What Risk Assessment Tools can be Used With Men Convicted of Child Sexual Exploitation Material (CSEM) Offenses? Recommendations From a Review of Current Research

Online Supplemental Materials

Table of Contents

Online Supplement A
Daubert Considerations for CSEM Risk Assessment Tools

This document supplements the following review article:

Helmus, L. M., Eke, A., W., & Seto, M. C. (in press). What risk assessment tools can be used with men convicted of Child Sexual Exploitation Material (CSEM) offenses? Recommendations from a review of current research. *Law and Human Behavior*.

Risk assessment tools are used in many different contexts (Helmus et al., 2022), not all of which will go before a judge. However, in any context, a case management or policy decision could be subject to a legal challenge debating the appropriateness of the risk assessment (e.g., tool selection, interpretation of tool, its link to policy/practice). Jurisdictions and settings may differ in evidentiary rules that are relevant to considering the risk tool.

In the United States, there are two common admissibility standards in court: *Frye v. United States* (1923) or *Daubert v. Merrell Dow Pharmaceuticals* (1993). The Frye test emphasizes scientific consensus to determine admissibility, often referred to as the "general acceptance test." The more nuanced *Daubert* standard includes general acceptance but also that the scientific method be tested, subjected to peer review, have a known error rate, and standards for use (Hilbert, 2019). Importantly, none of these Daubert standards reflect absolute thresholds, nor is any one particular standard necessary for admissibility. These standards are flexible and depend on the case, requiring judicial interpretation. For example, a "known" error rate does not mean a low error rate. Much of this judicial interpretation will involve considering the context of the decision at hand and the availability of alternatives. A tool with limited evidence demonstrating modest performance may be admissible if there is limited research and options for that type of task, but not if there is an established evidence base for alternatives. Additionally, lack of peer review may not preclude admissibility if the method is new and offers clear advantages over alternatives. However, lack of peer review may lead to a determination of inadmissibility if there are similar peer-reviewed options available. Note that many of the nuances in the Daubert standards (e.g., the absolute value of the known error rate) would go more to the weight given to the evidence as determined through cross-examination and presentation of contrary evidence, rather than its admissibility (*Daubert v. Merrell Dow Pharmaceuticals*, 1993).

Below, we summarize considerations for the Daubert criteria for the Child Sexual Exploitation Material (CSEM) risk tools reviewed. This is not meant to offer an ultimate opinion on Daubert admissibility as that would require a legal analysis specific to the context of a particular decision, and admissibility decisions may change as further evidence becomes available, not only for the risk tool in question, but for alternative tools. References cited in the table below are available in the full text of the article.

Some Technical Notes

- The information about known error rate below is referring to research on men charged/convicted of CSEM offenses, unless specified otherwise.
- AUC stands for the Area Under the Curve from Receiver Operating Characteristic Curve Analysis. Following Rice and Harris (2005), AUC values of .56, .64 and .71 denote small, moderate, and large effects, respectively. Harrell's C (referred to as C

below) is an analogous effect size taking into account varying lengths of follow-up, and effect sizes can be interpreted similarly as AUCs (see Helmus & Babchishin, 2017 for further discussion of interpreting these statistics).

-

Tool and Daubert Consideration	**Comments**
Child Pornography Offender Risk Tool (CPORT)	
Whether the technique in question can be and has been tested.	Yes.
Whether it has been subject to peer review and publication.	Yes. The development of the CPORT was published (Seto & Eke, 2015), as was a recent meta-analysis (Helmus et al., 2024), and the meta-analysis was updated in the current paper.
Its known or potential error rate.	For sexual recidivism, from the updated meta-analysis in the full text of this paper, it has a large AUC (.75) without Cohen (2023), which has a significantly lower effect size than all other studies combined; moderate to large AUC (.68 to .71) with Cohen (2023); moderate AUC for CSEM recidivism (AUC = .68). See the full paper for discussion of important differences in Cohen's (2023) application of the CPORT.
The existence and maintenance of standards controlling its operation.	Yes, there is a coding manual available (Eke et al., 2018) and training is available. Tool authors are available to answer questions via email.
Whether it has attracted widespread acceptance within a relevant scientific community.	All risk tools for CSEM populations are sufficiently new that it is difficult to determine "widespread" acceptance but CPORT is the most researched risk assessment tool for CSEM populations and would generally be considered the option with the most research evidence available. The CPORT is the only tool specifically designed to predict sexual recidivism among men convicted of CSEM offenses. We are not aware of published surveys of risk tool use for this population yet, but anecdotally, we are aware of its widespread use. It has currently been translated into Danish, Dutch, French, German, Norwegian, Portuguese, and Spanish.

