The predictive accuracy of the Risk Matrix 2000: A meta-analysis

Leslie Helmus[1,2], Kelly M. Babchishin[1,2], R. Karl Hanson[2]
[1]Carleton University
[2]Public Safety Canada

Abstract

Background. The Risk Matrix 2000 is a static actuarial tool designed for adult male sex offenders. It consists of three scales (Sex, Violence, and Combined) intended to assess risk for sexual, non-sexual violent, and any violent recidivism, respectively.
Methods. This meta-analysis identified 16 unique samples (from 14 studies) that examined the extent to which the Risk Matrix 2000 scales discriminate recidivists from non-recidivists.
Results/Conclusions. The three Risk Matrix 2000 scales significantly predicted all recidivism types (i.e., sexual, non-sexual violent, any violent, non-violent, and any recidivism). Effect sizes were generally comparable or higher than those found by Hanson and Morton-Bourgon (2009) for the Risk Matrix 2000 and for other actuarial scales designed for a similar purpose. The Sex scale provided the best predictive accuracy for sexual recidivism (d = .74). The Violence and Combined scales both predicted non-sexual violent recidivism and any violent recidivism with similarly large effect sizes. Although the scales were not designed to predict non-violent or any (including violent) recidivism, effect sizes for these outcomes were also moderate to large (d's exceeding .60). Effect sizes were significantly higher in the U.K. than in other countries, and low in samples of sex offenders preselected as unusually high risk or high need. These results support the use of the Risk Matrix in applied risk assessment practice. Additional research on the calibration of the Risk Matrix is needed.

Keywords: risk assessment, Risk Matrix 2000, sexual offenders, meta-analysis

The predictive accuracy of the Risk Matrix 2000: A meta-analysis

Risk assessment is one of the most ubiquitous tasks in the criminal justice system and informs nearly every decision made about an offender. Decades of research have led to the identification of risk factors that predict future reoffending (for meta-analyses of sexual reoffending, see Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005; Mann, Hanson, & Thornton, 2010; for meta-analyses of general reoffending, see Andrews & Bonta, 2010). No risk factor singularly determines recidivism; as such, empirically-informed risk assessment requires synthesis of multiple factors.

Numerous risk assessment scales have been developed for sex offenders. Although structured scales outperform unstructured clinical judgement, no particular risk scale performs consistently better than the others (Hanson & Morton-Bourgon, 2009). Evaluators interested in assessing the risk posed by sexual offenders must, therefore, choose from a variety of scales that were developed for slightly different purposes and will have different strengths or weaknesses (e.g., ease of coding, type of information required, type of recidivism designed to predict).

The Risk Matrix 2000 is a static scale developed to assess the risk of adult male sex offenders (Thornton, 2010; Thornton et al., 2003). Following Meehl's (1954) definition adopted by Hanson and Morton-Bourgon (2009), the Risk Matrix is an empirical-actuarial risk scale because the items were developed based on empirical research, there are mechanical guidelines for computing a total score based on the item codings, and each score can be linked to an empirically-derived estimate of recidivism probability (Thornton, 2010).

In other ways, however, the Risk Matrix 2000 differs from most static actuarial sex offender risk scales. Firstly, most actuarial scales are constructed using a development sample, whereas the Risk Matrix scale was developed based on items from an earlier scale (the SACJ-Min, described in Hanson & Thornton, 2000) as well as meta-analytic research on static risk factors for sex offenders (Hanson & Bussière, 1998). The absence of a large development sample precludes a parsimonious examination of all items similarly defined in the same dataset and examined for incremental predictive accuracy. Advantages of this construction strategy, however, is that the large and diverse body of research informing the scale construction may reduce reliance on idiosyncratic features of an individual sample, thereby enhancing generalizability of the scale.

Another unique feature of the Risk Matrix 2000 is that it contains three separate risk scales designed to capture different dimensions of recidivism risk. The Risk Matrix 2000/Sex scale is designed to predict sexual recidivism. The Risk Matrix 2000/Violence scale is designed to predict non-sexually violent recidivism. Both scales can also be combined into an overall scale (the Risk Matrix 2000/Combined), which is designed to predict any violent recidivism (sexual or non-sexual).

Coding forms for the three scales are included in the Appendix. The Risk Matrix 2000/Sex scale can be easily scored based on commonly available demographic and criminal history information. In the first step, three items are scored and used to assign offenders into one of four preliminary risk levels. In Step 2, four additional risk factors are scored, and the offender's risk level is increased for every two risk factors present. The offender's final risk category is labeled as either low, moderate, high, or very high.

The Risk Matrix 2000/Violence scale has only three items used to sort offenders into the same four risk categories. The Risk Matrix 2000/Combined scale is obtained by summing the risk category points for the Sex and Violence scales, which is then used to place offenders in the same four risk categories (low, moderate, high, and very high), representing risk for any violent recidivism.

Since its development, the Risk Matrix 2000 has permeated research and practice in the United Kingdom. It has been adopted by the police, probation, and prison services of England, Wales, Scotland, and Northern Ireland (National Policing Improvement Agency, 2007; Social Work Inspection Agency, HM Inspectorate of Constabulary for Scotland and HM Inspectorate of Prisons, 2009; see also http://www.publicprotectionni.com/index.php/risk).

A previous meta-analysis of sex offender risk scales (Hanson & Morton-Bourgon, 2009) identified 10 validation studies of the Risk Matrix 2000/Sex and found moderate predictive accuracy for sexual recidivism (mean weighted d = .67, 95% CI of .56 to .77, n = 2,755). For the prediction of violent (including sexual) recidivism, the Violence and Combined scales also demonstrated moderate predictive accuracy (d = .74 and .70, respectively), though these analyses were based on fewer studies (3 and 5, respectively). All three Risk Matrix scales showed moderate accuracy in predicting any recidivism (d's between .50 and .64). Although these results were promising, there was limited research on the performance of the Risk Matrix outside of the United Kingdom, and for specific types of samples (e.g., relatively routine, unselected samples of offenders). Additionally, Hanson

and Morton-Bourgon's (2009) meta-analysis was restricted to sexual, violent (including sexual), and any recidivism. Non-sexual violent recidivism, which is the outcome of interest for the Risk Matrix 2000/Violence scale, was not examined.

