for Risk Measures

| | Correlations (Sample Size) | | | Sample Size | <i>M</i> (<i>SD</i>) | Observed Range |
|--------------|----------------------------|--------------|-------------|-------------|------------------------|----------------|
| | Static-99R | Static-2002R | STABLE-2000 | | | |
| Static-99R | — | | | 764 | 2.4 (2.4) | –3, 10 |
| Static-2002R | .92 (710) | — | | 710 | 3.5 (2.5) | –2, 11 |
| STABLE-2000 | .40 (612) | .40 (571) | — | 616 | 4.9 (2.7) | 0, 12 |
| STABLE-2007 | .45 (611) | .44 (570) | .91 (615) | 615 | 7.4 (5.0) | 0, 26 |

Note. All correlations $p < .001$.

using the AUC (Swets, Dawes, & Monahan, 2000). AUCs can be interpreted as the probability that a randomly selected recidivist would have a higher risk score than a randomly selected non-recidivist. Hanley and McNeil's (1983) approach was used to compare AUCs between different groups.

Incremental predictive accuracy used Cox regression (Allison, 1984), which estimates relative predictive accuracy (hazard ratios) associated with one or more predictor variables from survival data with unequal follow-up times. For Cox regression, the offender's jurisdiction (i.e., province or CSC region) was used as a strata variable, allowing for separate baseline hazard functions.

Recidivism estimates for combining STABLE-2007 with either Static-99R or Static-2002R were calculated using life table survival analysis (Soothill & Gibbens, 1978). Although the median follow-up was 8 years, recidivism estimates are presented only up to 5 years because criminal events take considerable time to appear on official records (Bureau, 2015).

RESULTS

The overall recidivism rates were 10.8% ($n = 83$) for new sexual crimes, 12.9% ($n = 99$) for sexual crimes or sexually motivated breaches, 21.4% ($n = 164$) for any violent recidivism, 31.5% ($n = 242$) for any new crime, and 36.2% ($n = 278$) for any new transgression including breaches (all percentages based on a total sample of 768 offenders).

Descriptive statistics for the risk measures are presented in Table 1. The correlation between Static-99R and Static-2002R was .92, and between STABLE-2000 and STABLE-2007 was .91. The correlations between the STATIC and STABLE measures ranged from .40 to .45. The total scores of all the risk scales were significantly related to all outcomes (see Table 2). AUC values ranged from a low of .652 for STABLE-2000 predicting sexual recidivism to a high of .759 for Static-2002R predicting any recidivism (including breaches). In general, the STATIC risk scales (Static-99R; Static-2002R) showed somewhat better discrimination than the STABLE risk scales (STABLE-2000; STABLE-2007; $Mdn = .732$ vs. $Mdn = .677$).

Most STABLE-2007 items were positively associated with all types of recidivism (59 out of 65 comparisons; see Appendix A). The weakest effects were found for the emotional identification with children and deviant sexual interests, which were not significantly related to any of the recidivism outcomes. Five of the six STABLE-2000 items that were removed from STABLE-2007 (or revised) were not significantly related to either of the sexual recidivism outcomes. Readers are cautioned, however, that given the large number

TABLE 2: Discrimination of Risk Measures for Five Recidivism Outcomes

| | Sexual Recidivism | | Sexual (Including Breaches) | | Violent Recidivism | | Any Crime | | Any Recidivism (Includes Breaches) | | |
|------------------------------|-------------------|------|-----------------------------|------|--------------------|------|--------------|------|------------------------------------|------|--------------|
| | n | AUC | 95% CI | AUC | 95% CI | AUC | 95% CI | AUC | 95% CI | AUC | 95% CI |
| All assessments | | | | | | | | | | | |
| Risk measures | | | | | | | | | | | |
| Static-99R | 764 | .734 | [.674, .795] | .728 | [.672, .785] | .713 | [.670, .756] | .731 | [.694, .769] | .735 | [.698, .771] |
| Static-2002R | 710 | .733 | [.673, .792] | .731 | [.676, .787] | .732 | [.690, .775] | .754 | [.717, .791] | .759 | [.723, .795] |
| Stable-2000 | 616 | .652 | [.580, .723] | .661 | [.594, .728] | .666 | [.613, .719] | .681 | [.636, .726] | .695 | [.652, .738] |
| Stable-2007 | 615 | .669 | [.598, .740] | .689 | [.624, .753] | .673 | [.621, .724] | .687 | [.642, .731] | .707 | [.666, .749] |
| Risk categories | | | | | | | | | | | |
| Static-99R/STABLE-2007 | 611 | .732 | [.662, .802] | .742 | [.679, .805] | .690 | [.636, .743] | .712 | [.667, .757] | .722 | [.679, .765] |
| Static-2002R/STABLE-2007 | 570 | .718 | [.648, .787] | .729 | [.665, .793] | .714 | [.662, .766] | .737 | [.692, .781] | .748 | [.706, .790] |
| Completed assessments | | | | | | | | | | | |
| Risk measures | | | | | | | | | | | |
| Static-99R | 343 | .801 | [.721, .880] | .788 | [.714, .863] | .745 | [.680, .809] | .742 | [.685, .798] | .738 | [.682, .793] |
| Static-2002R | 322 | .804 | [.724, .884] | .798 | [.726, .870] | .765 | [.703, .827] | .762 | [.708, .817] | .759 | [.706, .813] |
| Stable-2000 | 343 | .729 | [.641, .818] | .747 | [.668, .825] | .706 | [.635, .777] | .728 | [.670, .786] | .726 | [.670, .782] |
| Stable-2007 | 343 | .757 | [.673, .841] | .784 | [.710, .859] | .721 | [.651, .790] | .746 | [.689, .803] | .747 | [.693, .802] |
| Risk categories | | | | | | | | | | | |
| Static-99R/STABLE-2007 | 343 | .811 | [.733, .889] | .812 | [.740, .884] | .742 | [.670, .814] | .733 | [.672, .794] | .725 | [.666, .784] |
| Static-2002R/STABLE-2007 | 322 | .809 | [.735, .883] | .814 | [.745, .884] | .774 | [.709, .839] | .761 | [.703, .819] | .757 | [.701, .814] |

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

of item-outcome comparisons (95), some of the variation in the predictive accuracy is simply due to chance.