Risk Matrix 2000/Sex	
Whether the technique in question can be and has been tested.	Yes.
Whether it has been subject to peer review and publication.	The development of the tool was not published in a peer-reviewed journal (Thornton et al., 2003), but a meta-analysis of it has been (Helmus et al., 2013), as well as the normative data (Lehmann et al., 2016).
Its known or potential error rate.	A recent meta-analysis of two studies found moderate accuracy in predicting sexual recidivism (AUC = .66), and moderate accuracy for predicting CSEM recidivism in a single sample (AUC = .67; Helmus et al., 2024).
The existence and maintenance of standards controlling its operation.	Risk Matrix 2000/S has a coding manual (Thornton et al., 2023) and training is available through www.saarna.org
Whether it has attracted widespread acceptance within a relevant scientific community.	Risk Matrix 2000/S has widespread acceptance for individuals with convictions for offline sexual offenses; the research on its application with CSEM populations is sufficiently new that it is difficult to gauge widespread acceptance.
OASys Sexual Reoffending Predictor – Indecent Image (OSP/I)	
Whether the technique in question can be and has been tested.	Yes.
Whether it has been subject to peer review and publication.	It is not published in peer reviewed journals, but the United Kingdom Ministry of Justice pays for external peer review for selected research reports, including research reports on the OSP/I.
Its known or potential error rate.	Difficult to determine given that the government reports (Howard & Wakeling, 2021; Craik et al., 2024) do not present effect size statistics for a subgroup of individuals with CSEM convictions. However, from the data presented in the reports, our main paper calculated small to moderate effect sizes (AUCs between .59 and .67).
The existence and maintenance of standards controlling its operation.	Available through the UK Ministry of Justice (HMPPS, 2023).
Whether it has attracted widespread acceptance within a relevant scientific community.	It is approved for use in prison and probation services in England and Wales, but we are not aware of acceptance among the scientific community beyond that. Its intent is more for

	triaging community supervision resources among the population of men convicted of sexual offenses.
Static-99R	
Whether the technique in question can be and has been tested.	Yes.
Whether it has been subject to peer review and publication.	Static-99R has a considerable published evidence base for predominantly offline sex offense samples (see Helmus et al., 2022 for review), but validations of it for CSEM samples are not yet published in peer reviewed journals, although we anticipate one study will be published soon (Eke et al., 2024).
Its known or potential error rate.	Small AUC for sexual (.63) and CSEM (.62) recidivism in two combined samples (Eke et al., 2024).
The existence and maintenance of standards controlling its operation.	Static-99R has a coding manual (Phenix et al., 2016) and training and implementation resources are available through www.saarna.org, with the possibility of submitting questions via email. As discussed in the full paper, the tool would likely benefit from some modifications to coding rules if the developers expand the appropriate populations for it to include men whose only sexual offense charges are for CSEM.
Whether it has attracted widespread acceptance within a relevant scientific community.	Static-99R has widespread acceptance within the scientific community for the broader population of men charged or convicted of sexual offenses (see Helmus et al., 2022 for detailed review), but does not yet have widespread acceptance for CSEM populations, unless they also have a Category A sex offense, as defined by the coding manual.
STABLE-2007	
Whether the technique in question can be and has been tested.	Yes
Whether it has been subject to peer review and publication.	The overall scale has been (Hanson et al., 2015), as well as a validation with a CSEM sample (Babchishin et al., 2023).
Its known or potential error rate.	Large effect size (Harrell's C ≥.78) for any sex or CSEM recidivism in one sample