Purpose of Current Meta-Analysis

The purpose of this study was to provide an updated meta-analytic summary of the predictive accuracy of the Risk Matrix 2000. Several validation studies have been produced since Hanson and Morton-Bourgon's (2009) meta-analysis, allowing for better estimates of the predictive accuracy of the scale across diverse samples and settings. Other outcomes were included (e.g., non-sexual violence) and moderator analyses were also explored.

This meta-analysis was, however, restricted to relative predictive accuracy (i.e., discrimination). This refers to the scale's ability to discriminate between recidivists and non-recidivists, regardless of the base rate of recidivism (in other words, discrimination assesses whether higher risk offenders are more likely to reoffend than lower risk offenders). Analyses of the absolute recidivism estimates of the scale (i.e., calibration) were excluded for several reasons. There is little consensus about how to measure calibration in the forensic psychology field and it is rarely reported in risk scale validation studies (for an exception and discussion, see Helmus, Hanson, Thornton, Babchishin, & Harris, 2012). Additionally, given that the base rate of recidivism is expected to change meaningfully over time, proper aggregation would require some standardization of the follow-up period across studies, which was not available at this time.

Method

Sample

To be included, studies had to include a sample of sex offenders, ratings on the Risk Matrix 2000 (either the Sexual, Violent, or Combined scales), and a follow-up period assessing either sexual, non-sexually violent, violent, non-violent, or any recidivism. We also required sufficient information to calculate the number of recidivists, non-recidivists, and the effect size (Cohen's d). All studies required a total sample size of at least 10, with at least one recidivist. One study in which the categories of the Risk Matrix were collapsed to produce a dichotomous measure of risk was excluded (Craissati, Webb, & Keen, 2005).

Studies were identified through numerous methods. Firstly, studies included in an earlier meta-analysis (Hanson & Morton-Bourgon, 2009) were obtained. Subsequently, Google Scholar was searched for all documents citing the paper describing the development of the scale (Thornton et al., 2003). Then, computer searches of PsychINFO, Google Scholar, Digital Dissertations, and UK Dissertations were conducted with the following key terms: Risk Matrix 2000, Risk-Matrix 2000, Risk Matrix-2000, RM2000, and RM-2000. Additionally, David Thornton (the developer of the scale) and a few of his colleagues reviewed the final study list for any known omissions.

As of March 15, 2013, our search yielded 14 eligible studies. One study included three distinct samples (Thornton et al., 2003), resulting in 16 unique samples. For two samples, some of the effect sizes were obtained from different articles and subsamples. The study from Barnett, Wakeling, and Howard (2010) contained a large sample of treated offenders from the United Kingdom with effect sizes for sexual, non-sexually violent, and violent recidivism. Effect sizes for any recidivism were obtained from a subset of offenders with both hands-on and internet sex offences (Wakeling, Howard, & Barnett, 2011). Additionally, Wilcox, Beech, Markall, and Blacker

(2009) provided an effect size for the Risk Matrix 2000/Sex predicting sexual recidivism among 27 developmentally delayed sex offenders. Effect sizes for the Risk Matrix 2000/Violence were obtained from a larger sample of 88 offenders, half of whom were developmentally delayed (from Blacker, Beech, Wilcox, & Boer, 2011).

Table 1 provides basic descriptive information for each sample. The total sample size was 11,284 and ranged from 51 to 4,946 (M = 705.2, SD = 1,163.9, Mdn = 385). Studies were produced between 2003 and 2013, with a median of 2009. Two samples were coded from unpublished data (Helmus, Hanson, & Babchishin, 2013; Lehmann et al., 2013). One study was a conference presentation (Harkins, Thornton, & Beech, 2009) and the remaining studies were published articles or book chapters.

The earliest year that offenders were released was 1978 (unknown in 6 studies) and the latest year of release was 2007 (unknown in 6 studies). Ten studies used convictions as the recidivism outcome, whereas six studies used arrests/charges or reliable reports that an offence had occurred. The average follow-up period for a study was 7.8 years (SD = 4.9, Mdn = 7.4, range of 2 to 19 years).

Table 1: Characteristics of Included Studies

Study #	Primary Reference	Largest n	Country	Sample	Sample type (preselection)	Earliest year released	Latest year released	Recidivism Criteria	Average follow-up (months)
1.01	Thornton et al. (2003)	647	U.K.	Prison treatment	Treatment			Convictions	24
1.02	Thornton et al. (2003)	429	U.K.	1979 prison releases	Routine	1979	1979	Convictions	228
1.03	Thornton et al. (2003)	311	U.K.	1980 prison releases	Routine	1980	1980	Convictions	120
2	Parent et al. (2011)	503	U.S.	Civil commitment assessment center	High Risk/need			Charges	60
3	Looman & Abracen (2010)	419	Canada	High-intensity treatment	High Risk/need			Convictions	85
4	Bengtson (2008)	304	Denmark	Forensic psychiatric evaluations	High Risk/need	1978	1992	Charges	194
5.1	Barnett et al. (2010)	4,946	U.K.	Prison and community	Routine		2007	Convictions	24
5.2	Walking et al. (2011)	1,326	U.K.	Prison and community	Routine		2007	Convictions	24
6		974	Scotland		Routine	1996	2001	Convictions	60

	Grubin (2011)			Prison releases					
7	Craissati et al. (2011)	216	U.K.	Community assessment and treatment	Routine	1994	2002	Charges	110
8	Helmus et al. (2013)	710	Canada	Community supervision	Routine	2001	2005	Charges	92
9.1	Wilcox et al. (2009)[1]	27	U.K.	Community treatment	Treatment	1994	2002	Convictions	60
9.2	Blacker et al. (2011)	88	U.K.	Community treatment	Treatment	1994	2002	Convictions	60
10	Beech & Ford (2006)	51	U.K.	Community residential treatment	Treatment			Convictions	24
11	Bates et al. (2004)	183	U.K.	Community treatment	Treatment	1995	1999	Charges	47
12	Kingston et al. (2008)	351	Canada	Outpatient clinic	Other	1982	1992	Charges	137
15	Lehmann et al. (2013)	940	Germany	All sex offences reported to police	Routine	1994		Convictions	110
16	Harkins et al. (2009)	212	U.K.	Prison and community treatment	Treatment			Convictions	120

[1]Half of the sample was developmentally delayed.