Completed cases displayed better predictive accuracy than incomplete cases (see Tables 2 and 3). Completed cases contained the Static-99, the Static-99 override, at least one STABLE-2000, and at least one ACUTE-2000. Of the 139 officers, 42 always submitted completed cases, 105 submitted at least one completed case, and 34 never submitted completed cases (median per officer was 50% completed cases). There was considerable officer consistency in the rate of submitting completed cases ($r^2 = .44$). To estimate this between-officer effect size, logistic regression was used with case completion as the dependent variables (0 = *incomplete*; 1 = *completed*) and officers as a categorical predictor (139 categories). The effect size was then computed as the square of the Pearson correlation between the predicted values and outcome (Hosmer, Lemeshow, & Sturdivant, 2013, §5.2.5). The proportion of completed cases was not significantly related to the number of cases each officer submitted ($r = -.149, p = .081$). The recidivism rates for completed cases were not significantly different from the rates for incomplete cases (odds ratios for the five outcomes ranged from .82 to .91 [with .5 added to each cell; Fleiss, Levin, & Paik, 2003], in the direction of lower recidivism rates for the complete cases; two-tailed Fisher's exact tests values ranging from $p = .29$ to $p = .67$).

For sexual recidivism, the differences in AUCs for the risk scales between the completed and incomplete cases were large and significant for all scales but Static-99R (for Static-99R, AUC of .80 vs. .68, difference = .116, 95% confidence interval [CI] of [-.001, .232]; for Static-2002R, .80 vs. .68, difference = .127, 95% CI of [.011, .243]; for STABLE-2000, .73 vs. .56, difference = .169, 95% CI of [.027, .310]; for STABLE-2007, .76 vs. .58, difference = .182, 95% CI of [.044, .319]). Even though the predictive accuracy of STATIC instruments was reduced in the incomplete cases, they were still significantly related to the outcome. In comparison, neither STABLE-2000 nor STABLE-2007 significantly predicted sexual recidivism for incomplete cases (AUCs of .56 and .58, respectively). In contrast, for the completed cases, the predictive accuracies of the STABLE-2000 and STABLE-2007 were moderate to large (AUCs of .73 and .76, respectively). Of the 13 STABLE items, 12 had a higher AUC for the completed cases. The probability that at least 12 of the 13 comparisons would favor completed cases was less than would be expected by chance (exact binomial test; $p = .002$).

Of the 343 completed cases, officers recommended override of the Static-99 risk score in 61 cases (17.8%). Risk was increased in 54 cases and decreased in 7 cases. The overrides had little relationship to recidivism (AUC for sexual recidivism = .554, 95% CI [.445, .663]). Although the Cox regression beta weights for the incremental effect over Static-99 were positive for all five recidivism outcomes (see Table 4), none were statistically significant. When the override value (-1, 0, +1) was added to the Static-99 total score, the AUCs for the adjusted scores were lower for four of the five outcomes (the exception being any criminal recidivism, where the AUC increased slightly from .715 to .721). The absolute values of all the changes were small (median change of -.005 AUC units). The same patterns were observed when Static-99R or Static-2002R was used as the control variable.

As can be seen in Table 5, STABLE-2007 scores added incrementally to four of the five recidivism outcomes after controlling for Static-99R and to all types of recidivism after controlling for Static-2002R. For sex crime recidivism, the incremental effect did not reach

TABLE 3: Comparing Discrimination for Completed Versus Incomplete Assessments (Sexual Recidivism)

| | Complete | | | Incomplete | | |
|------------------------------|----------|------|--------------|------------|------|--------------|
| | <i>n</i> | AUC | 95% CI | <i>n</i> | AUC | 95% CI |
| Static-99R | 343 | .801 | [.721, .880] | 421 | .685 | [.599, .771] |
| Stable-2000 | 343 | .729 | [.641, .818] | 273 | .561 | [.450, .672] |
| Stable-2007 | 343 | .757 | [.673, .841] | 272 | .575 | [.466, .684] |
| Static-2002R | 325 | .804 | [.724, .884] | 388 | .677 | [.593, .760] |
| Stable-2000/2007 items | | | | | | |
| Social influences | 343 | .644 | [.548, .740] | 272 | .541 | [.429, .652] |
| Hostility toward women | 343 | .694 | [.594, .794] | 273 | .490 | [.379, .601] |
| Rejection/loneliness | 343 | .645 | [.546, .743] | 273 | .590 | [.474, .706] |
| Lack of concern for others | 343 | .660 | [.561, .759] | 272 | .571 | [.458, .685] |
| Sexual preoccupation | 343 | .592 | [.490, .693] | 272 | .479 | [.368, .589] |
| Sex as coping | 343 | .631 | [.522, .739] | 272 | .519 | [.408, .629] |
| Cooperation with supervision | 343 | .592 | [.483, .701] | 273 | .572 | [.462, .682] |
| Impulsive acts | 343 | .650 | [.547, .752] | 273 | .652 | [.550, .753] |
| Poor problem-solving | 343 | .655 | [.559, .751] | 273 | .643 | [.540, .746] |
| Negative emotionality | 343 | .612 | [.509, .715] | 273 | .495 | [.380, .610] |
| Unique to Stable-2000 | | | | | | |
| Lovers | 343 | .665 | [.584, .745] | 273 | .496 | [.390, .602] |
| Identification with children | 343 | .497 | [.393, .601] | 272 | .429 | [.328, .530] |
| Deviant sexual interests | 343 | .592 | [.482, .701] | 273 | .464 | [.357, .571] |
| Sexual entitlement attitudes | 343 | .634 | [.532, .735] | 272 | .470 | [.362, .578] |
| Rape attitudes | 343 | .596 | [.491, .701] | 273 | .540 | [.427, .653] |
| Child molester attitudes | 343 | .469 | [.368, .571] | 272 | .451 | [.343, .560] |
| Unique to Stable-2007 | | | | | | |
| Lovers | 343 | .651 | [.566, .736] | 269 | .554 | [.448, .661] |
| Identification with children | 199 | .567 | [.406, .727] | 136 | .396 | [.241, .551] |
| Deviant sexual interests | 343 | .603 | [.499, .707] | 273 | .458 | [.344, .572] |