	(Babchishin et al., 2023).
The existence and maintenance of standards controlling its operation.	Yes. Coding manual is available (Fernandez et al., 2014), as well as guidance and normative data to combine the STABLE-2007 with a static risk tool such as the Risk Matrix 2000 (Brankley et al., 2017). Resources are available through www.saarna.org
Whether it has attracted widespread acceptance within a relevant scientific community.	STABLE-2007 has widespread acceptance for men convicted of sexual offenses (Brankley et al., 2021; Kelley et al., 2020) but its validation with CSEM samples is sufficiently new that it would not yet have widespread acceptance.
ACUTE-2007	
Whether the technique in question can be and has been tested.	Yes
Whether it has been subject to peer review and publication.	The overall scale has been (Hanson et al., 2015), as well as a validation with a CSEM sample (Babchishin et al., 2023).
Its known or potential error rate.	Large C (.71+) for any sex or CSEM recidivism in one sample (Babchishin et al., 2023).
The existence and maintenance of standards controlling its operation.	Yes. Coding manual is available (Fernandez et al., 2015), as well as guidance and normative data to combine with STABLE-2007 (Brankley et al., 2019) Resources are available through www.saarna.org
Whether it has attracted widespread acceptance within a relevant scientific community.	ACUTE-2007 has research support but its validation with CSEM samples is sufficiently new that it would not yet have widespread acceptance.
Post Conviction Risk Assessment (PCRA)	
Whether the technique in question can be and has been tested.	Yes
Whether it has been subject to peer review and publication.	The tool has been subject to peer review for general use (Lowenkamp et al., 2013), but the research studies of it with CSEM samples are not in peer-reviewed publications.
Its known or potential error rate.	Small (AUC = .61) for any sexual recidivism for the tool used without overrides (Cohen, 2023). However, as currently implemented, staff can and frequently do employ overrides, and one large study of individuals with sex offense convictions (not all CSEM) found that

	the overridden results demonstrated only chance accuracy in predicting sexual recidivism (AUC = .51; Cohen et al., 2020).
The existence and maintenance of standards controlling its operation.	There are standards available for its use from https://bja.ojp.gov/program/psrac/guidelines-post-conviction-rna
Whether it has attracted widespread acceptance within a relevant scientific community.	PCRA has widespread acceptance within the scientific community for its use in correctional systems to predict general recidivism among all individuals convicted of an offense. It does not yet have widespread acceptance for CSEM populations.
Level of Service Inventory – Ontario Revision (LSI-OR)	
Whether the technique in question can be and has been tested.	Yes
Whether it has been subject to peer review and publication.	The original development of this version of the LSI was not peer-reviewed, but subsequent research validations of it have been (for meta-analysis, see Olver et al., 2014). Its validation with a CSEM sample was in a dissertation (Pilon, 2016) that to our knowledge has not yet been published in a peer-reviewed journal, although dissertations do go through a peer review process to be successfully defended.
Its known or potential error rate.	Predictive accuracy was small for general recidivism (AUC = .62 to .63); it did not predict CSEM or sexual recidivism (AUCs below .50; Pilon, 2016). Note that in general, accuracy declines for the tool when overrides are used (Pilon, 2016; Wormith et al., 2012), particularly for individuals with sex offenses.
The existence and maintenance of standards controlling its operation.	There is a system of training and standards for the most recent version of the LSI scale (the Level of Service/Case Management Inventory) through Multi-Health Systems (https://storefront.mhs.com/collections/ls-cmi). We are not aware of what resources, training, and standards are available for this particular version.
Whether it has attracted widespread acceptance within a relevant scientific community.	The family of Level of Service Inventory tools (there are several variations) is probably the most widely accepted in the scientific community for the prediction of general

	recidivism among all individuals convicted of an offense (e.g., Olver et al., 2014). It does not yet have widespread acceptance for CSEM populations.
Offender Group Reconviction Scale – Version 3 (OGRS3)	
Whether the technique in question can be and has been tested.	Yes.
Whether it has been subject to peer review and publication.	The development of the original scale (Howard et al., 2009) is not published in peer reviewed journals, but the United Kingdom Ministry of Justice pays for external peer review for selected research reports. The validation of the OGRS3 with a CSEM sample was published in a peer-reviewed journal (Wakeling, Howard, & Barnett, 2011).
Its known or potential error rate.	Moderate for sexual recidivism (AUC = .65) and general recidivism (AUC = .70; Wakeling, Howard, & Barnett, 2011)
The existence and maintenance of standards controlling its operation.	Available through the UK Ministry of Justice.
Whether it has attracted widespread acceptance within a relevant scientific community.	It is approved for use in prison and probation services in England and Wales for the prediction of general recidivism, but we are not aware of acceptance among the scientific community beyond that or for its use in predicting sexual recidivism.

Online Supplement B
Additional Relevant Information on Some of the Risk Tools Reviewed

This document supplements the following review article:

Helmus, L. M., Eke, A., W., & Seto, M. C. (in press). What risk assessment tools can be used with men convicted of Child Sexual Exploitation Material (CSEM) offenses? Recommendations from a review of current research. *Law and Human Behavior*.