Procedure

Each study was coded separately by the first two authors, who then developed a consensus rating. Separate effect sizes were coded for each scale of the Risk Matrix 2000 (Sex, Violence, and Combined) and for each recidivism outcome (sexual, non-sexual violent, any violent, non-violent, and any recidivism). Consequently, one study could contribute up to 15 effect sizes. In total, 88 effect sizes were coded (33 for the Risk Matrix 2000/Sex, 30 for the Risk Matrix 2000/Violence, and 25 for the Risk Matrix 2000/Combined). In several studies, frequency data were provided so the effect sizes could be calculated directly from means and standard deviations. Studies were mixed in terms of whether they reported the Risk Matrix 2000/Combined scale as a sum of the scores on the Sex and Violence scales, or as the final 4-category risk level on the combined scale. Where frequency data were available, we calculated effect sizes for the four risk categories.

Overview of Analyses

Index of predictive accuracy. The effect size indicator used to summarize the relative predictive

accuracy of the Risk Matrix 2000 was the standardized mean difference (Cohen's *d*; Cohen, 1988). Cohen's *d* measured the difference in Risk Matrix scores (i.e., the risk categories) between recidivists and non-recidivists, relative to how much recidivists differ from each other, and how much non-recidivists differ from each other. As a heuristic for interpretation, Cohen (1988) suggested that a *d* of .20 is small, .50 is moderate, and .80 is large. Although AUCs were the most commonly reported effect size in the included studies, *d* was chosen as the summary effect size because its variance (which is used to weight studies in meta-analysis) is more easily defined, requiring only the number of recidivists and non-recidivists. With AUCs, different software programs (and even different versions of the same software) provide different estimates of the variance. As a guideline for comparing Cohen's *d* to AUCs, Cohen's *d*'s of .20, .50, and .80 are roughly comparable to AUCs of .56, .64, and .71, respectively (Rice & Harris, 2005).

Aggregation of findings. Findings across studies were aggregated using fixed-effect and random-effects meta-analysis weighted by the inverse of the variance (Borenstein, Hedges, Higgins, & Rothstein, 2009). Whereas the results of fixed-effect meta-analysis are conceptually restricted to the particular set of studies included in the meta-analysis, random-effects meta-analysis estimates effects for the population of which the current sample of studies is a part. In other words, fixed-effect models assume that all studies are sampling the same "true" effect, whereas random-effects models assume that there is a distribution of true effect sizes from which the studies are sampling. More specifically, random-effects meta-analysis incorporates variability across samples into the error term (reflecting the assumed distribution of true effect sizes), whereas fixed-effect meta-analysis does not. When variability across studies is low (Q < degrees of freedom), random-effects and fixed-effect meta-analysis produce identical results. As the variability across studies increases, the confidence intervals for random-effects meta-analysis exceed the fixed-effect estimates, and the random-effects method gives more weight to smaller studies.

Conceptually, random-effects models are often most appropriate (Borenstein et al., 2009; Schulze, 2007). On a practical level, however, the choice between them is less clear. When there is true variability in samples, the fixed-effect method is too liberal; when there is true homogeneity, the random-effects method is too conservative (Overton, 1998). Additionally, when the number of studies is small (e.g., less than 30), estimates of variability across samples are imprecise, which can affect the reliability of random-effects computations (Schulze, 2007).

To test the variability of findings across studies, we used Cochran's Q statistic and the I^2 statistic (Borenstein et al., 2009). The Q statistic provides a significance test for variability, whereas the I^2 is a measure of effect size for variability and can, therefore, be compared across analyses. The I^2 statistic describes the proportion of the overall variability (the Q) that is beyond what you would expect by chance from sampling error (Borenstein et al., 2009). Specifically, I^2 is (Q - df)/Q. For easier interpretation, I^2 was reported as a percentage. As a rough heuristic, I^2 values of 25%, 50%, and 75% can be considered low, moderate, and high variability, respectively (Higgins, Thompson, Deeks, & Altman, 2003).

Following Hanson and Bussière (1998), a finding was considered an outlier if it was the single extreme value and accounted for more than 50% of the total variability across studies (Q), and the total variability was significant. Further to the Hanson and Bussière (1998) rule, we also required that the study's effect had to have the largest weighted squared deviation from the average effect size. When outliers were identified, results are presented both with and without the outlier. Identifying outliers is imprecise when there are few studies; as such, at least four studies were required to determine whether one was an outlier.

Meta-analyses can be strongly influenced not only by outliers, but also by unusually large sample

sizes, which are almost conclusive in themselves. The sample size of one study (Barnett et al., 2010) was 4,946, which contributed roughly half the data in this meta-analysis. There is little guidance on dealing with large studies that dominate analyses. In this meta-analysis, non-significant variability across studies was interpreted as supporting the fixed-effect assumption that all studies are sampling the same effect, and no adjustments were needed. If variability across studies was significant, then the unusually large study would have more of a biasing effect. Consequently, if the Q statistic was significant, the weight of the largest study (Barnett et al., 2010) was reduced to be 150% of the next largest study. This assessment was conducted after examination for outliers.

Moderator analyses. Moderator analyses used fixed-effect tests of the $Q_{between}$. In this analysis, the overall variability among studies (measured by the Q statistic) can be partitioned into the variability within each level of the moderator and the variability between levels of the moderator ($Q_{between}$). The $Q_{between}$ can be tested for significance using a chi-square distribution where the degrees of freedom is equal to the number of categories in that moderator minus 1. The $Q_{between}$ can also be interpreted as the amount of overall variability (from the overall Q) that is explained by that moderator.

Results

Table 2 presents the mean weighted effect sizes for the Risk Matrix 2000/Sex, /Violence, and /Combined scales for the five recidivism outcomes (sexual, non-sexual violence, any violence, non-violent, and any recidivism). Each analysis had at least three studies and all effect sizes were statistically significant in both fixed-effect and random-effects analyses. Not surprisingly, for the prediction of sexual recidivism, the Risk Matrix 2000/Sex performed the best (mean $d = .74$ in both fixed-effect and random-effects models, $k = 15$, $n = 10,644$). Predictive accuracy of the Sex scale for the other outcomes was notably lower (mean d's varied between .37 and .60), as was the number of included studies. In all analyses of the Risk Matrix 2000/Sex, the variability across studies was not statistically significant.

