Sexual Abuse: A Journal of Research and Treatment

http://sax.sagepub.com

Long-Term Predictive Validity of the Risk Matrix 2000: A Comparison With the Static-99 and the Sex Offender Risk Appraisal Guide

Drew A. Kingston, Pamela M. Yates, Philip Firestone, Kelly Babchishin and John M. Bradford

The online version of this article can be found at:
http://sax.sagepub.com/cgi/content/abstract/20/4/466

Additional services and information for *Sexual Abuse: A Journal of Research and Treatment* can be found at:

Email Alerts: http://sax.sagepub.com/cgi/alerts

Subscriptions: http://sax.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations http://sax.sagepub.com/cgi/content/refs/20/4/466

Long-Term Predictive Validity of the Risk Matrix 2000

A Comparison With the Static-99 and the Sex Offender Risk Appraisal Guide

Drew A. Kingston
University of Ottawa, Ontario, Canada
Pamela M. Yates
Cabot Consulting and Research Services, Ottawa, Ontario, Canada
Philip Firestone
University of Ottawa, Ontario, Canada
Kelly Babchishin
Carleton University, Ottawa, Ontario, Canada
John M. Bradford
*Royal Ottawa Health Care Centre and University of Ottawa Institute
of Mental Health Research, Ontario, Canada*

The purpose of this study is to examine the predictive accuracy of the Risk Matrix 2000 on an independent sample of 351 sexual offenders, followed in the community for an average duration of 11.4 years (range 0-20 years, $SD = 4.4$ years). For comparison purposes, this study also examines the predictive accuracy of two other risk assessment instruments, specifically modified versions of the Static-99 and the Sex Offender Risk Appraisal Guide (SORAG). Results indicate that the Risk Matrix 2000 demonstrates convergent validity by correlating with the other risk assessment instruments. Moreover, the Risk Matrix 2000 is predictive of recidivism above chance levels, exhibiting medium to large effect sizes, although in general, the other two instruments, particularly the SORAG, are superior. Results also indicate differences in predictive validity when comparing 2-year, 5-year, and variable follow-up periods. Finally, a cumulative meta-analysis compares and integrates current findings with those obtained from the accumulation of previous studies, and new cumulative estimates are provided.

Keywords: *sex offender; recidivism; assessment; Risk Matrix 2000; Static-99; SORAG*

466

Risk assessment of sexual offenders has become increasingly important in recent years, and considerable progress has been made in developing risk assessment tools. Most commonly, assessing risk among offender populations is done for dispositional purposes, such as sentencing, civil commitment, and long-term community supervision. In these cases, decisions are made on the basis of the likelihood of recidivism, and resources are allocated accordingly to promote community safety. Equally important, however, is the application of such instruments for case management and treatment planning (Doren, 2002; Hanson & Yates, 2004; Langton et al., 2007; Yates & Kingston, 2007).

Several factors influence the selection of a particular assessment methodology, and the overall purpose guiding assessment directly informs the manner in which the evaluation is conducted (Doren, 2006). Current approaches to risk assessment can be classified on the basis of the selection of items used in evaluation and the decision-making processes employed (Barbaree, Langton, & Peacock, 2006; Doren, 2006). Recently, Hanson and Morton-Bourgon (2007) described four types of risk assessments, including empirical actuarial, conceptual actuarial, structured professional judgment, and unstructured clinical judgment. Although each methodology provides information that can be useful for different purposes (Doren, 2002; Yates & Kingston, 2007), actuarial instruments are most frequently used in the assessment of sexual offenders.

Current actuarial measures are empirically based assessment methods that, although some are theoretically driven, tend to be atheoretical in that they comprise items demonstrating a statistical, rather than conceptual, relationship with recidivism. For dispositional purposes in particular, actuarial instruments are predominantly employed, as they have accumulated a significant amount of empirical support (Hanson, 1998; Hanson & Morton-Bourgon, 2007; Quinsey, Harris, Rice, & Cormier, 1998, 2006). Examples of actuarial measures currently in use include the Rapid Risk Assessment of Sexual Offence Recidivism (Hanson, 1997), the Static-99 (Hanson & Thornton, 2000), the Sex Offender Risk Appraisal Guide (SORAG; Quinsey et al., 2006), and the Risk Matrix 2000 (RM2000; Thornton et al., 2003).

The first three of the above measures are widely used across various jurisdictions (Archer, Buffington-Vollum, Stredny, & Handel, 2006; Jackson & Hess, 2007) and have been subjected to extensive cross-validation that supports their use with sexual offenders. One relatively recent measure that has been widely adopted by criminal justice agencies throughout the United Kingdom is the RM2000 (Thornton et al., 2003). The RM2000 was intended to be an updated version of the Structured Anchored Clinical Judgement-Min (Grubin, 1998) and was developed to assess risk of sexual and violent reoffending among adult males convicted of sexual offenses. The instrument includes subscales designed to assess risk for sexual recidivism (RM: Sexual), nonsexual violent recidivism (RM: Violent), and sexual or violent recidivism (RM: Combined).

Thornton et al. (2003) validated the RM2000 on several samples from the United Kingdom. RM: Sexual was found to predict sexual recidivism better than chance (receiver operator characteristic [ROC] = .77 and .75 for the first and second cross-validation samples, respectively). Similarly, RM: Violent predicted violent recidivism better than chance in all three cross-validation samples (ROCs = .85, .80, and .78). Finally, the RM: Combined scale was associated with violent (sexual and nonsexual) recidivism in the first (ROC = .81) and second (ROC = .74) cross-validation samples.

Subsequent studies have typically reported significant predictive validity for the RM2000, although the rates have been generally lower than reported in the validation samples. In a sample of convicted sexual offenders (N = 85), Craig, Beech, and Browne (2006), compared the predictive accuracy of the RM2000 subscales (RM: Sexual and RM: Violent) to the Static-99 and the Sexual Violence Risk-20 (Boer, Hart, Kropp, & Webster, 1997) for 2-year, 5-year, and 10-year follow-up periods. Overall, results indicated that, although the RM2000 subscales performed better than the other two risk assessment tools for predicting sexual recidivism, none of the instruments were statistically significant in predicting this outcome. The RM: Violent subscale, however, was statistically associated with violent, sexual and violent, and general reoffending during each follow-up period (rs ranged from .17 to .48). In a community sample of rapists (N = 80) and child molesters (N = 230), Craissati and Beech (2005) reported that the RM: Sexual score predicted any type of recidivism (ROC = .70) or preoffense behavior (e.g., making contact with victim; ROC = .65) better than chance. Most recently, Looman and Abracen (2008) examined the subscales of the RM2000 and found each to be predictive of both sexual and violent (including sexual) recidivism.

