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

L. Maaike Helmus[1], Angela W. Eke[2], and Michael C. Seto[3]

[1] Department of Criminology, Simon Fraser University

[2] Criminal Behaviour Analysis Section, Ontario Provincial Police, Orillia, Ontario, Canada

[3] Institute of Mental Health Research, Royal Ottawa Health Care Group, Ottawa, Ontario, Canada

Manuscript submitted for publication, January 2024

Author Note

L. Maaike Helmus https://orcid.org/0000-0002-5032-2548

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

Michael C. Seto https://orcid.org/0000-0002-1212-3388

This work was completed on the traditional and unceded territories of the Coast Salish Peoples (Simon Fraser University), specifically the Squamish (Sḵwx̱wú7mesh Úxwumixw), Tsleil-Waututh (səl̓ilw̓ətaʔɬ) 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 W. Eke, Michael C. Seto, and L. Maaike Helmus are co-authors of the CPORT coding manual. They do not receive royalties for use of the CPORT, although Drs. Seto and Eke occasionally receive renumeration for trainings or consultations on the CPORT, they do not participate in court proceedings. L. Maaike Helmus is part of the development team for the Static-99R, STABLE-2007, ACUTE-2007, and Risk Matrix 2000, but does not receive royalties for the use of these tools. Dr. Helmus is also on the Board of Directors for Society for the Advancement of Actuarial Risk Needs Assessment (SAARNA), a Canadian nonprofit organization that promotes research, training, and implementation for actuarial risk tools. She occasionally receives renumeration for trainings and consultations, including on court cases.

Correspondence concerning this article should be addressed to L. Maaike Helmus, 10322 Saywell Hall, Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada, V5A 1S6. Email: maaike_helmus@sfu.ca

Abstract

Objective: Review current research on risk assessment tools with individuals convicted of child sexual exploitation materials (CSEM) offenses with recommendations for use in forensic, correctional, and legal settings. **Hypotheses**: Multiple tools would be defensible to use with individuals convicted of CSEM offenses. **Methods:** We discuss a minimum threshold of predictive accuracy to justify using a risk tool as an improvement on the typical level of accuracy expected from unstructured professional judgment. Then beyond this minimum threshold, we offer additional considerations that researchers and practitioners can use in evaluating and selecting risk tools. **Results:** We identified eight risk assessment tools with predictive accuracy research on individuals convicted of CSEM offenses: the Child Pornography Offender Risk Tool (CPORT), Risk Matrix 2000/Sex (RM2000/S), OASys Sexual Reoffending Predictor – Indecent Images (OSP/I), Static-99R, STABLE-2007, ACUTE-2007, Post Conviction Risk Assessment (PCRA), and the Level of Service Inventory – Ontario Revision (LSI-OR). We review each using the evaluation considerations. **Conclusions**: The CPORT, RM2000/S, STABLE-2007, and ACUTE-2007 (in conjunction with the STABLE) are all defensible tools to use for assessing risk of any sexual recidivism or CSEM recidivism specifically. There is preliminary evidence suggesting some support for Static-99R, but it may not be the ideal choice. The OSP/I consists of a single risk factor and considers risk of CSEM recidivism among all individuals convicted of sexual offenses, not only among individuals convicted of CSEM offenses. The PCRA and LSI-OR general recidivism risk tools have some empirical support in predicting general recidivism among CSEM samples (and sexual recidivism for the PCRA), with limitations noted. The use of multiple tools may have value in assessing risk and structuring management in CSEM cases, however how they are best combined for these samples is still unclear. We expect research in this area to continue to build rapidly.

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

Public Significance Statements:

Research on risk assessment with individuals convicted of Child Sexual Exploitation Materials (CSEM) offenses is advancing rapidly, with studies examining the predictive accuracy of risk tools specifically designed for individuals with CSEM offences, as well as tools designed for people convicted of any sexual offense, or any general offense. The CPORT, Risk Matrix 2000/Sex, STABLE-2007, and ACUTE-2007 (used along with the STABLE-2007) are all defensible to use for risk assessment with CSEM populations, with some support noted for Static-99R, OSP/I, the PCRA, and LSI-OR under some circumstances.

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

Evidence-based risk assessment tools should inform decisions related to managing the risk of individuals who have been convicted of an offense (Bonta & Andrews, 2023). One-size-fits-all approaches are rarely justified, economical, or efficacious. Higher risk individuals need higher dosages of treatment and supervision in order to reduce their likelihood of recidivism (Hanson et al., 2017); conversely, however, over-managing lower risk individuals is inefficient and can potentially worsen their outcome (Andrews & Dowden, 2006; Lovins et al., 2009; Lowenkamp & Latessa, 2005). Using a structured, empirically validated risk assessment tool is best practice (Bonta & Andrews, 2023), but it is not always obvious which tool is most suitable, or whether a tool is applicable to the individual in question.

Individuals convicted of child sexual exploitation material (CSEM)[1] offenses have become an increasingly sizeable group of individuals convicted of sexual offenses. The internet has provided more opportunities for accessing CSEM materials (Gannon et al., 2023), and CSEM charges have been increasing (e.g., Statistics Canada, 2021). Research has identified meaningful differences between those with CSEM offenses compared those with offline sexual offenses (for reviews see Babchishin et al., 2015, 2018; Henshaw et al., 2020). This raises important questions about how to conduct risk assessments with this population, such as whether existing sexual recidivism risk assessment tools will be applicable or whether specialized risk tools are needed.

The Child Pornography Offender Risk Tool (CPORT; Seto & Eke, 2015) was specifically designed to assess risk of sexual recidivism among men convicted of CSEM offenses, and it has

[1] Terminology varies across jurisdictions. Child pornography is the legal term in Canada and the United States, although Child Sexual Exploitation Materials (CSEM) is a preferred term in the empirical literature.

received the most attention in the research literature (Brown, 2022; Helmus et al., 2024). Recently, Scurich and Krauss (2023) reviewed available research on the CPORT, noting limitations such as small sample sizes of recidivists, considerable missing data, and concern around the applicability of the tool with rapidly changing technology in how CSEM offenses are committed. They concluded that "the existing research base is insufficient to justify the use of the CPORT in forensic settings" (p. 505). Although Scurich and Krauss (2023) note some valid limitations in the current literature, we disagree with their conclusions that the forensic use of the CPORT is not justified based on existing evidence. Their review does raise important questions about what constitutes sufficient evidence to justify using a risk tool for someone with a CSEM offense conviction.

The purpose of this review is to summarize current validation research on risk assessment tools with individuals convicted of CSEM offenses and offer our recommendations of which tools are fit to use in forensic, correctional, and legal settings. This review will hopefully support users in their consideration of which tools to use in which context, although we note that research is advancing quickly in this field and evaluators need to stay abreast of updates. Although we focus on predictive validity, this review was also written in a similar spirit as the articles in a recent special issue of the *Journal of Personality Assessment* focusing on psychometric properties and legal admissibility of a variety of psychological measures used in forensic evaluations (Neal et al., 2022).

We examine research on the following risk tools, which are described more fully below: The CPORT, Risk Matrix 2000/Sex (RM2000/S), OASys Sexual Reoffending Predictor – Indecent Image (OSP/I), Static-99R, STABLE-2007, ACUTE-2007, Post Conviction Risk Assessment (PCRA), and the Level of Service Inventory – Ontario Revision (LSI-OR). These are all the tools that to our knowledge have empirical research on their predictive validity with

men convicted of CSEM offenses, with one exception: We did not discuss one study examining the Offender Group Reconviction Scale 3 (Wakeling, Howard, & Barnett, 2011) as this tool was developed by His Majesty's Prison and Probation Service, but our understanding is that it is no longer in use. Before evaluating specific research on each of these tools, we will address broader issues about the type of evidence needed to justify using a risk tool, with particular attention to points that have been raised for risk assessment for CSEM populations.

This review does not discuss tools that have not been validated for predictive accuracy with CSEM samples. For example, there are several generic sexual recidivism risk tools that have been validated for men convicted of sexual offenses (for a summary, see Hanson & Morton-Bourgon, 2009) which could be informative for CSEM populations but do not yet have specific validation research (e.g., Sexual Violence Risk-20 Version 2 [SVR-20 V2; Boer et al., 2017], Risk for Sexual Violence Protocol Version 2 [RSVP V2; Hart et al., 2022], or the Violence Risk Scale: Sexual Offense version [VRS-SO; Olver et al., 2007]). Additionally, the Kent Internet Risk Assessment Tool – Version 2 (KIRAT) was designed to identify suspects with CSEM offenses who were most likely to have already committed a contact sexual offense, to prioritize investigation and risk management resources (Long et al., 2016). It was not designed to predict future reoffending, nor has it been validated for that purpose. The Estimated Risk for Internet Child Sexual Offending (ERICSO) is a newer measure that was developed to assess the risk of sexual recidivism among men convicted of CSEM offenses (Garrington et al., 2022). This measure was largely developed based on hypothesized factors relevant to individuals with CSEM convictions, informed by literature searches and surveys of professionals. The current version includes 46 items, but it is often unclear which responses to the items are meant to be indicative of higher risk. It appears more as a framework for data collection, whereby an evaluator can then form an opinion of the individual's risk to reoffend; the tool has yet to be validated.

What Evidence is Needed to Justify Using a Risk Tool?

There are no universally agreed upon thresholds of evidence needed to use a risk assessment tool, but some considerations and guidance are available. An important caveat in forensic, correctional, and legal contexts is that decisions must be made, even without evidence. It is not possible to postpone decisions for someone convicted of a CSEM offense until the research base has advanced to our satisfaction. Research takes time. Long-term follow-up data necessarily means that by the time the research is available, the sample was released a long time ago. We must make the best decisions we can with the research currently available. Furthermore, research is consistently advancing and suggesting refinements and updates to evidence-based practice, particularly as contexts change over time. Consequently, we must be able to adapt our recommendations and practices to reflect these advances in knowledge (Dawes et al., 1989).

Given a continually evolving evidence base and a context where decisions must be made on a routine basis, what constitutes acceptable evidence to be incorporated in practice depends on the alternative options (Helmus & Olver, 2023). There may be aspirational goals of what an ideal risk assessment tool should offer (e.g., see Hanson, 2009). However, when there are no existing validated tools available for a particular group or outcome, the minimum empirical threshold to justify its use is "Is it better than nothing" (i.e., unstructured professional judgment). As research advances, thresholds of evidence increase and shift to the question of "Is it better than existing alternatives?" (Helmus & Olver, 2023). The decision to use a risk tool is based not only on the evidence available for that tool, but also whether it is better than the alternative, and the potential harms of using the tool should be considered alongside the harms of not using the tool.

A Minimum Threshold of Evidence

Ultimately, risk assessment tools are designed to assess risk of recidivism, so the primary

criteria on which they should be evaluated is predictive accuracy (Hanson, 2022; Helmus & Babchishin, 2017). If the tool differentiates individuals based on their risk to reoffend, then it is defensible for case management decisions and resource allocation. Is there a bare minimum threshold of predictive accuracy? In the absence of a structured, empirically validated risk tool, decisions are made based on unstructured professional judgement. Results for unstructured judgement vary, but on average across nine systematic reviews, hovered around an AUC of .56 (Viljoen et al., 2021), which marks the minimum amount considered a "small" effect size (corresponding to a Cohen's *d* of .20; for comparison, AUCs of .64 and .71 are considered moderate and large effects, respectively; Rice & Harris, 2005). Consequently, AUCs above .56 would meet a threshold of 'better than the alternative' (unstructured professional judgement) and are a justifiable minimum threshold.

Other Considerations in Evaluating Risk Tools for Use

Minimally acceptable predictive accuracy is sufficient to justify the use of a risk tool when there are no alternative tools available. In other words, it is enough to say the tool can be used for risk assessment and to inform policy and practice. There are other considerations that should be taken into account when choosing between existing validated risk tools and evaluating the strengths and limitations of the tool. These inform how much confidence to have in the results of a tool, and it may vary depending on the context of the assessment.

Helmus and Olver (2023) recently outlined several considerations, summarized in Table 1. When there are multiple tools available that meet minimum thresholds for predictive accuracy, these considerations are helpful in selecting a tool. There is no expectation that all considerations will lead to the same choice of tool. This is one of several reasons why there are advantages to using multiple risk tools (Babchishin et al., 2012; Brankley et al., 2021; Olver et al., in press), including supporting different tasks (e.g., one to assess likelihood of recidivism and another to

frame individualized case management), or to add incrementally to the assessment of risk. Although using multiple risk tools is common, empirical and/or clinical guidance on how to resolve discrepant results is still in early development (Thornton & Helmus, 2022). These considerations include the match between the referral question and what is measured by the risk tool, volume/quality of research validations, availability and quality of recidivism estimates, interrater reliability, comprehensiveness, and empirical support for the risk factors, how the tool can inform treatment targets and assessments of change, and the quality of training/implementation resources available.

