

THE CPORT AND RISK MATRIX 2000 FOR MEN CONVICTED OF CHILD SEXUAL EXPLOITATION MATERIAL (CSEM) OFFENSES

A Predictive Accuracy Comparison and Meta-Analysis

L. MAAIKE HELMUS 

Simon Fraser University

ANGELA W. EKE 

Ontario Provincial Police

LINDA FARMUS

York University

MICHAEL C. SETO

Royal Ottawa Health Care Group

There is demand for valid risk assessment of individuals with child sexual exploitation material (CSEM) offenses. We compared the predictive performance of the Risk Matrix 2000/Sex (RM2000/S) and the Child Pornography Offender Risk Tool (CPORT) among 365 men convicted of CSEM offenses. In fixed 5-year follow-up analyses, the CPORT (area under the curve [AUC] = .73) had significantly higher predictive accuracy than the RM2000/S (AUC = .66) for any sexual recidivism. The predictive difference for CSEM recidivism was not statistically significant. A meta-analysis found the CPORT had large effects in predicting sexual recidivism (AUC = .75) and moderate accuracy for CSEM recidivism (AUCs = .65 and .66),

AUTHORS' NOTE: *We would like to thank the police officers who assisted with the original project and who investigated these offenses and we appreciate the research support and assistance of members of the Child Sexual Exploitation Unit and Criminal Behaviour Analysis Section of the Ontario Provincial Police. The current study sample contains protected information maintained by the Ontario Provincial Police and cannot be shared outside the service; however, requests for additional data analyses or verification can be submitted for review. The meta-analysis data sets (and syntax template) are available from Open Science Framework (https://osf.io/pn9d6/?view_only=da2fa684366645d9b6db5a14c72464c5). Syntax for analyses from the current study is not available (most analyses were conducted using point-and-click). This work was completed on the traditional and unceded territories of the Coast Salish Peoples (Simon Fraser University), specifically the Squamish (Sḵw̓xwú7mesh Úxwumixw), Tsleil-Waututh (səlilwətaʔ) and Musqueam (xʷməθkʷəy̓əm) Nations, also on the traditional territory of the Chippewas of Rama First Nation (OPP General Headquarters) and the traditional and unceded territory of Algonquin Anishnaabeg People (Royal Ottawa Health Care Group). Angela Eke, Michael Seto, and L. Maaïke Helmus are co-authors of the Child Pornography Offender Risk Tool (CPORT) coding manual. They do not receive royalties for use of the CPORT. Correspondence concerning this article should be addressed to L. Maaïke Helmus, Department of Criminology, Simon Fraser University, 10322 Saywell Hall, 8888 University Drive, Burnaby, BC V5A 1S6, Canada; e-mail: maaïke_helmus@sfu.ca.*

CRIMINAL JUSTICE AND BEHAVIOR, 202X, Vol. XX, No. X, Month 2023, 1–21.

DOI: 10.1177/00938548231208194

Article reuse guidelines: sagepub.com/journals-permissions



© 2023 International Association for Correctional and Forensic Psychology

while the RM2000/S had moderate accuracy in predicting any sexual recidivism (AUC = .66; insufficient studies of CSEM recidivism). Results suggest a tool developed specifically for CSEM offending, such as CPORT, may perform better at predicting any sexual recidivism than adapting a general sexual offending risk tool.

Keywords: child sexual exploitation materials; risk assessment; recidivism; sexual offenses; prediction

Risk assessment is an essential task in the criminal justice system. Interventions to manage or reduce risk to reoffend are most effective when they are proportionate and tailored to an individual's risk to reoffend, following the risk and need principles of the risk/need/responsivity model of effective correctional practices (Bonta & Andrews, 2017). Lower intensity interventions are not sufficient for high risk individuals and higher intensity interventions are inefficient and can have negative effects on low risk individuals (Andrews & Dowden, 2006; Lovins et al., 2009; Lowenkamp & Latessa, 2005). Consequently, while it is important to prioritize high risk individuals, we must ensure low risk people are not overmanaged.

The development, validation, and implementation of risk assessment tools to predict sexual offense recidivism have proliferated in the last 30 years (Hanson & Morton-Bourgon, 2009; Kelley et al., 2020). However, given the recency of widespread internet technologies and the length of time required to conduct recidivism studies with adequate follow-up, there is comparably less research and fewer risk assessment tools available for individuals convicted of internet sex offenses, particularly related to the possession and distribution of child sexual exploitation materials (CSEM; also sometimes referred to as child pornography or indecent images involving children). This gap in research and knowledge requires attention because CSEM offending in particular has increased rapidly (Allen, 2016; Statistics Canada, 2021).

RISK ASSESSMENT FOR MEN WITH CSEM OFFENSES

Virtually all empirically based risk assessment tools for sexual recidivism were developed with samples of men who had few to no CSEM offenses in their criminal history because the sampling timeframes predated the widespread adoption of the internet, which is itself linked to the rapid increase in CSEM-related offending (Martin, 2021; Seto, 2013). For example, the original Static-99 and Risk Matrix were developed using individuals released between 1959 and the early 1990s (Hanson & Thornton, 2000; Thornton et al., 2003). However, lack of inclusion in the development research does not necessarily require separate risk tools for this population. The 2014 Standards for Educational and Psychological Testing from the Joint Committee of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education (hereafter referred to as the Joint Committee) provide guidance for applying existing assessment tools to new populations. Although not every possible subgroup or unique case requires separate validation research, the Joint Committee (2014) emphasized that test development and validation must consider relevant subgroups (see Chapter 3).

The question then becomes whether individuals with CSEM offenses are a relevant subgroup requiring separate validation research and potentially unique risk tools. Early research found these individuals are often at lower risk to sexually reoffend than those who

committed contact sexual offenses (Seto, 2013). There is variation, however, as those with both CSEM and offline sexual offenses have higher rates of reoffense compared with individuals with only CSEM offenses (Babchishin et al., 2015, 2022; Seto et al., 2011). In addition, individuals with CSEM offenses also differ in demographic and psychological characteristics compared with those with contact sexual offenses, for example, being more likely to show evidence of atypical sexual interests such as pedophilia (for reviews, see Babchishin et al., 2015, 2018; Henshaw et al., 2020).

To our knowledge, only one risk assessment tool has been developed and validated to predict reoffense among individuals with CSEM offenses: the Child Pornography Offender Risk Tool (Seto & Eke, 2015). However, generic risk assessment tools have been examined for this offense group as well, including the Risk Matrix 2000 (Thornton et al., 2003, 2023). The Risk Matrix 2000 is routinely used in the United Kingdom and has been examined in several risk assessment studies following individuals who have committed CSEM offenses. Research on both tools will be discussed in turn. Overall, there is insufficient research on the applicability of generic sexual recidivism risk tools for those whose only sexual offenses are CSEM offenses. Moreover, for those with mixed online and offline sexual offenses, it is unknown whether CSEM-specific or generic sexual recidivism risk tools are more accurate.

CHILD PORNOGRAPHY OFFENDER RISK TOOL

The Child Pornography Offender Risk Tool (CPORT) is a risk assessment tool with seven items: age at the time of the index investigation (35 or younger), any prior criminal history, any failure on conditional release, any contact sexual offending, indication (admission or diagnosis) of sexual interest in children, more boy than girl content in child pornography, and more boy than girl content in other child-related materials (e.g., images of nude or partially clothed children). It can be scored primarily using criminal history and police investigative data (Eke et al., 2018). In the Canadian development study (Seto & Eke, 2015), the CPORT had a large effect size in predicting sexual recidivism for the overall sample (area under the curve [AUC] = .74, $n = 266$). In a follow-up study, Eke et al. (2019) collected data on 80 new cases. The new validation sample was too small to provide robust analyses on its own, but findings from the development and validation samples were comparable with regard to AUCs. Combining the development and validation samples ($N = 346$), the CPORT still demonstrated large effects in predicting sexual recidivism (AUC = .72 for all cases and .75 when restricted to cases with no missing items). Effect sizes were also meaningfully higher for cases with a contact sex offense compared to those without (AUC = .72 vs .68; these analyses allowed one missing item and use of the Correlates of Admission of Sexual Interest in Children [CASIC; Seto & Eke, 2017] scale to replace missing information for the pedophilic/hebephilic sexual interest item). Notably, separating cases by contact sex offense removes the CPORT contact sex offense item from the tool, as each group has no variability on this item, either being all yes or all no, reducing it to six items. Since the development and replication work on CPORT, there have been at least five external validation studies that we are aware of, with the three most recent scoring CPORT as recommended (Eke et al., 2018). Predictive validity results of these validation studies are presented and meta-analyzed in Study 2 of this article.