Table 2: Meta-analysis Results

Recidivism Type	Fixed-effect			Random-effects							
	M	95% CI		M	95% CI		Q	I^2	k	n	Studies
Risk Matrix-Sex											
Sexual	.740	.667	.812	.741	.661	.822	16.19	13.55	15	10,644	1.01, 1.02, 2, 3, 4, 5.1, 6, 7, 8, 9, 10, 11, 12, 15, 16
Non-sexual violence	.367	.292	.441	.367	.292	.441	2.41	0.00	4	7,099	2, 5.1, 8, 15

Violent	.484	.422	.545	.494	.407	.580	7.85	36.34	6	7,582	3, 5.1, 7, 8, 12, 15
Non-violent	.390	.298	.482	.410	.249	.572	5.68	64.78	3	2,153	2, 8, 15
Any recidivism	.604	.524	.685	.604	.524	.685	3.04	0.00	5	3,543	5.2, 7, 8, 12, 15
Risk Matrix-Violence											
Sexual	.257	.168	.346	.270	.135	.406	12.75*	52.94	7	7,947	2, 3, 5.1, 8, 9.2, 12, 15
Non-sexual violence	1.017	.954	1.080	.960	.782	1.139	55.15***	83.68	10	9,836	1.01, 1.02, 1.03, 2, 4, 5.1, 6, 8, 9.2, 15
Adjusted Study 5.1	.980	.911	1.049	.961	.786	1.135	48.05***	81.27	10	9,836	1.01, 1.02, 1.03, 2, 4, 5.1, 6, 8, 9.2, 15
Violent	.805	.740	.871	.766	.554	.979	32.87***	87.83	5	7,291	3, 5.1, 8, 12, 15
Outlier removed	.904	.830	.979	.891	.803	.980	3.47	13.55	4	6,351	3, 5.1, 8, 12
Non-violent	.708	.616	.801	.872	.511	1.233	36.96***	91.88	4	2,231	2, 8, 9.2, 15
Any recidivism	.730	.645	.816	.750	.562	.938	13.51**	77.80	4	3,317	5.2, 8, 12, 15
Outlier removed	.647	.547	.747	.663	.526	.800	3.40	41.24	3	2,617	5.2, 12, 15
Risk Matrix-Combined											
Sexual	.552	.460	.643	.552	.457	.647	5.39	7.30	6	7,859	2, 3, 5.1, 8, 12, 15
Non-sexual violence	.830	.755	.906	.774	.539	1.009	22.08***	86.42	4	7,089	2, 5.1, 8, 15
Outlier removed	.923	.837	1.009	.905	.787	1.023	2.69	25.72	3	6,149	2, 5.1, 8

Violent	.807	.748	.867	.796	.717	.876	10.00	30.00	8	8,277	1.01, 1.02, 3, 4, 5.1, 8, 12, 15
Non-violent	.648	.555	.742	.715	.415	1.015	18.88***	89.40	3	2,143	2, 8, 15
Any recidivism	.784	.698	.869	.789	.637	.940	8.71*	65.57	4	3,317	5.2, 8, 12, 15
Outlier removed	.718	.617	.818	.721	.603	.838	2.57	22.25	3	2,617	5.2, 12, 15

The Risk Matrix 2000/Violence was specifically designed to predict non-sexual violence, and had the largest effect size for that outcome (after adjusting the largest study weight, mean fixed-effect d = .98 and random-effects d = .96, k = 10, n = 9,836). The variability across studies, however, was significant and large. Specifically, 84% of the observed variability was more than would be expected by chance. The Violence scale also had a large effect size in predicting violent (including sexual) recidivism (after removing a statistical outlier, fixed-effect d = .90 and random-effects d = .89). The Violence scale also demonstrated good predictive accuracy for non-violent and any recidivism (d's between .65 and .87), with significant variability in the former. Predictive accuracy for sexual recidivism was significant but poor (fixed-effect d = .26 and random-effects d = .27), and the variability in effect sizes across studies was also significant.

Analyses of the Risk Matrix 2000/Combined were based on the smallest number of studies (between 3 and 8, depending on the outcome). Although the scale was designed to predict any violent recidivism (including sexual), effects sizes were highest for the prediction of non-violent recidivism (with outlier removed, fixed-effect d = .92 and random-effects d = .90), though effect sizes were also large for any violent recidivism (fixed-effect d = .81 and random-effects d = .80). The Combined scale also demonstrated at least moderate predictive accuracy for the remaining recidivism outcomes, with the lowest accuracy found for predicting sexual recidivism (d = .55 in both fixed-effect and random-effects analyses).

In the analyses of the Violence scale and the Combined scale, two outliers were identified. For violent recidivism, a large routine sample from Germany (Lehmann et al., 2013) was an outlier with the lowest effect sizes (for the Violence scale, d = .47; for the Combined scale, d = .66), although both effect sizes were roughly in the moderate range. For any recidivism, a large routine sample of Canadian offenders (Helmus et al., 2013) was an outlier with the largest effect sizes (both the Violence and Combined scales had a d of .96).

Moderator Analyses

Table 3 presents the results of moderator analyses. Moderator analyses were conducted for two sets of effect sizes: for the Risk Matrix 2000/Sex predicting sexual recidivism and for the Risk Matrix 2000/Violence predicting non-sexual violent recidivism. There were insufficient studies to provide meaningful analyses with the Risk Matrix 2000/Combined. Additionally, moderator analyses were only presented if there were at least 3 studies in each level of the moderator.

Table 3: Moderator Analyses

Recidivism Type	Fixed-effect			Random-effects			Q	I^2	k	n	Studies
	M	95% CI		M	95% CI		Q	I^2	k	n	Studies
RM-Sex and Sexual Recidivism											
Publication Bias	.740	.667	.812	.741	.661	.822	16.19	13.55	15	10,644	
Unpublished	.752	.620	.885	.788	.589	.987	3.80	47.38	3	1,862	8, 15, 16
Published	.734	.648	.820	.732	.638	.826	12.34	10.89	12	8,782	1.01, 1.02, 2, 3, 4, 5.1, 6, 7, 9.1, 10, 11, 12
$Q_{Between}$							**0.05**				
Sample Type	.744	.670	.817	.747	.662	.832	15.83	17.90	14	10,358	
Routine	.768	.678	.857	.770	.674	.865	5.61	10.80	6	8,012	1.02, 5.1, 6, 7, 8, 15
Treatment Need	.951	.699	1.202	.951	.699	1.202	3.39	0.00	5	1,120	1.01, 9.1, 10, 11, 16
High risk/need	.598	.446	.750	.598	.446	.750	0.42	0.00	3	1,226	2, 3, 4
$Q_{Between}$							**6.41***				
Country	.740	.667	.812	.741	.661	.822	16.19	13.55	15	10,644	
England and Scotland	.829	.724	.934	.829	.724	.934	7.82	0.00	9	7,482	1.01, 1.02, 5.1, 6, 7, 9.1, 10, 11, 16
Other	.658	.558	.758	.658	.558	.758	2.99	0.00	6	3,162	2, 3, 4, 8, 12, 15
$Q_{Between}$							**5.38***				
Recidivism Criteria	.740	.667	.812	.741	.661	.822	16.19	13.55	15	10,644	
Charges	.675	.552	.797	.675	.552	.797	2.86	0.00	6	2,202	