Some jurisdictions offer more specific guidance regarding standards for admissibility of a scientific method in court cases, which reflect one context where risk tools may be used (Helmus et al., 2022). 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; see also Table 1). Importantly, none of these Daubert standards are 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. Imprecision in these standards often 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).

Author Allegiance and Financial Incentive Biases

One topic related to the quality of research evidence (a consideration in Table 1) that deserves special mention is the possibility of biases when a developer of a tool is involved in its

validation research. This is one example of a potential conflict of interest that can introduce implicit or explicit biases in how data are collected, analyzed, and reported, and in extreme cases, could increase the likelihood of academic fraud. Ideally, all researchers should be impartial, evaluate competing hypotheses with equal rigor, accurately report their findings, and acknowledge when the evidence supports changes to existing practice or tools. Room for bias increases when the researcher has some kind of interest in the tool or method being evaluated. This is often a financial interest, but can include an intellectual interest (e.g., evaluating one's own theory).

Room for biases may depend on the nature of the financial interest or other considerations. For example, some risk tools are proprietary, whereby the developers receive royalties for their use and therefore may have a considerable financial interest in the widespread use of the tool. Even if the tool is free to use, the developer may still indirectly benefit from its success, in terms of renumeration for trainings or consultations regarding the tool. Beyond developers and researchers, any expert may have financial interests insofar as they receive renumeration for testifying for or against the use of a risk tool.

The extent to which allegiance biases influence research has been debated. Early meta-analyses found significantly larger effect sizes for studies co-authored by a tool developer for multiple risk tools, including the LSI family of tools (Andrews et al., 2011) and Static-99 and VRAG (Blair et al., 2008). Blair and colleagues (2008) concluded that allegiance bias was the most likely explanation for the findings (see Harris et al., 2010, for a response to this article). Andrews and colleagues (2011), however, concluded that author involvement was likely an indicator of the quality of the implementation as opposed to bias. There have been two recent meta-analyses of Static-99R validation studies, one including all available studies (k = 56; Helmus et al., 2022) and another examining field validity studies (k = 15; Helmus, Hanson et al.,

2021). In both meta-analyses, validations with one of the tool developers as a co-author did not have higher predictive accuracy. Instead, significantly higher accuracy was observed in field implementations that reported following a certified training protocol for users of the tool (Helmus, Hanson et al., 2021), and in studies that cited the coding manual (Helmus et al., 2022). This suggests quality of implementation is the more likely explanation for previous findings related to author involvement.

Are Validation Studies Needed in the United States to Apply a Tool There?

Scurich and Krauss (2023) concluded that it was not appropriate to use the CPORT in the United States, partly because it had not yet been validated in the United States at that time. Broadly, they cautioned that we should not assume risk tools developed in one location (often in Canada) apply in the United States or other parts of the world. Although this seems to be a sensible caveat, it sets a high standard to require that a risk tool must be validated in each new jurisdiction before it can be applied there, especially in jurisdictions that lack the necessary infrastructure or resources to conduct this type of research (e.g., centralized national criminal records). This raises issues of which characteristics (e.g., of the setting, or the individual being assessed) necessitate separate validation research before a tool can be used. In other words, when should we assume generalizability of research findings, versus requiring separate validation (for further discussion of this issue, see Helmus & Olver, 2023).

To support their conclusion that the lack of validation studies in the United States should preclude its use there, Scurich and Krauss (2023) quoted Helmus et al. (2011), who said that it is "essential to validate actuarial risk tools before they are routinely used in countries other than those in which they were developed" (p. 64). This quote was from a book chapter exploring factors that could impact international generalizability of actuarial risk tools. Helmus et al. (2011) concluded that generalizability across Western countries (including the United States)

seemed generally well-supported, but more caution was warranted in applying actuarial risk tools in other regions of the world, such as Africa, Asia, and Latin America, given the larger differences in laws, law enforcement practices, and social/cultural contexts. For example, countries such as Canada, the United States, Australia, and New Zealand are former dominions of the United Kingdom, share common law origins, have English as a primary official language, and are relatively high-income countries; in sum, they are WEIRD countries (meaning Westernized, Educated, Industrialized, Rich, and Democratic, see Henrich et al., 2010).

In addition, although we agree with Helmus et al. (2011) that international validations are helpful and would be considered as part of the quality of the research evidence (in Table 1), we disagree with the strict interpretation of that quote that such validation is necessary before applying a risk tool in a country outside its development frame. We do so for two reasons, one being that the Helmus et al. (2011) quote cited by Scurich and Krauss (2023) predated the *Standards for Educational and Psychological Testing* published jointly by the American Psychological Association, American Educational Research Association, and National Council on Measurement in Education (hereafter referred to as Joint Committee, 2014). These standards offer more nuanced, authoritative, and updated guidance on this issue (more on this below). Secondly, more than a decade of research on international and cultural generalizability has accrued since Helmus et al. (2011), much of it conducted or meta-analyzed by Helmus and colleagues (Ahmed & Helmus, 2023; Ahmed et al., 2023; Helmus, Babchishin, & Hanson, 2013; Helmus et al., 2022, 2023; Lee et al., 2023; Olver et al., in press). Simply put, the 2011 quote selected and relied on by Scurich and Krauss (2023) is outdated.

Should risk assessment tools developed outside the United States be applied in the United States before validation research is published? That depends. Following the Joint Committee (2014) standards, separate validation studies are warranted where sample sizes are adequate, and

there is sufficient theory or research to suggest differential predictive accuracy for that subgroup. It is not sufficient to assert a difference between countries (e.g., sentencing policies or legal definitions of CSEM offenses); of relevance is whether those differences would reasonably threaten the performance of the tool. So, is there theory or research to suggest that risk tools developed in other Western democratic countries will not be applicable in the United States? Not really. That would imply that the underlying risk factors for recidivism in the United States differ from other WEIRD countries. This is unlikely based on existing evidence (Bonta & Andrews, 2023).

We are not aware of research or theory to expect meaningful differential prediction in the United States compared to other Western democratic countries (for further discussion of when country or race/ethnicity may be a relevant consideration, see Helmus & Olver, 2023). In the sexual recidivism risk assessment field, meta-analyses indicate that risk tools that work in these countries tend to work in the United States as well, although predictive accuracy tends to be highest in the United Kingdom, followed by Canada, followed by the United States (Hanson & Morton-Bourgon, 2009; Helmus et al., 2022). Based on our experience working with sexual recidivism datasets across diverse countries, we suspect this is not due to fundamental differences in what predicts recidivism, but rather is an artifact of the quality and comprehensiveness of criminal records, which are necessary for accurately measuring detected recidivism. Despite some methodological variations, criminal records tend to be better quality and more comprehensive in the United Kingdom, followed by Canada, then by the United States (Helmus et al., 2011). Based on existing theory and research, we believe it is reasonable to presume that risk tools validated in other Western democratic countries will generalize to the United States, although there may be a small drop in predictive accuracy, likely attributable to difficulties reported in gathering national recidivism data in research from the United States (for

challenges with U.S. criminal records, see Laudon, 1986). This observed drop in predictive accuracy is generally not sufficient to fall below minimal standards of acceptable predictive accuracy.

Which Risk Tools Are Defensible to Use for CSEM Populations?

This section will review the research support for various risk tools for individuals convicted of CSEM offenses. For each tool, we will provide a brief description, summary of the research evidence, and evaluate whether the tool is defensible for use with CSEM populations, following the considerations outlined in Table 1, including information that relates to the Daubert criteria for legal admissibility. Table 2 contains a brief summary of the conclusions from this section.

Tools Designed to Predict Sexual (Including CSEM) Recidivism

CPORT

The CPORT is a risk assessment tool developed in Canada (Seto & Eke, 2015; Eke et al., 2018). It was designed to assess risk of any sexual recidivism among adult men convicted of a child pornography offense. 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). It can be scored primarily using criminal history and police investigative data (Eke et al., 2018). Total scores can range from 0 to 7. The CPORT has good interrater reliability in both research and field contexts (Hermann et al., 2019; Savoie et al., 2021; Seto & Eke, 2015).

Eke and colleagues (2019) used logistic regression to provide preliminary 5-year estimates for the outcomes of any sexual recidivism and child pornography recidivism. However,

given the small sample sizes (e.g., 40 and 29 recidivists, respectively), the authors recommended caution in reporting these probabilities in evaluation reports until further research is available. Given additional research (summarized below), it is possible that recidivism estimates can be produced by combining the results of newer studies, but these are not yet released.

Recently, Helmus and colleagues (2023) meta-analyzed the available research on the CPORT. They found six studies (including the development sample), with five reporting on predictive accuracy for any sexual recidivism (mean weighted AUC = .75, $p < .05$, $N = 1{,}411$), and five reporting on predictive accuracy for any new CSEM offense (mean weighted AUC = .65 to .66, depending on the analytic method used, $p < .05$, $N = 1{,}376$). Studies were from Canada, Scotland, Spain, Iceland, and New Zealand; one study from the United States was released after this meta-analysis was published, and will be incorporated into the findings below.[2] Effect sizes for predicting any sexual recidivism were large, and were larger even than the typical accuracy found for Static-99R among individuals with offline sexual offenses (Helmus et al., 2022). Effect sizes were also consistent across studies, with no more variability in accuracy than what would be expected by chance. This is particularly encouraging given that the validations spanned multiple countries and included two studies that were missing two or more items on the CPORT (Black, 2018; Pilon, 2016). Given the substantial amount of missing information, this could be considered an underestimate of the tool's predictive accuracy, since missing information usually reduces accuracy.

The predictive effects were lower when examining CPORT's ability to predict CSEM recidivism as a subset of any sexual recidivism (Helmus et al., 2024), but still moderate and comparable with predictive accuracy seen for Static-99R for offline offenders (Helmus et al.,

[2] Also since this meta-analysis was conducted, there is a new conference presentation (Eke, 2022) providing preliminary results from a research study by Eke and colleagues that included a small sample of 136 cases in Québec, Canada a where CPORT significantly predicted any sexual recidivism (Cox regression hazard ratio = 1.52). Given the small sample size, this study would have a trivial impact if integrated in the meta-analysis.

2022). There was significant variability in accuracy across studies, although there were too few studies for moderator analyses. This suggests that CPORT is predicting CSEM recidivism specifically, but not quite as well or consistently as it predicts any sexual recidivism. This could reflect greater challenges in measuring CSEM recidivism, given that its detection is more heavily dependent on investigative techniques, resources, and priorities.

Despite the consistency in predictive accuracy for sexual recidivism across Canada, Scotland, Iceland, and New Zealand,[3] Scurich and Krauss (2023) argued that it should not be applied in the US given the lack of American validation studies. Reflecting the rapidly evolving nature of research in this field, one month after Scurich and Krauss's (2023) article was accepted for publication, a validation study of the CPORT in the United States was published (Cohen, 2023), countering that particular objection to the use of the tool.

The U.S. validation study of the CPORT (Cohen, 2023) is unique because of its large sample size (N = 5,768) and because it is not a validation of the CPORT scored by human raters using the scoring manual. Instead, natural language processing and machine learning techniques were used to extract data from extensive file information. The information used for this activity did not fully comport with the CPORT coding rules; hence, this is a modified or approximated version of the CPORT. For example, admission of interest in children was extracted from self-report, not interviews with police, and gender preferences was based on self-report rather than objective information about the CSEM collection. The quality of the approximation is dependent on the availability and quality of the information in the files, the impact of the deviations of the scoring, and how well the natural language processing technique is able to appropriately capture how trained raters would interpret the meaning of the words (e.g., jargon, jurisdiction-specific language, slang, context clues).

[3] There was one study from Spain, but it analyzed CSEM recidivism only, not any sexual recidivism (Soldino et al., 2021).

Examining fixed 5-year follow-up data on rearrests for a sexual offense, the CPORT significantly predicted sexual recidivism with a small effect size (AUC = .62, 95% CI [.58, .65]) among 5,768 men starting federal post-conviction community supervision between 2010 to 2016 (4.5% sexual recidivism rate; Cohen, 2023). All CPORT items significantly predicted sexual recidivism, except the two boy preference items, although this was likely the information that was most difficult to approximate using data mining techniques. Additionally, the item for pedohebephilic interests was examined separately for pedophilic and hebephilic interests, with only the former significantly predicting sexual recidivism.

Another unique feature about this American study by Cohen (2023) was that the sample appeared unusually low risk, even for CSEM cases. The average approximated CPORT score was 1.44 (*SD* = 1.13),[4] whereas other studies with complete information generally found means between 1.8 and 2.0 (see Helmus et al., 2024), with the exception of a small sample from Spain with an even lower average score (0.8; Soldino et al., 2021). This is relevant because AUCs are meaningfully impacted by heterogeneity in risk scores (Howard, 2017). Lower AUCs are expected when the distribution of CPORT scores is skewed towards lower (or higher) risk scores, as in this study, where 82% of the sample scored 0, 1, or 2, and only 18% scored between 3-7.

