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

Thomas H. Cohen[1]
Probation and Pretrial Services Office
Administrative Office of the U.S. Courts

OVER THE LAST two decades, the number of persons placed on federal supervision for Child Sexual Exploitation Material (CSEM)[2] offenses has increased exponentially. The surge in CSEM supervisees can be attributed to technological changes that allow for easier access to sexually explicit materials on the internet and federal laws and enforcement mechanisms that have resulted in growing numbers of persons convicted of CSEM under federal sentencing statutes (Faust & Motivans, 2015; U.S. Sentencing Commission [USSC], 2012, 2021; Wolak et al., 2005, 2009). Specifically, the use of various technologies, including peer-to-peer networks, texting and instant messaging, cloud-based hosting services, social media platforms, and chatrooms, has created the context in which the typical person convicted of CSEM offenses will have generated voluminous collections of graphical images, including those of very young children (USSC, 2012, 2021). Moreover, federal legislation, particularly the Prosecutorial Remedies and Other Tools to End the Exploitation of Children Today Act of 2003 (The PROTECT Act), has resulted in increased penalties for persons convicted of CSEM through the addition of new enhancements and mandatory minimums to the federal sentencing guidelines (USSC, 2021). The PROTECT Act also gave federal judges the discretion to impose life supervision terms on persons convicted of federal sex offenses (Faust & Motivans, 2015; USSC, 2012). In addition to these technological and legislative changes, numerous regional taskforces and specialized units have been established by the U.S. Department of Justice to prosecute persons engaged in CSEM offenses (Wolak et al., 2005).

These trends combined have resulted in substantial increases in the number of persons prosecuted, incarcerated, and (most importantly for this research) placed on federal post-conviction supervision for CSEM offenses[3] (Faust & Motivans, 2015; U.S. Sentencing Commission [USSC], 2012, 2021). Faust and Motivans (2015) report that the number of persons placed on federal post-conviction supervision for sex offenses increased by 1,400 percent, from 321 supervisees in 1994 to 4,714 supervisees, in 2013. Much of this increase could be attributed to the prosecution of persons charged with CSEM offenses (i.e., possession, receipt, distribution, or production of child pornography). Moreover, persons convicted of CSEM offenses are increasingly being sentenced to lengthy post-conviction supervision terms in the federal system. The average term imposed on nearly 4,700 CSEM supervisees placed on federal post-conviction supervision during fiscal years 2010 through 2016 was about 98 months (see Table 1). In comparison, the average term imposed on federal supervisees in 2010 was about 43

[1] Thomas H. Cohen, Senior Social Science Analyst, Administrative Office of the U.S. Courts, Washington, D.C. The author would like to thank Scott VanBenschoten, Michelle Spidell, Stacy Merolla, Vignesh Narayanaswamy, Ann Smith, and Mariana Madrid for their helpful suggestions and comments. This publication benefited from the careful editing of Ellen Fielding. Direct correspondence to Thomas H. Cohen, Administrative Office of the U.S. Courts, One Columbus Circle NE, Washington, D.C. 20544 (email: thomas_cohen@ao.uscourts.gov).

[2] In prior research of federal sex offenders conducted by the Federal Probation and Pretrial Services Office (PPSO), the term "child pornography offender" was used to refer to persons placed on supervised release for possessing, receiving, distributing, or producing child pornography. Given the efforts to discourage the use of the word "offender," the term CSEM or CSEM supervisee was substituted for child pornography offender.

[3] Federal post-conviction supervision refers to persons sentenced to a term of community supervision following a period of imprisonment within the Federal Bureau of Prisons (18 U.S.C. §3583). Probation refers to persons sentenced to a period of federal supervision without any imposed incarceration sentence (18 U.S.C. §3561).

months (USSC, 2012).

The growth of CSEM supervisees presents serious challenges to the federal supervision system. Prior research shows many CSEM supervisees initially being designated as low risk to reoffend according to the federal Post Conviction Risk Assessment (PCRA) instrument (Cohen & Spidell, 2016; Cohen 2018); however, these persons tend to be placed through overrides into the highest supervision levels (Cohen et al., 2020). Officers' propensity to override CSEM supervisees stems from concerns about whether these persons have histories of, or are likely to engage in, offline contact sexual behavior with children (DeLisi et al., 2016). A meta-analysis focusing on the backgrounds of CSEM persons, for example, found that about 12 percent had an official arrest or conviction record of contact sexual behavior, but 55 percent admitted through self-reporting that they had prior sexual contact with children (Seto et al., 2011). Moreover, the risk tool officers used to gauge the likelihood of recidivism for federal supervisees is not calibrated to measure sexual deviance, ascertain the presence of non-official contact sex behavior, or assess the risk of sexual recidivism for sex offenders generally or CSEM supervisees in particular (Cohen & Spidell, 2016; Cohen 2018).