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

conventional standards of significance ($p = .081$, two-tailed). When the analysis was restricted to the completed cases, however, STABLE-2007 was significantly incremental to both STATIC measures for all recidivism outcomes. For the completed cases, the median incremental hazard ratio was 1.08, meaning that for each 1 unit increase in STABLE-2007 total score, offenders would be expected to have an 8% increase in the rate of recidivism, averaged across the follow-up period. Given that STABLE-2007 and the STATIC measures were incremental to each other, we created rules for combining both measures into an overall evaluation of risk (see Appendix B). The combination rules start with the STATIC risk categories and then adjust up one category if the STABLE-2007 score is high and adjust down one category if the STABLE-2007 score is low. Low and high STABLE-2007 scores were defined as ± 1 *SD* from the mean in routine samples of sexual offenders (i.e., 0-3 = *low*, 4-11 = *moderate*, 12 or higher = *high*; Fernandez et al., 2014). Exceptions to the rules were as follows: No category was lower than *low*, no category was higher than *high* for Static-2002R, and offenders with high Static-99R scores (6+) remained in the high category even when their STABLE-2007 score was low. Of the three offenders with high Static-99R scores and low STABLE-2007 scores, two sexually reoffended. There were no offenders with high Static-2002R scores and low STABLE-2007 scores.

TABLE 4: Incremental Predictive Accuracy of Professional Override (-1, 0, +1)

| Recidivism Type | Controlling for Static-99 | | | AUC | | | |
|-----------------|-------------------------------|--------------|----------|-----------------------|--------------|---------------------------------------|--------------|
| | <i>N</i> recidivist/ Total | Hazard Ratio | <i>p</i> | Static-99 (0 to 9) | 95% CI | Static-99 + Override (-1 to 10) | 95% CI |
| Sex | 34/313 | 1.056 | .90 | .809 | [.736, .882] | .804 | [.730, .879] |
| Any sex | 42/317 | 1.046 | .90 | .787 | [.716, .857] | .779 | [.707, .852] |
| Violent | 67/326 | 1.303 | .36 | .730 | [.662, .798] | .724 | [.655, .792] |
| Any crime | 104/337 | 1.442 | .12 | .715 | [.656, .774] | .721 | [.662, .780] |
| Any recidivism | 120/338 | 1.192 | .44 | .711 | [.654, .768] | .710 | [.653, .767] |

Note. Total sample size of 343 for the AUC analyses, with the number of recidivists identical to those in the Cox regression analyses (using jurisdiction as strata). AUC = area under the receiver operating characteristic curve; CI = confidence interval.

TABLE 5: Incremental Predictive Accuracy of STABLE-2007

| Recidivism Type | All Assessments | | | | | | Completed Assessments | | | | | |
|-----------------|-------------------------------|-------|----------|---------------------------------|-------|----------|-------------------------------|-------|----------|---------------------------------|-------|----------|
| | Controlling for Static-99R | | | Controlling for Static-2002R | | | Controlling for Static-99R | | | Controlling for Static-2002R | | |
| | <i>N</i> recidivist/ Total | HR | <i>p</i> | <i>N</i> recidivist/ Total | HR | <i>p</i> | <i>N</i> recidivist/ Total | HR | <i>p</i> | <i>N</i> recidivist/ Total | HR | <i>p</i> |
| Sex | 64/572 | 1.049 | .081 | 64/538 | 1.057 | .049 | 34/313 | 1.082 | .048 | 34/303 | 1.085 | .044 |
| Any sex | 78/580 | 1.075 | .003 | 77/545 | 1.077 | .004 | 42/317 | 1.134 | .001 | 42/307 | 1.132 | .001 |
| Violent | 125/600 | 1.054 | .007 | 123/544 | 1.053 | .011 | 67/326 | 1.075 | .010 | 66/306 | 1.069 | .020 |
| Any crime | 189/609 | 1.047 | .003 | 186/569 | 1.050 | .002 | 104/337 | 1.077 | .001 | 102/317 | 1.078 | .001 |
| Any recidivism | 224/610 | 1.066 | <.001 | 220/569 | 1.068 | <.001 | 120/338 | 1.086 | <.001 | 118/317 | 1.084 | <.001 |

Note. HR = hazard ratio.

The discrimination for the combined STATIC/STABLE risk categories are displayed in Table 2. The AUCs were mostly moderate to large, ranging from .69 to .75 for all assessments (*Mdn* = .73) and .72 to .81 for the completed assessments (*Mdn* = .77). Readers should note that the AUC values for risk categories are expected to be lower than for continuous scores because precision is lost when data are clumped.

DISCUSSION

The current study examined the extent to which community supervision officers could assess the criminogenic needs of sexual offenders. When officers used a structured approach to identifying risk-relevant propensities (STABLE-2000/STABLE-2007), their scores were significantly related to sexual, violent, and any recidivism. Furthermore, STABLE-2007 provided incremental information over Static-99R and Static-2002R (well-validated risk tools based on criminal history and demographic information) for most types of recidivism.