A recent meta-analysis (Hanson & Morton-Bourgon, 2007) also provided some support for the RM2000. Specifically, RM: Sexual demonstrated a large mean effect size (d = 0.82, k = 6, 95% confidence interval [CI] = .68-.97) in predicting sexual recidivism, and RM: Violent demonstrated a large mean effect size (d = 0.98, k = 3, 95% CI = 0.81-1.16) in predicting violent recidivism. RM: Sexual demonstrated only a small mean effect size (d = 0.56, k = 3, 95% CI = .35-.78) in predicting any criminal recidivism.

The RM2000 has not been extensively validated on diverse samples (i.e., outside the United Kingdom) and, as such, requires additional research prior to use in applied contexts and in other jurisdictions. This study examined the predictive accuracy of a slightly modified version of the RM2000 in a Canadian sample of sexual offenders and represents an independent cross-validation of this measure. This is, to our knowledge, one of only two studies (with the other being Looman & Abracen, 2008) to evaluate this measure with a sample of sexual offenders outside the United Kingdom. In addition, data on two other risk assessment instruments (Static-99 and SORAG) were analyzed to allow for comparisons across the scales. Portions of the data pertaining to the Static-99 and SORAG have been published previously (Nunes,

Firestone, Bradford, Greenberg, & Broom, 2002). However, in contrast to Nunes et al. (2002), this investigation had a longer follow-up period (11.4 years vs. 7.3 years) and recorded every offense during the follow-up period rather than only the first reoffense. In other words, in Nunes et al., if an offender had any other conviction after the first incident of recidivism, this was not recognized as further evidence of recidivism, resulting in an underestimation of actual recidivism rates.

Method

Participants

All participants were assessed between 1982 and 1992 at the Royal Ottawa Hospital Sexual Behaviors Clinic, a predominantly outpatient clinic that conducts assessments on men and women with problematic sexual behaviors and interests. The 351 participants were adult males who were convicted of contact sexual offenses and assessed just prior to or just after sentencing. In terms of the index offense, 150 (43%) had offended against a related child victim (i.e., biological child, stepchild, niece, grandchild, or sibling), 130 (37%) had offended against an unrelated child victim, and 71 (20%) had offended against an unrelated adult female. Participants who had offended against both related and unrelated victims or against both adults and children were not available in this study's database. RM2000 scores were available for all participants, but given our utilization of an archival database, a smaller proportion of individuals (n = 192) had Static-99 and SORAG scores available.

The average age of the sample was 36.8 years (range: 18-73; SD = 11.4), and the average education level was 10.2 years (SD = 3.2 years). Of the 351 participants, 21.9% had previous charges or convictions for sexual offenses, 38.5% had previous charges or convictions for violent (including sexual) offenses, and 55% had previous charges or convictions for any criminal offenses. Average scores on both the Static-99 (M = 1.92, SD = 1.89) and the SORAG (M = 2.58, SD = 1.68) were low. All participants signed a consent form at the time of assessment, permitting use of their data for research, which was conducted in compliance with the internal review board of the hospital.

Measures

Static-99. The Static-99 (Hanson & Thornton, 1999, 2000) is a brief actuarial instrument designed to predict the long-term probability of sexual recidivism among adult male sexual offenders. The 10 items in this measure are summed to yield four risk categories: low (0-1); medium-low (2-3); medium-high (4-5); and high (6-12;

Hanson & Thornton, 2000). The scale has consistently shown high reliability and validity in numerous independent studies (Barbaree, Seto, Langton, & Peacock, 2001; de Vogel, de Ruiter, van Beek, & Mead, 2004; Ducro & Pham, 2006; Hanson, 2001; Nunes et al., 2002; Sjöstedt & Långström, 2001), as well as in recent meta-analyses (e.g., Hanson & Morton-Bourgon, 2007).

SORAG. The SORAG was developed by integrating variables associated with violent recidivism, extrapolated from the Violence Risk Appraisal Guide (Quinsey et al., 1998, 2006), with the addition of items associated with sexual offending (e.g., deviant sexual preference). The measure consists of 14 items measuring criminal history, demographic information, early behavior problems, and psychiatric diagnoses and is particularly comprehensive. Compared with the Static-99 and RM2000, however, it is relatively resource intensive to complete. SORAG scores can range from –27 to +51, which can be divided into nine equal-sized risk "bins." The predictive validity of the SORAG for both sexual and violent recidivism has been supported in a variety of studies (e.g., Dempster, 1998; Hanson & Morton-Bourgon, 2007; Harris & Rice, 2003; Looman, 2006).

RM2000. The RM2000 (Thornton et al., 2003) was developed to assess risk for sexual and violent reoffending among adult males convicted of sexual offenses. The RM2000 utilizes a stepwise approach to risk classification. In assessing risk for sexual aggression, offender age, number of sexual appearances, and number of criminal appearances are considered. Next, four aggravating factors (male victim, stranger victim, marital history, and noncontact offense history) are examined, with the presence of two to three risk factors elevating risk by one category and the presence of all four risk factors raising the risk rating by two categories. The risk scores and categories for the sexual subscale (RM: Sexual) are as follows: 0 (low risk), 1-2 (medium risk), 3-4 (high risk), and 5-6 (very high risk). A second subscale (RM: Violent) is designed to assess risk for violent recidivism and is comprised of three items: age, violent appearances, and prior convictions for burglary. The risk scores and categories for RM: Violent are as follows: 0-1 (low risk), 2-3 (medium risk), 4-5 (high risk), and 6-8 (very high risk). The combination of risk categories (both RM: Sexual and RM: Violent) is tabulated to produce an overall level of risk (on a 0-to-6 scale) intended for predicting sexual or other types of violence (see Thornton et al., 2003, for a more complete description of scoring procedures).