There may be repetition with content in this Supplement and in the review article, but this version has additional information about some of the risk tools reviewed.

CPORT

It has seven items, with one point for each item that applies: age at the time of the index investigation (35 or younger), any prior criminal history, any failure on conditional release, any contact sexual offending, 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). Table 1 summarizes studies on the predictive accuracy of the CPORT that were used in the meta-analysis reported in text.

Risk Matrix 2000 – Sex (RM2000/S)

The RM2000/S includes seven static risk factors assessed across two steps of coding: age, sexual crime court appearances, general crime court appearances, male victim, stranger victim, never lived with a lover for two years, and noncontact sex offense. After scoring, the individual is placed in one of four risk levels: below average risk, average risk, above average risk, or well above average risk (Thornton et al., 2023).

Thornton and colleagues tended to revise the coding manual regularly to respond to frequently asked questions, so it is not easy to identify exactly when some revisions to the coding rules would have been applied during field scoring in the United Kingdom (personal communication with D. Thornton, June 13, 2024). Much of the current coding rules for those with CSEM offenses was developed in coding manual versions dating back at least as early as 2007 (personal communication with D. Thornton, June 13, 2024), although further elaborations for the male victim item was added in 2023 (Thornton et al., 2023). This additional guidance has yet to be applied in new validation studies of the RM2000/S, although it was written around 2023 consultations with Angela Eke based on her extensive research and coding experience in this content area, including the operationalization of this item used in Helmus et al.'s (2024) validation study.

OASys Sexual Reoffending Predictor – Indecent Image (OSP/I)

In the OSP/I as originally developed (Howard & Wakeling, 2021), there are four risk levels: low risk (for individuals with no CSEM offenses and without multiple sanctions for contact sex offenses against children); moderate risk (no CSEM offenses but multiple sanctions for contact sex offenses against children); high risk (one CSEM sanction); and very high risk (multiple sanctions for CSEM offenses). For operational reasons, the OSP/I was revised to collapse the first two levels into a single low risk level (reflecting all individuals without a CSEM offense). In other words, the OSP/I distinguishes between those with no CSEM sanctions,

one CSEM sanction, and multiple CSEM sanctions.

To develop the OSP/I, Howard and Wakeling (2021) analyzed data for 2,728 men discharged from prison between 2003 and 2008 who had a current or prior conviction for a sexual offense, of whom 593 had a conviction for a CSEM offense. In a previous version of the OSP/I with four risk levels (see Online Supplement B), it did not significantly predict new contact sexual offenses during the average 4.5-year follow-up (Harrell's *C* = .49; note that Harrell's *C* is analogous to AUCs with survival data with varying follow-ups and can be interpreted similarly; Helmus & Babchishin, 2017) but did significantly predict new CSEM convictions (*C* = .74) in the entire sample. There were only 47 cases in the top risk category, where survival analyses estimated that roughly one third (32%) would have a new CSEM conviction in 5 years, compared to 9%, 3% and 1%, in descending order of risk category.

Howard and Wakeling (2021) did not examine how the three-level OSP/I that has been adopted in practice predicted recidivism, nor did they examine how well the tool predicted recidivism for individuals who had a CSEM offense. Using the information they presented in Table 3 of their report on the number of individuals in each risk category and the estimated 5-year CSEM recidivism rates from survival analysis, it was possible to create a dataset to calculate AUCs. Note that this will not match their dataset perfectly as they used Harrell's *C* analyses to incorporate the varying follow-up, whereas this approximation is based on AUCs from estimated fixed follow-ups of 5 years. But it allows us to gauge changes in predictive accuracy for the three-level OSP/I adopted in practice, and for the subgroup of individuals with CSEM convictions. Whereas their analysis of the full sample and four-level OSP/I had a Harrell's *C* of .74 for CSEM recidivism, the AUC from the data provided in Table 3 indicated an AUC of .76 (95% CI [.71, .82]) for the full sample. Combining the two lowest levels to create the three-level version, the AUC showed a small drop (.74, 95% CI [.68, .80]). When examining the group of 593 individuals who had a CSEM conviction, the AUC was lower but statistically significant (AUC = .59, 95% CI [.51,.67]).