There was considerable variability in the predictive accuracy of the risk measures based on whether or not the case assessments were complete. Specifically, when officers completed all the measures requested of them (static, stable, acute, override), the predictive

accuracies of the STATIC and STABLE assessments were high, and higher than when officers submitted incomplete cases. Among incomplete cases, the STATIC measures (Static-99R, Static-2002R) continued to be moderately accurate. In contrast, neither STABLE-2000 nor STABLE-2007 significantly predicted the intended outcome (sexual recidivism) among the incomplete cases.

The substantial difference in predictive accuracy for STABLE-2007 for the complete and incomplete cases (AUC values of .76 vs. .58, respectively) cannot be interpreted with confidence because it was not expected. Our working hypothesis is that it was related to the conscientiousness with which the officers engaged in the risk assessment process. Although risk assessment has become a core task of community supervision, it is not clear that all officers have the skills and motivation to provide valid assessments. Miller and Maloney (2013), for example, found that only half of the probation officers in their sample were substantive compliers with risk/needs assessments, with the remaining officers divided between bureaucratic compliers and cynical compliers. Consequently, it is likely that a non-trivial portion of the officers in the current study did not fully engage in the assessment process.

The largest differences in predictive accuracy for the completed/incomplete cases were for the sexual crime specific items that required psychological inference (see Table 3). Whereas the effect sizes for the completed cases were moderate (AUC > .60) for hostility toward women, sex as coping, sexual entitlement attitudes, and deviant sexual interests, the effect sizes for these items hovered near chance levels (.46-.52) for the incomplete cases. In contrast, the predictive accuracies for cooperation with supervision, impulsive acts, and poor cognitive problem solving were similar for the complete and incomplete cases (significant AUCs ranging from .57 to .66). Given that these latter items are relevant to all probationers (sexual or non-sexual), it is likely that most officers were already skilled at identifying these features prior to the training provided in the current project.

Overall, we found the predictive accuracy of the STABLE risk tools to be similar to that observed in other settings. The AUC values we found for STABLE-2000 predicting sexual recidivism (.65 for the full sample; .73 for completed cases) bounded those of Saum (2006; AUC = .68) and Webb, Craissati, and Keen (2007; AUC = .72). Eher et al. (2012) found a somewhat lower value (.62).

For STABLE-2007 predicting sexual recidivism, we observed AUC values of .67 overall and .76 for complete cases. Eher's research group observed intermediate AUC values for STABLE-2007 predicting sexual recidivism of .71 for Austrian prisoners and .71 for German forensic psychiatric patients (Eher et al., 2012; Eher et al., 2013). Sowden (2013), however, found only small effects for STABLE-2007 predicting sexual recidivism among Canadian federal sexual offenders attending a high-intensity treatment program (AUC of .56 for pre-treatment scores; .58 for post-treatment scores). It was unclear whether Sowden had the necessary information to score STABLE-2007 as the rater reliability in her study was only ICC = .38 for the pre-treatment scores and ICC = .56 for the post-treatment scores.

The significant incremental effect of STABLE-2007 scores over STATIC scores contrasts with the failure of unstructured professional overrides to improve risk prediction. At the time that this study was designed (2000), unstructured adjustments to empirical actuarial risk tools were considered a plausible approach to risk assessment (Hanson, 1998, p. 53).

There is now professional consensus that structured risk assessments are more accurate than unstructured risk assessments (Hanson, 2009; Heilbrun, Yasuhara, & Shah, 2010; Skeem & Monahan, 2011). Furthermore, since 2000, professional opinion has become increasingly disillusioned with post hoc adjustments to actuarial risk tools (Campbell & DeClue, 2010; Hanson & Morton-Bourgon, 2009; Wormith et al., 2012), a trend supported by the current findings. Post hoc adjustments likely fail because evaluators have difficulty judging whether particular characteristics are risk-relevant, and, if characteristics are risk-relevant, the extent to which they have been already captured by the existing items.

According to Kahneman and Klein (2009), mechanical risk tools outperform professional judgment when (a) the number of relevant variables is large, (b) the effect of any particular variable is small, and (c) we do not receive rapid feedback concerning the accuracy of our decisions. All three conditions apply to risk assessment with sexual offenders. As observed in the current study and previous meta-analyses (Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005), no variable was strongly related to sexual recidivism. It is often years between the assessment date and the date at which recidivism is known—if at all. Consequently, it should not be surprising that the evaluators in the current study, like many before them (Bengtson, & Långström, 2007; Cocozza, & Steadman, 1976; Quinsey & Maguire, 1986), failed to demonstrate intuitive expertise for offender risk assessment.

LIMITATIONS

Empirically derived risk assessment instruments gain credibility through replication in diverse, independent samples. Although the final version of STABLE-2007 has been validated in independent samples (Eher et al., 2012, 2013), the construction of STABLE-2007 was based on a single sample. Consequently, there is the risk of overfitting, that is, selecting variables that do not generalize to other samples. A related concern is that potentially important variables may have been missed. For example, attitudes variables included in STABLE-2000 were dropped from STABLE-2007 because they were unrelated to recidivism in previous versions of this dataset; however, subsequent research in other settings has found the STABLE-2000 child molester attitudes item to significantly predict sexual recidivism among child molesters (Helmus, Hanson, Babchishin, & Mann, 2013). Consequently, we believe that the propensities targeted by STABLE-2007 are credible but are unlikely to be ideal.

Large sample sizes are needed for the development of empirically derived risk tools. In agreement with Vergouwe, Steyerberg, Eijkemans, and Habbeman (2005), we believe that approximately 100 events and 100 non-events are required for stable estimates. Compared with this standard, the sample size of recidivists in the current study was barely adequate for evaluating the prediction of sexual crime recidivism (83 recidivists), and insufficient for reliable analyses of subgroups, or incremental effects.

Risk measures may also perform differently with offenders of different ethnic background (Babchishin, Blais, & Helmus, 2012; Långström, 2004). Analyses of an earlier version of this dataset found that STABLE-2007 worked poorly for offenders of Aboriginal heritage (Helmus, Babchishin, & Blais, 2012). Consequently, further research is needed before STABLE-2007 can be used confidently with Aboriginal sexual offenders.