Although most studies have typically reported on a portion of the subscales of the RM2000 (e.g., Craig et al., 2006), we included every subscale contained in the measure. In other words, this study included the RM: Sexual subscale (scored on a 1-to-4 scale, representing the categories low risk to very high risk), the two RM: Violent scales (scored both on a 0-to-8 scale for the RM: Violent *score*, and on a 1-to-4 scale for the RM: Violent *level*, representing the categories low risk to very high risk), and the RM:

Combined scale (scored on a 0-to-6 scale). Similar to the other two risk assessment instruments, average scores on the RM: Sexual ($M = 1.83$, $SD = 0.78$), RM: Violent score ($M = 1.85$, $SD = 1.58$), RM: Violent level ($M = 1.72$, $SD = 0.80$), and RM: Combined score ($M = 1.55$, $SD = 1.43$) were low among our sample.

Procedure

All instruments were scored without knowledge of recidivism outcome and were based on information gathered prior to an individual's release date. Scoring of the risk assessment measures adhered to their respective coding guidelines (Hanson & Thornton, 1999; Quinsey et al., 1998, 2006; Thornton, 2000; Thornton et al., 2003); however, given our use of a preexisting database, there were slight coding deviations. Specifically, for the RM2000, the "age at next opportunity to offend" and "age on release" items were modified, such that we scored these items as "age at time of assessment." For both the Static-99 and the SORAG, there were several deviations involving demographic information and offense characteristics, which have been described in detail elsewhere (see Nunes et al., 2002, for a more complete description). For example, the "never married" items on both the Static-99 and the SORAG were scored based on whether the offender reported that he had ever been married or lived in a common-law relationship, regardless of the amount of time he had cohabited with his partner. Items that were unavailable in our database pertaining to the SORAG ("elementary school maladjustment") and both the Static-99 and RM2000 ("noncontact sex offense convictions") were scored as zero, consistent with Nunes et al. (2002).

Given these deviations, all three measures are best described as modified versions of the original scales. Although the adjective *modified* may not always be used throughout this article, any reference to the RM2000, Static-99, and SORAG pertains to these modified versions. It should be noted that such modifications have not been found in other research to detract from these instruments' predictive validity (Kingston, Fedoroff, Firestone, Curry, & Bradford, 2008; Nunes et al., 2002).

Recidivism Analyses

The dependent measures in this study were organized in a cumulative hierarchical manner, including specific categories of recidivism and a comprehensive category that included all types of recidivism. This classification method is similar to that of other studies (Firestone, Kingston, Wexler, & Bradford, 2006; Kingston, Firestone, Wexler, & Bradford, 2008; Rice, Quinsey, & Harris, 1991) and allows for the inclusion of sexually motivated offenses that may not have resulted in a conviction for a sexual offense, such as in plea bargains (Rice, Harris, Lang, & Cormier, 2006).

Recidivism data were obtained from the Canadian Police Information Centre (CPIC), a national database of criminal arrests and convictions. Recidivism categories included the following: (a) any criminal recidivism, which was used as the comprehensive overall measure of recidivism and included any charge or conviction noted in the CPIC report (i.e., sexual, violent, or nonviolent); (b) violent (including sexual) recidivism, which was defined as any charge or conviction for a violent or sexual offense (e.g., assault, assault causing bodily harm, invitation to sexual touching); and (c) sexual recidivism, defined as any charge or conviction for a sexual offense (e.g., invitation to sexual touching). Recidivism was calculated to include all offenses, regardless of when these offenses occurred during the follow-up period, to capture all recidivism evident during the follow-up period. As such, it is important to note that individuals were not deleted from analysis following the first reoffense committed.

In contrast to the above, analyses involving fixed follow-up periods included only the first incident of recidivism, rather than all incidents of recidivism. Follow-up time and opportunity to reoffend were dependent on individuals' initial assessment date at the clinic, which varied between March 1982 and November 1992. The *at-risk* period was calculated as the latest of three possible incidents: (a) date of conviction, (b) date of assessment, or (c) date of release if incarcerated. The at-risk period ended either when a new charge or conviction was incurred or at the study end date if no recidivism was evident. It should be stressed that because recidivism in this study reflected only detected charges or convictions, recidivism rates were likely underestimates of actual reoffense rates because some men who reoffended would not have been apprehended.

Statistical Analyses

The primary index of predictive accuracy was the area under the curve (AUC) of the ROC. AUC values are an improvement over other common indices of predictive accuracy (e.g., Pearson product–moment correlation coefficient) as they are unaffected by recidivism base rates or selection ratios (Rice & Harris, 1995, 2005; Swets, 1986). AUC values, which can range from 0 to 1, can be interpreted as the probability that a randomly selected recidivist has a higher score on a particular instrument than a randomly selected nonrecidivist. Values of 1 represent perfect predictive accuracy, whereas values of .5 indicate that an instrument is performing at chance levels of predictive accuracy. For descriptive purposes, minimum AUC values of .56, .64, and .71 are described as small, medium, and large, respectively (Rice & Harris, 2005).

In this investigation, ROC analyses were carried out with the Statistical Package for the Social Sciences (SPSS) Version 15. To detect differences between two AUC values, we determined the critical ratio *z*, the ratio of a difference score to the standard error of the difference score, using the formula shown below. This method is more appropriate than other methods (e.g., calculating the standard error of the

difference between two areas) when both modalities are correlated (Hanley & McNeil, 1983), as was the case in this study.

$$z = \frac{A_1 - A_2}{\sqrt{SE_1^2 + SE_2^2 - 2rSE_1SE_2}}$$

where A_1 and SE_1 refer to the AUC value and standard error of the first modality, A_2 and SE_2 refer to the AUC value and standard error for the second modality, and r refers to the correlation between the two constructs under investigation. A Bonferroni correction was applied to correct for spurious findings (i.e., based on 30 possible comparisons), and thus, statistical significance was set at $p < .001$.