RISK MATRIX 2000

The Risk Matrix 2000 (RM2000) is an actuarial static risk scale for adult men convicted of sexual offenses (Thornton et al., 2003, 2023). The RM2000 consists of three scales: Sex (designed to predict sexual recidivism), Violent (designed to predict nonsexual violent recidivism), and Combined (designed to predict any violent recidivism, including sexual). This study will consider only the Sex scale (RM2000/S), which has seven items across two steps of coding, resulting in placement in one of four risk levels. Unlike Static-99R, which is not intended to be used for individuals whose sole sexual offense history relates to possessing or distributing CSEM materials (Phenix et al., 2016), RM2000 does not have this exclusion and the coding manual includes guidance for scoring CSEM cases (Thornton et al., 2023).

Three studies have examined predictive accuracy of the RM2000/S for internet CSEM offenses in the United Kingdom (Barnett et al., 2010; Elliott et al., 2019; Wakeling et al., 2011). However, there is considerable overlap among these samples, with Wakeling et al. (2011) subsuming most (if not all) of the other two samples. Similar to the CPORT (Seto & Eke, 2015), although the RM2000/S predicted reasonably well for the full sample with internet offenses, accuracy was notably reduced when divided into subgroups based on noninternet sexual offending history. This demonstrates the effects of reduced heterogeneity on accuracy (Howard, 2017).

In addition, RM2000/S only distinguishes between four risk levels: below average, average, above average, and well above average. The group with internet sex offenses solely had virtually no cases at the two highest risk levels, and the group with contact and noncontact sexual offenses had virtually no cases in the low risk level. This supports Babchishin et al.'s (2015) meta-analytic findings that individuals with mixed online/offline sex offenses are meaningfully higher risk than individuals with online sex offenses only.

PURPOSE OF CURRENT STUDY

The CPORT is the only risk tool specifically developed to predict sexual recidivism among men convicted for CSEM offenses. It is new and purpose-built; however, research on the CPORT is ongoing and stable recidivism norms have not yet been established. In addition, some of the items needed to score the CPORT (e.g., related to CSEM and other child content; see "Measures" section below) may not be routinely available. This has created challenges for missing data in validation studies, although it is possible that over time the increased use of the tool may motivate police and corrections staff to report this information more regularly. The RM2000 is a preexisting tool that can and is being applied to these individuals. It already comes with a history of research and normative data (Helmus et al., 2013; Lehmann et al., 2016). However, it is necessary to separately validate this tool for CSEM offenses (see Joint Committee, 2014).

There are seven predictive accuracy studies of the CPORT (including the development study) and only one (albeit large) validation of the RM2000/S in a CSEM sample. The effect size for the RM2000/S is higher than some studies on the CPORT and lower than others. However, it is hard to directly compare the accuracy of the two risk assessment tools because they were from different studies with potentially important differences between the samples or in the methodology. For example, validation studies from Europe tend to have significantly higher predictive accuracy compared with studies from North America

(Hanson & Morton-Bourgon, 2009; Helmus et al., 2022). This may be due to higher quality criminal records in Europe (Helmus et al., 2011). In addition, many CPORT studies had substantial missing information.

This study provides the first direct comparison of the predictive accuracy of the CPORT and the RM2000/S among individuals with convictions for CSEM offenses. We hypothesized that the CPORT would have higher accuracy because it is a specialized tool. We also examined two analytic approaches that have been recently used to compare recidivism prediction scales or models (Bayesian information criterion [BIC] and the DeLong test; described in the “Methods” section). We also explored whether the recidivism norms for the RM2000/S (Lehmann et al., 2016), which were not developed on CSEM samples, are applicable to this sample of individuals with CSEM offenses; no hypotheses were made. Finally, we provided a cumulative meta-analysis of the CPORT and RM2000/S validation research to date with CSEM populations.

METHOD

SAMPLE

This study used 365 cases from the combined CPORT development and validation samples from Seto and Eke (2015) and Eke et al. (2019). Data were from 10 police services in the most populous province in Canada (Ontario) and included regional, municipal, and provincial police service data. The sample consisted of adult (18 years of age or older) males convicted of one or more CSEM¹ offense(s) (i.e., possession, accessing, distributing, or making/production). To be included in the current sample, the case had to have sufficient information available to score it on both the CPORT (following the missing data rules in the coding manual) and the RM2000/S. Conviction dates ranged between 1993 and 2010, with the vast majority occurring from 2000 onward. Most of the sample (99%) had at least one index charge for possession of child pornography, over a third (37%) had distribution charges, and fewer had production charges (21%) or accessing (21%) charges. Although many of the charges for production involved direct victimization of a child (e.g., taking images during contact sexual offenses), production charges can also be laid for transferring material from one electronic storage device to another. The average age at index investigation was 38.1 ($SD = 12.9$, range = 18–76) years and the majority of the individuals in the sample were White (specific information relating to race and ethnicity was not collected in the original study).

Individuals were classified into one of two groups based on their offense history: The child sexual exploitation materials/no-contact group (CSEM/NC; $n = 283$) included individuals who had not committed any known contact sexual offenses. For the vast majority of this group, their only sex offenses were for CSEM offenses, but a small number also had noncontact sex offenses, such as online sexual communication with a child or exhibitionism ($n = 10$). The mixed group (CSEM + Contact; $n = 82$) had both CSEM and contact sexual offenses in their criminal record. Individuals who directly victimized children as part of their CSEM production offenses were included in the CSEM + Contact group. Both groups also included individuals with nonsexual offenses in their criminal history. This distinction between CSEM/NC and CSEM + contact is important given previous research on meaningful differences between the two groups, and also because those with contact sexual offenses are within the typical samples of generic sexual recidivism risk tools.

MEASURES

CPORT

The seven CPORT items are scored dichotomously, as yes or no. CPORT total scores can therefore range from 0 to 7, with a maximum of one missing item (Eke et al., 2018). When information on admission/diagnosis of sexual interest in children is unavailable, this item can be replaced by a score of 3 or higher on the Correlates of Admission of Sexual Interest in Children (Seto & Eke, 2017). As reviewed earlier, CPORT predicts sexual recidivism with moderate to high accuracy in most studies. In addition, it has good interrater reliability in both research and field contexts (Eke et al., 2019; Hermann et al., 2019; Savoie et al., 2021; Seto & Eke, 2015).

Risk Matrix 2000/Sex

Scoring the RM2000/S involves two steps; in the first step, three items are scored: age at commencement of risk to reoffend, sexual crime court appearances, and general crime court appearances. Based on these scores, the individual is assigned one of four preliminary risk levels: below average, average, above average, and well above average. Then in Step 2, four aggravating items are scored dichotomously: male victim, stranger victim, never lived with a lover for 2 years, and noncontact sex offense. Stranger victim is not scored on the basis of CSEM materials (Thornton et al., 2023). Noncontact sex offense is only scored for internet offenses if the individual also has an offline sex offense. Finally, male victims are only scored on the basis of CSEM materials if there is evidence that the individual deliberately sought boys in the material. In this study, we operationalized this based on whether they had more boy than girl content in their CSEM collection. After Step 2, the individual's initial risk level is increased by one category for every two aggravating factors that apply. The scale predicts well among diverse samples of men convicted of primarily offline sex offenses (Helmus et al., 2013).

RECIDIVISM

Data on recidivism were collected from two main sources: police occurrence reports and a national database of criminal charges and convictions maintained by the Canadian Police Information Centre (CPIC), a service of the Royal Canadian Mounted Police (RCMP). Recidivism was defined as a new charge or conviction for any sexual offense (including CSEM, noncontact sexual offenses such as exhibitionism, and contact sexual offenses) and any CSEM offense. Follow-up began at the date of first release from the index CSEM charge(s) (e.g., release on bail, release at conviction) and ended at the date when criminal records were checked (summer of 2012 for the development sample and summer 2015 for the validation sample), or date of death, whichever was sooner. Time in custody (e.g., time in jail for the index or any subsequent offense) was subtracted, so follow-up time represented the person's opportunity to offend while residing in the community ($M = 8.1$ years, $SD = 2.4$ years; range = 0.3–17.6 years).

PROCEDURE

The CPORT and RM2000/S were scored from police files that included criminal history records, police occurrence reports, interview notes or transcription, police officer notes,

forensic computer analysis reports, details about the size and content of the pornographic and child material, and in most cases, either videos or transcripts of police interviews. The categorization of the CSEM materials was obtained from forensic and police notes. Permission to access case file information was obtained from the participating police services. This research was approved by the institutional review board of the Royal Ottawa Health Care Group.

TRANSPARENCY AND OPENNESS

The current study sample contains protected information maintained by the Ontario Provincial Police and cannot be shared outside the service; however, requests for additional analyses and data verification can be submitted for review for accommodation on-site. The meta-analysis data sets (and syntax template) are available from Open Science Framework (https://osf.io/pn9d6/?view_only=da2fa684366645d9b6db5a14c72464c5). Materials needed to score CPORT are available from ResearchGate and materials to score the Risk Matrix 2000 are available from www.saarna.org.