TABLE 1.
Descriptive Statistics of Study Sample

Descriptive factors	n	Percent or mean
Race/Ethnicity		
White, not Hispanic	5,113	88.8 %
Hispanic, any race	377	6.6
Black	157	2.7
Asian or Pacific Islander	76	1.3
American Indian or Alaska Native	30	0.5
Other	6	0.1
Average age	5,768	46.1
Type of supervision		
Term of supervised release	5,615	97.4 %
Probation	153	2.7
Most serious conviction offense/a		
Possession of child pornography	3,806	66.0 %
Distribution/receipt/transportation of child pornography	1,483	25.7
Sexual exploitation of children	462	8.0
Transfer obscene material to minors	118	2.1
Production of child pornography	39	0.7
Obscenity	6	0.1
PCRA risk levels		
Low	4,323	75.0 %
Low/moderate	1,234	21.4
Moderate	190	3.3
High	21	0.4
Average PCRA score	5,768	4.1
Supervision time imposed (Months)/b	4,695	97.7
Rearrested for any sex offense within 60 months of initial PCRA assessment	262	4.5 %
Number of supervisees	5,768	

Note: Includes 5,768 male supervisees convicted of online sex offenses placed on federal supervision between fiscal years 2010 through mid-2016 whose rearrest activity could be tracked for 60 months.
PCRA = Post Conviction Risk Assessment
a\Will not sum to 5,768 because supervisees can be convicted of multiple offenses.
b\Post-conviction supervision time imposed available for 81% of the CSEM study population.

The combination of growing numbers of persons on federal supervision for CSEM offenses, concerns about the frequency with which this population has engaged in unrecorded contact sexual behavior, and issues with using the current risk tool employed by federal probation officers to assess the risk of sexual recidivism (i.e., the PCRA) gave rise to an initiative by the Administrative Office of the U.S. Courts, Probation and Pretrial Services Office (PPSO), to construct a risk tool that could gauge the likelihood of sexual recidivism for the CSEM population.

This article documents PPSO's efforts to construct a risk tool that could be used on persons placed on federal post-conviction supervision for CSEM offenses. Initially, the article will delve into federal policies for supervising persons convicted of CSEM offenses and contrast those policies with an examination of how the CSEM population is actually being supervised. Next, it will detail PPSO's attempts to build a risk tool based on the Child Pornography Offender Risk Tool (CPORT) (see Eke et al., 2018; 2019), PCRA (see Johnson et al., 2011; Lowenkamp et al., 2013; 2015) and FBI criminal history records to assess the risk of sexual recidivism for CSEM supervisees. Specifically, the paper will describe the methods, data, and principal findings stemming from PPSO's efforts to use the CPORT and an amalgamation of fields obtained from the CPORT, PCRA, and FBI criminal history files to predict sexual recidivism for CSEM supervisees. Additional work involving the use of machine learning to gauge the likelihood of sexual reoffending for CSEM supervisees will also be detailed. Ultimately, as will be shown, none of these efforts were successful in creating a risk tool that officers could use for CSEM supervisees. The paper will conclude by discussing the implications of PPSO's efforts to build a CSEM-specific risk tool and suggest possible alternatives for future research.

Federal Policies and Practices for Supervising Persons Convicted of CSEM Offenses

PPSO has responded to the growing number of persons convicted of CSEM offenses under federal supervision and the concerns that some CSEM supervisees might be involved in hands-on offending by issuing guidance for federal officers charged with supervising these persons. Under current policy, officers are instructed to use information gleaned from both the PCRA and other sources of information, including presentence reports, polygraphs, and psychosexual evaluations, to conduct an initial risk assessment evaluation. The PCRA is a dynamic actuarial instrument developed for federal probation officers that classifies supervisees into a matrix containing 12 risk categories (Administrative Office of the U.S. Courts (AO), 2018: see page 14). These categories provide crucial information about a supervisee's likelihood of committing any or violent offenses both during and after the supervisee has completed the supervision term (for more information about the PCRA, see Johnson et al., 2011; Lowenkamp et al., 2013; Lowenkamp et al., 2015; Luallen et al., 2016; Serin et al., 2016).