Offenders with developmental delays may also have distinct risk factors that are not addressed by STABLE-2007. Hanson, Sheahan, and VanZuylen's (2013) meta-analysis concluded that Static-99 works as intended with DD sexual offenders. The small number of DD offenders in the current sample (8 sexual recidivists out of a total of 38) limits our ability to make strong statements concerning the utility of STABLE-2007 with sexual offenders with developmental delays.

DIRECTIONS FOR RESEARCH

STABLE-2007 is intended to assess risk-relevant propensities, that is, offender characteristics that can be used to guide interventions and community safety planning with sexual offenders. Although there is reasonable evidence that the items in STABLE-2007 are related to recidivism risk (Hanson & Harris, 2013), further research is needed concerning whether the STABLE items are dynamic risk factors (Kraemer et al., 1997). Currently, there is some evidence that STABLE items change during the course of treatment (Nunes, Babchishin, & Cortoni, 2011; Sowden, 2013); however, there is no research demonstrating that changes in STABLE items are related to changes in recidivism risk. Furthermore, little is known about how the use of STABLE-2007, or any other risk tool, influences community supervision practices. Based on research on community supervision for general offenders (Bonta, Rugge, Scott, Bourgon, & Yessine, 2008), officers require more than valid risk tools to meaningfully influence recidivism risk.

IMPLICATIONS FOR PRACTICE

The reliability and predictive accuracy of STABLE-2007 were sufficient to justify its use in applied risk assessment with adult male sexual offenders. The use of any evidence-based risk tool, of course, requires a preliminary judgment concerning the similarity of the case-at-hand to the participants in the samples for which the tool was intended and validated. The greater the differences, the greater the inferences required to interpret the results. Without research evidence, however, it is difficult for evaluators to know the extent to which any observed differences on factors external to the risk scheme actually make a difference in the offender's overall risk.

Given that there were consistent incremental effects of Static-99R and Static-2002R over STABLE-2007, evaluators should combine STABLE-2007 with a static risk tool when conducting comprehensive risk evaluations. Tables with recidivism estimates based on combining STABLE-2007 with Static-99R, Static-2002R, or Risk Matrix-2000 Sex (Helmus, Hanson, Babchishin, & Thornton, 2015) are available upon request.

The results also suggest that care is needed when implementing STABLE-2007, given that a substantial portion of the current sample submitted STABLE assessments that had little or no relationship to recidivism. High fidelity implementation requires officer training, effective methods for motivating evaluators, ongoing supervision, and systemic support from senior managers (for more guidance on implementation and fidelity issues, see Fernandez et al., 2014).

APPENDIX A

DISCRIMINATION OF STABLE-2000 AND STABLE-2007 ITEMS FOR THE FULL SAMPLE

| | n | Sexual Recidivism | | Sexual (Including Breaches) | | Violent Recidivism | | Any Crime | | Any Recidivism (Includes Breaches) | |
|-------------------------------|-----|-------------------|--------------|-----------------------------|--------------|--------------------|--------------|-----------|--------------|------------------------------------|--------------|
| | | AUC | 95% CI | AUC | 95% CI | AUC | 95% CI | AUC | 95% CI | AUC | 95% CI |
| Stable-2000/2007 items | | | | | | | | | | | |
| Social influences | 615 | .596 | [.523, .670] | .610 | [.541, .679] | .631 | [.573, .688] | .611 | [.561, .660] | .603 | [.555, .650] |
| Hostility toward women | 616 | .600 | [.522, .677] | .597 | [.526, .667] | .612 | [.555, .670] | .596 | [.546, .646] | .601 | [.553, .648] |
| Rejection/loneliness | 616 | .619 | [.543, .694] | .622 | [.553, .691] | .566 | [.510, .622] | .584 | [.536, .632] | .593 | [.546, .639] |
| Lack of concern for others | 615 | .619 | [.544, .695] | .606 | [.536, .675] | .624 | [.568, .680] | .620 | [.571, .669] | .620 | [.573, .667] |
| Sexual preoccupation | 615 | .539 | [.464, .615] | .555 | [.485, .624] | .522 | [.465, .579] | .538 | [.488, .588] | .553 | [.505, .601] |
| Sex as coping | 615 | .579 | [.501, .657] | .589 | [.518, .660] | .503 | [.445, .561] | .526 | [.476, .576] | .538 | [.490, .586] |
| Cooperation with supervision | 616 | .584 | [.506, .661] | .595 | [.524, .666] | .633 | [.576, .691] | .614 | [.565, .664] | .630 | [.583, .677] |
| Impulsive acts | 616 | .653 | [.580, .725] | .678 | [.612, .743] | .705 | [.652, .758] | .690 | [.643, .738] | .712 | [.667, .756] |
| Poor problem-solving | 616 | .649 | [.579, .719] | .664 | [.601, .728] | .680 | [.630, .731] | .684 | [.638, .729] | .700 | [.658, .743] |
| Negative emotionality | 616 | .556 | [.479, .634] | .581 | [.511, .651] | .595 | [.538, .653] | .591 | [.541, .641] | .607 | [.560, .654] |
| Unique to STABLE-2000 | | | | | | | | | | | |
| Lovers | 616 | .589 | [.522, .656] | .580 | [.517, .644] | .551 | [.498, .604] | .588 | [.541, .634] | .599 | [.554, .645] |
| Identification with children | 615 | .464 | [.391, .538] | .519 | [.448, .589] | .477 | [.422, .533] | .473 | [.425, .522] | .489 | [.442, .536] |
| Deviant sexual interests | 616 | .531 | [.453, .609] | .561 | [.489, .632] | .494 | [.437, .551] | .499 | [.450, .549] | .508 | [.460, .555] |
| Sexual entitlement attitudes | 615 | .558 | [.482, .634] | .560 | [.490, .630] | .572 | [.514, .629] | .583 | [.533, .633] | .586 | [.538, .633] |
| Rape attitudes | 616 | .569 | [.492, .646] | .556 | [.485, .627] | .597 | [.540, .654] | .592 | [.542, .642] | .579 | [.532, .627] |
| Child molester attitudes | 615 | .461 | [.387, .535] | .476 | [.406, .546] | .444 | [.389, .499] | .440 | [.392, .489] | .450 | [.404, .497] |
| Unique to STABLE-2007 | | | | | | | | | | | |
| Lovers | 612 | .607 | [.539, .675] | .609 | [.545, .673] | .581 | [.526, .635] | .611 | [.564, .658] | .617 | [.572, .662] |
| Identification with children | 335 | .489 | [.374, .603] | .547 | [.444, .650] | .493 | [.410, .577] | .486 | [.415, .556] | .503 | [.436, .569] |
| Deviant sexual interests | 616 | .532 | [.453, .611] | .559 | [.488, .630] | .476 | [.418, .535] | .480 | [.430, .530] | .489 | [.441, .537] |