OVERVIEW OF ANALYSES

Discrimination and Calibration

Discrimination and calibration are two types of predictive accuracy that can be examined for risk assessment tools (also sometimes referred to as relative and absolute prediction, respectively; Helmus & Babchishin, 2017). Discrimination examines how well the tool distinguishes recidivists from nonrecidivists (i.e., the extent to which higher risk scores are associated with higher likelihoods of recidivism). There are several statistics commonly used to assess discrimination; we reported AUCs as well as Harrell's C and hazard ratios from Cox regression analyses. The AUC from receiver operating characteristic curve (ROC) analyses can range between 0 and 1, with values between .50 and 1 indicating positive predictive accuracy (higher scoring individuals are more likely to recidivate than lower scoring individuals), values of .50 indicating no predictive accuracy, and values below .50 indicating negative predictive accuracy. AUCs of .56, .64, and .71 were considered small, moderate, and large effect sizes, respectively, as they roughly correspond to Cohen's *d* values of .20, .50, and .80 (Rice & Harris, 2005).

We also used Cox regression (Singer & Willett, 2003), which accounts for varying follow-up periods. Cox regression provides hazard ratios, quantifying increases in recidivism with each one-point increase on the risk scale, averaged across time. Hazard ratios, however, cannot easily be compared across scales unless they have the same possible range of scores, because scales with more points are expected to have smaller differences in the outcome between adjacent values. We, therefore, also reported Harrell's C values (Harrell et al., 1996), which were derived from the Cox regression model. Harrell's C is an analogue of AUCs for survival data, and can be interpreted in the same way (e.g., .56, .64, and .71 reflecting small, moderate, and large values).

Calibration refers to the correspondence between observed and predicted rates of recidivism. We could not examine calibration of the CPORT because we used the same sample used to report preliminary CPORT recidivism estimates. Analyses of the calibration of the RM2000/S were conducted with the E/O index (Gail & Pfeiffer, 2005; Rockhill et al., 2003),

comparing with the estimates obtained from Lehmann et al. (2016). The E/O index is the ratio of the predicted or expected number of recidivists (E) divided by the observed number of recidivists (O; Method M₀ from Viallon et al., 2009). The E/O index is both a significance test and a measure of effect size. If the predicted numbers of recidivists perfectly matches the observed number, the E/O index will be 1. Values below 1 mean that the RM2000/S underestimated recidivism, and values above 1 reflect overestimation. Ninety-five percent confidence intervals that do not include 1 indicate significant differences between observed and predicted rates. For further explanation and calculation examples of the E/O Index, see Hanson (2017). Analyses were run in either SPSS version 20.0 or R.

Comparing the Discrimination of CPORT and RM2000/S

A key purpose of this article was to compare the predictive accuracy of the CPORT and RM2000/S, but the optimal analysis for this is unclear. In the last 10 years, many researchers have compared AUCs for risk scales using the Delong test, which accounts for the correlation between the two tools (Delong et al., 1988; for further examples on the use of this test, see Babchishin et al., 2012; Eher et al., 2016; Helmus et al., 2019; Wakeling et al., 2011). More recently, the BIC has been used to compare regression models, for example, to examine changes in risk over time (Babchishin & Hanson, 2020; Hanson et al., 2021; Helmus et al., 2021; Lloyd et al., 2020). For Cox regression models, the $BIC = -2 LL + [k \times \ln(n)]$, where k is the number of parameters and n is the number of recidivists (Raftery, 1995; Volinsky & Raftery, 2000). The BIC compares non-nested models, with smaller BIC values suggesting better fitting models. BIC differences (absolute value) of 0 to 2, 2 to 6, 6 to 10, and 10 and higher, respectively, represent “weak,” “positive,” “strong,” and “very strong” evidence of differences in model fit (Gordon, 2012).

A key advantage of the Delong test is that it accounts for the correlation between the tools being compared, which maximizes statistical power. A drawback is that it is primarily a null hypothesis significance test (i.e., there are no benchmarks for interpreting the differences in AUCs). The BIC includes benchmarks for interpretation of the magnitude (i.e., effect size) of differences in model fit. However, like many benchmarks, they may have some arbitrariness to them and are not meant to be applied blindly. Both are strongly influenced by the number of recidivists (which are small in many of our analyses, which could lead to substantial fluctuations). Consequently, we explored comparisons using both techniques.

Meta-Analysis

Validation studies of CPORT and RM2000/S with CSEM samples were obtained by searching both scale names in Google Scholar and PsycINFO. In addition, the developers of both scales were asked to provide a list of all validations they were aware of, and an additional scholar who specializes in CSEM research was also contacted to see whether they were aware of anything we missed. This field of research is sufficiently small that it is unlikely any of these scholars would be unaware of any published research, but it is possible that some unpublished research was missing, although most of the CPORT validation studies are unpublished dissertations that have been shared or discovered by CPORT authors.

TABLE 1: CPORT and Risk Matrix 2000/S Total Scores and Subgroup Comparisons

Scale/Level	Full sample		CSEM/NC		CSEM + contact		Cohen's <i>d</i> or <i>h</i> ^a [95% CI]
	<i>N</i>	<i>M</i> (<i>SD</i>) or %	<i>N</i>	<i>M</i> (<i>SD</i>) or %	<i>N</i>	<i>M</i> (<i>SD</i>) or %	
CPORT total	365	1.98 (1.57)	283	1.54 (1.23)	82	3.49 (1.67)	1.45 [1.18, 1.72]
RM2000/S	365		283		82		
Below average	146	40.0	130	46.0	16	19.5	0.58 [0.33, 0.82]
Average	150	41.1	124	43.8	26	31.7	0.25 [0.005, 0.50]
Above average	58	15.9	28	9.9	30	36.6	-0.66 [-0.91, -0.41]
Well above average	11	3.0	1	0.4	10	12.2	-0.59 [-0.83, -0.34]

Note. CPORT = Child Pornography Offender Risk Tool; RM2000/S = Risk Matrix 2000/S; CSEM = child sexual exploitation materials; NC = noncontact; CI = confidence interval.

^aFor comparisons of the proportions in risk levels between the CSEM/NC and CSEM + Contact samples, Cohen's *h* effect size for two proportions is reported; for comparisons of means, Cohen's *d* is reported (Cohen, 1992).

Meta-analyses followed the formulae of Borenstein et al. (2021). Although random-effects analyses are often conceptually preferable, they are unstable when the number of studies is below 30 (Schulze, 2007); consequently, we reported both, but primarily relied on fixed-effect analyses. The primary drawback of fixed-effect meta-analysis is that it assumes studies are measuring the same common effect size and variability across studies is not incorporated into the error term, often resulting in unrealistically narrow confidence intervals. Variability in findings across studies was reported using Cochran's *Q* statistic and the *I*² effect size statistic. *I*² values of 25%, 50%, and 75% have been proposed as reflecting low, moderate, and high variability, respectively (Higgins et al., 2003).

Overlap With Previous Research

Fixed 5-year follow-up AUCs for the CSEM/NC and CSEM + Contact subgroups were reported in Eke et al. (2019), with minor fluctuations due to ongoing data cleaning. CPORT analyses for the full group are similar to what is reported in Eke et al. (2019) but are following the new rules for handling missing data (Eke et al., 2018). Cox regression, Harrell's *C*, all analyses of the RM2000/S, and comparisons between the CPORT and the Risk Matrix have not been reported elsewhere. The meta-analysis has also not been reported elsewhere.

RESULTS

Table 1 presents the mean CPORT scores and distributions among RM2000/S levels for the full sample, as well as the CSEM/NC and CSEM + Contact group. CPORT scores were significantly higher for the CSEM + Contact group compared with the CSEM/NC group (3.5 vs. 1.5), with a large effect size (Cohen's *d* = 1.45). This is not surprising given that everyone in the contact group would have an extra point for having a contact sexual offense. However, the difference in means was two points, indicating that they scored an additional point higher on average, based on risk factors beyond their contact offense.