											2, 4, 7, 8, 11, 12
Convictions	.774	.685	.864	.788	.670	.907	11.68	31.49	9	8,442	1.01, 1.02, 3, 5.1, 6, 9.1, 10, 15, 16
$Q_{Between}$							**1.65**				
RM-Violent and NSV Recidivism[a]											
Country[b]	.974	.902	1.047	.951	.754	1.148	47.82***	83.27	9	9,136	
England and Scotland	1.151	1.060	1.242	1.140	1.024	1.256	6.82	26.74	6	7,389	1.01, 1.02, 1.03, 5.1, 6, 9.2
Other	.662	.542	.783	.662	.542	.783	0.86	0.00	3	1,747	2, 4, 15
$Q_{Between}$							**40.14***				
Recidivism Criteria	.980	.911	1.049	.961	.786	1.135	48.05***	81.27	10	9,836	
Charges	.849	.706	.991	.833	.604	1.063	5.05	60.40	3	1,507	2, 4, 8
Convictions	1.020	.941	1.099	1.022	.794	1.251	38.75***	84.52	7	8,329	1.01, 1.02, 1.03, 5.1, 6, 9.2, 15
$Q_{Between}$							**4.25***				

Note. A minimum of 3 studies in each level of the moderator was required to present analyses. NSV = non-sexual violence;
[a]Moderator analyses of the Risk Matrix 2000/Violence used the adjusted weight for study 5.
[b]Study 8 was removed as an outlier.

There was no significant difference in the effect sizes from published and unpublished studies (for the Risk Matrix Sex, $Q_{between}$ = 0.05, df = 1, p = .823). There was a significant moderator effect for the type of sample (for the Risk Matrix Sex, $Q_{between}$ = 6.43, df = 2, p = .040). For the three samples that had been preselected as unusually high risk or high need[1], the effect sizes were the lowest (d = .60 in both fixed-effect and random-effects analyses) and consistent. In routine samples (i.e., relatively unselected), the effect size was similar to the average of all studies (d .74). Effect sizes were highest in treatment samples (d = .951 in both fixed-effect and random-effects analyses).

Analyses of both the Sex and Violence scales found significantly higher effect sizes in studies from the United Kingdom compared to other countries. For the Risk Matrix 2000/Violence, analyses are presented without an outlier study from Canada (Helmus et al., 2013). If this higher effect size was

included, the moderator would still be significant ($Q_{between}$ = 31.97, *df* = 1, *p* < .001).

For the Risk Matrix 2000/Sex, effect sizes did not differ based on whether the study used charges or convictions as the recidivism criteria ($Q_{between}$ = 1.65, *df* = 1, *p* = .199). For the Risk Matrix 2000/Violence, however, this moderator was significant ($Q_{between}$ = 4.25, *df* = 1, *p* = .039), with higher effect sizes found among studies using convictions as the recidivism criteria. These results are difficult to interpret, however, because they are based on fewer studies (only 3 studies used charges as the recidivism outcome, compared to 6 studies in the analyses of the Sex scale) and are confounded with both the Country and Sample Type moderators. Specifically, two of the three studies using charges were from outside the U.K. and were preselected as high risk/need, and both these factors were associated with lower overall effect sizes. The third study using charges was a routine sample but was also outside the U.K.

Discussion

This study found that the Risk Matrix 2000 (including the Sex, Violence, and Combined scales) significantly predicted sexual, non-sexual violent, any violent, non-violent, and any recidivism. In general, the effect sizes in the current meta-analysis were comparable or higher than those found by Hanson and Morton-Bourgon (2009) for the Risk Matrix 2000 and for the average of other actuarial scales designed for a similar purpose (i.e., designed to predict sexual or violent recidivism). As predicted, the Sex scale provided the best predictive accuracy for sexual recidivism (approaching a large effect size, *d* = .74). The Violence scale and the Combined scale both predicted non-sexual violent recidivism and any violent recidivism with similarly large effect sizes. Although the scales were not designed to predict non-violent or any (including violent) recidivism, the effect sizes for these outcomes were also moderate to large (*d*'s exceeding .60).

The current findings are consistent with a large body of previous research indicating that the predictors of general criminality (e.g., the Central 8; Andrews & Bonta, 2010) are applicable to a variety of other outcomes, including violent recidivism (Campbell, French, & Gendreau, 2009) and even institutional misbehaviour (Gendreau, Goggin, & Law, 1997). In developmental criminology, the onset of non-sexual interpersonal crimes is often observed in individuals following an entrenched trajectory of defiant and antisocial behaviour (Loeber, 1990). In contrast, there are some relatively distinct features associated with sexual criminality. Where items related to deviant sexual interests (e.g., unrelated, boy victims) are risk factors for sexual crime, sexual deviancy has little or no relationship to non-sexual criminality (Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005).

The effect sizes in samples that were preselected to be high risk or need were significantly smaller than those observed in other samples. Previous research has found that sample of offenders screened for unusual measures (e.g., high intensity treatment, civil commitment) generally have higher levels of both risk and need, even after controlling for static actuarial scales (Hanson & Thornton, 2012). Although other risk scales has not found this effect (i.e., Static-99R, Static-2002R; Helmus, Hanson et al., 2012), the same sample-type effect (reduced accuracy for high risk/need samples) has been found for several of the individual items in Static-99R/2002R (Helmus & Thornton, 2012), many of which are similar to the Risk Matrix items. The reduced accuracy likely reflects restriction of range in the underlying constructs associated with recidivism. In other words, the items of the risk scales are intended as indicators of the broader constructs related to recidivism; samples using more detailed assessments to pre-select offenders based on other (and potentially better) indicators of these constructs would reduce the range on the underlying constructs, thereby reducing predictive accuracy.