Static-99R

Static-99R (Hanson & Thornton, 2000; Helmus, Thornton, et al., 2012) has 10 static risk factors assessing demographic, criminal history, and victim demographic information. Total scores can range from -3 to 12, placing the individual into one of five risk levels: I - very low (-3 to -2)–, II - below average (-1 to 0)– III - average (1 to 3)– IVa - above average (4 to 5) and– IVb - well above average (6+; Hanson et al., 2017).

Category A sex offenses include the most common contact and noncontact offenses against direct and identifiable victims (Phenix et al., 2016), including sexual assault of adults or children, exhibitionism, or voyeurism. Excluded are some offenses without a sexual motivation (e.g., public urination), without a victim (e.g., prostitution involving consenting adults), and some offenses that were sufficiently distinct from the original development research for the authors to sanction its use (e.g., CSEM offenses, with the exception of creating CSEM with a child, which is a Category A offense because it involves interaction with an identifiable victim). Static-99R can therefore be used with individuals with a CSEM offense only if they also have a charge or conviction for a Category A sex offense in their record.

Although Static-99R has many risk factors similar to the RM2000/S, the biggest hindrance in validation research with CSEM populations is that the current coding rules do not provide useful guidance for assessing individuals whose only sexual offenses involve CSEM. For example, three of the items pertain to victims of direct sexual offenses (any unrelated victim,

any stranger victim, and any male victim), and one item gives a point for a noncontact sexual offense conviction, which could include a CSEM conviction. For the victim items, the coding rules specify that children depicted in CSEM are not counted. This makes sense for the unrelated/stranger victim items, which conceptually are not applicable to CSEM behavior, although future research could identify some risk relevance in rare cases where the CSEM includes children known to the individual. Excluding CSEM images makes less sense for the male victim item because a preference for boy content in CSEM is a valid risk factor (Seto & Eke, 2015). In the context of CSEM images where it is easy to download many images at once, a single image depicting a male is not a good indicator of this risk factor, which generally reflects atypical sexual interests.

The risk factor for noncontact sexual offense convictions is more problematic if applied to CSEM populations. CSEM offenses are counted as noncontact, warranting a point on this item. However, this is uninformative and misleading if applied to individuals whose only sex offenses are for CSEM. This item is based on the development studies (Hanson & Thornton, 2000) and on a previous meta-analysis indicating that individuals with noncontact sex offenses were more likely to sexually reoffend than those with contact sex offenses (Hanson & Bussière, 1998). However, these samples predated the widespread availability of the internet, so the noncontact sexual offenses in this research would not include internet offending. The vast majority of noncontact offenses would be related to activities such as exhibitionism, voyeurism, indecent phone calls, or stealing items for sexual gratification. There may have been a few CSEM offenses in this research, but it would have been committed offline (e.g., purchased or traded with others, involving physical items such as photographs, videos, books, or magazines). Regardless, the research behind this noncontact offending item explicitly excludes online CSEM offenses. CSEM offenses are included in the current coding rules (Phenix et al., 2016) only for individuals who have a Category A sex offense in addition to their CSEM offense. Applying this tool to CSEM populations, it would not make sense to give this point to all individuals (following a strict interpretation of the coding manual) for two reasons. Firstly, it cannot serve as a risk factor for CSEM individuals if there is no variability (i.e., they all have the point). Secondly, it increases the risk of all individuals convicted of CSEM offenses relative to those with other types of sex offenses, which is inconsistent with existing research on sexual recidivism among these groups (for a review, see Helmus, 2023).

STABLE-2007

The STABLE-2007 (Fernandez et al., 2014; Hanson et al., 2007) total scores range from 0 to 26 for offenders with a victim under 14 years old, and 0 to 24 for others. Total scores of 0-3, 4-11, and 12+ reflect low, moderate, and high levels of dynamic need.

The STABLE-2007 was developed by revising the STABLE-2000 based on a prospective field validity study across Canada (Hanson et al., 2007). One early study on the STABLE-2000 examined 73 men with internet sexual offenses and found preliminary evidence that the STABLE-2000 predicted probation failure, which included new internet offenses in two cases (Webb et al., 2007).