In the overall sample, more than 80% of the men scored Below Average or Average risk on the RM2000/S. Comparing the CSEM/NC and CSEM + Contact groups, there were significantly more CSEM/NC men in Below Average and Average groups, and significantly

TABLE 2: Comparing AUCs From the CPORT and Risk Matrix 2000/Sex Using the Delong Test, and Fixed 5-Year Follow-Up

Scale	N	AUC	95% CI	Difference between the AUCs			
				Difference	95% CI	Z	p
All cases	339						
Any sexual recidivism	<i>n</i> recid = 40						
CPORT Total		.73	[.64, .82]				
RM2000/S		.66	[.56, .76]	.07	[0.001, 0.148]	1.98	.048
CSEM recidivism	<i>n</i> recid = 29						
CPORT Total		.74	[.64, .85]				
RM2000/S		.67	[.56, .79]	.07	[-0.013, 0.148]	1.65	.099
CSEM/NC	273						
Any sexual recidivism	<i>n</i> recid = 23						
CPORT Total		.68	[.56, .80]				
RM2000/S		.56	[.43, .69]	.12	[0.009, 0.230]	2.12	.034
CSEM recidivism	<i>n</i> recid = 17						
CPORT Total		.69	[.54, .84]				
RM2000/S		.59	[.45, .75]	.10	[-0.031, 0.222]	1.48	.139
CSEM + contact	66						
Any sexual recidivism	<i>n</i> recid = 17						
CPORT		.72	[.56, .87]				
RM2000		.74	[.58, .90]	.03	[-0.150, 0.098]	-0.41	.683
CSEM recidivism	<i>n</i> recid = 12						
CPORT		.77	[.59, .95]				
RM2000		.73	[.55, .92]	.03	[-0.078, 0.146]	0.60	.551

Note. Effect sizes in bold are statistically significant ($p < .05$). AUC = areas under the curve; CPORT = Child Pornography Offender Risk Tool; CI = confidence interval; RM2000/S = Risk Matrix 2000/S; CSEM = child sexual exploitation materials; NC = noncontact.

fewer in the Above Average and Well Above Average groups. In the overall sample ($n = 365$), 15.9% reoffended with any type of sexual offense, and 11.2% reoffended with a CSEM offense. Consistent with their higher risk scores, the CSEM + Contact group ($n = 82$) had sexual (28.0%) and CSEM (20.7%) recidivism rates over twice as high as the CSEM/NC group (12.4% sexual recidivism and 8.5% CSEM recidivism).

COMPARING THE CPORT AND THE RISK MATRIX 2000/SEX

CPORT and RM2000/S scores were strongly and positively correlated ($r = .73, p < .0001$), suggesting good convergent validity. Table 2 presents AUCs for fixed 5-year follow-ups and the corresponding Delong test of AUC differences for the CPORT and RM2000/S for sexual and CSEM recidivism, for the overall sample and broken down by the two subgroups. For the full sample, the CPORT had large AUCs ($> .71$), whereas the RM2000/S had moderate AUCs (between .66 and .67). The difference in AUCs was significant for any sexual recidivism but not CSEM recidivism. AUCs were consistently high for the CSEM + Contact group (ranging between .72 and .77) for both scales, but for the CSEM/NC group, the CPORT had moderate effects and the RM2000/S had only small effects.

For the CSEM + Contact group, which was the group the RM2000/S was designed for, the RM2000/S had a slightly higher AUC than the CPORT for sexual recidivism (AUC = .74 vs. .72, respectively), but the difference was not significant. In all other analyses in

TABLE 3: Comparing CPORT and Risk Matrix 2000/Sex Based on BIC From Cox Regression Models

Predictors	Any sexual recidivism					Any CSEM recidivism				
	<i>N</i> recid	Harrell's C	Hazard ratio [95% CI]	BIC	BIC difference	<i>N</i> recid	Harrell's C	Hazard ratio (95% CI)	BIC	BIC difference
Full sample (<i>N</i> = 365)										
CPORT	58	.69	1.54 [1.33, 1.77]	633.6	-7.57	41	.74	1.71 [1.45, 2.03]	433.8	-11.21
RM2000/S	58	.64	2.15 [1.59, 2.92]	641.2		41	.69	2.51 [1.76, 3.58]	445.0	
CSEM/Non-contact (<i>N</i> = 283)										
CPORT	35	.61	1.34 [1.04, 1.71]	378.0	-3.23	24	.70	1.62 [1.22, 2.15]	251.6	-3.74
RM2000/S	35	.55	1.39 [0.86, 2.23]	381.2		24	.62	2.07 [1.19, 3.60]	255.3	
CSEM + contact (<i>N</i> = 82)										
CPORT	23	.69	1.64 [1.27, 2.14]	180.1	-0.14	17	.73	1.90 [1.38, 2.62]	128.9	-7.00
RM2000/S	23	.69	2.64 [1.55, 4.53]	180.2		17	.68	2.52 [1.37, 4.64]	135.9	

Note. CPORT = Child Pornography Offender Risk Tool; BIC = Bayesian information criterion; CI = confidence interval; RM2000/S = Risk Matrix 2000/S; CSEM = child sexual exploitation materials; NC = noncontact.

Table 2, the AUC for the CPORT was higher. It was a small difference for the CSEM + Contact group (AUC differences of .03), but in all other comparisons, the differences in AUCs were more pronounced, ranging between .07 and .12. In addition to sexual recidivism for the full group, the only other AUC difference that was statistically significant was predicting sexual recidivism among CSEM/NC individuals.

Table 3 presents the Cox regression results, which include the Harrell's C effect size based on survival data. These results were similar to the AUCs based on fixed follow-ups in Table 2. Here, however, the focus is on the BIC as an indicator of model fit, and differences between BICs are a comparison of the fit of the CPORT versus the RM2000/S. For the full sample, CPORT was a better fit than the RM2000/S; the difference in model fit was strong for any sexual recidivism, and very strong for CSEM recidivism. For the CSEM/NC sample, the differences between model fit again favored the CPORT. For the CSEM + Contact sample, the difference between the CPORT and RM2000/S was almost nonexistent for any sexual recidivism, but strong for the prediction of CSEM recidivism, again favoring the CPORT.

CALIBRATION OF THE RM2000/S

Table 4 presents the 5-year sexual recidivism rates per RM2000/S risk level for the full sample and subgroups, alongside the recidivism norms for the tool from Lehmann et al. (2016). Recidivism rates observed in the current sample were generally higher than expected rates. Overall, RM2000/S significantly underestimated recidivism for the full sample and the CSEM/NC and CSEM + Contact subsamples, all by roughly the same amount, with E/O indices between .57 and .58; this indicates that expected recidivism rates were 57% to 58% of what was observed, or conversely, observed recidivism rates were nearly twice as high as expected. For example, for the full sample, the scale

TABLE 4: RM2000/S Calibration Analyses With Fixed 5-Year Follow-Up Data

Risk Matrix category	Observed data			Expected data		E/O index	[95% CI]
	<i>N</i>	<i>N</i> recid (O)	% recid	Predicted recid (%)	Predicted recid <i>n</i> (E)		
Full sample							
Below average	137	11	8.0	4.4	6.0	.55	 [.30, .99]
Average	138	12	8.7	10.0	1.4	.12	 [.07, .20]
Above average	54	8	14.8	21.4	11.6	1.44	[.72, 2.89]
Well above average	10	9	90.0	40.0	4.0	.44	 [.23, .85]
Total	339	40			23.0	.57	 [.42, .78]
CSEM/NC							
Below average	127	9	7.1	4.4	5.6	.22	 [.06, .88]
Average	117	10	8.5	10.0	1.2	.11	 [.03, .42]
Above average	28	3	10.7	21.4	6.0	1.11	[.46, 2.67]
Well above average	1	1	100.0	40.0	0.4	.45	 [.23, .90]
Total	273	23			13.2	.58	 [.36, .93]
CSEM/C							
Below average	10	2	20.0	4.4	0.4	.62	 [.32, 1.19]
Average	21	2	9.5	10.0	0.2	.12	 [.06, .22]
Above average	26	5	19.2	21.4	5.6	2.00	[.64, 6.19]
Well above average	9	8	88.9	40.0	3.6	.40	[.06, 2.84]
Total	66	17			9.8	.57	 [.38, .86]

Note. Expected values were obtained from Lehmann et al. (2016). CSEM/NC refers to the CSEM/Noncontact group and CSEM/C refers to the CSEM + Contact group. RM2000/S = Risk Matrix 2000/S; Recid = recidivism; CSEM = child sexual exploitation materials; NC = noncontact. Bolded values denote statistically significant E/O index.

predicted 23 sexual recidivists but there were 40. The underestimation was roughly consistent for all risk levels except the above average risk level, where there was slight (non-significant) overestimation of recidivism.

CUMULATIVE META-ANALYSIS OF CPORT AND RM2000/S VALIDATION STUDIES

Table 5 summarizes the replication studies of CPORT and RM2000/S with men with CSEM offenses, including the current sample (which overlaps with Eke et al., 2019 and Seto & Eke, 2015). Where multiple effect sizes were reported, we coded effects based on the CPORT recommendations for how much missing information was allowed (Eke et al., 2018), with the exception of Soldino et al. (2021) and Eke et al. (2019), where we used all cases because restricting the sample to those with complete information severely reduced the sample size. Especially given the small number of replication studies, there could be a concern that including the development study (Seto & Eke, 2015) might inflate the findings. However, the effect size from the development study was the median value; consequently, there was no need to remove it as it would not meaningfully impact the analysis. Instead, effect sizes from the current study were used to replace Seto and Eke (2015) and Eke et al. (2019) as this study contains both samples, with the most updated and cleaned data set.