Moderator analyses also found that effect sizes were significantly lower outside the U.K. However, the effect sizes were still significant and moderate in all countries. One possible explanation for the large effects in the U.K. is that the scale was developed there. This explanation is not sufficient, however, given that Hanson and Morton-Bourgon (2009) found that all risk scales predict better in the U.K. than in most other countries. Another possible explanation is the quality of criminal record-keeping. There is some evidence that the criminal records maintained in the United Kingdom are more reliable than records in other countries (see Helmus, Hanson, & Morton-Bourgon, 2011). Completeness of criminal records would improve both the quality of information on which the assessments are based as well as the reliability of the outcome measures (i.e., recidivism).

A limitation of this meta-analysis is that the number of available studies was lower than desired, particularly for some analyses. Stable estimates using random-effects models require at least 30 studies (Schulze, 2007); in contrast, the largest analysis in the current study contained 15 studies. For the Violence and Combined scales, and for all moderator analyses, the number of studies was even smaller. The modest number of studies reduces power for moderator analyses and also makes it difficult to reliably detect variability and outliers. Of the two studies identified as outliers (Helmus et al., 2013; Lehmann et al., 2013), both were large studies with credible effect sizes and were notably different from the small number of other studies available. It is possible that with the accumulation of additional studies, they will no longer appear extreme.

Another limitation is that we were unable to evaluate the quality of the risk assessment ratings in the validation studies. Predictive accuracy should be enhanced when scales are scored correctly by conscientious and properly trained evaluators using complete data and who have access to ongoing training and support (Bonta, Bogue, Crowley, & Motiuk, 2001; Fernandez, Harris, Hanson, & Sparks, 2012; Flores, Lowenkamp, Holsinger, & Latessa, 2006). One large study of the routine field use of static actuarial scales for sex offenders has demonstrated substantially lower predictive accuracy than what is found in most studies where the scale was scored by researchers (Boccaccini, Murrie, Caperton, & Hawes, 2009), highlighting the potential for widespread implementation of a scale to go awry. Analyses from the Dynamic Supervision Project (one of the samples included in this meta-analysis) have examined the role of conscientiousness, which was liberally defined as community supervision officers who submitted all the data that were requested of them. Predictive accuracy for several risk scales was substantially higher for "conscientious officers" (Hanson, Harris, Scott, & Helmus, 2007), and this effect was also found with the Risk Matrix (for example, the AUC for the Risk Matrix 2000/Sex was .70 in the overall sample, but .76 for the conscientious officers; Helmus et al., 2013)[2]. Given the widespread routine use of the Risk Matrix in the United Kingdom, these findings highlight the critical importance of quality control in risk assessment practices.

Implications for Research

An important gap in the current research base pertains to the validation of the absolute recidivism estimates of the scale (i.e., calibration). Although analyses of calibration are relatively new in forensic assessment research and there is no clear consensus on the best statistics for these analyses, there are examples of studies using statistics such as the chi-square, E/O index, and the Hosmer-Lemeshow test for this purpose (Babchishin, Hanson, & Helmus, 2012; Duwe & Freske, 2012; Harris et al., 2003; Helmus, Thornton, Hanson, & Babchishin, 2012; Lehmann, Hanson, & Babchishin, 2013; Montana et al., 2012). Meta-analyses of calibration, however, require that researchers report similar outcome criteria for standardized follow-up periods. The standardized follow-up period should be sufficiently long that the base rate is high enough for reliable recidivism estimates, and the outcome criteria should be clearly defined and easily accessible in diverse settings. As well, it is desirable that researchers report outcomes and follow-up periods that match

those used in the user guidance and training materials.

For the Risk Matrix 2000, the training materials report recidivism outcomes for convictions at 2, 4, 5, 10, and 15 years (Thornton, 2010). This is more time points than most decision-makers would need. Given the differences in base rates, it may be desirable for researchers to focus on standardized follow-up periods of 5 and 10 years for sexual recidivism, and 2 and 5 years for non-sexual violence and any violent recidivism. For any recidivism, it would be possible to also include 1 year recidivism rates.

Our current recommendation for reporting calibration is to use the E/O index (Rockhill, Byrne, Rosner, Louie, & Colditz, 2003; Viallon, Ragusa, Clavel-Chapelon, & Bénichou, 2009). It is preferable to other statistics (such as the chi-square or Hosmer-Lemeshow test) because it is a measure of effect size, is relatively easy to interpret, and its variance can be easily computed. If researchers consistently reported this statistic, it would be possible to meta-analyze the effect sizes even if the follow-up period differed across studies (within each study, however, calculations of the E/O index would need to match one of the follow-up periods provided for absolute recidivism estimates).Calculations of the E/O index are simplest when there is a fixed follow-up period for the complete sample; alternative estimation procedures for survival data, however, are available (Viallon et al., 2009).

Implications for Practice

The Risk Matrix 2000 has moderate to large predictive accuracy among sex offenders and can be used with confidence in applied risk assessment practice. Of the Risk Matrix 2000 scales, the Risk Matrix 2000/Sex is the preferred scale for assessing the risk for sexual recidivism. The Violence and Combined scales are intended for predicting non-sexually violent and any violent recidivism, respectively, and work as intended for these outcomes. Our analyses, however, found that both the Violence and Combined scales are similarly effective for both types of violent recidivism, and also demonstrate good predictive accuracy for non-violent and any recidivism. The relative predictive accuracy of the Risk Matrix is at least as good as other risk assessment scales available for sex offenders. Increased confidence can be placed in the scale when it is used as intended by the developers and in Western developed countries. Lower predictive accuracy is expected in samples that have been heavily preselected on risk-relevant features. This meta-analysis was restricted to an examination of relative predictive accuracy (i.e., does the scale discriminate between recidivists and non-recidivists?). Currently, little is known about the stability and generalizability of the absolute recidivism estimates of the Risk Matrix 2000.

Notes

[1] For further explanation of this variable, see Helmus (2009) and Static-99R resource materials available from www.static99.org.
[2] The complete sample (including both conscientious and non-conscientious officers) was used in the current meta-analysis.