In addition to the above, we determined how the predictive accuracy of the RM2000 in this study compared with the mean effect sizes reported in the most recent accumulation of previous studies (see Hanson & Morton-Bourgon, 2007) using cumulative meta-analysis (Lau, Schmid, & Chambers, 1995) and the procedures and formulas described by Hanson and Broom (2005) to calculate new cumulative estimates. For this analysis, we utilized Cohen's d as the index of predictive accuracy to facilitate comparison with Hanson and Morton-Bourgon (2007). The d statistic was determined on the basis of the relationship to the corresponding AUC value (see Rice & Harris, 2005), and it is generally suggested that d values of 0.20 are small, 0.50 are medium, and 0.80 are large (Cohen, 1988, 1992).

Results

Recidivism Rates

The overall recidivism rates in this sample were 20.5%, 33.6%, and 43.9% for sexual, violent (including sexual), and any criminal recidivism, respectively. The follow-up period ranged up to 20 years, with an average time-at-risk of 11.4 years ($SD = 4.4$ years). In terms of sexual offender type, the sexual, violent (including sexual), and criminal recidivism rates were as follows: intrafamilial child molesters, 12%, 24.7%, and 33.3%, respectively; extrafamilial child molesters, 26.9%, 40%, and 49.2%, respectively; and, rapists, 26.8%, 40.8%, and 56.3%, respectively.

Intercorrelations Among Risk Assessment Instruments

To examine the relationship between the RM2000, Static-99, and SORAG, Pearson product–moment correlation coefficients were calculated. As indicated in Table 1, all subscales of the RM2000 were significantly correlated with one another. The amount of variability accounted for across RM2000 subscales ranged from 32%

Table 1
Correlation Matrix for Risk Assessment Measures

Measure	RM: S	RM: V Score	RM: V Level	RM: C	Static-99	SORAG
RM: S	—	.57 (351)	.62 (351)	.90 (351)	.76 (192)	.61 (192)
RM: V score		—	.95 (351)	.85 (351)	.50 (192)	.64 (192)
RM: V level			—	.90 (351)	.54 (192)	.68 (192)
RM: C				—	.72 (192)	.71 (192)
Static-99					—	.69 (192)

Note: Numbers in parentheses are *N* of sample. RM: S = Risk Matrix: Sexual; RM: V score = Risk Matrix: Violent score; RM: V level = Risk Matrix: Violent level; RM: C = Risk Matrix: Combined score; SORAG = Sex Offender Risk Appraisal Guide.
All correlations significant at $p < .001$.

(RM: Sexual and RM: Violent score; $r^2 = .32$) to 90% (RM: Violent level and RM: Violent score $r^2 = .90$). The RM2000 subscales were also highly correlated with the Static-99, accounting for 25% (RM: Violent score and Static-99; $r^2 = .25$) to 58% (RM: Sexual and Static-99; $r^2 = .58$) of the variance. Similarly, the RM2000 was significantly correlated with the SORAG, accounting for 37% (RM: Sexual and SORAG; $r^2 = .37$) to 50% (RM: Combined score and SORAG; $r^2 = .50$) of the variance. Finally, the Static-99 was significantly correlated with the SORAG, accounting for 48% of the variance ($r^2 = .48$).

Predictive Validity

Table 2 presents the AUC values and 95% CIs for the subscales of the RM2000 and the other two risk assessment instruments. The formula indicated above was utilized to compare AUC values both across measures and for the RM2000 subscales, with the criterion for statistical significance set at $p < .001$ to correct for multiple comparisons. AUC values ranged from medium to large using the aforementioned criteria. All values were statistically significant, indicating predictions that were better than chance.

For sexual recidivism, predictive accuracy ranged from medium (RM: Violent score and RM: Violent level; ROCs = .64, 95% CI = .57-.71) to large (SORAG; ROC = .77, 95% CI = .69-.85). The Static-99 was significantly more accurate than the RM: Sexual in predicting sexual recidivism ($z = 4.23, p < .0001$), RM: Violent level ($z = 3.13, p < .001$), and RM: Combined score ($z = 3.60, p < .0001$). Similarly, the SORAG demonstrated significantly higher predictive accuracy than all RM subscales ($zs > 3.72, ps < .0001$). No other significant differences were found in the prediction of sexual recidivism.

In predicting violent (including sexual) recidivism, predictive accuracy ranged from medium (RM: Sexual; ROC = .65, 95% CI = .60-.71) to large (SORAG;

Table 2
Predictive Accuracy of the RM2000, Static-99, and SORAG

| | Type of Recidivism | | | | | |
| | Sexual | | Violent (Including Sexual) | | Any Criminal | |
Instrument	AUC	95% CI	AUC	95% CI	AUC	95% CI
RM: S	.65	.59-.72	.65	.60-.71	.69	.63-.74
RM: V score	.64	.57-.71	.70	.64-.76	.74	.68-.79
RM: V level	.64	.57-.71	.69	.63-.74	.72	.67-.77
RM: C	.66	.60-.73	.69	.63-.75	.73	.67-.78
Static-99	.76	.68-.84	.70	.62-.77	.72	.64-.79
SORAG	.77	.69-.85	.76	.69-.83	.80	.73-.86

Note: AUC = area under the curve; RM: S = Risk Matrix: Sexual; RM: V score = Risk Matrix: Violent score; RM: V level = Risk Matrix: Violent level; RM: C = Risk Matrix: Combined score; SORAG = Sex Offender Risk Appraisal Guide.
All AUCs significant at $p < .001$.

ROC = .76, 95% CI = .69-.83). The SORAG was statistically superior when compared with the RM: Sexual ($z = 3.42$, $p < .0001$). No other significant differences were found in the prediction of violent (including sexual) recidivism.

Finally, for criminal recidivism, predictive accuracy ranged from medium (RM: Sexual; ROC = .69, 95% CI = .63-.74) to large (SORAG; ROC = .80, 95% CI = .73-.86). The SORAG demonstrated significantly higher AUC values than the RM: Sexual ($z = 4.15$, $p < .0001$) and RM: Violent level ($z = 3.33$, $p < .0001$). No other significant differences were found in the prediction of criminal recidivism.

It has been suggested that clinicians may be particularly concerned with assessing risk within relatively short time periods (Bengtson & Långström, 2007; Harris et al., 2003). Therefore, to provide a direct comparison with the original cross-validation samples described in Thornton et al. (2003), we examined the predictive accuracy of the RM2000 subscales using a 2-year fixed follow-up period. Also, for illustrative purposes, we provided data as it pertains to a 5-year fixed follow-up period. As such, we considered only offenders who had at least 2 years or 5 years of opportunity to reoffend (depending on the specific analysis), with recidivists being those individuals who reoffended within that specified period. Individuals who reoffended after the relevant period were counted as nonrecidivists.