For the prediction of any sexual recidivism, there were five CPORT studies with a large weighted average AUC of .75 (95% confidence interval [CI] = [.71, .79], $N = 1,411$; see Table 6). The variability in predictive accuracy across studies was no more than what would

TABLE 5 Studies of CPORT and Risk Matrix 2000 Included in Cumulative Meta-Analysis

Study	Location	Sampling timeframe	CPORT <i>M</i>	<i>SD</i>	Follow-up (years)	<i>n</i> recid / total	Recid rate (%)	AUC	[95% CI]
CPORT—Any sexual recidivism									
Black (2018)	New Zealand	1998–2014	1.27	1.24	7.6	?/547	—	.77	[.71, .82]
Eke et al. (2019)	Canada	2006–2010	1.77	1.50	5.0	12/80	15.0	.70	[.54, .86]
Gunnarsdóttir (2019) ^a	Iceland	2000–2014	1.79	1.15	5.0	?/106	—	.75	[.62, .89]
Pilon (2016)	Canada	2010–2011	—	—	3.2	8/279	2.9	.56	[.32, .79]
Savoie et al. (2021)	Scotland	2010–2013	1.91	1.29	5.0	14/140 ^b	10.0	.77	[.67, .87]
Seto & Eke (2015)	Canada	1993–2006	1.94	1.57	5.0	28/266	11.0	.74	[.63, .84]
<i>Current study</i> ^c	Canada	1993–2010	1.98	1.57	8.1	40/339	11.8	.73	[.64, .82]
CPORT—CSEM recidivism									
Black (2018)	New Zealand	1998–2014	1.27	1.24	7.6	71/547	13.0	.77	[.71, .82]
Gunnarsdóttir (2019) ^d	Iceland	2000–2014	1.79	1.15	5.0	12/106	11.3	.62	[.53, .70]
Pilon (2016)	Canada	2010–2011	—	—	3.2	7/279	2.5	.52	[.27, .77]
Savoie et al. (2021)	Scotland	2010–2013	1.91	1.29	5.0	11/140	7.9	.73	[.61, .85]
Soldino et al. (2021)	Spain	2009–2013	0.8	0.93	5.0	6/304	2.0	.56	[.51, .62]
<i>Current study</i> ^c	Canada	1993–2010	1.98	1.57	8.1	29/346	8.4	.74	[.64, .85]
Risk Matrix 2000—Any sexual recidivism									
Wakeling et al. (2011)	U.K.	Up to 2007	—	—	2.0	31/994	3.1	.67	[.57, .77]
<i>Current Study</i>	Canada	1993–2010	—	—	8.1	40/339	11.8	.66	[.56, .76]

Note. Where possible we coded effect sizes to three significant figures, but sometimes articles reported less than that. When 5.0 years is reported for the follow-up period, a fixed length of follow-up was used for all cases. Pilon (2016) was missing Items 6 and 7 for CPORT data; Black (2018) was missing Items 5, 6, and 7. The current data set used charges as the recidivism outcome and Soldino et al. (2021) used arrests; the remaining studies used convictions. CPORT = Child Pornography Offender Risk Tool; AUC = area under the curve; CI = confidence interval; CSEM = child sexual exploitation materials.

^aThis effect size was not in the dissertation but was obtained by personal communication (H. Gunnarsdóttir, personal communication, December 16, 2021). ^bTwo of the 14 recidivism incidents were for technical breaches of their sexual offense supervision order, and not necessarily for committing a new sexual offense. ^cNote that the current study subsumes the samples from Eke et al. (2019) and Seto and Eke (2015). The meta-analysis reported in text replaces those studies. ^dLog odds ratios and their 95% confidence interval limits were transformed to Cohen's *d* (Sánchez-Meca et al., 2003) and then to AUCs (Ruscio, 2008).

TABLE 6: Meta-Analysis Results for CPORT and RM2000

Predictor	Fixed-effect model		Random-effects model		<i>k</i>	<i>N</i>	<i>Q</i>	<i>I</i> ²
	AUC	[95% CI]	AUC	[95% CI]				
Any sexual recidivism								
1. Age—35 or younger at investigation	.60	[.55, .65]	.60	[.55, .65]	4	1,313	1.12	0.0
2. Prior criminal history	.70	[.65, .74]	.67	[.58, .75]	4	1,313	6.46	53.6
3. Any failure on conditional release	.68	[.63, .73]	.68	[.60, .76]	4	1,313	5.10	41.1
4. Any contact sex offense	.58	[.53, .64]	.58	[.53, .64]	4	1,313	3.07	2.4
5. Pedo/hebephilic interests	.60	[.51, .68]	.58	[.48, .69]	3	712	2.44	17.9
6. More boy child pornography	.56	[.46, .65]	.47	[.18, .77]	2	479	5.78	82.7
7. More boy nudity/other material	.54	[.44, .62]	.45	[.09, .79]	2	361	9.51	89.5
CPORT total score	.75	[.71, .79]	.75	[.71, .79]	5	1,411	3.44	0.0
Risk Matrix 2000/Sex	.66	[.59, .74]	.66	[.59, .74]	2	1,340	0.02	0.0
CSEM recidivism								
1. Age—35 or younger at investigation	.60	[.56, .64]	.60	[.54, .65]	5	1,617	5.83	31.4
2. Prior criminal history	.63	[.59, .67]	.63	[.56, .71]	5	1,617	10.20	60.8
3. Any failure on conditional release	.60	[.56, .64]	.62	[.52, .73]	5	1,617	18.43	78.3
4. Any contact sex offense	.54	[.50, .58]	.54	[.49, .59]	5	1,616	4.91	18.5
5. Pedo/hebephilic interests	.55	[.49, .62]	.56	[.47, .64]	4	863	4.25	29.4
6. More boy CSEM	.59	[.53, .64]	.56	[.43, .70]	3	734	7.27	72.5
7. More boy nudity/other child material	.56	[.50, .62]	.52	[.35, .69]	3	530	10.94	81.7
CPORT Total Score	.66	[.63, .70]	.65	[.55, .76]	5	1,376	29.65	86.5

Note. CPORT = Child Pornography Offender Risk Tool; AUC = area under the curve; CI = confidence interval; CSEM = child sexual exploitation materials.

be expected from sampling error ($Q = 3.44, p = .487, I^2 = 0\%$). For the prediction of CSEM recidivism, there were also five studies, with a moderate effect (AUC = .66, 95% CI = [.63, .70]), and significant and large variability in findings across studies ($Q = 29.65, p < .001, I^2 = 86.5\%$). Specifically, 86% of the variability across studies was above and beyond what would be expected from sampling error. Unfortunately, there was an insufficient number of studies to conduct moderator analyses to explain the variability or to examine possible outliers.

Table 6 also presents the meta-analytic results for the individual items of the CPORT. Item data were available from the current study as well as Black (2018), Pilon (2016), Savoie et al. (2021), and for CSEM recidivism, also Soldino et al. (2021). As noted in Table 5, however, some samples were missing information on some items, so sample sizes fluctuate dramatically for these analyses. For any sexual recidivism, Items 1 through 4 had data from four samples and demonstrated significant predictive accuracy. For Items 5, 6, and 7, sample sizes dropped considerably. Item 5 (pedo/hebephilic interests) was a significant predictor in the fixed-effect model but not the random-effects model. Items 6 and 7 (related to preferences for boys in CSEM and other child content) only had data from two samples (the current study and Savoie et al., 2021). These items significantly predicted in the current sample but had negative predictive accuracy in Savoie et al.'s (2021) sample. These items were not significantly predictive in the aggregated analysis. For CSEM recidivism, Items 1 through 3 significantly predicted recidivism in both fixed-effect and random-effects

analyses. Item 4 (any contact sex offense) did not quite reach statistical significance and had lower effect sizes ($AUC = .54$) compared with any sexual recidivism ($AUC = .58$). Items 5 through 7 did not significantly predict CSEM recidivism, although sample sizes were much reduced. In addition, as per the total score for CSEM recidivism, four of the seven items had significant variability in predictive accuracy across samples.

There were only two nonoverlapping studies of the RM2000/S, but combining them yielded an average AUC of .66 (95% $CI = [.59, .74]$, $k = 2$, $N = 1,333$), and the sample size from these two studies is roughly the same as the meta-analyses of the CPORT. Both studies had very similar effect sizes (.66 vs. .67); consequently, the variability between these two results were within what would be expected from sampling error ($Q < 0.1$, $I^2 = 0\%$). Results were not meta-analyzed for the RM2000/S items as they were not presented in the previous RM2000/S study (Wakeling et al., 2011).

DISCUSSION

This study was the first to directly compare the CPORT and the RM2000/S in the same sample and also used two different techniques to compare the predictive accuracy of the scales, the Delong test and BIC differences. Both the CPORT and the RM2000/S demonstrated large effect sizes in the fixed 5-year follow-up analyses for the overall sample; effect sizes dropped a little for the Harrell's C analyses based on survival data. All but one effect size favored the CPORT (survival analyses of any sexual recidivism with the CSEM + Contact group slightly favored the RM2000/S), although differences in AUC s ranged considerably, from .03 to .12.