It is possible to integrate the results of Cohen's (2023) in the meta-analysis from Helmus et al. (2023), although a few caveats should be offered. Meta-analyses can be influenced not only by an outlier study (an unusually high or low result that meaningfully impacts the findings), but also by an unusually large study sample in comparison to the others, such that it can have undue influence on the aggregate findings, even if it is not an outlier (Helmus, Hanson, et al., 2013).

[4] The mean was calculated based on the distribution of scores presented in Table 5 of Cohen (2023). Unfortunately, 46 cases out of 5,768 scored 5, 6, or 7 on the CPORT and were clumped together as a score of 5+. Given the small number in this group, it is unlikely to substantively impact the mean CPORT score. Assuming all 46 cases score 7 (a highly improbable situation), this would increase the mean CPORT score from 1.44 to 1.46.

Cohen's (2023) study meets both criteria; it is an outlier according to the criteria from Hanson and Bussière (1998), and it also exerts undue influence on the findings because the sample size is more than ten times larger than the next largest study. Study weights are determined by the inverse of the variance (Borenstein et al., 2009), which takes into account more than just the total sample size. The study weights for the six initial studies ranged from 71 to 1,270, whereas the weight for the Cohen (2023) study was 3,135. Given these issues, it would be customary in a meta-analysis to present the results both without Cohen (2023), which is already presented by Helmus et al., (2023), summarized above, as well as with Cohen (2023).

The average weighted AUC for predicting sexual recidivism without Cohen (2023) was .75 (Helmus et al., 2024). Integrating Cohen (2023), the average weighted AUC dropped to .68 in fixed-effect analyses (95% CI [.65, .70]) and .71 in random-effects analyses (95% CI [.63,.79]; $k = 6$, $n = 7{,}179$). Whereas there was no significant variability in predictive accuracy across the studies summarized by Helmus et al. (2023), variability across studies is statistically significant once Cohen (2023) is incorporated ($Q = 27.74$, $df = 5$, $p < .001$) and large ($I^2 = 82\%$; for more information on interpreting meta-analysis results, see Borenstein et al., 2009, or Helmus, Hanson, et al., 2013). The effect size from Cohen (2023) was significantly lower than the meta-analytic results of the other five studies ($Q_{change} = 24.3$, $df = 1$, $p < .001$), confirming that it is an outlier. It is unclear if this is related to methodological differences, such as the novel data mining methods used to approximate CPORT scores, reliance on self-report, or a statistical impact of restriction of range in scores in this unusually low risk sample. Likely, all these factors contributed to the lower accuracy observed.

In addition to providing further support for the predictive accuracy of the CPORT, the Cohen (2023) study also demonstrates the possibilities of using machine learning and language analysis techniques for risk assessment. This is a promising avenue for research as it allows for

efficient processing of vast amounts of data, including massive sample sizes. However, the limitations of this technique must be better explored. For example, an important question is how well the CPORT scores from this technique would match the consensus rating of properly trained scorers who relied on the same information.

Although developed and validated for men convicted of CSEM offenses, a new study from Germany found that the CPORT is also promising among non-justice-involved men voluntarily seeking treatment for sexual interest in children (von Franqué et al., 2023). The CPORT was scored based primarily on self-report information, so this would have required some deviations from the coding manual (i.e., there was no investigation related to an index offense). Among 132 men who self-reported having accessed CSEM at some point in their lifetime (none had ever been officially detected by the criminal justice system), the CPORT score using the CASIC to rate the item for pedohebephilic interests significantly predicted self-reported future contact sex offenses against a child (AUC = .69; 3.8% offended) and self-reported CSEM use (AUC = .63; 39.4% reoffended) during an average of 28 months at risk. If clinical diagnosis was used to score the pedohebephilic item in the CPORT, the AUC dropped slightly and was not quite statistically significant (AUC = .68 for child sexual abuse, p = .054; AUC = .57 for CSEM recidivism, p = .052).

Is the CPORT Defensible to Use? Yes. The CPORT is defensible to use for assessing risk of any sexual recidivism or specifically CSEM recidivism among adult men convicted of CSEM offenses (see Table 2 for summary). Including or excluding the results of Cohen (2023), which represents a statistical outlier, the CPORT demonstrates moderate to large predictive accuracy, comparable to how well sexual recidivism risk tools generally predict among those with non-CSEM sexual offenses (Hanson & Morton-Bourgon, 2009; Helmus et al., 2022). Although there is no persuasive theory or evidence that risk factors should predict differently in

the United States compared to the other studies where CPORT was validated, and considering that allegiance to scoring should improve performance, if interpretation is restricted solely to the Cohen (2023) study of approximated CPORT scores, the AUC is in the upper range of a small effect size (AUC = .62). This is still a meaningful improvement over unstructured professional judgement (e.g., AUCs ~ .56, Viljoen et al., 2021), particularly when considering the other benefits of the tool (e.g., demonstrated interrater reliability, clear coding manual).

The CPORT provides preliminary information on expected recidivism rates from one study with high quality recidivism data from multiple sources, although the authors advised against using it in applied reports until more research is available (Eke et al., 2019). Given the accumulation of new research, it is likely that updated CPORT recidivism estimates will be available in the near future. Several studies have found strong interrater reliability. The tool currently contains a detailed coding manual (Eke et al., 2018), and training is available from the tool developers.

The CPORT includes a reasonable sampling of static risk factors that would be expected to primarily assess general antisociality and pedohebephilic sexual interests. These correspond to established dimensions of risk for sexual recidivism among men with offline sex offenses (Brouillette-Alarie et al., 2016), and would likely be able to provide some recommendations for treatment. For example, scores across items would likely indicate whether the individual might be referred for treatment for general criminality (e.g., if they have risk factors for criminal history or conditional release violations), paraphilic sexual interests (e.g., if they have risk factors for contact sex offenses, pedohebephilic interests, or more boy than girl content), or neither. The CPORT cannot currently be used to assess changes in risk over time.

As summarized in the sections above and in Table 2, the CPORT could meet the Daubert criteria for legal admissibility. The most difficult prong of the Daubert criteria to assess is

widespread acceptance within the scientific community. This is difficult to assess because this area of research evolves so quickly. The CPORT is the only tool specifically designed to predict sexual recidivism among men convicted of CSEM offenses, and it has received more research attention than any other CSEM offense-specific risk tool. We are not aware of published surveys of risk tool use for this population yet (and they would be difficult to conduct given the rapidly changing evidence base), but anecdotally, we are aware of its widespread use. It has currently been translated into Dutch, Danish, French, German, Norwegian, Portuguese, and Spanish.

Risk Matrix 2000 – Sex (RM2000/S)

The RM2000/S is an actuarial static risk tool for adult men convicted of sexual offenses (Thornton et al., 2003, 2023). 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). Normative data are available for the RM2000/S, including sexual recidivism estimates (Lehmann et al., 2016; Thornton et al., 2023). There are also structured rules and recidivism estimates available for combining the RM2000/S with the STABLE-2007 (Brankley et al., 2017). Training and resources for RM2000/S are available from www.saarna.org.

Meta-analyzing 16 samples of men convicted primarily of offline sexual offenses across the United Kingdom, Denmark, Germany, Canada, and the United States, the RM2000/S predicted sexual recidivism with a moderate effect size (Cohen's d = .74; Helmus, Babchishin, & Hanson, 2013; this corresponds to an AUC of .70, following Rice & Harris, 2005). One study found excellent interrater reliability of RM2000/S scores in two field samples (Wakeling, Mann, & Milner, 2011).

The tool was originally developed using individuals released from prison in the early 1990s in England and Wales (Thornton et al., 2003), which predated widespread use of the internet and the ability to access CSEM online. Despite the lack of inclusion of men with online offenses in the development sample, adjustments to the coding manual over time have provided scoring guidance to apply the tool to men convicted of CSEM offenses (Thornton, 2017; Thornton et al., 2023). As examples, the stranger victim item is not scored on the basis of individuals in CSEM images (as that would automatically increase the score of almost everyone with a CSEM offense), the noncontact sex offense item is only scored for CSEM offenses if there is also a contact sex offense, and there is more nuanced guidance regarding how to score the male victim item based on CSEM activity (Thornton et al., 2023). These coding rule adjustments have been tested (see Helmus et al., 2024) and have face validity based on existing research on the risk posed by individuals with CSEM offenses; if all CSEM offenses were automatically counted as noncontact and towards the victim items, then virtually all individuals with CSEM offenses would be considered above average risk for sexual recidivism, which is out of step with existing research on their recidivism rates compared to individuals with offline sexual offenses (Helmus, 2023).

Three studies from the United Kingdom examined predictive accuracy of the RM2000/S for internet CSEM offenses (Barnett et al., 2010; Elliott et al., 2019; Wakeling, Howard, & Barnett, 2011). However, there is considerable overlap among these samples, with Wakeling, Howard, & Barnett (2011) subsuming most (if not all) of the other two samples. Consequently, only the Wakeling, Howard, and Barnett study will be discussed. Examining 994 men convicted of CSEM offenses across a fixed 2-year follow-up period, the RM2000/S significantly predicted sexual recidivism (AUC = .67). Recently, Helmus and colleagues (2023) examined the RM2000/S in the Canadian CPORT development and validation sample used by Eke and

colleagues (2019). In this study, the scoring rules were applied based on the above. Across 339 men with a fixed 5-year follow-up, the RM2000/S significantly predicted any sexual recidivism (AUC = .66) and CSEM recidivism (AUC = .67). Combining the Canadian and UK results, the RM2000/S predicted any sexual recidivism with moderate accuracy (AUC = .66, 95% CI [.59,.74]). In a recent field study, scores on the RM2000/S were strongly correlated with te CPORT total scores (r = .52; Azizian et al., 2023).

Given research suggesting low sexual recidivism rates of men with CSEM offenses (Babchishin et al., 2015, 2018; Seto et al., 2011), one concern we had heard raised was that applying recidivism estimates from a generic sexual offense risk assessment tool may overestimate the risk of men with CSEM offenses. Helmus et al. (2023) found the opposite: estimated recidivism rates from the RM2000/S were significantly lower than the observed recidivism rates in the Canadian sample. It is unclear why the observed recidivism rates were higher than expected, though the most likely explanation is methodological differences in how recidivism was defined and measured compared to the normative data (systematic access to police occurrence reports and national data; Helmus et al., 2024). It may also be a reflection of other general findings that recidivism probability estimates for sexual recidivism risk tools are less stable across samples than the results of relative predictive accuracy metrics, such as AUCs (for a brief review, see Helmus, 2018). Consequently, it is unclear whether the RM2000/S truly underpredicts sexual recidivism for those with CSEM offenses, or whether the sample used by Helmus et al. (2023) had unusually good quality recidivism data reflecting a higher recidivism rate. Nonetheless, the validation study suggested that the RM2000/S does not overestimate sexual recidivism rates with CSEM populations.

Is the RM2000/S Defensible to Use? Yes. The RM2000/S is a risk assessment tool designed to predict sexual recidivism among men convicted of a sexually motivated offense. It is

defensible to use for assessing risk of any sexual recidivism or specifically CSEM recidivism among adult men convicted of CSEM offenses (see Table 2 for summary). Although there are only two non-overlapping validation studies of the RM2000/S with CSEM samples, one study was particularly large (Wakeling, Howard, & Barnett, 2011), and both studies found moderate predictive accuracy (AUC of roughly .66), which is comparable to predictive accuracy observed for generic sexual recidivism risk tools among men who have committed offline offenses.

There are recidivism estimates for the RM2000/S used alone (Lehmann et al., 2016) and in combination with the STABLE-2007 (Brankley et al., 2017). These norms were developed on samples of men who committed primarily offline sexual offenses, and their applicability to men with CSEM offenses is not well established, with initial results suggesting the RM2000/S norms may underpredict sexual recidivism. The RM2000/S has strong interrater reliability, with training and implementation resources available from the SAARNA website.

When the RM2000/S is used in conjunction with the Risk Matrix 2000 – Violence tool (RM2000/V) (Thornton et al., 2023), it includes a reasonable sampling of static risk factors assessing age, general antisociality, and atypical sexual interests. Although the risk factors are static in nature, the items can be organized in treatment need domains, indicating the likelihood that the individual has substantial treatment needs related to general antisociality or atypical sexuality (Fernandez et al., 2023; Thornton, 2022). Consequently, although the RM2000 cannot assess changes in risk (with the exception of aging effects, which can be incorporated), it can make reasonable and empirically defensible recommendations for treatment. The items included in the RM2000/S are similar to items in Static-99R, which are well-supported static risk factors among men who committed primarily offline sexual offenses (Helmus & Thornton, 2015). In terms of the Daubert criteria for legal admissibility (see summary in Table 2), the RM2000/S has widespread acceptance within the scientific community, although we note that there is more

support for its acceptance among populations primarily of men who committed offline sexual offenses. However, it was used for CSEM-offending populations for quite some time in the United Kingdom and has also been used within clinical practices in the United States (Azizian et al., 2023) and the new validation study from Canada provides further support for its use.