To compare our results with the original cross-validation samples described in Thornton et al. (2003), we utilized the RM: Sexual, RM: Violent level, and RM: Combined score in relation to observed sexual and violent (including sexual) recidivism rates (see Tables 3, 4, and 5). A Fisher's exact test was used to compare the fixed 2-year recidivism rates between the present sample and those reported in Thornton et al. (2003). As shown in these tables, individuals classified within the

Table 3
Sexual Recidivism Rates for RM: S Risk Categories
for 2-Year and 5-Year Fixed Follow-Up Periods

| | Sexual Recidivism | | |
| | 2 Years | | 5 Years |
RM: S Risk Categories	Present Sample % (Nos.)	Thornton et al., 2003 % (Nos.)	Present Sample % (Nos.)
I: Low risk	1.5 (2 of 131)	0.9 (2 of 215)	3.9 (5 of 129)
II: Medium risk	5.1 (6 of 118)	1.3 (4 of 298)*	13.7 (14 of 102)
III: High risk	4.9 (3 of 61)	5.7 (6 of 105)	11.5 (6 of 52)
IV: Very high risk	33.3 (1 of 3)	17.2 (5 of 29)	66.7 (2 of 3)
Total	3.8 (12 of 313)	2.6 (17 of 647)	9.4 (27 of 286)

Note: RM: S = Risk Matrix: Sexual; Recidivism rates were based on the first incident of recidivism. A Fisher's exact test was used to identify potential differences between 2-year recidivism rates in the present sample and Thornton et al. (2003).
*$p < .05$.

Table 4
Violent (Including Sexual) Recidivism Rates for RM:
V Risk Categories for 2-Year and 5-Year Fixed Follow-Up Periods

| | Violent (Including Sexual) Recidivism | | |
| | 2 years | | 5 years |
RM: V Risk Categories	Present Sample % (Nos.)	Thornton et al., 2003 % (Nos.)	Present Sample % (Nos.)
I: Low risk	2.6 (4 of 154)	0 (0 of 297)*	4.7 (7 of 150)
II: Medium risk	9.2 (11 of 120)	2.8 (6 of 218)*	24.1 (26 of 108)
III: High risk	7.4 (2 of 27)	3.0 (3 of 100)	25.0 (5 of 20)
IV: Very high risk	8.3 (1 of 12)	18.8 (6 of 32)	37.5 (3 of 8)
Total	5.8 (18 of 313)	2.3 (15 of 647)**	14.3 (41 of 286)

Note: RM: S = Risk Matrix: Sexual; Recidivism rates were based on the first incident of recidivism. A Fisher's exact test was used to identify potential differences between 2-year recidivism rates in the present sample and Thornton et al. (2003).
*$p < .05$. **$p < .01$.

lowest risk categories tended to reoffend at lower rates than individuals classified at higher risk levels. However, the trend toward increasing recidivism rates in relation to higher risk levels was not entirely linear, as was demonstrated in the original cross-validation samples. It is important that, as shown in the relevant tables, when

Table 5
Violent (Including Sexual) Recidivism Rates for
RM: C Scores for 2-Year and 5-Year Fixed Follow-Up Periods

	Violent (Including Sexual) Recidivism		
	2 years		5 years
RM: C scores	Present Sample % (Nos.)	Thornton et al., 2003 % (Nos.)	Present Sample % (Nos.)
0	1.6 (2 of 124)	0 (0 of 7)	3.3 (4 of 122)
1	6.5 (2 of 31)	0 (0 of 35)	13.8 (4 of 29)
2	9.3 (7 of 75)	1 (1 of 74)	26.5 (18 of 68)
3	8.2 (5 of 61)	10 (6 of 61)	17.6 (9 of 51)
4	6.3 (1 of 16)	16 (8 of 51)	41.7 (5 of 12)
5	0 (0 of 5)	24 (9 of 37)	0 (0 of 3)
6	100 (1 of 1)	46 (5 of 11)	100 (1 of 1)
Total	5.8 (18 of 313)	10.5 (29 of 276)*	14.3 (41 of 286)

Note: RM: S = Risk Matrix: Sexual; recidivism rates were based on the first incident of recidivism. A Fisher's exact test was used to identify potential differences between 2-year recidivism rates in the present sample and Thornton et al. (2003).
*$p < .05$.

a liberal criterion for statistical significance was used, there were few statistically significant differences between the rates in this study and those reported in Thornton et al. (2003), and there were no differences in observed recidivism rates when multiple comparisons were corrected for (i.e., if statistical significance was set at $p < .003$).

With regard to predictive accuracy, the AUC value for the RM: Sexual in predicting sexual recidivism at a fixed 2-year follow-up was equal to the value produced when an open-ended time frame was used but was no longer statistically significant (ROC = .65, 95% CI = .50-.80). In terms of violent recidivism, the RM: Violent level produced a lower but significant AUC value than when a variable time frame was employed (ROC = .64, 95% CI = .52-.76). Finally, the RM: Combined score produced a lower AUC value than the original estimate but remained predictive above chance levels (ROC = .66, 95% CI = .55-.77).

With regard to the fixed 5-year follow-up period, predictive accuracy of the RM: Sexual in predicting sexual recidivism was approximately equal to the rate produced when an open-ended time frame was used and was statistically significant (ROC = .66, 95% CI = .55-.76). In terms of violent recidivism, the RM: Violent produced higher and statistically significant levels of predictive accuracy when compared with the variable time frame analyses (ROC = .71, 95% CI = .63-.79). Similarly, the RM: Combined score produced a higher AUC value when compared with the original estimate and was statistically significant in predicting this outcome (ROC = .71, 95% CI = .64-.79).