Delong tests showed that the CPORT significantly outperformed the RM2000/S in predicting any sexual recidivism for the overall sample and the CSEM/NC sample. Comparing BICs, all comparisons suggested the CPORT model was a better model than the RM2000/S, except for sexual recidivism among those with contact offenses. The differences in model fit were considered strong or very strong for predicting any sexual and CSEM recidivism among the full sample, and CSEM recidivism among the CSEM + Contact sample. The stronger effect sizes for the CPORT could be because it was specially developed for CSEM perpetrators, or possibly because it has a larger range of scores within which to distinguish risk, whereas the RM2000/S has only four levels (and the range within those levels tends to be restricted among CSEM samples).

For evaluators using these risk tools for CSEM cases, the group of most interest is those who do not have any offline sexual offenses because other widely used risk tools such as Static-99R are not applicable in these cases.² Effect sizes were meaningfully lower for this subgroup, but still significant, and in the fixed follow-up analyses for CPORT, moderate in magnitude and comparable with Static-99R (see Helmus et al., 2022). It is difficult to interpret these subgroup differences. While the CSEM/NC group is often of particular interest, separating them out from men with CSEM and contact offenses is restricting the range in risk and will necessarily reduce effect sizes. In this sample, the CSEM + Contact group scored two points higher than the CSEM/NC group on the CPORT, where only one point of which could be explained by the item for contact sex offenses. So for a broad and heterogeneous sample of people with CSEM convictions, CPORT does well at discrimination in predicting sexual recidivism. When looking at a narrow subgroup with considerably less heterogeneity, such as those with no other criminal history, it becomes harder to distinguish risk, but in the context of the full group, their lower risk scores and reduced variability is informative in and of itself.

METHODS FOR COMPARING PREDICTIVE ACCURACY

Both the Delong and BIC tests revealed meaningful or significant differences between the CPORT and the Risk Matrix, but not necessarily in the same comparisons. This may partly be because the Delong tests used fixed 5-year follow-ups, which tended to yield cleaner and stronger effects than the Cox regression survival analyses used for the BIC comparisons. Nonetheless, the heuristics for interpreting BIC differences appeared to identify more differences as meaningful. Given the BIC is used to examine magnitude of model fit differences, it may be sensitive to large and meaningful differences that do not have sufficient statistical power to reach significance in the Delong test. The Delong tests, however, take into account the correlation between the two tools. Until more research is conducted comparing the two approaches, there may be some benefit in reporting both for direct scale comparisons. Minimally however, it is important to recognize that the analytic approach taken will impact the conclusions.

CALIBRATION ISSUES IN CSEM RISK ASSESSMENT

Given the low recidivism rates of men with CSEM offenses (Babchishin et al., 2015, 2018; Seto et al., 2011), a concern about using generic sexual offense risk assessment tools is that they may overestimate the risk of men with CSEM offenses, leading to violations of the risk principle of effective correctional practice (Bonta & Andrews, 2017). The current sample found the opposite; observed recidivism rates in this sample were almost twice the predicted recidivism rates from the RM2000/S norms. It is not clear why this was the case. It could not be attributable to the number of people in the higher risk group with mixed CSEM and contact offenses, as this evidence of poor calibration was found in the CSEM/NC group as well. The current study defined recidivism as new charges (with a large majority known to have ended in convictions) for sexual offenses, whereas the recidivism norms for the RM2000/S are for new convictions only. This methodological difference is unlikely to account for a large difference in calibration, based on previous research demonstrating that differences in recidivism rates across studies were not consistently and meaningfully explained by the use of charges or convictions as the recidivism outcome (Helmus, 2009).

The current sample may have a higher sexual recidivism rate in large part because of our access to multiple sources of good quality recidivism information. For example, we had access to outcome data through police occurrence reports from a large group of police services. The benefit of police occurrence reports is that they include charges at the time they are laid. We also had information from a national database of convictions. However, there may be lag time between a conviction being registered in court and it being recorded on the national system. As well, some convictions may be removed from the national database if the individual received a pardon for their offense(s), although reference to these offenses may still appear in the police databases. Furthermore, not all convictions are included in the national database. This study had an 11.8% sexual recidivism rate after 5 years, which is slightly higher than the base rate of recidivism among men charged or convicted for non-CSEM sexual offenses in previous research (9.1%; Hanson et al., 2018). Savoie et al. (2021) had a somewhat similar high 5-year sexual recidivism rate (10.0%) in their CPORT validation study, but Pilon (2016) had a recidivism rate of 2.9%, based on a more restrictive definition of new convictions solely within the province of Ontario. Consequently, it is unclear whether the RM2000/S truly underpredicts sexual recidivism for those with CSEM offenses,

or whether this sample has an unexpectedly high recidivism rate. Minimally, however, the results should mitigate concerns that the scale's recidivism probabilities will overpredict recidivism. Although those with CSEM offenses as a group may (in most studies) have lower sexual recidivism rates than those with offline sex offenses, their risk level will likely take this into account, especially given that the RM2000/S has clear coding rules for handling these cases.

SUMMARY OF META-ANALYTIC FINDINGS

Finally, we conducted a meta-analysis of existing CPORT and RM2000/S validation studies among men with CSEM offenses. Comparisons between the two tools are less useful in these meta-analyses because the samples and methods differ across studies. Nonetheless, the results support the use of both tools. CPORT had large effect sizes in predicting any sexual recidivism and moderate effect sizes in predicting CSEM recidivism. The RM2000/S had moderate effect sizes in predicting any sexual recidivism, with insufficient data to test CSEM recidivism. The CPORT has more validation studies available, although some have considerable missing information (Black, 2018; Pilon, 2016) and the scoring of the items were modified in Pilon (2016). The RM2000/S has only two nonoverlapping studies, but one is quite large, with nearly 1,000 men with CSEM offenses, and both yielded very similar effects, with AUCs between .66 and .67. As an important point for comparison, although not large, the meta-analytic average effects for the CPORT and the RM2000/S for this population are similar to or greater than a recent meta-analysis of the predictive accuracy of Static-99R, the most widely used sexual offense risk tool (fixed-effect AUC = .68, $k = 56$; Helmus et al., 2022).

There was considerable overlap in the studies available to test CPORT's accuracy in predicting CSEM and any sexual recidivism. Two interesting patterns emerged across these findings. Predictive accuracy was consistent across samples for any sexual recidivism but not for CSEM recidivism. Effect sizes were also meaningfully larger for predicting any sexual recidivism (AUC = .75) compared with CSEM recidivism (AUC = .66). Any sexual recidivism is a broader outcome and includes CSEM offenses, and it was the outcome being predicted by the development sample, so in some ways it is understandable that the scale may be maximized for this outcome. This is also convenient as sexual recidivism risk assessments of individuals with CSEM convictions are often concerned with any sexual recidivism.

The lower accuracy and greater variability in findings for CSEM recidivism may also reflect challenges in measuring this outcome reliably. Although all sexual offenses are underreported to police, CSEM is unique in that its detection is heavily reliant on the investigative techniques and resources of police agencies, political priorities, and jurisdictional variation in prosecution practices. This could simply make it a harder outcome to predict reliably and accurately. In contrast, however, once it is identified/prosecuted, the digital evidence inherent in modern CSEM offending makes a conviction highly likely.

Finally, another possible explanation for the lower accuracy for predicting CSEM recidivism compared with any sexual recidivism may be a statistical artifact. Lower base rate outcomes are harder to predict, and many effect size measures are influenced by the base rate of the dichotomous outcome, such as recidivism. The effect sizes we used (AUCs and Harrell's Cs) are less impacted by base rates than correlations, but are not unaffected (Babchishin & Helmus, 2016).

STRENGTHS AND LIMITATIONS

This study is the first to directly compare two risk assessment tools in the same sample of individuals involved with CSEM. In addition, the data in Study 1 included limited missing information, high-quality recidivism information, and a reasonably good length of follow-up (5+ years). There are also limitations to the study and constraints on generality. The current sample contained adult men, predominantly White, residing in Ontario, Canada. The meta-analysis supported the consistency of Risk Matrix 2000 and CPORT in predicting sexual recidivism in some other similar countries (industrialized, high-income, relatively educated). Generalizability outside these countries is unknown. Cross-cultural validity analyses should be conducted where sample sizes are sufficient, and more research is needed across countries. Women who sexually offend are meaningfully different than men who sexually offend (Cortoni, 2018) and these tools are not recommended for women without additional research. We also operationalized the male victim item of the RM2000/S based on a greater proportion of male versus female content, as opposed to the guidance in the coding manual, which focuses on evidence of deliberately searching for male content. This may have underestimated the prevalence of this risk factor.