ACUTE-2007

The ACUTE-2007 (Fernandez et al., 2015; Hanson et al., 2007) is a dynamic risk assessment tool for adult males convicted of sexual offenses. It assesses risk factors that can change more rapidly than the STABLE-2007 items. The seven items are assessed on a scale from

0-3 and include victim access, hostility, sexual preoccupation, rejection of supervision, emotional collapse, change in social support, and substance use. Total scores can therefore range from 0 to 21. ACUTE-2007 is intended to be scored at each contact in the community (e.g., weekly, monthly). The items are considered to represent high risk situations (access to victims) or the current expressions of stable risk-relevant factors (e.g., sexual preoccupation) (Fernandez et al., 2015). The ACUTE-2007 predicts sexual, violent, and any recidivism among individuals convicted of sexual offenses (Hanson et al., 2007; Nitsche et al., 2022). Change in ACUTE-2007 total scores are associated with changes in the likelihood of sexual, violent, and general recidivism (Babchishin & Hanson, 2020; Lee et al., 2023). There are no recidivism estimates for ACUTE-2007 scores because it is meant to guide treatment and supervision efforts and to identify periods of increased or decreased risk. Formal interrater reliability analyses are not available, but a verification exercise in the development sample suggested good agreement for ACUTE-2007 items (Hanson et al., 2007). Training and resources for the STABLE-2007 are available from www.saarna.org.

Online Supplement B, Table 1

Studies of CPORT for Meta-Analysis

	Location	Sampling Timeframe	CPORT *M*	*SD*	Recidivism Criteria	Follow-up (Years)	*n* recid / Total	Recid rate (%)	AUC	[95% CI]
CPORT – Any Sexual Recidivism										
Black (2018)	New Zealand	1998-2014	1.27	1.24	Conviction	7.6	?/547	-	.77	[.71, .82]
Cohen (2023)	U.S.	2010-2016	1.44	1.13	Arrest	5.0	262/5,768	4.5	.62	[.58, .65]
Eke et al. (2024)	Canada (QC)	2010-2021	2.15	1.36	Charge	5.1	9/136	6.6	.70	[.55, .84]
Gunnarsdóttir (2019)[a]	Iceland	2000-2014	1.79	1.15	Conviction	5.0	?/106	-	.75	[.62, .89]
Helmus et al. (2024)[c]	Canada (ON)	1993-2010	1.98	1.57	Charge	5.0	40/339	11.8	.73	[.64, .82]
Pilon (2016)	Canada (ON)	2010-2011	-	-	Conviction	3.2	8/279	2.9	.56	[.32, .79]
Savoie (2021)	Scotland	2010-2013	1.91	1.29	Conviction	5.0	14/140[b]	10.0	.77	[.67, .87]
Seiser et al. (2024)	Austria	2002-2016	2.87	1.52	Conviction	8.5	24/128	18.8	.72	[.62, .82]
CPORT – CSEM Recidivism										
Black (2018)	New Zealand	1998-2014	1.27	1.24	Conviction	7.6	71/547	13.0	.77	[.71, .82]
Eke et al. (2024)	Canada (QC)	2010-2021	2.15	1.36	Charge	5.1	8/136	5.9	.67	[.51, .82]
Gunnarsdóttir (2019)[d]	Iceland	2000-2014	1.79	1.15	Conviction	5.0	12/106	11.3	.62	[.53, .70]
Helmus et al. (2024)	Canada (ON)	1993-2010	1.98	1.57	Charge	8.1	29/339	8.6	.74	[.64, .85]
Pilon (2016)	Canada (ON)	2010-2011	-	-	Conviction	3.2	7/279	2.5	.52	[.27, .77]
Savoie (2021)	Scotland	2010-2013	1.91	1.29	Conviction	5.0	11/140	7.9	.73	[.61, .85]
Seiser et al. (2024)	Austria	2002-2016	2.87	1.52	Conviction	8.5	24/128	18.8	.72	[.62, .82]
Soldino et al. (2021)	Spain	2009-2013	0.8	0.93	Arrest	5.0	6/304	2.0	.56	[.51, .62]

Note. CPORT = Child Pornography Offender Risk Tool; CSEM = child sexual exploitation materials. ON = Ontario. QC = Quebec. This table includes information from the meta-analysis presented by Helmus et al. (2024), with the addition of Cohen (2023), Eke et al. (2024), and Seiser et al. (2024). 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 dataset used charges as the recidivism outcome and Soldino et al. (2020) used arrests; the remaining studies used convictions.

[a] This effect size was not in the dissertation but was obtained by personal communication (H. Gunnarsdóttir, personal communication, December 16, 2021).

[b] 2 of the 14 recidivism incidents were for technical breaches of their sexual offense supervision order, and not necessarily for committing a new sexual offense.

[c] This study subsumes the samples from Eke et al., 2019, and Seto & Eke, 2015.

[d] Log 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).