Cumulative Meta-Analysis of the RM2000

To compare and integrate the current findings with results obtained from the most recently available accumulation of previous studies (Hanson & Morton-Bourgon, 2007), we conducted a cumulative meta-analysis (Lau et al., 1995) utilizing the procedures and formula outlined by Hanson and Broom (2005). As indicated earlier, Hanson and Morton-Bourgon (2007) found that the RM: Sexual subscale demonstrated a large mean effect size in predicting sexual recidivism and a medium mean effect size in predicting any criminal recidivism. They also found a large mean effect size for the RM: Violent in predicting violent recidivism. By comparison, our results demonstrated medium effect sizes between RM: Sexual and sexual recidivism (d = 0.55, 95% CI = .32-.82) and any recidivism (d = 0.71, 95% CI = .47-.91). Finally, in terms of violent recidivism, the RM: Violent level, in our study, revealed a medium mean effect size (d = 0.71, 95% CI = .47-.91).

Compared with previous findings, the association between RM: Sexual and sexual recidivism was not statistically different in our study (Q_Δ = 3.36, df = 1, p = .067), nor was there a significant difference between our findings for RM: Sexual and previous findings concerning the prediction of any criminal recidivism (Q_Δ = .92, df = 1, p = .337). Similarly, the association between RM: Violent and violent recidivism was not statistically different in our study when compared with the accumulation of previous studies (Q_Δ = 3.55, df = 1, p = .059). The new cumulative effect size between RM: Sexual and sexual recidivism was large (d = 0.75, 95% CI = .63-.87), and when the increase in variability (Q_Δ = 3.36) was added to the old variability (Q_{old} = 4.50), the total variability was no more than would be expected by chance (Q_{new} = 7.86, df = 6, p = .249). The new cumulative effect size statistic for the RM: Sexual in predicting any criminal recidivism was medium (d = 0.63, 95% CI = .48-.78), and when the increase in variability (Q_Δ = .92) was added to the old variability (Q_{old} = 1.21), the total variability was no more than would be expected by chance (Q_{new} = 2.13, df = 3, p = .546). Finally, the new cumulative effect size statistic for the RM: Violent and violent recidivism was large (d = 0.88, 95% CI = 0.74-1.02), and when the increase in variability (Q_Δ = 3.55) was added to the old variability (Q_{old} = 2.11), the total variability was no more than would be expected by chance (Q_{new} = 5.66, df = 3, p = .129).

Discussion

The purpose of this study was to examine the long-term predictive accuracy of the RM2000 among an independent sample of sexual offenders and to compare such results with other commonly used risk assessment measures (i.e., the Static-99 and SORAG). This study employed an extended average follow-up period of approximately 12 years and represents one of only two cross-validation studies of the

RM2000 to be conducted using a North American sample. In addition, a cumulative meta-analysis was conducted to integrate results from this study in the context of the extant literature, and new cumulative estimates were provided. Finally, analyses involving fixed follow-up periods were conducted to facilitate both comparison with the original validation studies of the measure (Thornton et al., 2003) and the accuracy of predictive estimates over shorter time frames with which clinicians may be particularly concerned (Bengtson & Långström, 2007; Harris et al., 2003).

Results provided preliminary support overall for the RM2000. The convergent validity of the measure was supported, as demonstrated by significant associations with both the Static-99 and the SORAG. Specifically, large, positive correlations between RM2000 scores and the other two risk assessment measures were indicated, accounting for between 25% and 57% of the variance. However, these associations may have resulted, in part, from similar items contained within these measures.

The predictive accuracy of the RM2000 was also supported and ranged from medium to large, depending on the specific subscale utilized and the outcome measure examined. All RM2000 subscales and the total combined scale significantly predicted all types of recidivism. For sexual recidivism, AUCs were moderate; for violent recidivism and any criminal recidivism, AUCs ranged from medium to large. These findings support the predictive validity of the RM2000 with an independent, North American sample of sexual offenders. In comparison with the original development and validation samples (Thornton et al., 2003), the predictive accuracy of the RM2000 was generally lower in this study, although readers should note that the difference was within the range expected by chance. Nonetheless, it is possible that difference could arise from jurisdictional and setting differences, such as population stability and availability of recidivism data in the United Kingdom. Findings from this study were more consistent with subsequent cross-validation studies than they were with the original Thornton et al. (2003) findings. For example, the present results for RM: Sexual were similar to Craissati and Beech (2005) in predicting any criminal reoffenses, to Craig et al. (2006) with respect to the prediction of sexual recidivism (RM: Sexual and RM: Violent score), and to findings reported in Looman and Abracen (2008) for the RM: Violent scale and RM: Combined score in predicting sexual recidivism.

The somewhat higher AUC values for the RM2000 scores in predicting any criminal recidivism, compared with sexual or violent recidivism, suggest that the RM2000 may be particularly suited for this type of outcome, although this will need to be tested in larger and more diverse samples. Nevertheless, this is consistent with research and theory suggesting that sexual offense risk is multidimensional in nature. Indeed, at least two distinct dimensions of risk among sexual offenders have been identified (antisociality/general criminality and sexual deviance), actuarial risk assessment instruments often produce different results on these dimensions, and research indicates different pathways to offending that are associated with different risk dimensions

(Barbaree et al., 2006; Doren, 2004; Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005; Knight & Sims-Knight, 2003; Lussier, 2005; Roberts, Doren, & Thornton, 2002; Yates & Kingston, 2006). Taken together, results suggest that the RM2000 may be a useful measure of the general criminality dimension of risk.

Although some investigations have found equivalence across actuarial measures for the prediction of sexual recidivism (e.g., Knight & Thornton, 2006; Langton et al., 2007), this study found that all RM2000 scales were less accurate than the SORAG and Static-99 for the prediction of sexual recidivism. For other outcomes (i.e., violent, criminal), all the significant differences favored the SORAG over the RM2000 scales. Even though the current study supported the validity of the RM2000, it may be that its predictive accuracy could be lower than the Static-99 and SORAG in some instances, depending on the specific outcome measure and subscale examined. Of note, the present results are similar to other findings suggesting variability in predictive accuracy across individual studies (e.g., Hanson & Morton-Bourgon, 2007; Seto, 2005).