Similar to most studies on individuals convicted of CSEM offenses, the sample size is not particularly large. Aggregating the results through meta-analysis improves our understanding of predictive accuracy, but introduces additional limitations. A meta-analysis is only as good as the quality of the studies included, which typically vary in terms of sample size and quality of data available for risk tool scoring and recidivism information. There were only two studies available on the RM2000/S, and some studies on the CPORT had considerable missing information. However, the missing information suggests that if anything, the current meta-analysis would offer a conservative estimate of its predictive accuracy. In addition, there were too few studies on the CPORT overall for any meaningful moderator analyses.

CONCLUSION

If sufficient information is available, the current study suggests the CPORT may be preferable to use. The RM2000/S is empirically defensible to use with men with CSEM offenses, but may underestimate recidivism; further research is needed to replicate this finding. Future research could also explore the applicability of other risk tools (e.g., Static-99R, STABLE-2007) for those with CSEM offenses.

ORCID iDs

L. Maaik Helmus  <https://orcid.org/0000-0002-5032-2548>

Angela W. Eke  <https://orcid.org/0000-0002-4115-9662>

NOTES

1. Canadian legislation uses the terminology “child pornography” but we use CSEM (child sexual exploitation material) throughout the article as a term that directly recognizes the exploitative nature of the content.

2. Rather than have a strictly defined CSEM-only group, the current data set included 10 individuals who also had either luring or exhibitionism offenses; these individuals would be within the sampling frame appropriate to use scales such as Static-99R. However, they were retained in the CSEM/no-contact (NC) group because the primary distinction in CSEM research is whether those with CSEM offenses have also committed a contact sex offense against a child.

REFERENCES

- Allen, M. (2016). *Police-reported crime statistics in Canada, 2015* (Report No. 85-002-X). Statistics Canada, Canadian Centre for Justice Statistics. http://www.statcan.gc.ca/access_acces/alternative_alternatif.action?l=eng&loc=/pub/85-002-x/2016001/article/14642-eng.pdf
- Andrews, D. A., & Dowden, C. (2006). Risk principle of case classification for reduced recidivism: A meta-analytic investigation. *International Journal of Offender Therapy and Comparative Criminology*, *50*(1), 88–100. <https://doi.org/10.1177/0306624X05282556>
- Babchishin, K. M., Eke, A. W., Lee, S. C., Lewis, N., & Seto, M. C. (2022). Applying offending trajectory analyses to men adjudicated for child sexual exploitation material offenses. *Criminal Justice and Behavior*, *49*(8), 1095–1114. <https://doi.org/10.1177/00938548211040849>
- Babchishin, K. M., & Hanson, R. K. (2020). Monitoring changes in risk of reoffending: A prospective study of 632 men on community supervision. *Journal of Consulting and Clinical Psychology*, *88*(10), 886–898. <https://doi.org/10.1037/ccp0000601>
- Babchishin, K. M., Hanson, R. K., & Helmus, L. (2012). Even highly correlated measures can add incrementally to predicting recidivism among sex offenders. *Assessment*, *19*(4), 442–461. <https://doi.org/10.1177/1073191112458312>
- Babchishin, K. M., Hanson, R. K., & VanZuylen, H. (2015). Online child pornography offenders are different: A meta-analysis of the characteristics of online and offline sex offenders against children. *Archives of Sexual Behavior*, *44*(1), 45–66. <https://doi.org/10.1007/s10508-014-0270-x>
- Babchishin, K. M., & Helmus, L. M. (2016). The influence of base rates on correlations: An evaluation of proposed alternative effect sizes with real-world dichotomous data. *Behavior Research Methods*, *48*(3), 1021–1031. <https://doi.org/10.3758/s13428-015-0627-7>
- Babchishin, K. M., Merdian, H. L., Bartels, R. M., & Perkins, D. (2018). Child sexual exploitation material offenders: A review. *European Psychologist*, *23*(2), 130–143. <https://doi.org/10.1027/1016-9040/a000326>
- 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*, *22*(4), 443–470. <https://doi.org/10.1177/1079063210384274>
- Black, C. (2018). *Predicting recidivism among an adult male child sexual abuse imagery offender population with the Child Pornography Offender Risk Tool Short Version (CPORT-SV): A New Zealand Validation Study* [Master's thesis, University of Canterbury, UC Research Repository]. <http://doi.org/10.26021/6729>
- Bonta, J., & Andrews, D. A. (2017). *The psychology of criminal conduct* (6th ed.). Routledge. <https://doi.org/10.4324/9781315677187>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2021). *Introduction to meta-analysis* (2nd ed.). Wiley.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Cortoni, F. (2018). *Women who sexually abuse: Assessment, treatment, & management*. Safer Society Press.
- Delong, E. R., Delong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, *44*(3), 837–845. <https://doi.org/10.2307/2531595>
- Eher, R., Schilling, F., Hansmann, B., Pumberger, T., Nitschke, J., Habermeyer, E., & Mokros, A. (2016). Sadism and violent reoffending in sexual offenders. *Sexual Abuse*, *28*(1), 46–72. <https://doi.org/10.1177/1079063214566715>
- Eke, A. W., Helmus, L. M., & Seto, M. C. (2018). *Scoring guide for the Child Pornography Offender Risk Tool (CPORT): Version 2*. Unpublished manual.
- Eke, A. W., Helmus, L. M., & Seto, M. C. (2019). A validation of the Child Pornography Offender Risk Tool (CPORT). *Sexual Abuse*, *31*(4), 456–476. <https://doi.org/10.1177%2F1079063218762434>
- Elliott, I. A., Mandeville-Norden, R., Rakestrow-Dickens, J., & Beech, A. R. (2019). Reoffending rates in a U.K. community sample of individuals with convictions for indecent images of children. *Law and Human Behavior*, *43*(4), 369–382. <https://doi.org/10.1037/lhb0000328>
- Gail, M. H., & Pfeiffer, R. M. (2005). On criteria for evaluating models of absolute risk. *Biostatistics*, *6*(2), 227–239. <https://doi.org/10.1093/biostatistics/kxi005>
- Gordon, R. (2012). *Applied statistics for the social and health sciences*. Routledge.
- Gunnarsdóttir, H. O. (2019). *Risk assessment of convicted child pornography offenders in Iceland 2000–2017* [Master's thesis, Reykjavík University]. Open Access Theses and Dissertations. <https://hdl.handle.net/1946/32926>
- Hanson, R. K. (2017). Assessing the calibration of actuarial risk scales: A primer on the E/O index. *Criminal Justice and Behavior*, *44*(1), 26–39. <https://doi.org/10.1177/0093854816683956>
- Hanson, R. K., Harris, A. J. R., Letourneau, E., Helmus, L. M., & Thornton, D. (2018). Reductions in risk based on time offense-free in the community: Once a sexual offender, not always a sexual offender. *Psychology, Public Policy, and Law*, *24*(1), 48–63. <https://doi.org/10.1037/law0000135>
- 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), 1–21. <https://doi.org/10.1037/a0014421>
- Hanson, R. K., Newstrom, N., Brouillette-Alarie, S., Thornton, D., Robinson, B. B. E., & Miner, M. H. (2021). Does reassessment improve prediction? A prospective study of the Sexual Offender Treatment Intervention and Progress Scale