Author Note

References

*Studies with an asterisk were included in the meta-analysis

1. Andrews, D. A., & Bonta, J. (2010). The psychology of criminal conduct (5th ed.). Newark, NJ: LexisNexus/Anderson.
2. Babchishin, K. M., Hanson, R. K., & Helmus, L. (2012). Even highly correlated measure can add incrementally to predicting recidivism among sex offenders. Assessment, 19, 442-461. doi:10.1177/1073191112458312
3. *Barnett, G. D., Wakeling, H. C., & Howard, P. D. (2010). An examination of the predictive validity of the Risk Matrix 2000 in England and Wales. Sexual Abuse: A Journal of Research and Treatment, 22, 443-470. doi:10.1177/1079063210384274
4. *Bates, A., Falshaw, L., Corbett, C., Patel, V., & Friendship, C. (2004). A follow-up study of sex offenders treated by Thames Valley Sex Offender Groupwork Programme, 1995-1999. Journal of Sexual Aggression, 10, 29-38. doi:10.108013552600410001667724
5. *Beech, A., & Ford, H. (2006). The relationship between risk, deviance, treatment outcome and sexual reconviction in a sample of child sexual abusers completing residential treatment for their offending. Psychology, Crime & Law, 12, 685-701. doi:10.1080/10683160600558493
6. *Bengtson, S. (2008). Is newer better? A cross-validation of the Static-2002 and the Risk Matrix 2000 in a Danish sample of sexual offenders. Psychology, Crime & Law, 14, 85-106. doi:10.1080/10683160701483104
7. *Blacker, J., Beech, A. R., Wilcox, D. T., & Boer, D. P. (2011). The assessment of dynamic risk and recidivism in a sample of special needs sexual offenders. Psychology, Crime & Law, 17, 75-92. doi:10.1080/10683160903392376
8. Boccaccini, M. T., Murrie, D. C., Caperton, J. D., Hawes, S. W. (2009). Field validity of the Static-99 and MnSOST-R among sex offenders evaluated for civil commitment as sexually violent predators. Psychology, Public Policy, and Law, 15, 278-314. doi.10.1037/a0017232
9. Bonta, J., Bogue, B., Crowley, M., & Motiuk, L. (2001). Implementing offender classification systems: Lessons learned. In G. A. Bernfeld, D. P. Farrington, & A. W. Leschied (Eds.), Offender rehabilitation in practice: Implementing and evaluating effective programs (pp. 227-246). Chichester, England: John Wiley & Sons.
10. Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Introduction to meta-analysis. Chichester, West Sussex, U.K.: Wiley.
11. Campbell, M., French, S., & Gendreau, P. (2009). The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. Criminal Justice and Behavior, 36, 567-590. doi:10.1177/0093854809333610
12. Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd.). Hillsdale, NJ: Lawrence Erlbaum.
13. *Craissati, J., Bierer, K., & South, R. (2011). Risk, reconviction and 'sexually risky behaviour' in sex offenders. Journal of Sexual Aggression, 17, 153-165. doi.10.1080/13552600.2010.490306
14. Craissati, J., Webb, L., & Keen, S. (2005). Personality disordered sex offenders. Unpublished manuscript.
15. Duwe, G., & Freske, P. J. (2012). Using logistic regression modeling to predict sexual recidivism: The Minnesota Sex Offender Screening Tool-3 (MnSOST-3). Sexual Abuse: A Journal of Research and Treatment, 24, 350-377. doi:10.1177/1079063211429470
16. Fernandez, Y., Harris, A. J. R., Hanson, R. K., & Sparks, J. (2012). STABLE-2007 coding manual - revised 2012. Unpublished report. Ottawa, ON: Public Safety Canada.
17. Flores, A. W., Lowenkamp, C. T., Holsinger, A. M., & Latessa, E. J. (2006). Predicting outcome with the Level of Service Inventory-Revised: The importance of implementation

integrity. Journal of Criminal Justice, 34, 523-529. doi:10.1016/j.jcrimjus.2006.09.007

18. Gendreau, P., Goggin, C. E., Law, M. A. (1997). Predicting prison misconducts. Criminal Justice and Behaviour, 24, 414-431. doi:10.1177/0093854897024004002

19. *Grubin, D. (2011). A large-scale evaluation of Risk Matrix 2000 in Scotland. Sexual Abuse: A Journal of Research and Treatment, 23, 419-433. doi:10.1177/1079063211411309

20. 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.66.2.348

21. 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, ON: Public Safety Canada. Available from http://www.publicsafety.gc.ca/res/cor/rep/_fl/crp2007-05-en.pdf

22. 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

23. 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

24. 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

25. Hanson, R. K., & Thornton, D. (2012, October). Preselection effects can explain group differences in sexual recidivism base rates in Static-99R validation studies. Paper presented at the 31st Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Denver, CO.

26. *Harkins, L., Thornton, D., & Beech, A. (2009, September). The use of dynamic risk domains assessed using psychometric measures to revise relative risk assessment using RM 2000 and Static 2002. Paper presented at the 28th Annual Conference of the Association for the Treatment of Sexual Abusers, Dallas, TX.

27. Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumière, M. L., Boer, D., & Lang, C. (2003). A multi-site comparison of actuarial risk instruments for sex offenders. Psychological Assessment, 15, 413-425. doi:10.1037/1040-3590.15.3.413

28. Helmus, L. (2009). Re-norming Static-99 recidivism estimates: Exploring base rate variability across sex offender samples (Master's thesis). Available from ProQuest Dissertations and Theses database. (UMI No. MR58443)

29. *Helmus, L., Hanson, R. K., & Babchishin, K. M. (2013). The Risk Matrix 2000: Validation and Combination with STABLE-2007. Unpublished manuscript.

30. Helmus, L., Hanson, R. K., & Morton-Bourgon, K. E. (2011). International comparisons of the validity of actuarial risk tools for sexual offenders, with a focus on Static-99. In D. P. Boer, L. A. Craig, R. Eher, M. H. Miner, & F. Pfafflin (Eds.), International perspectives on the assessment and treatment of sexual offenders: Theory, practice, and research (pp. 57-83). Wiley.

31. 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 & Behavior, 39, 1148-1171. doi:10.1177/0093854812443648

32. 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

33. Helmus, L., & Thornton, D. (2012, October). Performance of individual items of Static-99R and Static-2002R. Paper presented at the 31st Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Denver, CO.

34. Higgins, J., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. British Medical Journal, 327, 557-560. doi:10.1136/bmj.327.7414.557

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

36. *Knight, R. A., & Thornton, D. (2007). Evaluating and improving risk assessment schemes for sexual recidivism: A long-term follow-up of convicted sexual offenders (Document No. 217618). Submitted to the U.S. Department of Justice.

37. *Lehmann, R. J. B., Gallasch-Nemitz, F., Biedermann, J., & Dahle, K.-P. (2013). [Local validation of sexual offender risk assessment instruments for 940 sexual offenders from Berlin, Germany] Unpublished raw data.