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

The United Kingdom's Ministry of Justice recently developed a brief static risk assessment tool (the OSP/I) to assess risk of a CSEM offense among adult males with a current or previous conviction for a sexually motivated offense (His Majesty's Prison and Probation Service [HMPPS], 2023; Howard & Wakeling, 2021). The OSP/I is not so much a tool as it is a single item reflecting the individual's sex offense history (it captures whether someone with a sex offense conviction has zero, one, or two or more sanctions for CSEM offenses). In the OSP/I as originally developed, 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 its original version with four risk levels, the OSP/I 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 2 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 2 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]).

Is the OSP/I Defensible to Use? Partially (see Table 2 for summary). The OSP/I is defensible to use among all men convicted of sexual offenses to make a reasonable differentiation of risk of CSEM recidivism. Rather than a risk assessment tool which combines multiple risk factors, it is a single risk factor that indicates whether someone with a sex offense

conviction has zero, one, or two or more sanctions for CSEM offenses. If applied to assess risk among individuals with a CSEM conviction, it simply differentiates two groups: those with and without multiple CSEM sanctions. This is a valid risk factor and simple to score, a key goal highlighted by Howard and Wakeling (2021). However, its reliance on a single factor means it will not provide guidance for areas to focus on in case management decisions. Additionally, of those with CSEM convictions, only 8% had multiple sanctions for CSEM (Howard & Wakeling, 2021), so applied to CSEM populations, the OSP/I mainly identifies a small subset of individuals (<10%) who are higher risk for CSEM recidivism compared with the vast majority of men convicted of CSEM offenses. Those with multiple CSEM sanctions are higher risk than those with only one sanction, and the expected 2-year reconviction rates for a CSEM offense are 6% and 3% for these two groups, respectively (HMPPS, 2023). Of note, the risk levels of the OSP/I (low, moderate, and high risk) must be carefully interpreted in line with its intended purpose: it is meant to assess the risk of CSEM recidivism among all individuals convicted of sexual offenses, not only among individuals convicted of CSEM offenses. Since someone with a CSEM conviction cannot be considered low risk on the OSP/I, it is not in line with the risk principle of effective correctional practices (Bonta & Andrews, 2023). Interrater reliability for this risk factor is not reported, but there is scoring guidance available, and the scoring is likely to be highly reliable since it relies simply on a review of sexual offense history (HMPPS, 2023).

In terms of the Daubert criteria for legal admissibility (see summary in Table 2), the OSP/I research has been published in government documents in the United Kingdom (HMPPS, 2023; Howard & Wakeling, 2021). This is not peer-reviewed in a scientific journal, but the United Kingdom Ministry of Justice pays for external peer review for selected research reports, including the Howard and Wakeling (2021) report. The OSP/I has therefore undergone a peer review process that is roughly analogous to scientific journals, although the final decision to

publish the report remains with the government. It is used in the United Kingdom, and is quite new, so does not necessarily have widespread acceptance in the scientific community, beyond the peer review process it underwent.

Static-99R

Static-99R (Hanson & Thornton, 2000; Helmus, Thornton, et al., 2012) is an actuarial risk assessment tool designed to predict sexual recidivism among adult males charged or convicted of "Category A" sex offenses, defined below (Phenix et al., 2016). Static-99R 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). A recent review concluded that Static-99R typically has excellent interrater reliability and moderate predictive accuracy for sexual recidivism among its intended population (weighted AUCs around .68 to .69; $k = 56$; Helmus et al., 2022).

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. The coding manual

acknowledges that the tool could be modified for individuals whose sexual offenses are restricted to CSEM if there is validation research to support this.

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).

Consequently, if applied to a sample of individuals whose sexual offenses are exclusively for CSEM, the tool effectively has only six items instead of ten (they would all score 1 for noncontact, and 0 on the three victim items if unmodified). If those with CSEM offenses are integrated into a larger sample that has other types of sexual offenses, then two of the items (noncontact sex offenses and male victims) are not likely to be meaningfully assessing the same risk factor across these groups, which complicates interpretation of the total score in comparison to the full sample.

If Static-99R is to be effectively assessed and validated among individuals with CSEM offenses, modifications to the coding rules are needed. But there are no official modifications proposed by the tool developers, so researchers must make their own decisions. The modifications applied to the RM2000/S (Thornton et al., 2023) are a sensible option: noncontact

would be scored for individuals with dual types of sex offenses, male victims would be scored if the evidence suggested deliberate searching for CSEM depicting boys, and stranger and unrelated victims would not be scored based on CSEM offenses. But other modifications could also be plausible, such as substituting the CPORT item regarding boy CSEM content for the male victim item.

So far only one study has examined the predictive accuracy of Static-99R with CSEM populations, across two samples (see the conference presentation by Eke, 2022), scoring the tool following the existing coding rules (e.g., CSEM images would not be considered in scoring victim items). Eke and colleagues examined Static-99R with 348 cases from the main CPORT development and validation sample from Ontario (Eke et al., 2019) as well as 136 new cases from Québec (Eke, 2022). Static-99R showed convergent validity with the CPORT, with correlations of .69 and .47 in the two samples, respectively. Static-99R significantly predicted any sexual, contact sexual, and CSEM recidivism for the sample from Ontario (AUCs were presented as a graph and appear moderate, large, and small, respectively for those outcomes). Of these 348 cases, however, nearly one third (n = 108) also had a Category A sex offense and therefore fit the inclusion criteria of Static-99R. Among the smaller sample size of 240 individuals whose sex offenses consisted solely of CSEM (i.e., are currently outside the sampling frame of Static-99R), Static-99R did not significantly predict any sexual, contact sexual, or CSEM recidivism.

In the second sample of 136 CSEM cases from Québec reported by Eke (2022), Static-99R significantly predicted any sexual recidivism (Hazard ratio from Cox regression = 1.25, 95% CI [1.07, 1.49]). Again though, restricting the sample to individuals with CSEM offenses who were outside the sampling frame of Static-99R (sample size not reported), Static-99R did not significantly predict sexual recidivism, although the effect size was quite similar (Hazard

ratio = 1.22, 95% CI [.99, 1.52]).

Is Static-99R Defensible to Use? Partially (see Table 2 for summary, and a later section will provide further guidance about comparing options available). Technically, it meets minimum criteria for predictive accuracy for any sexual, contact sexual, and CSEM recidivism among individuals convicted of CSEM offenses. However, there are sufficient caveats and limitations to this research that give us hesitation. Most obviously, these data are not published yet, and the conference presentation does not provide exact AUC values. Most importantly, however, the research on individuals with CSEM offenses draws an important distinction between those with dual offenses (e.g., CSEM and a Category A sex offense) and those whose sex offenses are restricted to CSEM (Babchishin et al., 2018). Those with dual sex offenses are higher risk, which is unsurprising: by definition, they must have committed more than one sex offense and demonstrated some versatility in their sexual offending. If they have committed any type of Category A sex offense (Phenix et al., 2016), they are already within the sampling frame of Static-99R and many other sexual recidivism risk tools. It is the group whose sex offenses are specific to CSEM that are of primary concern in the field, as they are not in the typical sampling frame of sexual recidivism risk tools.

In Eke and colleagues' samples (Eke, 2022), among these subgroups of individuals whose sex offenses are restricted to CSEM (who are not part of the appropriate population for Static-99R), Static-99R did not significantly predict recidivism. However, some caveats should also be noted here. The sample sizes for these subgroups are quite small, and there would be insufficient sexual recidivists for adequate statistical power, particularly in the CSEM-only subgroup. Additionally, by reducing the heterogeneity of the CSEM group, we would expect reduced accuracy compared to the full CSEM group based on restriction of range effects (Howard, 2017). In other words, the current research is still too preliminary to know much about

how Static-99R predicts among this subgroup.

Consequently, broadly defining the population of interest as individuals convicted of CSEM offenses, who may also have contact or noncontact sexual offenses that would make them eligible for scoring, Static-99R appears defensible to use to assess relative risk of sexual recidivism (as well as contact and CSEM recidivism, separately). However, for the narrower group of individuals with CSEM offenses who fall outside the sampling frame of the 2016 Static-99R coding rules (Phenix et al., 2016), we are not yet confident that Static-99R adequately distinguishes risk within this group. Until further research is available to update our recommendations, we would advise against using Static-99R for an individual with CSEM offenses who does not also have a charge or conviction for a Category A sex offense in their history.

There are recidivism estimates available for Static-99R (Helmus, Lee, et al., 2021), but insufficient research on their applicability to individuals with CSEM offenses. The data presented by Eke (2022) suggested the Static-99R recidivism norms would underpredict recidivism, but it is difficult to draw conclusions based on only one study given the amount of variability in recidivism rates across samples (Helmus, Hanson et al., 2012). These analyses used the sample from Eke et al. (2019), which had unusually good quality recidivism data and perhaps not coincidentally also had higher recidivism rates compared to many of the studies included in the Static-99R recidivism norms (for further discussion of how the methods may affect calibration with this CSEM sample, see Helmus et al., 2024). As of yet, we do not have Static-99R recidivism estimates specific to CSEM samples, or sufficient research available to evaluate the generalizability of the existing norms.

Across the broader literature, Static-99R has strong interrater reliability, with training and implementation resources available from www.saarna.org. In fact, the research base, resources,

and specificity of the coding rules for Static-99R is generally much more advanced than any other sexual recidivism risk tool. It has a strong sampling of static risk factors assessing age, general antisociality, and atypical sexual interests, with a meta-analysis supporting these items as defined and measured by Static-99R (Helmus & Thornton, 2015). It cannot assess changes in risk, and the underlying constructs may give some indication of broad treatment need domains (see Brouillette-Alarie et al., 2016), but it is not ideal for assessing treatment targets.

Based on the limitations discussed above, we do not think Static-99R can meet Daubert criteria for legal admissibility for men with exclusively CSEM sexual offenses, at least not until the results from Eke (2022) are peer reviewed and published, and ideally with larger sample sizes available. Alternately, aggregating the results from the Ontario and Québec samples through meta-analysis may give sufficient statistical power for the group whose sex offenses are restricted to CSEM.

STABLE-2007

The STABLE-2007 (Fernandez et al., 2014; Hanson et al., 2007) is an empirically derived risk tool assessing stable dynamic risk factors relevant to treatment and supervision of individuals convicted of a sexually motivated offense. It should be assessed based on a detailed interview and review of available file information. The tool has 13 items scored on a scale of 0 to 2. 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. Changes in STABLE-2007 scores predict sexual recidivism, with the most recent assessment generally being most informative (Lee et al., 2023). The STABLE-2007 is a widely used measure for dynamic risk (Kelley et al., 2020). It has good field interrater reliability (ICC = .86; Fernandez & Helmus, 2017) and a recent meta-analysis found moderate accuracy in predicting sexual recidivism (AUC = .67; Brankley et al., 2021). Ideally, it should be combined

with a static risk tool, such as Static-99R or the RM2000/S (Brankley et al., 2017). Training and resources for the STABLE-2007 are available from www.saarna.org.

The STABLE-2007 was developed by revising the STABLE-2000 based on a prospective field validity study across Canada (Hanson et al., 2007). Given the limited representation of men with CSEM offenses in the development sample and the lack of validation research, the 2014 coding manual indicated that the tool should not be used with men whose only sexual offense convictions were for CSEM (Fernandez et al., 2014). 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).

The STABLE-2007 was recently validated in a large field study from British Columbia, Canada (Babchishin et al., 2023). A sample of 308 men with any CSEM offense (further subdivided into a group of 228 men whose only sexual offenses were for CSEM and 80 men with both CSEM and contact sexual offenses) were compared to 1,042 men with contact sexual offenses against children. New charges or convictions for sexual offenses were examined for a follow-up period averaging roughly four years for the CSEM groups and 3 years for the contact group. Recidivism outcomes examined including any sexual recidivism, CSEM recidivism, or contact sexual recidivism (although statistical power was very limited for specialized sexual recidivism outcomes among some subgroups). Among the overall group of men with any CSEM offense, and the subgroups of CSEM-exclusive and mixed, STABLE-2007 total scores had large effect sizes in predicting any sexual, CSEM, and contact sexual recidivism (Harrell's *C*s of .75 and above), although the effects were nonsignificant for predicting contact sexual recidivism among CSEM-exclusive men, and for the subtypes of sexual recidivism among the mixed group (Babchishin et al., 2023). Effect sizes tended to be larger than the group with contact sexual

offenses against children.

It was not possible to assess whether the recidivism norms for STABLE-2007 applied to the CSEM group, because the recidivism estimates are derived from combining the STABLE-2007 with the Static-99R. Instead, using logistic regression, 3-year sexual recidivism rates for the median STABLE-2007 score (7) were compared for the group that had any contact sexual offense and the CSEM-exclusive group (Babchishin et al., 2023). There was no significant difference in the recidivism rates between groups, controlling for STABLE-2007 scores. This provides preliminary evidence that score-wise recidivism estimates are likely to be applicable to men whose only sexual offenses are for CSEM.