- (SOTIPS). *International Journal of Offender Therapy and Comparative Criminology*, 65(16), 1775–1803. <https://doi.org/10.1177/0306624X20978204>
- Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior*, 24(1), 119–136. <https://doi.org/10.1023/A:1005482921333>
- Harrell, F. E., Lee, K. L., & Mark, D. B. (1996). Tutorial in biostatistics multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, 15(4), 361–387. [https://doi.org/10.1002/\(SICI\)1097-0258\(19960229\)15:4<361::AID-SIM168>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4)
- Helmus, L. M. (2009). *Re-norming Static-99 recidivism estimates: Exploring base rate variability across sex offender samples* [Master's thesis, Carleton University]. Carleton University Research Virtual Environment. <https://repository.library.carleton.ca/concern/etds/7m01bm20j>
- Helmus, L. M., & Babchishin, K. M. (2017). Primer on risk assessment and the statistics used to evaluate its accuracy. *Criminal Justice and Behavior*, 44(1), 8–25. <https://doi.org/10.1177/0093854816678898>
- Helmus, L. M., Babchishin, K. M., & Hanson, R. K. (2013). The predictive accuracy of the Risk Matrix 2000: A meta-analysis. *Sexual Offender Treatment*, 8(2), 1–20.
- Helmus, L. M., 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, R. Eher, L. A. Craig, M. H. Miner, & F. Pfäfflin (Eds.), *International perspectives on the assessment and treatment of sexual offenders: Theory, practice, and research* (pp. 57–84). John Wiley. <http://doi.org/10.1002/9781119990420.ch4>
- Helmus, L. M., Hanson, R. K., Murrie, D. C., & Zaborauckas, C. L. (2021). Field validity of Static-99R and STABLE-2007 with 4,433 men serving sentences for sexual offences in British Columbia: New findings and meta-analysis. *Psychological Assessment*, 33(7), 581–595. <https://doi.org/10.1037/pas0001010>
- Helmus, L. M., Johnson, S., & Harris, A. J. R. (2019). Developing and validating a tool to predict placements in administrative segregation: Predictive accuracy with inmates, including indigenous and female inmates. *Psychology, Public Policy, and Law*, 25(4), 284–302. <https://doi.org/10.1037/law0000201>
- Helmus, L. M., Kelley, S. M., Frazier, A., Fernandez, Y. M., Lee, S. C., Rettenberger, M., & Boccaccini, M. T. (2022). Static-99R: Strengths, limitations, predictive accuracy meta-analysis, and legal admissibility review. *Psychology, Public Policy and Law*, 28(3), 307–331. <https://doi.org/10.1037/law0000351>
- Henshaw, M., Darjee, R., & Clough, J. A. (2020). Online child sexual offending. In I. Bryce & W. Petherick (Eds.), *Child sexual abuse* (pp. 85–108). Elsevier.
- Hermann, C., Lucenti, L., & Eke, A. W. (2019). *CPORT and CASIC: Exploring consistency in scoring between probation and parole officers*. Ontario Ministry of the Solicitor General.
- Higgins, J., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, 327, 557–560. <https://doi.org/10.1136/bmj.327.7414.557>
- Howard, P. D. (2017). The effect of sample heterogeneity and risk categorization on Area Under the Curve (AUC) predictive validity metrics. *Criminal Justice and Behavior*, 44(1), 103–120. <https://doi.org/10.1177/0093854816678899>
- Joint Committee on Standards for Educational and Psychological Testing of the American Educational Research Association, American Psychological Association and National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Kelley, S. M., Ambroziak, G., Thornton, D., & Barahal, R. M. (2020). How do professionals assess sexual recidivism risk? An updated survey of practices. *Sexual Abuse*, 32(1), 3–29. <https://doi.org/10.1177/1079063218800474>
- Lehmann, R. J. B., Thornton, D., Helmus, L. M., & Hanson, R. K. (2016). Developing non-arbitrary metrics for risk communication: Norms for the Risk Matrix 2000. *Criminal Justice and Behavior*, 43(12), 1661–1687. <https://doi.org/10.1177/0093854816651656>
- Lloyd, C. D., Hanson, R. K., Richards, D. K., & Serin, R. C. (2020). Reassessment improves prediction of criminal recidivism: A prospective study of 3,421 individuals in New Zealand. *Psychological Assessment*, 32(6), 568–581. <http://doi.org/10.1037/pas0000813>
- Lovins, B., Lowenkamp, C. T., & Latessa, E. J. (2009). Applying the risk principle to sex offenders: Can treatment make some sex offenders worse? *The Prison Journal*, 89(3), 344–357. <https://doi.org/10.1177/0032885509339509>
- Lowenkamp, C. T., & Latessa, E. J. (2005). Increasing the effectiveness of correctional programming through the risk principle: Identifying offenders for residential placement. *Criminology and Public Policy*, 4(2), 263–290. <https://doi.org/10.1111/j.1745-9133.2005.00021.x>
- Martin, M. (2021). *Computer and internet use in the United States: 2018 (ACS-49)*. United States Census Bureau. www.census.gov/content/dam/Census/library/publications/2021/acs/acs-49.pdf
- Phenix, A., Fernandez, Y., Harris, A. J. R., Helmus, M., Hanson, R. K., & Thornton, D. (2016). *Static-99R coding rules: Revised-2016* (Correction Research User Report 2017–R012). Public Safety Canada. <https://www.publicsafety.gc.ca/cnt/rsrsc/pblctns/sttc-2016/sttc-2016-en.pdf>
- Pilon, A. (2016). *The predictive validity of general and offense-specific risk assessment tools for child pornography offenders' reoffending* [Master's thesis, University of Saskatchewan]. University of Saskatchewan's Research Archive. <https://harvest.usask.ca/handle/10388/ETD-2016-01-2414>
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25, 111–163. <https://doi.org/10.2307/271063>

- 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*(5), 615–620. <https://doi.org/10.1007/s10979-005-6832-7>
- 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*(9), 856–861. [https://doi.org/10.1016/S0895-4356\(03\)00124-0](https://doi.org/10.1016/S0895-4356(03)00124-0)
- Ruscio, J. (2008). A probability-based measure of effect size: Robustness to base rates and other factors. *Psychological Methods, 13*(1), 19–30. <https://doi.org/10.1037/1082-989X.13.1.19>
- Sánchez-Meca, J., Chacón-MoscOSO, S., & Marín-Martínez, F. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods, 8*(4), 448–467. <https://doi.org/10.1037/1082-989X.8.4.448>
- Savoie, V., Quayle, E., Flynn, E., & O'Rourke, S. (2021). Predicting risk of reoffending in persons with child sexual exploitation material offense histories: The use of Child Pornography Offender Risk Tool in a Scottish population. *Sexual Abuse, 34*, 568–596. <https://doi.org/10.1177/10790632211047190>
- Schulze, R. (2007). Current methods for meta-analysis: Approaches, issues, and developments. *Zeitschrift für Psychologie/ Journal of Psychology, 215*(2), 90–103. <https://doi.org/10.1027/0044-3409.215.2.90>
- Seto, M. C. (2013). *Internet sex offenders*. American Psychological Association. <https://doi.org/10.1037/14191-000>
- Seto, M. C., & Eke, A. W. (2015). Predicting recidivism among adult male child pornography offenders: Development of the Child Pornography Offender Risk Tool (CPORT). *Law and Human Behavior, 39*(4), 416–229. <https://doi.org/10.1037/lhb0000128>
- Seto, M. C., & Eke, A. W. (2017). Correlates of admitted sexual interest in children among individuals convicted of child pornography offenses. *Law and Human Behavior, 41*(3), 305–313. <https://doi.org/10.1037/lhb0000240>
- Seto, M. C., Hanson, R. K., & Babchishin, K. M. (2011). Contact sexual offending by men with online sexual offenses. *Sexual Abuse, 23*(1), 124–145. <https://doi.org/10.1177/1079063210369013>
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. Oxford University Press.
- Soldino, V., Carbonell-Vayá, E. J., & Seigfried-Spellar, K. C. (2021). Spanish validation of the Child Pornography Offender Risk Tool. *Sexual Abuse, 33*(5), 503–528. <https://doi.org/10.1177/1079063220928958>
- Statistics Canada. (2021). *Police-reported crime for selected offences, Canada, 2020*. <https://www150.statcan.gc.ca/n1/daily-quotidien/210727/t001a-eng.htm>
- Thornton, D., Fernandez, Y. M., & Helmus, L. M. (2023). *Scoring guide for Risk Matrix 2000 S & V Scales: International version*. <https://saarna.org/>
- 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.), *Sexually coercive behavior: Understanding and management* (pp. 225–235). New York Academy of Sciences.
- Viallon, V., Ragusa, S., Clavel-Chapelon, F., & Bénichou, J. (2009). How to evaluate the calibration of a disease risk prediction tool. *Statistics in Medicine, 28*(6), 901–916. <https://doi.org/10.1002/sim.3517>
- Volinsky, C. T., & Raftery, A. E. (2000). Bayesian information criterion for censored survival models. *Biometrics, 56*(1), 256–262. <https://doi.org/10.1111/j.0006-341X.2000.00256.x>
- 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, 23*(1), 146–168. <https://doi.org/10.1177%2F1079063210375974>

L. Maaïke Helmus, PhD, is an assistant professor in the criminology department at Simon Fraser University, and an adjunct professor at Carleton University. She is also the vice-president (research) for the Society for the Advancement of Actuarial Risk Needs Assessment (SAARNA). She has presented and published extensively on developing and validating risk assessment tools, primarily for sexual recidivism, but also for intimate partner violence recidivism, and for subgroups such as different racial/ethnic groups.

Angela W. Eke, PhD, is the research coordinator for the Criminal Behaviour Analysis Section of the Ontario Provincial Police. Eke's areas of research include intimate partner violence and sexual offending, with a focus on risk assessment. Eke provides training and consultations for police and other criminal justice stakeholders and is an adjunct research professor at Carleton University. In 2017, Eke was appointed a Member of the Order of Merit of the Police Forces (Canada).

Linda Farmus is a PhD candidate in the quantitative methods program at York University. Her research focuses on equivalence testing.

Michael C. Seto, PhD, is a clinical and forensic psychologist currently serving as the forensic research director for the Royal Ottawa Health Care Group. He is also a full professor in psychiatry at the University of Ottawa, with cross-appointments at the University of Toronto and Carleton University. Michael has presented and published extensively on pedophilia, sexual offending against children, and online sexual offending.