As a result of such variability in individual studies, a cumulative meta-analysis was used to integrate our findings with previous meta-analytic results (Hanson & Morton-Bourgon, 2007), resulting in new cumulative estimates. Compared with these previous findings, the current cumulative meta-analysis revealed a lower effect size for RM: Sexual in the prediction of sexual recidivism and a lower effect size for the RM: Violent level in the prediction of violent recidivism. By contrast, the new cumulative estimate for RM: Sexual in predicting any criminal recidivism was higher in our study when compared with the estimate obtained from previous studies. However, none of these cumulative estimates were statistically different from those reported in Hanson and Morton-Bourgon (2007). Similar to the previous meta-analysis, findings from this study indicated that the RM: Sexual subscale demonstrated a large mean effect size in predicting sexual recidivism and a medium effect size in predicting any recidivism. Finally, the effect size for RM: Violent level in predicting violent recidivism was large.

This study also examined the predictive accuracy of the RM2000 at 2-year and 5-year fixed follow-up periods to evaluate the predictive accuracy over shorter periods that may be of interest to clinicians who consider predictive accuracy in the context of risk management (Bengtson & Långström, 2007; Harris et al., 2003). As indicated previously, the former time period was specifically selected to facilitate direct comparisons with the samples described in Thornton et al. (2003). At first glance, recidivism rates for the RM2000 risk categories and scores for the various outcome measures did not follow a linear trend, as demonstrated by Thornton et al. However, after correcting for multiple comparisons, there were no statistical differences between recidivism rates for the present sample and those reported in Thornton et al.

With regard to short-term accuracy (i.e., a 2-year fixed follow-up period), findings indicated that neither the RM: Sexual nor the RM: Violent level significantly predicted

sexual or violent recidivism, respectively, perhaps because lower statistical power led to less precise prediction estimates (wider confidence intervals). Conversely, the RM: Combined score attained significance in the prediction of violent recidivism, albeit with a lower AUC value. By contrast, all findings were statistically significant at the fixed 5-year follow-up period. These results suggest that, although the RM2000 Combined score may predict violent (including sexual) recidivism over a shorter follow-up period, the other subscales exhibited reduced predictive accuracy, although as the time frame increased, so did the accuracy of the subscales.

Several limitations must be considered when interpreting these results and are similar to those reported in another study using portions of the same data set (Nunes et al., 2002). Specifically, this study used modified versions of the RM2000, Static-99, and SORAG because of the reliance on a preexisting database, resulting in missing data and deviations from the scoring protocols for the instruments. Although certain modifications may have lowered risk scores (i.e., the lack of data pertaining to noncontact sexual offenses), other deviations may have slightly raised risk scores (i.e., coding age at time of assessment rather than at time of release). These deviations must be taken into consideration when interpreting our results. In addition, we were unable to evaluate interrater reliability for any of the measures in this study as all instruments were coded on the basis of previously entered file data. The predictive power of all three instruments may have been improved had we been able to strictly adhere to the scoring guidelines. In general, increased accuracy is expected given no missing items and high rater reliability (Harris & Rice, 2003).

It is also noted that the present sample was comprised of a relatively low-risk group, as indicated by scores on the Static-99 and SORAG, as well as by the distribution of scores on the RM2000. Given the characteristics of the present sample, predictive accuracy of the RM2000 may improve with the application of the measure to a sample of offenders who are at relatively higher risk and who demonstrate a broader range of obtained risk scores. The sample was also somewhat selective in that it was composed solely of offenders presenting to a forensic psychiatric hospital for assessment, although this does not appear to have affected the predictive validity of the other assessment instruments included in this study.

In summary, this study provided independent validation for the RM2000 in a North American sample in a forensic hospital setting. The measure, which is easier to implement and score by comparison with some other actuarial risk assessment measures, was found to have strong convergent validity with other measures and good predictive validity for sexual, violent, and any criminal recidivism. Because of the limitations of this study, it is suggested that future research be conducted with an unmodified version of the RM2000, with demonstrated interrater reliability, on a higher risk sample of sexual offenders demonstrating a greater distribution of obtained scores on the measure.

References

Archer, R. P., Buffinton-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment*, *87*, 84-94.

Barbaree, H. E., Langton, C. M., & Peacock, E. J. (2006). Different actuarial risk measures produce different risk rankings for sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, *18*, 423-440.

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*, 490-521.

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.

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, BC: British Columbia Institute Against Family Violence.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.

Craig, L. A., Beech, A. R., & Browne, K. D. (2006). Cross-validation of the Risk Matrix 2000 Sexual and Violent scales. *Journal of Interpersonal Violence*, *21*, 612-633.

Craissati, J., & Beech, A. R. (2005). Risk prediction and failure in a complete urban sample of sex offenders. *Journal of Forensic Psychiatry & Psychology*, *16*, 24-40.

Dempster, R. J. (1998). *Prediction of sexually violent recidivism: A comparison of risk assessment instruments.* Unpublished Masters Thesis, Simon Fraser University, Burnaby, BC, Canada.

de Vogel, V., de Ruiter, C., van Beek, D., & Mead, G. (2004). Predictive validity of the SVR-20 and Static-99 in a Dutch sample of treated sex offenders. *Law and Human Behavior*, *28*, 235-251.

Doren, D. (2002). *Evaluating sex offenders: A manual for civil commitment and beyond.* Thousand Oaks, CA: Sage.

Doren, D. M. (2004). Toward a multidimensional model for sexual recidivism risk. *Journal of Interpersonal Violence*, *19*, 835-856.

Doren, D. M. (2006). Recidivism risk assessments: Making sense of controversies. In W. L. Marshall, Y. M. Fernandez, L. E. Marshall, & G. A. Serran (Eds.), *Sexual offender treatment: Controversial issues* (pp. 3-15). West Sussex, UK: John Wiley.

Ducro, C., & Pham, T. (2006). Evaluation of the SORAG and the Static-99 on Belgian sex offenders committed to a forensic facility. *Sexual Abuse: Journal of Research and Treatment*, *18*, 15-26.

Firestone, P., Kingston, D. A., Wexler, A., & Bradford, J. M. (2006). Long-term follow-up of exhibitionists: Psychological, phallometric, and offence characteristics. *Journal of the American Academy of Psychiatry and the Law*, *34*, 349-359.

Grubin, D. (1998). *Sex offending against children: Understanding the risk* (Police Research Series Paper 99). London: Home Office.

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.