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

APPENDIX B

GUIDELINES FOR COMBINING STABLE-2007 WITH STATIC-99R/2002R

TABLE B1: Combining STABLE-2007 With Static-99R

| Static-99R Category | STABLE-2007 Category | Combined STATIC/STABLE Priority Category |
|----------------------|----------------------|--|
| Low (1 or lower) | Low | Low |
| | Moderate | Low |
| | High | Moderate-Low |
| Moderate-Low (2, 3) | Low | Low |
| | Moderate | Moderate-Low |
| | High | Moderate-High |
| Moderate-High (4, 5) | Low | Moderate-Low |
| | Moderate | Moderate-High |
| | High | High |
| High (6+) | Low | High |
| | Moderate | High |
| | High | Very High |

TABLE B2: Combining STABLE-2007 With Static-2002R

| Static-2002R Category | STABLE-2007 Category | Combined STATIC/STABLE Priority Category |
|-----------------------|----------------------|--|
| Low (2 or lower) | Low | Low |
| | Moderate | Low |
| | High | Low–Moderate |
| Low–Moderate (3, 4) | Low | Low |
| | Moderate | Low–Moderate |
| | High | Moderate |
| Moderate (5, 6) | Low | Low–Moderate |
| | Moderate | Moderate |
| | High | Moderate–High |
| Moderate–High (7, 8) | Low | Moderate |
| | Moderate | Moderate–High |
| | High | High |
| High (9+) | Low | Moderate–High |
| | Moderate | High |
| | High | High |

NOTE

1. ACUTE-2000/ACUTE-2007 is a structured rating scale to measure short-term changes in risk during community supervision. For further information, see Hanson, Harris, Scott, and Helmus (2007) and Babchishin (2013).

REFERENCES

- Allison, P. D. (1984). *Event history analysis: Regression for longitudinal event data*. Beverly Hills, CA: SAGE.
- Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). New Providence, NJ: LexisNexis.
- Andrews, D. A., Bonta, J., & Wormith, J. S. (2010). The Level of Service (LS) assessment of adults and older adolescents. In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 199-225). New York, NY: Routledge.
- Babchishin, K. M. (2013). *Sex offenders do change on risk-relevant propensities: Evidence from a longitudinal study of the ACUTE-2007* (Doctoral thesis). Carleton University, Ottawa, Ontario, Canada.
- Babchishin, K. M., Blais, J., & Helmus, L. (2012). Do static risk factors predict differently for Aboriginal sex offenders? A multi-site comparison using the original and revised Static-99 and Static-2002 scales. *Canadian Journal of Criminology and Criminal Justice*, *54*, 1-43. doi:10.3138/cjccj.2010.E.40
- Babchishin, K. M., Hanson, R. K., & Helmus, L. (2012). Even highly correlated measures can add incrementally to actuarial risk prediction. *Assessment*, *19*, 442-461. doi:10.1177/1073191112458312
- Beech, A. R. (1998). A psychometric typology of child abusers. *International Journal of Offender Therapy and Comparative Criminology*, *42*, 319-339. doi:10.1177/0306624X9804200405
- Bengtson, S., & Långström, N. (2007). Unguided clinical and actuarial assessment of re-offending risk: A direct comparison with sex offenders in Denmark. *Sexual Abuse: A Journal of Research and Treatment*, *19*, 135-153. doi:10.1177/107906320701900205
- 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, Canada: The British Columbia Institute Against Family Violence.
- 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.
- Bonta, J., Rugee, T., Scott, T.-L., Bourgon, G., & Yessine, A. (2008). Exploring the black-box of community supervision. *Journal of Offender Rehabilitation*, *47*, 248-270. doi:10.1080/10509670802134085
- Bureau, B. (2015, March 10). RCMP database remains out of date, police and prosecutors say. *CBC News*. Retrieved from <http://www.cbc.ca/news/politics/rcmp-database-remains-out-of-date-police-and-prosecutors-say-1.2989397>
- Campbell, T. W., & DeClue, G. (2010). Flying blind with naked factors: Problems and pitfalls in adjusted-actuarial sex-offender risk assessment. *Open Access Journal of Forensic Psychology*, *2*, 75-101.
- Cocozza, J. J., & Steadman, H. J. (1976). The failure of psychiatric predictions of dangerousness: Clear and convincing evidence. *Rutgers Law Review*, *29*, 1084-1101.