Based on these results, Babchishin and colleagues (2023) indicated that STABLE-2007 total scores were defensible to use for men convicted of CSEM offenses to assess relative risk and to direct treatment and risk management resources. Consequently, an FAQ was released for the STABLE-2007, indicating that the tool could be applied to men convicted of CSEM offenses, replacing original user guidance for the tool (SAARNA, 2023[5]). The FAQ document advises that the emotional congruence with children item should be rated for men with CSEM offenses, given its strong predictive accuracy in the study from Babchishin and colleagues (2023).

Is the STABLE-2007 Defensible to Use? Yes. The STABLE-2007 is a risk assessment tool designed to predict sexual recidivism among men convicted of a sexually motivated offense. It is defensible to use for assessing risk of any sexual recidivism or specifically CSEM recidivism among adult men convicted of CSEM offenses, including men whose sexual offending history is limited to CSEM offenses (see Table 2 for summary). Recidivism estimates are not provided for STABLE-2007 without combining it with a static risk tool, such as the RM2000/S (Brankley et al., 2017). Both the RM2000/S and STABLE-2007 have been validated

[5] The first author of this review is also on the Board of Directors for SAARNA and was therefore involved in the release of this FAQ.

for CSEM samples in different studies, but as of yet there is no study examining the combination of the tools in men convicted of CSEM offenses. Given the encouraging findings of Babchishin et al. (2023), however, it is reasonable to presume the combined risk levels and recidivism estimates for RM2000/STABLE-2007 are informative for CSEM populations.

The STABLE-2007 includes a fairly comprehensive selection of psychologically meaningful dynamic risk factors for sexual recidivism (Brankley et al., 2021; Brouillette-Alarie & Hanson, 2015; Mann et al., 2010) and is surprisingly robust to missing information (Perley-Robertson et al., in press). It can inform treatment recommendations and can assess changes in risk, with the most recent assessment providing the most accurate results (Babchishin & Hanson, 2020; Lee et al., 2023). However, no studies have yet examined changes in risk over time among men convicted of CSEM offenses.

Regarding the Daubert criteria for legal admissibility, the STABLE-2007 has widespread acceptance for its use among the broader population of men convicted of sexually motivated offenses (Kelley et al., 2020), but the evidence to support its accuracy for individuals with CSEM offenses has only been available for a few months at the time of writing. Consequently, we expect its use to increase for this group.

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.

Like the STABLE-2007, the ACUTE-2007 was recently validated in British Columbia (Babchishin et al., 2023). Although the ACUTE-2007 is meant to be reassessed regularly, this study examined only the first ACUTE-2007 assessment. The analyses of the ACUTE-2007 included 297 men with any CSEM offense (further subdivided into a group of 222 men whose only sexual offenses were for CSEM and 75 men with both CSEM and contact sexual offenses) compared to 920 men with contact sexual offenses against children.

Among the overall group of men with any CSEM offense, and the subgroups of CSEM-exclusive and mixed, ACUTE-2007 total scores generally had large effect sizes in predicting any sexual, CSEM, and contact sexual recidivism (Harrell's *C*s of .71 and above, with the exception that the ACUTE-2007 predicted CSEM recidivism for the small mixed group with a *C* of .61; Babchishin et al., 2023). Not all analyses were statistically significant, but the ACUTE-2007 did significantly predict sexual and CSEM recidivism for the overall CSEM group and the CSEM-exclusive group. Effect sizes tended to be larger than the group with contact sexual offenses against children. In fact, in the logistic regression analyses of 3-year sexual recidivism rates, the

relative predictive accuracy of the ACUTE-2007 was significantly higher for the CSEM-exclusive group compared to the group that had any contact offense against a child (including both a contact and CSEM offense). Although statistical power was limited, the ACUTE-2007 items were promising for the CSEM groups, especially compared to the group with contact offenses against children. Similar to the results for the STABLE-2007, there was no significant difference in the recidivism rates between groups, controlling for ACUTE-2007 scores.

As with the STABLE-2007, Babchishin and colleagues (2023) concluded that the ACUTE-2007 was defensible to use for men convicted of CSEM offenses to assess relative risk and to direct treatment and risk management resources. The FAQ noted for STABLE-2007 also includes the ACUTE-2007 indicating both are defensible to use for men convicted of CSEM offenses, replacing original user guidance for the tool (SAARNA, 2023). The FAQ document provides further examples of how to score the ACUTE-2007 for CSEM offenses, with particular attention to unique considerations for victim access, given that the risk is related to unsupervised use of the internet (SAARNA, 2023). Although internet use is relevant for other online offenses beyond CSEM (e.g., internet luring), it had not been discussed in the original ACUTE-2007 coding manual, so these updates may be applicable more broadly than to CSEM populations.

Is the ACUTE-2007 Defensible to Use? Yes. The ACUTE-2007 is defensible to use for assessing risk of any sexual recidivism or specifically CSEM recidivism among adult men convicted of CSEM offenses, including men whose sexual offending history is limited to CSEM offenses (see Table 2 for summary). The ACUTE-2007 should be used in conjunction with STABLE-2007, and given the research base for CSEM individuals is the same for both tools, the general comments for the STABLE-2007 are applicable here as well. Recidivism estimates are not applicable for the ACUTE-2007 as it is meant to signal short-term fluctuations in risk.

The ACUTE-2007 includes a good selection of dynamic risk factors that can change

rapidly over time. In fact, few sexual recidivism risk tools contain anything like this, so the ACUTE-2007 fills an important gap in ongoing community supervision. It is particularly well suited to assessing changes in risk (Babchishin & Hanson, 2020; Lee et al., 2023), although no studies have yet examined changes in risk over time among men convicted of CSEM offenses. In terms of the Daubert criteria for legal admissibility (see summary in Table 2), similar to the STABLE-2007, the ACUTE-2007 has widespread acceptance for its use among the broader population of men convicted of sexually motivated offenses, but the evidence to support its accuracy for individuals with CSEM offenses has only been available for a few months at the time of writing. Consequently, we expect its use to increase for this group.

Tools Designed to Predict General/Any Recidivism

There are two general recidivism risk assessment tools that have been validated with CSEM samples, both of which include static and dynamic risk factors. We will discuss the research available on them below. Although they each have a large evidence base in the general risk assessment literature, we do not provide a comprehensive discussion of the content and structure of these tools because they were not intended to assess any risk factors specific to sexual offending, much less CSEM offending.

Post-Conviction Risk Assessment (PCRA)

The PCRA (Lowenkamp et al., 2013) is an actuarial risk assessment tool designed in the United States to be used with federal probationers. It includes 15 items assessing central risk domains of criminal history, education/employment, substance use, peers, and attitudes. Total scores can range from 0 to 18, placing individuals into one of four risk levels: low (0 to 5), low-moderate (6 to 9), moderate (10 to 12), and high (13+; Cohen et al., 2020). It had a large effect size (AUC = .71) in a validation sample, and high levels of interrater reliability (agreement of 87% and above across sample cases; Lowenkamp et al., 2013).

Several studies have examined the PCRA with individuals on federal supervision for sex offenses, including CSEM offenses (Cohen, 2018, 2023; Cohen et al., 2020; Cohen & Spidell, 2016). These studies contain largely overlapping samples, although they vary in the sampling timeframe, follow-up, and types of analyses reported. Overall, we considered them as one overarching study, with multiple documents reporting different analyses. The most comprehensive analysis of the predictive accuracy of the PCRA comes from Cohen (2023), which is the same sample used in the American validation study of the CPORT, cited earlier. The PCRA had a significant AUC of .61 (95% CI [.58, .64]) for predicting 5-year rearrests for sexual offenses (n = 5,768), which was very similar to the AUC of .62 for the approximated version of the CPORT.

Although the PCRA had significant and similar predictive accuracy to the CPORT, this program of research has noted several additional challenges with using the PCRA in practice for individuals convicted of sexual offenses and CSEM offenses, given that the PCRA was designed to predict general and not sexual recidivism (Cohen, 2023). Firstly, the focus on risk factors not specific to sexual offending limits the extent to which the PCRA can meaningfully distinguish risk among individuals with CSEM offenses, given their generally low antisociality (Babchishin et al., 2015). For example, roughly 75% of individuals with CSEM offenses score low risk, 21.4% score low-moderate risk, 3.3% score moderate risk, and only 0.4% score high risk on the PCRA (Cohen, 2023). When looking at a subsample of individuals without any contact sexual offenses, it is further skewed, with roughly 81% scoring low risk (Cohen & Spidell, 2016).

These distributions may reflect a truly low risk for general recidivism, but is not particularly helpful for case management decisions, particularly if the goal is to focus on risk for sexual recidivism. Cohen (2023) notes that this is a limitation of using a tool designed for general recidivism when the actual interest is on sexual recidivism, but notably, individuals from

similar timeframes convicted of sexual assault have more spread out risk distributions on the PCRA (with roughly 30%, 43%, 18%, and 9% scoring low, low-moderate, moderate, and high risk, respectively; Cohen & Spidell, 2016).

Another and related practical concern with using a generic risk tool like the PCRA with individuals convicted of CSEM offenses pertains to the use of overrides. Evaluators using the PCRA are permitted to override the risk level if they believe it does not accurately capture the risk posed by the individual. Given consistent research that overriding actuarial tools degrades accuracy (for review, see Helmus, 2021), evaluators are encouraged to use overrides rarely and that it must be justified (Cohen et al., 2020). However, one of the suggested justifications in the policy is because the person is convicted of a sexual offense (Cohen et al., 2020). In a comprehensive study examining 259,571 people initially assessed on the PCRA between 2012 and 2017, Cohen et al. (2020) found that staff would override the PCRA results in 10.6% of cases overall, but more than half of individuals with an index sexual offense had their results overridden. In all cases where the reason for override was because of a sexual offense conviction, risk level was increased (never decreased), and typically increased to high risk, regardless of the starting risk level. In other words, increases of 2 or even 3 risk levels were common, solely based on someone having a sexual offense conviction. Among the 12,321 cases who were overridden to a higher risk level because of a sexual offense, the original PCRA demonstrated moderate accuracy in predicting 2-year rearrests for any recidivism (AUC = .66), but the overridden result demonstrated no accuracy whatsoever (AUC = .51).

Overall, the PCRA does appear to predict sexual recidivism among those with CSEM offenses, however the option to override the risk level can cause issues in practice so a more constrained approach is important. Given that its purpose is to assess risk for general recidivism, it does not include any risk factors specific to sexual offending, and does not provide much

meaningful differentiation among CSEM populations for risk management purposes (i.e., they tend to almost all score as low risk). Consequently, the tool is not particularly useful unless the referral question is concerned about risk for general recidivism and even then, it is important to focus on unadjusted PCRA scores.

Level of Service Inventory (LSI)

The Level of Service Inventory (LSI) family of tools includes several iterations of dynamic actuarial risk assessment tools developed to assess the Central 8 risk factors for general recidivism (Bonta & Andrews, 2023). The LSI tools are widely used and well established for predicting general recidivism (Olver et al., 2014), with good interrater reliability (Labrecque et al., 2017; Rocque & Plummer-Beale, 2004). The Level of Service Inventory – Ontario Revision (LSI-OR; Girard & Wormith, 2004) is a variation used in Ontario that is similar to the Level of Service/Case Management Inventory (LS/CMI; Andrews et al., 2004) utilized in other jurisdictions. The LSI-OR includes 102 items assessing the Central 8 risk domains, as well as additional criminogenic needs, institutional factors, other potential issues, and responsivity considerations (Wormith & Hogg, 2012). The LSI-OR can generate a risk/needs total score, as well as a final risk level out of five (ranging from very low to very high risk). Similar to the PCRA, staff can override the risk level, and again the use of overrides degrades predictive accuracy, particularly for individuals convicted of sexual offenses (Wormith et al., 2012). In a recent study, the LS/CMI was modestly and non-significantly correlated with the CPORT (r = .14), with stronger correlations found between the criminal history subscale of the LS/CMI and the CPORT criminal history items (r's between .21 to .44; Azizian et al., 2023).

Pilon (2016) examined the predictive accuracy of the LSI-OR among 279 men convicted of CSEM offenses, with an average follow-up of 3.2 years. This sample likely had some overlap with the Ontario sample examined by Seto and Eke (2015) and Eke et al. (2019), although the

degree of overlap is expected to be small. Pilon (2016) also examined a modified version of the CPORT with the boy content items missing. The modified CPORT was moderately correlated with the LSI-OR risk/needs total score ($r = .32$, $p < .001$). Unlike the PCRA results above, individuals with CSEM offenses had a more even distribution across the five initial risk levels (16%, 46%, 29%, 7%, and 1% respectively for very low, low, moderate, high, and very high risk levels). Staff tended to override the risk results to higher levels, though the final risk level results did show meaningful spread as well (7%, 20%, 29%, 35%, and 9% respectively).