38. Lehmann, R. J. B., Hanson, R. K., Babchishin, K. M., Gallasch-Nemitz, F., Biedermann, J., Dahle, K.-P. (2013). Interpreting multiple scales for sex offenders: Evidence for averaging. Psychological Assessment, 25, 1019-1024. doi:10.1037/a0033098.

39. Loeber, R. (1990). Development and risk factors of juvenile antisocial behavior and delinquency. Clinical Psychology Review, 10, 1-41. doi:10.1016/0272-7358(90)90105-J

40. *Looman, J., & Abracen, J. (2010). Comparison of measures of risk for recidivism in sexual offenders. Journal of Interpersonal Violence, 25, 791-807. doi:10.1177/0886260509336961

41. 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

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

43. Montana, S., Thompson, G., Ellsworth, P. J., Lagan, H., Helmus, L., & Rhoades, C. J. (2012). Predicting relapse for Catholic clergy sex offenders: The use of Static-99. Sexual Abuse: A Journal of Research and Treatment, 24, 575-590. doi:10.1177/1079063212445570

44. National Policing Improvement Agency. (2007). Guidance on protecting the public: Managing sexual offenders and violent offenders. Report.

45. Overton, R. C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. Psychological Methods, 3, 354-379. doi:10.1037//1082-989X.3.3.354

46. *Parent, G., Guay, J.-P., & Knight, R. A. (2011). An assessment of long-term risk of recidivism by adult sex offenders: One size doesn't fit all. Criminal Justice and Behavior, 38, 188-209. doi:10.1177/0093854810388238

47. 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. doi:10.1007/s10979-005-6832-7

48. Rockhill, B., Byrne, C., Rosner, B., Louie, M. M., & Colditz, G. (2003). Breast cancer risk prediction with a log-incidence model: Evaluation of accuracy. Journal of Clinical Epidemiology, 56, 856-861. doi:10.1016/S0895-4356(03)00124-0

49. Schulze, R. (2007). Current methods for meta-analysis: Approaches, issues, and developments. Zeitschrift für Psychologie / Journal of Psychology, 215, 90-103. doi:10.1027/0044-3409.215.2.90

50. Social Work Inspection Agency, HM Inspectorate of Constabulary for Scotland, and HM Inspectorate of Prisons. (2009). Multi-agency inspection: Assessing and managing offenders who present a high risk of serious harm 2009. Retrieved from http://www.scotland.gov.uk/Resource/Doc/275852/0082871.pdf

51. Thornton, D. (2010). Scoring guide for Risk Matrix 2000.10/SVC. Unpublished document.

52. *Thornton, D., Mann, R., Webster, S., Blud, L., Travers, R., Friendship, C., & Erikson, M. (2003). Distinguishing and combining risks for sexual and violent recidivism. In R. A. Prentky, E. S. Janus, & M. C. Seto (Eds.), Annals of the New York Academy of Sciences: Vol. 989. Sexually coercive behavior: Understanding and management (pp. 225-235). New York: New York Academy of Sciences.
53. Villian, V., Ragusa, S., Clavel-Chapelon, F., & Bénichou, J. (2009). How to evaluate the calibration of a disease risk prediction tool. Statistics in Medicine, 28, 901-916. doi:10.1002/sim.3517
54. *Wakeling, H. C., Howard, P., & Barnett, G. (2011). Comparing the validity of the RM2000 Scales and OGRS3 for predicting recidivism by internet sexual offenders. Sexual Abuse: A Journal of Research and Treatment, 23, 146-168. doi:10.1177/1079063210375974
55. *Wilcox, D., Beech, A., Markall, H. F., & Blacker, J. (2009). Actuarial risk assessment and recidivism in a sample of UK intellectually disabled sexual offenders. Journal of Sexual Aggression, 15, 97-106. doi:10.1080/13552600802578577

Appendix: Risk Matrix Coding Forms

Important Note: Evaluators and researchers interested in using the scale should receive training and consult the full coding manual for scoring guidelines and instructions (Thornton, 2010).This Appendix was adapted from Thornton (2010), with the author's permission.

Instructions for Scoring Risk Matrix 2000/Sex

This scale involves two steps. In Step 1, the following items are scored.

Age	18-24 = 2 points; 25-34 = 1 point; Older = 0 points
Sexual Appearances	1 = 0 points; 2 = 1 point; 3,4 = 2 points; 5+ = 3 points
Criminal Appearances	4 or Less = 0 points; 5 or more = 1 point

Points accumulated across these three items are then turned into the following four risk categories:

Points	Label
0	Low
1-2	Medium
3-4	High
5-6	Very High

In Step 2, the following items are scored:

Aggravating Factors	Scoring
Male Victim of Sex Offence	No = 0; Yes = 1
Stranger Victim of Sex Offence	No = 0; Yes = 1
Single (absence of 2 year co-habitation)	No = 0; Yes = 1
Non-Contact Sex Offence	No = 0; Yes = 1

Offenders are increased by one category if Step Two score = 2 or 3. They are increased two categories if score = 4. Do not change Step One Risk Category if score = 0 or 1. Offenders already in the highest risk category cannot be increased further.

Instructions for Scoring Risk Matrix 2000/Violence

This scale involves the following items:

Risk Factor	Points Assigned
Age	18 to 24 = 3 points; 25 to 34 = 2 points; 35 to 44 = 1 point
Violent Appearances	0 = 0 points; 1 = 1 point; 2,3 = 2 points; 4+ = 3 points
Burglary	None = 0 points; Any = 2 points

Based on their scores, offenders are placed in one of the following four risk categories:

Points	Label
0-1	Low
2-3	Medium
4-5	High
6+	Very High

Instructions for Scoring Risk Matrix 2000/Combined

Offenders are assigned points based on their risk categories for the Sex and Violence scales:

S or V Scale Risk Category	Low	Medium	High	Very High
C Points assigned for S scale	0	1	2	3
C points assigned for V scale	0	1	2	3

Summed points on the Sex and Violence scales are used to assign categories on the Combined scale:

Score on C Scale	Label
0	Low
1	Medium
2	Medium
3	High
4	High
5	Very High
6	Very High

Author address

Leslie Helmus
Loeb Building
Carleton University
1125 Colonel By Drive
Ottawa, ON, K1S 5B6
leslie_helmus@carleton.ca