- 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: A Journal of Research and Treatment*, 24, 5-28. doi:10.1177/1079063211403164
- Eher, R., Rettenberger, M., Gainersdorfer, K., Haubner-MacLean, T., Matthes, A., Schilling, F., & Mokros, A. (2013). Über die Treffsicherheit standardisierter Risikoeinschätzungsverfahren bei aus der Maßregel entlassenen Sexualstraf Tätern [About the accuracy of standardized risk assessment procedures for the evaluation of discharged sex offenders]. *Forensische Psychiatrie, Psychologie, Kriminologie*, 7, 264-272. doi:10.1007/s11757-013-0212-9
- Fernandez, Y., Harris, A. J. R., Hanson, R. K., & Sparks, J. (2014). *STABLE-2007 coding manual: Revised 2014*. Unpublished manual, Public Safety Canada, Ottawa, Ontario.
- Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions* (3rd ed.). Hoboken, NJ: Wiley.
- Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, 148, 839-843. doi:10.1148/radiology.148.3.6878708
- Hanson, R. K. (1998). What do we know about sexual offender risk assessment? *Psychology, Public Policy and Law*, 4, 50-72. doi:10.1037/1076-8971.4.1-2.50
- Hanson, R. K. (2009). The psychological assessment of risk for crime and violence. *Canadian Psychology/Psychologie canadienne*, 20, 172-182. doi:10.1037/a0015726
- 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. doi:10.1037/0022-006X.73.6.1154
- Hanson, R. K., & Harris, A. J. R. (2000). Where should we intervene? Dynamic predictors of sex offense recidivism. *Criminal Justice and Behavior*, 27, 6-35. doi:10.1177/0093854800027001002
- Hanson, R. K., & Harris, A. J. R. (2001). A structured approach to evaluating change among sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, 13, 105-122. doi:10.1023/102660030448
- Hanson, R. K., & Harris, A. J. R. (2013). Criminogenic needs of sexual offenders on community supervision. In L. Craig, L. Dixon, & T. Gannon (Eds.), *What works in offender rehabilitation: An evidence based approach to assessment and treatment* (pp. 421-435). Chichester, UK: Wiley-Blackwell.
- 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* (User Report No. 2007-05). Ottawa, Ontario: Public Safety Canada. Retrieved from <http://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/ssssng-rsk-sxl-ffndrs/index-eng.aspx>
- Hanson, R. K., & Morton-Bourgon, K. (2004). *Predictors of sexual recidivism: An updated meta-analysis* (Corrections User Report No. 2004-02). Ottawa, Ontario: Public Safety and Emergency Preparedness Canada. Retrieved from <http://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/2004-02-prdctrs-sxl-rcdvs-m-pdtd/index-eng.aspx>
- 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. doi:10.1037/0022-006X.73.6.1154
- 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. doi:10.1037/a0014421
- Hanson, R. K., Sheahan, C. L., & VanZuylen, H. (2013). Static-99 and RRASOR predict recidivism among developmentally delayed sexual offenders: A cumulative meta-analysis. *Sexual Offender Treatment*, 8(1), 1-14.
- 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. doi:10.1023/A:1005482921333
- Hanson, R. K., & Thornton, D. (2003). *Notes on the development of Static-2002* (Corrections Research User Report No. 2003-01). Ottawa, Ontario: Department of the Solicitor General of Canada.
- Hare, R. D. (2003). *Manual for the Psychopathy Checklist—Revised* (2nd ed.). Toronto, Ontario, Canada: Multi-Health Systems.
- Harris, A. J. R., & Hanson, R. K. (2003, August). The Dynamic Supervision Project: Improving the community supervision of sex offenders. *Corrections Today*, 65, 60-64.
- Harris, A. J. R., & Hanson, R. K. (2010). Clinical, actuarial, and dynamic risk assessment of sexual offenders: Why do things keep changing? *Journal of Sexual Aggression*, 16, 296-310. doi:10.1080/13552600.2010.494772
- Harris, A. J. R., Phenix, A., Hanson, R. K., & Thornton, D. (2003). *Static-99 coding rules: Revised 2003*. Ottawa, Ontario: Solicitor General Canada. Retrieved from www.static99.org/pdfdocs/static-99-coding-rules_e.pdf
- Hart, S. D., & Boer, D. P. (2010). Structured professional judgment guidelines for sexual violence risk assessment: The Sexual Violence Risk-20 (SVR-20) and Risk for Sexual Violence Protocol (RSVP). In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 269-294). New York, NY: Routledge.
- Heilbrun, K., Yasuhara, K., & Shah, S. (2010). Violence risk assessment tools: Overview and critical analysis. In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 1-17). New York, NY: Routledge.
- Helmus, L., Babchishin, K. M., & Blais, J. (2012). Predictive accuracy of dynamic risk factors for aboriginal and non-aboriginal sex offenders: An exploratory comparison using STABLE-2007. *International Journal of Offender Therapy and Comparative Criminology*, 56, 856-876. doi:10.1177/0306624X11414693
- Helmus, L., Hanson, R. K., Babchishin, K. M., & Mann, R. E. (2013). Attitudes supportive of sex offending predict recidivism: A meta-analysis. *Trauma, Violence, & Abuse*, 14, 34-53. doi:10.1177/1524838012462244