Neither the LSI-OR risk/needs score, initial risk level, or final risk levels significantly predicted new convictions for CSEM, any sexual offense, or violent recidivism (in fact, most AUCs were below .50; Pilon, 2016). Notably, there were very few recidivists (7 for CSEM recidivism, 8 for sexual, and 3 for violent), limiting the power of this study. The LSI-OR risk/needs score and initial risk level did, however, significantly predict any recidivism (25/279 reoffended; AUCs = .63 and .62, respectively; the final risk level after the override did not, AUC = .57). In this sample the modified CPORT also did not significantly predict CSEM, sexual, or violent recidivism (AUCs between .44 to .56), but did significantly predict general recidivism (AUC = .68).

Given the small sample size and few recidivists, conclusions about the LSI-OR or other LSI tools are preliminary until more research is available. Based on Pilon's (2016) results, the tool should not be used to predict sexual or CSEM recidivism among CSEM offending populations. It can, however, be used to assess risk of general recidivism, and similar to the PCRA, with cautions around the use of overrides. The LSI tools are also helpful in informing case management and treatment targets.

Discussion

Ten years ago, Seto (2013) published a book on individuals who had committed internet

sex offenses, when there was little research available on predictors of recidivism for this population, and only one published validation of a risk tool with a very small sample size (*n* = 73), suggesting that the STABLE-2007 may have some possible utility with this group (Webb et al., 2007). Within a decade, we have gone from no validated risk tools for predicting sexual recidivism among men convicted of CSEM offenses, to having several options to choose from. In fact, several important studies were published in 2023 (Babchishin et al., 2023; Cohen, 2023; Helmus et al., 2024), indicating that the evidence base is expanding rapidly.

In this paper, we reviewed eight risk assessment tools that have been validated for predicting recidivism among men convicted of CSEM offenses; six were designed to predict sexual recidivism (or CSEM recidivism specifically), and two were designed to predict general recidivism. Only one tool, the CPORT, was developed specifically for men convicted of CSEM offenses. Based on current evidence, and the framework of evaluation considerations in Table 2, we think the CPORT, RM2000/S, STABLE-2007, and ACUTE-2007 (in conjunction with the STABLE-2007) are all defensible tools to use for assessing risk of any sexual recidivism or CSEM recidivism specifically among men convicted of CSEM offenses. There is preliminary evidence suggesting that Static-99R is applicable for the broader population of men charged or convicted of CSEM offenses (some of whom are already eligible because of other sexual offenses they have committed), although there are likely better options based on current research (see below). The OSP/I is not a risk tool so much as it is a single (relevant) risk factor and is not helpful in distinguishing levels of risk among men with CSEM histories, given that roughly 90% will be classified in the same level. The PCRA and LSI-OR general recidivism risk tools have some empirical support in predicting general recidivism among CSEM samples (and sexual recidivism for the PCRA), but only when they are used without overrides, which is not consistent with current user guidance for the tools. They are also not informative about risk factors specific

to sexual recidivism; modifications that included sexual risk factors specifically might add to their overall predictive ability.

Although we believe there is sufficient evidence to support using the CPORT, RM2000/S, and STABLE/ACUTE-2007, the research base is still smaller compared to men who have committed offline sexual offenses. This raises important issues about how much evidence is enough to use a tool, particularly when applying it to a setting that is not identical to one of the existing validation studies (for further discussion of this, see Helmus & Olver, 2023). As pointed out in the Joint Committee's (2014) Standards for Psychological Testing, where sample sizes are sufficient, validation research should be conducted for relevant subgroups where there is theory or evidence to suggest differential predictive validity.

Additionally though, even if there are some grounds to suspect differential validity, we must also consider the alternative. If a decision must be made with potentially serious consequences for individual liberty and/or public safety, is it better to presume the validation research on these or other structured tools will generalize to one's setting, or to rely on unstructured clinical judgement? Existing meta-analytic evidence tends to suggest remarkable universality of many risk factors (Bonta & Andrews, 2023), including those for sexual recidivism (Ahmed et al., 2023; Hanson & Morton-Bourgon, 2004, 2009; Helmus & Thornton, 2015; Helmus et al., 2022). Even where differential validity is found, such as for Indigenous men, the evidence suggests enough generalization that using the tools is often still preferable to not using them (Olver et al., in press). Now that there are validation studies available for several sexual recidivism risk tools showing fairly similar results, it is reasonable to assume these tools will apply in most settings, unless there is credible theory or evidence to suggest otherwise.

The available validation studies are currently restricted to Western democratic countries (WEIRD: Western, Educated, Industrialized, Rich, Democratic) that are similar in their CSEM

laws and share similarities in criminal justice processes and systems. Further research is particularly needed in non-Western countries, particularly where cultural and legal norms around CSEM materials are meaningfully different (e.g., Russia, where production/distribution is illegal but possession is not; International Centre for Missing and Exploited Children, 2023) or high speed access to the internet is less common (large parts of Southeast Asia, South America, and Africa). Preliminary evidence suggests that the predictive accuracy of Static-99R generalizes surprisingly well to Asian samples in wealthy countries (Ahmed et al., 2023; Ahmed & Helmus, 2023; Lee et al., 2023), although the risk/need profiles and treatment targets may be meaningfully different. This gives reason for cautious optimism regarding sexual recidivism risk tools with Asian populations, although less is known about other parts of the world.

We do, however, think that there are sufficient theoretical and empirical grounds to doubt that this research will generalize to women or juveniles convicted of CSEM offenses and we do not recommend any of these tools for assessing sexual recidivism risk with these populations. Women convicted of sexual offenses are a minority overall (Cortoni et al., 2017), and this gender discrepancy is even larger for CSEM offending (Seto, 2013). In addition, there is sufficient evidence to suggest their motivations and risk factors for sexually offending are meaningfully different (Cortoni, 2018). For example, Static-99R has not been found to generalize to women (Marshall et al., 2021). There is also considerable research establishing important risk-relevant differences between juveniles and adults who commit sexual offenses, such that they should not be assessed and managed the same way (Chaffin, 2008). Only one study so far examined juveniles with CSEM offenses, and this study found a very low rate of sexual recidivism (Aebi et al., 2014). Furthermore, there are theoretical, developmental, and logistical reasons why producing, distributing, and possessing sexual images of minors would not be indicative of atypical sexual interests among juveniles in the same way as it would be for adults.

Considerations in Choosing a Risk Tool

Based on the current research, it is defensible to use the CPORT, RM2000/S, or the STABLE-2007 (with the ACUTE-2007 as a possible adjunct) in assessing risk of any sexual or specifically CSEM recidivism among men who have been convicted of CSEM offenses. Static-99R may be defensible, and the PCRA or LSI-OR may be defensible in some situations (e.g., particularly for general recidivism). Returning to Table 1, the choice of which tools to use depends on multiple considerations. The CPORT has clear advantages as it is the only tool specifically developed for men convicted of CSEM offenses, it has the largest body of available research, and its average effect sizes are among the highest. This may change as new tools are developed, or more validation studies on generic sexual recidivism risk tools are conducted. For jurisdiction-wide implementation of risk assessment policies and processes, there are logistical advantages to applying the same risk tool (e.g., RM2000/S or Static-99R) for all individuals with sexual offenses. Additionally, it is often the case that the optimal tool or subset of a tool differs depending on the outcome of interest (Babchishin et al., 2016; Hanson & Morton-Bourgon, 2009). Consequently, it could be that CSEM recidivism may be best predicted by the CPORT, and contact sexual recidivism may be better predicted by Static-99R. This is consistent with preliminary results presented by Eke (2022). Tools like the PCRA or LSI may be best for predicting general recidivism among individuals with CSEM offenses. Thus, different tools may be preferred for different purposes. The Static-99R might be preferred when the decision concerns risk of sexual violence and thus a focus on contact sexual recidivism (as in Sexually Violent Predator cases in the United States). The CPORT might be preferred when the concern is about new CSEM offenses, as when a decision is being made by a court or by a probation or parole officer regarding access to the internet. The PCRA might be preferred when the concern is about any new offenses, as when a decision is being made about early release from prison by a

parole board. Access to the information required to score a tool will also play a role in selection.

In many contexts, it is helpful to assess dynamic risk factors to identify treatment targets and to assess changes in risk. As noted by Babchishin et al. (2023), dynamic risk assessment may be particularly helpful among CSEM populations given their generally limited criminal history. Although the RM2000/S has some ability to inform broad domains of treatment needs (Fernandez et al., 2023; Thornton, 2022), the STABLE/ACUTE-2007 are the most advantageous for this purpose. However, given the relative importance of criminal history and static risk factors (Bonta & Andrews, 2023; Brankley et al., 2021), dynamic risk assessments should generally be used in conjunction with a static risk tool, such as the CPORT or RM2000/S (for further discussion of selecting between the CPORT or RM2000/S, see Helmus et al., 2024). There are already structured rules available to combine STABLE-2007 with the RM2000/S (Brankley et al., 2017). These structured rules were developed based on data from men with primarily offline sexual offenses (Helmus et al., 2015) but there is no theory or evidence to suggest the combination rules for two validated tools with CSEM populations should differ from the broader combination rules. It is highly likely, however, that research on combining static and dynamic tools within the same sample of men convicted of CSEM offenses will start to appear in the next few years, if not sooner.

We also hope to see studies examining the predictive accuracy of changes on dynamic risk assessment tools like the STABLE-2007 and ACUTE-2007, more validation studies of the CPORT and other risk assessment tools, more risk assessment studies from non-Western countries, studies following women or juveniles who have committed CSEM offenses, and research with structured professional judgment tools that are sometimes used when assessing individuals who have committed offline sexual offenses, and used to inform case management strategies in CSEM cases.

References

Aebi, M., Plattner, B., Ernest, M., Kaszynski, K., & Bessler, C. (2014). Criminal history and future offending of juveniles convicted of the possession of child pornography. *Sexual Abuse*, *26*(4), 375-390. https://doi.org/10.1177/1079063213492344

Ahmed, S., & Helmus, L. M. (2023). Comparing Indian and White men charged or convicted of sexual offenses on the Static-99R and STABLE-2007. *Psychology, Crime & Law*. Advance online publication. https://doi.org/10.1080/1068316X.2023.2227313

Ahmed, S., Lee, S. C., & Helmus, L. M. (2023). Predictive accuracy of Static-99R across different racial/ethnic groups: A meta-analysis. *Law and Human Behavior*, *47*(1), 275-291. https://doi.org/10.1037/lhb0000517

Andrews, D. A., Bonta, J., & Wormith, J. S. (2004). *The Level of Service/Case Management Inventory (LS/CMI)' User's Manual*. Multi-Health Systems.

Andrews, D. A., Bonta, J., Wormith, J. S., Guzzo, L., Brews, A., Rettinger, J., & Rowe, R. (2011). Sources of variability in estimates of predictive validity: A specification with Level of Service general risk and need. *Criminal Justice and Behavior*, *38*(5), 413-432. https://doi.org/10.1177/0093854811401990

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

Azizian, A., Eke, A. W., Farmus, L., Scott, S., & Seto, M. C. (2023). *Convergent and divergent validity of the Child Pornography Offender Risk Tool (CPORT) in a clinical sample from California* [Manuscript submitted for publication]. Department of Criminology, California State University, Fresno.

Babchishin, K. M., Dibayula, S., McCulloch, C., Hanson, R. K., & Helmus, L. M. (2023). ACUTE-2007 and STABLE-2007 predict recidivism for men adjudicated for child sexual exploitation material offending. *Law and Human Behavior, 47*(5), 606-618. https://doi.org/10.1037/lhb0000540

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., & Blais, J. (2016). Less is more: Using Static-2002R subscales to predict violent and general recidivism among sexual offenders. *Sexual Abuse*, *28*(3), 187-217. https://doi.org/10.1177/1079063215569544

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., 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

Blair, P. R., Marcus, D. K., Boccaccini, M. T. (2008). Is there an allegiance effect for assessment instruments? Actuarial risk assessment as an exemplar. *Clinical Psychology: Science and Practice*, *15*(4), 346-360. https://doi.org/10.1111/j.1468-2850.2008.00147.x

Boer, D. R., Hart, S. D., Kropp, P. R., & Webster, C. D. (2017). *Manual for Version 2 of the Sexual Violence Risk-20, Structured professional guidelines for assessing and managing risk of sexual violence*. Protect International Risk and Safety Services Inc.

Bonta, J., & Andrews, D. A. (2023). *The psychology of criminal conduct* (7th ed.). Routledge.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* Wiley & Sons, Ltd. https://doi.org/10.1002/9780470743386

Brankley, A. E., Babchishin, K. M., & Hanson, R. K. (2021). STABLE-2007 demonstrates predictive and incremental validity in assessing risk-relevant propensities for sexual offending: A meta-analysis. *Sexual Abuse*, *33*(1), 34-62. https://doi.org/10.1177/1079063219871572

Brankley, A., Helmus, L. M., & Hanson, R. K. (2017). *STABLE-2007 evaluator workbook: Updated recidivism rates (includes combinations with Static-99R, Static-2002-R, and Risk Matrix 2000)* [Unpublished report]. Public Safety Canada.