Hanson, R. K. (1997). The development of a brief actuarial risk scale for sexual offence recidivism (User Report No. 1997-04). Ottawa: Department of the Solicitor General of Canada. Retrieved September 3, 2008, from http://ww2.ps-sp.gc.ca/publications/corrections/199704_e.pdf

Hanson, R. K. (1998). What do we know about sex offender risk assessments? *Psychology, Public Policy, and the Law*, 4, 50-72.

Hanson, R. K. (2001). *Note on the reliability of Static-99 as used by the California Department of Mental Health evaluators* (unpublished report). Sacramento: California Department of Mental Health.

Hanson, R. K., & Broom, I. (2005). The utility of cumulative meta-analysis: Application to programs for reducing sexual violence. *Sexual Abuse: A Journal of Research and Treatment*, *17*, 357-373.

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

Hanson, R. K., & Morton-Bourgon, K. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology*, *73*, 1154-1163.

Hanson, R. K., & Morton-Bourgon, K. (2007). *The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis.* Ottawa: Public Safety Canada. Retrieved September 3, 2008, from http://www.ps-sp.gc.ca/res/cor/rep/_fl/crp2007-01-en.pdf

Hanson, R. K., & Thornton, D. (1999). *Static-99: Improving actuarial risk assessments for sex offenders* (User Report 99-02). Ottawa: Department of the Solicitor General of Canada.

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

Hanson, R. K., & Yates, P. M. (2004). Sexual violence: Risk factors and treatment. In M. Eliasson (Ed.), *Anthology on interventions against violent men* (pp. 151-166). Acta Universitatis Upsaliensis, Uppsalla Women's Studies B: Women in the Humanities: 3. Uppsala, Sweden: Uppsala Universitet.

Harris, G. T., & Rice, M. E. (2003). Actuarial assessment of risk among sex offenders. In. R. A. Prentky, E. S. Janus, & M. C. Seto (Eds.), *Sexually coercive behavior: Understanding and management* (pp. 198-210). New York: New York Academy of Sciences.

Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumiere, M. L., Boer, D., & Lang, C. (2003). A multisite comparison of actuarial risk instruments for sex offenders. *Psychological Assessment*, *15*, 413-425.

Jackson, R. L., & Hess, D. T. (2007). Evaluation for civil commitment of sex offenders: A survey of experts. *Sexual Abuse: A Journal of Research and Treatment*, *19*, 409-448.

Kingston, D. A., Fedoroff, P., Firestone, P., Curry, S., & Bradford, J. M. (2008). Pornography use and sexual aggression: The impact of frequency and type of pornography use on recidivism among sexual offenders. *Aggressive Behavior*, *34*, 341-351.

Kingston, D. A., Firestone, P., Wexler, A., & Bradford, J. M. (2008). Factors associated with recidivism among intra-familial child molesters. *Journal of Sexual Aggression*, *14*, 3-18.

Knight, R. A. & Sims-Knight, J. E. (2003). The developmental antecedents of sexual coercion against women: Testing alternative hypotheses with structural equation modelling. In R. A. Prentky, E. S. Janus, & M. C. Seto (Eds.), *Sexually coercive behavior: Understanding and management* (Vol. 989, pp. 72-85). New York: Annals of the New York Academy of Sciences.

Knight, R. A., & Thornton, D. (2006). *Evaluating and improving risk assessment schemes for sexual recidivism: A long-term follow-up of convicted sexual offenders*. US Department of Justice. Retrieved September 3, 2008, from http://www.ncjrs.gov/pdffiles1/nij/grants/217618.pdf

Langton, C. M., Barbaree, H. E., Seto, M. C., Peacock, E. J., Harkins, L., & Hansen, K. T. (2007). Actuarial assessment of risk for reoffence among adult sex offenders: Evaluating the predictive accuracy of the Static-2002 and five other risk assessment instruments. *Criminal Justice and Behavior*, *34*, 37-59.

Lau, J., Schmid, C. H., & Chambers, T. C. (1995). Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care. *Journal of Clinical Epidemiology*, *48*, 45-57.

Looman, J. (2006). Comparison of two risk assessment instruments for sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, *18*, 193-206.

Looman, J., & Abracen, J. (2008). *Comparison of measures of risk for recidivism in sexual offenders*. Manuscript submitted for publication.

Lussier, P. (2005). The criminal activity of sexual offenders in adulthood: Revisiting the specialization debate. *Sexual Abuse: A Journal of Research and Treatment*, *17*, 269-291.

Nunes, K. L., Firestone, P., Bradford, J. M., Greenberg, D. M., & Broom, I. (2002). A comparison of modified versions of the Static-99 and the Sex Offender Risk Appraisal Guide. *Sexual Abuse: A Journal of Research and Treatment*, *14*, 253-269.

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

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk* (2nd ed.). Washington, DC: American Psychological Association.

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

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

Rice, M. E., Harris, G. T., Lang, C., & Cormier, C. (2006). Violent sex offences: How are they best measured from official records? *Law and Human Behavior*, *30*, 525-541.

Rice, M. E., Quinsey, V. L., & Harris, G. T. (1991). Sexual recidivism among child molesters released from a maximum security psychiatric institution. *Journal of Consulting and Clinical Psychology*, *59*, 381-386.

Roberts, C. F., Doren, D. M., & Thornton, D. (2002). Dimensions associated with assessments of sex offender recidivism risk. *Criminal Justice and Behavior*, *29*, 569-589.

Seto, M. C. (2005). Is more better? Combining actuarial risk scales to predict recidivism among adult sex offenders. *Psychological Assessment*, *17*, 156-167.

Sjöstedt, G., & Langström, N. (2001). Actuarial assessment of sex offender recidivism risk: A cross-validation of the RRASOR and the Static-99 in Sweden. *Law and Human Behavior*, *25*, 629-645.

Swets, J. A. (1986). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin*, *99*, 100-117.

Thornton, D. (2000). *Scoring guide for Risk Matrix: 2000.* Unpublished manuscript.

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.

Yates, P. M., & Kingston, D. A. (2006). Pathways to sexual offending: Relationship to static and dynamic risk among treated sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, *18*, 259-270.

Yates, P. M., & Kingston, D. A. (2007). *A companion text to the casebook of sexual offending for use in scoring the Risk for Sexual Violence Protocol (RSVP).* Victoria, British Columbia, Canada: Pacific Psychological Assessment Corporation.