- Helmus, L., Hanson, R. K., Babchishin, K. M., & Thornton, D. (2015). Sex offender risk assessment with the Risk Matrix 2000: Validation and guidelines for combining with the STABLE-2007. *Journal of Sexual Aggression, 21*, 136-157. doi:10.1080/13552600.2013.870241
- Helmus, L., Hanson, R. K., Thornton, D., Babchishin, K. M., & Harris, A. J. R. (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*, 1148-1171. doi:10.1177/0093854812443648
- Helmus, L., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2012). 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, 24*, 64-101. doi:10.1177/1079063211409951
- Helmus, L. M. D., & Hanson, R. K. (2007). Predictive validity of the Static-99 and Static-2002 for sex offenders on community supervision. *Sexual Offender Treatment, 2*(2), 1-14.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Hoboken, NJ: Wiley.
- Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist, 64*, 515-526. doi:10.1037/a0016755
- Kelley, S. M., Barahal, R. M., Thornton, D., & Olson, G. (2015, April). *How do professionals assess sexual recidivism risk? An international survey of current practices*. Paper presented at the joint conference of the Massachusetts Adolescent Sex Offender Coalition and the Massachusetts Chapter of the Association for the Treatment of Sexual Abusers, Marlborough, MA.
- Kraemer, H. C., Kazdin, A. E., Offord, D. R., Kessler, R. C., Jensen, P. S., & Kupfer, D. J. (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry, 54*, 337-343. doi:10.1001/archpsyc.1997.01830160065009
- Långström, N. (2004). Accuracy of actuarial procedures for assessment of sexual offender recidivism risk may vary across ethnicity. *Sexual Abuse: A Journal of Research and Treatment, 16*, 107-120. doi:10.1177/107906320401600202
- Looman, J., & Abracen, J. (2012, October). *Long-term follow-up of two groups of sex offenders*. Presentation at the Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Denver, CO.
- 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. doi:10.1177/1079063210366039
- McGrath, R. J., Cumming, G. F., Burchard, B. L., Zeoli, S., & Ellerby, E. (2010). *Current practices and emerging trends in sexual abuser management: The Safer Society 2009 North American Survey* (ISBN: 978-1-884444-85-2). Brandon, VT: Safer Society Press.
- McGrath, R. J., Lasher, M. P., & Cumming, G. F. (2012). The Sex Offender Treatment Intervention and Progress Scale (SOTIPS): Psychometric properties and incremental predictive validity with Static-99R. *Sexual Abuse: A Journal of Research and Treatment, 24*, 431-458. doi:10.1177/1079063211432475
- Miller, J., & Maloney, C. (2013). Practitioner compliance with risk/needs assessment tools: A theoretical and empirical analysis. *Criminal Justice and Behavior, 40*, 716-736. doi:10.1177/0093854812468883
- Nafekh, M., & Motiuk, L. L. (2002). *The Statistical Information on Recidivism, Revised 1 (SIR-R1) Scale: A psychometric examination*. Ottawa, Ontario: Correctional Service of Canada.
- Nunes, K. L., Babchishin, K. M., & Cortoni, F. (2011). Measuring treatment change in sexual offenders: Clinical and statistical significance. *Criminal Justice and Behavior, 38*, 157-173. doi:10.1177/0093854810391054
- Olver, M. E., Beggs Christofferson, S. M., Grace, R. C., & Wong, S. C. (2014). Incorporating change information into sexual offender risk assessments using the Violence Risk Scale–Sexual Offender version. *Sexual Abuse: A Journal of Research and Treatment, 26*, 472-499. doi:10.1177/1079063213502679
- Olver, M. E., Wong, S. C., Nicholaichuk, T., & Gordon, A. (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. doi:10.1037/1040-3590.19.3.318
- Quinsey, V. L., & Maguire, A. (1986). Maximum security psychiatric patients: Actuarial and clinical prediction of dangerousness. *Journal of Interpersonal Violence, 1*, 143-171. doi:10.1177/088626086001002002
- Saum, S. (2006). *A comparison of an actuarial risk prediction measure (Static-99) and a stable dynamic risk prediction measure (Stable-2000) in making risk predictions for a group of sexual offenders* (Doctoral dissertation). Available from ProQuest Dissertations. (3255539)
- Skeem, J. L., & Monahan, J. (2011). Current directions in violence risk assessment. *Current Directions in Psychological Science, 20*, 38-42. doi:10.1177/0963721410397271
- Soothill, K. L., & Gibbens, T. C. N. (1978). Recidivism of sexual offenders. *British Journal of Criminology, 18*, 267-276.
- Sowden, J. N. (2013). *Examining the relationship of risk, treatment readiness, and therapeutic change to recidivism in a sample of treated sex offenders* (Doctoral thesis). University of Saskatchewan, Saskatoon.
- Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1*, 1-26. doi:10.1111/1529-1006.001
- Thornton, D. (2002). Constructing and testing a framework for dynamic risk assessment. *Sexual Abuse: A Journal of Research and Treatment, 14*, 139-153. doi:10.1177/107906320201400205

- Thornton, D., & Knight, R. A. (2015). Construction and validation of the SRA-FV need assessment. *Sexual Abuse: A Journal of Research and Treatment, 27*, 360-375. doi:10.1177/10790632135111120
- Vergouwe, Y., Steyerberg, E. W., Eijkemans, M. J. C., & Habbeman, J. D. F. (2005). Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *Journal of Clinical Epidemiology, 58*, 475-483. doi:10.1016/j.jclinepi.2004.06.017
- Webb, L., Craissati, J., & Keen, S. (2007). Characteristics of Internet child pornography offenders: A comparison with child molesters. *Sexual Abuse: A Journal of Research and Treatment, 19*, 449-465. doi:10.1007/s11194-007-9063-2
- Wormith, J. S., Hogg, S., & Guzzo, L. (2012). The predictive validity of a general risk/needs assessment inventory on sexual offender recidivism and an exploration of the professional override. *Criminal Justice and Behavior, 39*, 1511-1538. doi:10.1177/0093854812455741

R. Karl Hanson, PhD, CPsych, is a research manager with Public Safety Canada and an adjunct research professor in the Psychology Department of Carleton University, Ottawa. His research focuses on the psychological assessment of the risk-relevant propensities of offenders in the criminal justice system.

Leslie-Maaike Helmus, PhD, is a researcher with Public Safety Canada in Ottawa. Her research interests focus on the assessment and treatment of sexual offenders. She has been part of the development team for several structured risk assessment tools, including the STABLE-2007/ACUTE-2007, Static-99R, Static-2002R, and BARR-2002R.

Andrew J. R. Harris, PhD, CPsych, is in private clinical and consulting practice and is director of the Forensic Assessment Group (www.offenderrisk.com). He speaks on the history of prison architecture, risk assessment for violent, sexual, and pornography offenders, psychopathy, high-risk offenders, and sexual offenders with prominent cognitive delays.