Brouillette-Alarie, S., Babchishin, K. M., Hanson, R. K., & Helmus, L.-M. (2016). Latent constructs of the Static-99R and Static-2002R: A three-factor solution. *Assessment*, *23*(1), 96-111. https://doi.org/10.1177/1073191114568114

Brouillette-Alarie, S., & Hanson, R. K. (2015). Comparaison de deux mesures d'évaluati anadirisque de récidive des délinquants sexuels [Comparison of two measures of recidivism risk assessment of sexual offenders]. *Canadian Journal of Behavioural Science / anadiennedienne des sciences du comportement*, *47*(4), 292-304. https://doi.org/10.1037/cbs0000019

Brown, S. J. (2022). Assessing the risk of users of child sexual exploitation material committing further offences: A scoping review. *Journal of Sexual Aggression*. Advance online publication. https://doi.org/10.1080/13552600.2022.2104394

Chaffin, M. (2008). Our minds are made up - don't confuse us with the facts: Commentary on policies concerning children with sexual behavior problems and juvenile sex offenders. *Child Maltreatment, 13*(2), 110-121. https://doi.org/10.1177/1077559508314510

Cohen, T. H. (2018). Predicting sex offender recidivism: Using the federal Post Conviction Risk Assessment instrument to assess the likelihood of recidivism among federal sex offenders. *Journal of Empirical Legal Studies, 15*(3), 456-481. https://doi.org/10.1111/jels.12184

Cohen, T. H. (2023). Building a risk tool for persons placed on federal post-conviction supervision for child sexual exploitation material offenses: Documenting the federal system's past, current, and future efforts. *Federal Probation, 87*(1), 19-31. https://www.uscourts.gov/federal-probation-journal/2023/06/building-risk-tool-persons-placed-federal-post-conviction

Cohen, T. H., Lowenkamp, C. T., Bechtel, K., & Flores, A.W. (2020). Risk assessment overrides: Shuffling the risk deck without any improvements in prediction. *Criminal Justice and Behavior*, *47*(12), 1609-1629. https://doi.org/10.1177/0093854820953449

Cohen, T. H., & Spidell, M. C. (2016). How dangerous are they? An analysis of sex offenders under federal post-conviction supervision. *Federal Probation*, *80*(2), 21-32. https://www.uscourts.gov/federal-probation-journal/2016/09/how-dangerous-are-they-analysis-sex-offenders-under-federal-post

Cortoni, F. (2018). *Women who sexually abuse: Assessment, treatment, & management*. Safer Society Press.

Cortoni, F., Babchishin, K. M., & Rat, C. (2017). The proportion of sexual offenders who are female is higher than thought: A meta-analysis. *Criminal Justice and Behavior, 44*(2), 145-162. https://doi.org/10.1177/0093854816658923

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993), 113 S. Ct. 2786, 125 L. Ed. 2d 469.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243*(4899), 1668-1674. https://doi.org/10.1126/science.2648573

Eke, A. W. (2022, October 26-29). Predictive accuracy of the Static-99R in two samples of men adjudicated for child sexual exploitation material offending. In K. M. Babchishin (Chair), *Risk assessment with unique groups: Child pornography and other non-contact sexual offenses* [Symposium]. Association for the Treatment of Sexual Abusers 41st Annual Research & Treatment Conference, Los Angeles, CA, United States.

Eke, A. W., Helmus, L. M., & Seto, M. C. (2018). *Scoring guide for the Child Pornography Offender Risk Tool (CPORT): Version 2*.

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

Fernandez, Y., Gotch, K., Hanson, R. K., & Harris, A. J. R. (2015). *ACUTE-2007 coding manual – Revised 2015* [Unpublished report]. Public Safety Canada.

Fernandez, Y., Harris, A. J. R., Hanson, R. K., & Sparks, J. (2014). *STABLE-2007 coding manual – Revised 2014* [Unpublished report]. Public Safety Canada.

Fernandez, Y. M., & Helmus, L. M. (2017). A field examination of the inter-rater reliability of the Static-99 and STABLE-2007 scored by Correctional Program Officers. *Sex Offender Treatment, 12*(2), 1-9.

Fernandez, Y. M., Helmus, L. M., & Thornton, D. (2023, September 28-30). *Using Risk Matrix to identify treatment needs* [Paper presentation]. Association for the Treatment of Sexual Abusers 42nd Annual Research & Treatment Conference, Aurora, CO, United States.

Frye v. United States, 293 F. 1013, 34 A.L.R. 145 (D.C. Cir. 1923).

Gannon, C., Blokland, A. A. J., Huikuri, S., Babchishin, K. M., & Lehmann, R. J. B. (2023). Child sexual abuse material on the darknet. *Forensische Psychiatrie, Psychologie, Kriminologie, 17*, 353-365. https://doi.org/10.1007/s11757-023-00790-8

Garrington, C., Kelty, S. F., Rickwood, D., & Boer, D. (2022). The development of the ERICSO – A proposed instrument for internet child abuse material offender assessment. *The*

Journal of Forensic Practice, *24*(4), 453-470. https://doi.org/10.1108/JFP-07-2022-0033

Girard, L., & Wormith, J. S. (2004). The predictive validity of the Level of Service Inventory-Ontario Revision on general and violent recidivism among various offender groups. *Criminal Justice and Behavior*, *31*(2), 150-181. https://doi.org/10.1177/0093854803261335

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. (2009). The psychological assessment of risk for crime and violence. *Canadian Psychology/Psychologie Canadienne*, *50*(3), 172-182. https://doi.org/10.1037/a0015726

Hanson, R. K. (2022). *Prediction statistics for psychological assessment.* American Psychological Association. https://doi.org/10.1037/0000275-000

Hanson, R. K., Babchishin, K. M., Helmus, L. M., Thornton, D., & Phenix, A. (2017). Communicating the results of criterion-referenced prediction measures: Risk categories for the Static-99R and Static-2002R sexual offender risk assessment tools. *Psychological Assessment*, *29*(5), 582-597. https://doi.org/10.1037/pas0000371

Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sex offender recidivism studies. *Journal of Consulting and Clinical Psychology*, *66*(2), 348-362. https://doi.org/10.1037/0022-006X.66.2.348

Hanson, R. K., Harris, A. J. R., Scott, T.-L., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project* (Corrections Research User Rep. No. 2007-05). Public Safety Canada.

Hanson, R. K., & Morton-Bourgon, K. (2004). *Predictors of sexual recidivism: An updated meta-analysis* (User Report 2004-02). Public Safety and Emergency Preparedness Canada. https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/2004-02-prdctrs-sxl-rcdvsm-pdtd/2004-02-prdctrs-sxl-rcdvsm-pdtd-eng.pdf

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., & 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

Harris, G. T., Rice, M. E., & Quinsey, V. L. (2010). Allegiance or fidelity? A clarifying reply. *Clinical Psychology: Science and Practice, 17*(1), 82-89. https://doi.org/10.1111/j.1468-2850.2009.01197.x

Hart, S. D., Kropp, P. R., Watt, K. A., Darjee, R, Davis, M. R., Klaver, J., Laws, D. R., & Logan, C. (2022). *RSVP-V2: Version 2 of the Risk for Sexual Violence Protocol, Structured professional judgement guidelines for assessing and managing risk for sexual violence*. Protect International Risk and Safety Services Inc.

Helmus, L. M. (2018). Sex offender risk assessment: Where are we and where are we going*? Current Psychiatry Reports*, *20(*46), 1-9. https://doi.org/10.1007/s11920-018-0909-8

Helmus, L. M. (2021). Estimating the probability of sexual recidivism among men charged or convicted of sexual offenses: Evidence-based guidance for applied evaluators. *Sexual Offending: Theory, Research, and Prevention, 16*, 1-24. https://doi.org/10.5964/sotrap.4283

Helmus, L. M. (2023). Recidivism rates of men charged/convicted of child pornography offenses. *The ATSA Forum Newsletter, 35*(2). http://newsmanager.commpartners.com/atsa/issues/2023-03-15/2.html

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., 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., Eke, A. W., Farmus, L., & Seto, M. C. (2024). The CPORT and Risk Matrix 2000 for men convicted of child sexual exploitation material (CSEM) offenses: A predictive accuracy comparison and meta-analysis. *Criminal Justice and Behavior*, *51*(1), 3-23. https://doi.org/10.1177/00938548231208194

Helmus, L., Hanson, R. K., Babchishin, K. M., & Mann, R. E. (2013). Attitudes supportive of sexual offending predict recidivism: A meta-analysis. *Trauma, Violence, & Abuse, 14*(1), 34-53. https://doi.org/10.1177/1524838012462244

Helmus, L., Hanson, R. K., Babchishin, K. M., & Thornton, D. (2015). Sex offender risk assessment with the Risk Matrix 2000: Validation and guidelines for combining with the STABLE-2007. *Journal of Sexual Aggression*, *21*(2), 136-157. https://doi.org/10.1080/13552600.2013.870241

Helmus, L., Hanson, R. K., & Morton-Bourgon, K. E. (2011). International comparisons of the validity of actuarial risk tools for sexual offenders, with a focus on Static-99. In D. P. Boer, 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 & Sons, Ltd. http://doi.org/10.1002/9781119990420.ch4

Helmus, L. M., Hanson, R. K., Murrie, D. C., & Zabarauckas, C. L. (2021). Field validity of Static-99R and STABLE-2007 with 4,433 men serving sentences for sexual offenses in British Columbia: New findings and meta-analysis. *Psychological Assessment*, *33*(7), 581-595. https://doi.org/10.1037/pas0001010

Helmus, L., Hanson, R. K., Thornton, D., Babchishin, K. M., & Harris, A. J. R. (2012). Absolute recidivism rates predicted by Static-99R and Static-2002R sex offender risk assessment tools vary across samples: A meta-analysis. *Criminal Justice and Behavior*, *39*(9), 1148-1171. https://doi.org/10.1177/0093854812443648

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

Helmus, L. M., Lee, S. C., Phenix, A., Hanson, R. K., & Thornton, D. (2021). *Static-99R and Static-2002R Evaluators' Workbook*. Society for the Advancement of Actuarial Risk Needs Assessment (SAARNA). https://saarna.org/static-99/#

Helmus, L. M., & Olver, M. E. (2023). *What evidence is needed to use a recidivism risk tool in court?* [Unpublished manuscript]. Department of Criminology, Simon Fraser University.

Helmus, L. M., & Thornton, D. (2015). Stability and predictive and incremental accuracy of the individual items of Static-99r and Static-2002r in predicting sexual recidivism: A meta-analysis. *Criminal Justice and Behavior*, *42*(9), 917-937. https://doi.org/10.1177/0093854814568891

Helmus, L., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2012). Improving the predictive accuracy of Static-99 and Static-2002 with older sex offenders: Revised age weights. *Sexual Abuse*, *24*(1), 64-101. https://doi.org/10.1177/1079063211409951

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behaviour and Brain Sciences*, *33*(2-3), 61-83. https://doi.org/10.1017/S0140525X0999152X

Henshaw, M., Darjee, R., & Clough, J. A. (2020). Online child sexual offending. In I. Bryce & W. Petherick (Eds.), *Child sexual abuse: Forensic issues in evidence, impact, and management* (1st ed., pp. 85-108). Academic Press. https://doi.org/10.1016/B978-0-12-819434-8.00005-2

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.

Hilbert, J. (2019). The disappointing history of science in the courtroom: *Frye*, *Daubert*, and the ongoing crisis of junk science in criminal trials. *Oklahoma Law Review*, *71*(3), 759-821. https://digitalcommons.law.ou.edu/olr/vol71/iss3/3/

His Majesty's Prison and Probation Service. (2023). *OASys Sexual reoffending Predictor (OSP): Guidelines for practitioners, Version 3*. Public Protection Group. https://assets.publishing.service.gov.uk/media/64c24bfe7aea5b0014075045/osp-guidance-practitioners.pdf

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

Howard, P., & Wakeling, H. (2021). *Comparing two predictors of sexual recidivism: The Risk Matrix 2000 and the OASys Sexual Reoffending Predictor.* HM Prison & Probation

Service, Ministry of Justice Analytical Series. https://www.gov.uk/government/publications/comparing-2-predictors-of-sexual-recidivism-rm2000-and-osp

International Centre for Missing and Exploited Children. (2023). *Child sexual abuse material: Model legislation & global review* (10th Ed.). https://cdn.icmec.org/wp-content/uploads/2023/10/CSAM-Model-Legislation_10th-Ed-Oct-2023.pdf

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. https://www.apa.org/science/programs/testing/standards

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

Labrecque, R. M., Campbell, C. M., Elliott, J., King M., Christmann, M., Page, K., McVay, J., & Roller, K. (2018). An examination of the inter-rater reliability and rater accuracy of the Level of Service/Case Management Inventory. *Corrections, 3*(2), 105-118. https://doi.org/10.1080/23774657.2017.1323253

Laudon, K. C. (1986). Data quality and due process in large interorganizational record systems. *Communications of the ACM*, *29*(1), 4-11. https://doi.org/10.1145/5465.5466

Lee, S. C., Babchishin, K. M., Mularczyk, K. P., & Hanson, R. K. (2023). Dynamic risk scales degrade over time: Evidence for reassessments. *Assessment*. Advance online publication. https://doi.org/10.1177/10731911231177227

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

Long, M., Alison, L., Tejeiro, R., Hendricks, E., & Giles, S. (2016). KIRAT: Law enforcement's prioritization tool for investigating indecent image offenders. *Psychology, Public Policy, and Law*, *22(*1), 12-21. https://doi.org/10.1037/law0000069

Lovins, B., Lowenkamp, C. T., & Latessa, E. J. (2009). Applying the risk principle to sex of fenders: Can treatment make some sex offenders worse? *The Prison Journal*, *89*(3), 344-357. https://doi.org/10.1177/0032885509339509

Lowenkamp, C. T., Johnson, J. L., Holsinger, A. M., VanBenschoten, S. W., & Robinson C. R. (2013). The federal Post Conviction Risk Assessment (PCRA): A construction and validation study. *Psychological Services*, *10*(1), 87-96. https://doi.org/10.1037/a0030343

Lowenkamp, C. T., & Latessa, E. J. (2005). Increasing the effectiveness of correctional programming through the risk principle: Identifying offenders for residential placement. *Criminology & Public Policy, 4*(2), 263-290. https://doi.org/10.1111/j.1745-9133.2005.00021.x

Mann, R. E., Hanson, R. K., & Thornton, D. (2010). Assessing risk for sexual recidivism: Some proposals on the nature of psychologically meaningful risk factors. *Sexual Abuse*, *22*(2), 191-217. https://doi.org/10.1177/1079063210366039

Marshall, E., Miller, H. A., Cortoni, F., & Helmus, L. M. (2021). The Static-99R is not valid for women: Predictive validity in 739 females who have sexually offended. *Sexual Abuse*, *33*(6), 631-653. https://doi.org/10.1177/1079063220940303

Neal, T. M. S., Sellbom, M., & de Ruiter, C. (2022). Personality assessment in legal contexts:

Introduction to the special issue. *Journal of Personality Assessment, 104*(2), 127-136. https://doi.org/10.1080/00223891.2022.2033248

Nitsche, K., Etzler, S., Balas, J., Eher, R., & Rettenberger, M. (2022). A field study of acute dynamic risk assessment in individuals convicted of sexual offenses. *Psychological Assessment, 34*(6), 583-594. https://doi.org/10.1037/pas0001123

Olver, M. E., Stockdale, K. C., Helmus, L. M., Woods, P., Termeer, J., & Prince, J. (in press). Too risky to use, or too risky not to? Lessons learned from over 30 years of research on forensic risk assessment with Indigenous persons. *Psychological Bulletin.*

Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2014). Thirty years of research on the Level of Service scales: A meta-analytic examination of predictive accuracy and sources of variability. *Psychological Assessment, 26*(1), 156-176. https://doi.org/10.1037/a0035080

Olver, M. E., Wong, S. C. P., Nicholaichuk, T., & Gordon, A. (2007). The validity and reliability of the Violence Risk Scale-Sexual Offender version: Assessing sex offender risk and evaluating therapeutic change. *Psychological Assessment, 19*(3), 318-329. https://doi.org/10.1037/1040-3590.19.3.318

Perley-Robertson, B., Babchishin, K. M., & Helmus, L. M. (in press). The effect of missing item data on the predictive accuracy of correctional risk assessment tools. *Assessment.*

Phenix, A., Fernandez, Y., Harris, A. J. R., Helmus, M., Hanson, R. K., & Thornton, D. (2016). *Static-99R coding rules: Revised – 2016* (Research report: 2017–R012). Public Safety Canada. https://www.publicsafety.gc.ca/cnt/rsrcs/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. http://hdl.handle.net/10388/ETD-2016-01-2414

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

Rocque, M., & Plummer-Beale, J., (2014). In the eye of the beholder? An examination of the inter-rater reliability of the LSI-R and YLS/CMI in a correctional agency. *Journal of Criminal Justice, 42*(6), 568-578. https://doi.org/10.1016/j.jcrimjus.2014.09.011

SAARNA (2023-06-29). *STABLE-2007 and ACUTE-2007 FAQ and Other Updates*. SAARNA : Author. Available on www.saarna.org (to certified STABLE/ACUTE users only).

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*(5), 568-596. https://doi.org/10.1177/10790632211047190

Scurich, N., & Krauss, D. A. (2023). Risk assessment of child-pornography-exclusive offenders. *Law and Human Behavior, 47*(4), 499-509. https://doi.org/10.1037/lhb0000537

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), 416-429. https://doi.org/10.1037/lhb0000128

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

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 offenses, Canada, 2020. https://www150.statcan.gc.ca/n1/daily-quotidien/210727/t001a-eng.htm*

Thornton, D. (2017). *Scoring guide for Risk Matrix 2000 S & V Scales: International Version.*

Thornton, D. (2022). *2022 evaluator guide for Risk Matrix 2000 S & V Scales: International Version.* www.saarna.org

Thornton, D., Fernandez, Y. M., & Helmus, L. M. (2023). *Scoring guide for Risk Matrix 2000 S & V Scales: International Version.* www.saarna.org

Thornton, D., & Helmus, L. M. (2022, June). *Using multiple actuarial instruments to assess sexual recidivism risk*. Online training for Society for the Advancement of Actuarial Risk Needs Assessment (SAARNA).

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.

Viljoen, J. L., Vargen, L. M., Cochrane, D. M., Jonnson, M. R., Goossens, I., & Monjazeb, S. (2021). Do structured risk assessments predict violent, any, and sexual offending better than unstructured judgment? An umbrella review. *Psychology, Public Policy, and Law*, *27*(1), 79-97. https://doi.org/10.1037/law0000299

von Franqué, F., Bergner-Koether, R., Schmidt, S., Pellowski, J. S., Peterson, J. H., Hakak, G., & Briken, P. (2023). Individuals under voluntary treatment with sexual interest in minors: What risk do they pose? *Frontiers in Psychiatry, 14*, Article 1277225. https://doi.org/10.3389/fpsyt.2023.1277225

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

Wakeling, H. C., Mann, R. E., & Milner, R. J. (2011). Interrater reliability of Risk Matrix 2000/s. *International Journal of Offender Therapy and Comparative Criminology, 55*(8), 1324-1337. https://doi.org/10.1177/0306624X11386933

Webb, L., Craissati, J., & Keen, S. (2007). Characteristics of internet child pornography offenders: A comparison with child molesters. *Sexual Abuse, 19*(4), 449-465. https://doi.org/10.1007/s11194-007-9063-2

Wormith, J. S., & Hogg, S. M. (2012). *The predictive validity of Aboriginal offender recidivism with a general risk/needs assessment inventory*. University of Saskatchewan. https://cfbsjs.usask.ca/documents/research/research_papers/LSI-OR%20Aboriginal%20Paper%20w%20Abstract.pdf

Wormith, J. S., Hogg, S., & Guzzo, L. (2012). The predictive validity of a general risk/needs assessment inventory on sexual offender recidivism and an exploration of the professional override. *Criminal Justice and Behavior*, *39*(12), 1511-1538. https://doi.org/10.1177/0093854812455741

Table 1

Considerations in Using a Risk Tool from Helmus & Olver (2023)

Source	Consideration or Criterion	Comments
Helmus & Olver (2023) Considerations for General Applied Use	**Bare Minimum**	
	Predictive accuracy (discrimination)	Statistically significant with AUC of .56+
	Additional Considerations	
	Closest match between tool and referral question/context	What outcome(s) does it predict? Does it inform relevant case management decisions? Is the type of information needed for this tool available in this context?
	Volume/quality of research	Number, quality, consistency of research studies.
	Availability and quality of recidivism estimates	Not necessary for all risk assessment decisions. Ideally based on 100+ recidivists, but lower can still be informative. Stability across samples/jurisdictions should not be assumed.
	Interrater reliability	Consistency across raters.
	Comprehensive and empirically supported risk factors	Are all items empirically supported? Comprehensiveness (taking into account point of diminishing returns). Plausibility of latent constructs as psychologically meaningful causal factors.
	Inform treatment targets and assessments of change	Not necessary for all decisions. Static risk tools provide some information, but stronger evidence comes from dynamic factors and research on change.
	Quality of training/implementation resources	Comprehensiveness and availability of coding manual and training; norms and resources/templates for communicating results; ability to ask questions of development team; mechanism for revising tool as new research becomes available.
Daubert Criteria for Legal Admissibility	Whether the technique in question can be and has been tested.	
	Whether it has been subject to peer review and publication.	
	Its known or potential error rate.	
	The existence and maintenance of standards controlling its operation.	
	Whether it has attracted widespread acceptance within a relevant scientific community.	

Table 2

Evaluation of Risk Tools Validated for CSEM Samples

Consideration or Criterion	CPORT	RM2000/S	OSP/I	Static-99R	STABLE-2007	ACUTE-2007	PCRA	LSI-OR
Purpose Designed For								
Outcome	Sex recid	Sex recid	CSEM recid	Sex recid	Sex recid	Sex recid	General recid	General recid
Population (offense type)	CSEM	All sex offenses	All sex offenses	(Nearly) all sex offenses	(Nearly) all sex offenses	(Nearly) all sex offenses	All offenses	All offenses
Type of risk factors	Static	Static	Static	Static	Dynamic	Dynamic	Static/dynamic	Static/dynamic
Defensible to Use?	Yes	Yes	Partially (single dichotomous risk factor for CSEM individuals)	Partially (for CSEM group as whole; too little research on non-dual offenders)	Yes	Yes, with STABLE-2007	No	For general recidivism only (without override)
Predictive accuracy (discrimination)	Large AUC (.75) without outlier; moderate to large (.68 to .71) with outlier; moderate AUC for CSEM recidivism	Moderate AUC (.66 to .67)	Estimated as small AUC of .59	Roughly small to moderate (exact values not reported)	Large *C* (.78+) for any sex or CSEM recidivism.	Large *C* (.71+) for any sex or CSEM recidivism.	Small (.61) for any sexual recidivism	Small (.62 to .63 for general recidivism); AUCs below .50 for CSEM and sexual recidivism
Closest match between tool and referral question	Predicts any sex and CSEM recidivism	Predicts any sex and CSEM recidivism	Predicts CSEM recidivism	Tentatively, predicts any sex, contact, and CSEM recid among full CSEM sample	Predicts any sex and CSEM recidivism	Predicts any sex and CSEM recidivism	Predicts any sexual recidivism	Predicts general recidivism only
Volume/quality of research	Most of any tool for CSEM; 7 predictive validity studies	2 predictive validity studies (one quite large)	1 study	2 small samples from 1 conference presentation	1 field validity study	1 field validity study	1 study	1 small study with few sexual recidivists

Availability and quality of recidivism estimates	Preliminary	Generic; not CSEM population	Preliminary	Generic; not CSEM population	Combined with RM2000; Not tested for CSEM group	Not available	Preliminary (from Cohen, 2023)	For general recidivism only for general offending samples
Interrater reliability	Strong	Strong	Not reported but should be strong	Strong	Strong	Good	Strong	Good
Comprehensive and empirically supported risk factors	Good sampling of static risk factors	Good sampling of static risk factors	No. Single risk factor.	Good sampling of static risk factors	Good sampling of dynamic risk factors	Good sampling of acute risk factors	Good sampling of general risk factors, not sexual	Good sampling of general risk factors, not sexual
Inform treatment targets and assessments of change	Indirectly only (general antisociality and atypical sexuality)	Some guidance for treatment need indicators	No	Indirect only (general antisociality and atypical sexuality)	Yes	Assesses more rapidly fluctuating changes	Yes	Yes
Quality of training / implementation resources	Good	Good	Adequate	Good	Good	Good	Good	Good
Daubert Criteria for Legal Admissibility								
Can be and has been tested.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subject to peer review and publication.	Yes	Yes	Yes (UK government)	Not yet (conference)	Yes	Yes	Not for CSEM	No (dissertation)
Its known or potential error rate.	See above	See above	See above	Approximated	See above	See above	See above	See above
Standards controlling its operation.	Good	Good	Good	Yes (but coding rules not ideal)	Good	Good	Good	Good
Widespread acceptance within a relevant scientific community.	Yes. Most frequently used and researched.	Yes, though the research support is more recent	Adopted by UK government	As yet, only if also have Category A offense	Broadly yes; for CSEM, it's new	Broadly yes; for CSEM, it's new	Yes for general recidivism only	Yes for general recidivism only