Although the PCRA provides crucial information about a supervisee's propensity for reoffending, it is not geared towards CSEM supervisees, constructed to assess their likelihood of sexual reoffending, or designed to measure sexual deviance. Additionally, nearly all CSEM supervisees (97 percent) are classified as low or low/moderate risk according to the PCRA (Cohen & Spidell, 2016). The lack of any official contact record for many CSEM supervisees, combined with their low-risk classification status, initially produced a policy in which officers were required to place all CSEM supervisees into the highest supervision levels regardless of their original risk classification (Cohen & Spidell, 2016; Cohen, 2018). Officers responded to this policy by applying overrides as a means of supervising nearly all CSEM supervisees at the highest supervision levels (Cohen et al., 2016; 2020).

This policy underwent a revision several years ago. Specifically, officers are no longer required to place all CSEM supervisees into the highest supervision levels through overrides. Rather, in December 2017, the policy was changed to acknowledge the risk principle and account for variations of risk within this population. As a result, officers are now encouraged to consider a combination of factors when designating the levels of supervision intensity. At the onset of supervision, officers may have limited case information and rely on known recidivism rates of CSEM individuals and suggested PCRA risk levels. However, during the course of supervision, as information related to the risk and needs of the case change, officers should respond by adjusting supervision levels as necessary.[4]

While policy no longer mandates that officers place CSEM supervisees into higher risk categories, evidence suggests that officers continue to use overrides to elevate the supervision levels for these persons. An examination of nearly 6,900 CSEM supervisees who received PCRA assessments between fiscal years 2017-21 showed officers overriding 96 percent of these persons and placing nearly all of them into the moderate or high supervision categories (data not shown). Moreover, officers typically keep these supervisees in the highest supervision categories even after multiple assessments. The continued use of supervision overrides for CSEM supervisees and the intensity of resources and staff directed at CSEM supervision provided the impetus for PPSO to develop an actuarial tool that could be used to supervise this specific population of sex offenders. The remainder of this paper details PPSO's effort to use the CPORT and a combination of CPORT, PCRA, and criminal history risk factors to construct a tool that could be used to assess the risk of sexual recidivism for CSEM supervisees.

Using the CPORT to Assess the Risk of Sexual Recidivism for CSEM Supervisees

In order to address the challenges inherent in supervising persons convicted of CSEM offenses, PPSO decided to attempt to assess whether the CPORT alone, or in combination with the PCRA and criminal history files, could be used to provide officers with a means of accurately gauging a CSEM supervisee's risk of engaging in sexual recidivism. PPSO selected the CPORT because of a growing literature showing its efficacy in differentiating the risk of sexual recidivism for persons convicted of CSEM offenses (Black, 2018; Eke et al., 2018, 2019; Pilon, 2016; Savoie et al., 2022; Seto & Eke, 2015; Soldino et al., 2021). Specifically, the CPORT was created to gauge the risk of any sexual recidivism among a population of adult males convicted of CSEM offenses (Eke et al., 2018, 2019; Seto & Eke, 2015; Soldino et al., 2021).[5] This risk instrument was originally constructed using a sample of 266 males convicted of CSEM offenses in Canada whose arrest activity could be followed for a period of five years and then validated on an additional sample of 80 men (Seto & Eke, 2015; Eke et al., 2019; Soldino et al., 2021). The CPORT's developers showed that this tool was effective at predicting any sexual recidivism (Area Under the Curve (AUC =. 74)) as well as sexual recidivism for CSEM subpopulations with histories of contact sexual behavior (AUC = .80) or backgrounds of general criminal activity not involving contact sexual offending (AUC = .69) (Seto & Eke, 2015). The tool's predictive efficacy, however, degraded when predicting sexual recidivism for CSEM subpopulations with only a history of child pornography offenses (AUC = .63) (Seto & Eke, 2015).

Subsequent CPORT studies showed the tool manifesting mixed effectiveness in terms of its ability to predict sexual recidivism for persons convicted of CSEM offenses. In a study conducted on 141 adult CSEM males in Scotland, the authors found that the CPORT significantly predicted various forms of reoffending behavior, including recidivism for any offenses (AUC = .81), any sexual offenses (AUC = .78), and CSEM offenses (AUC = .74) (Savoie et al., 2022). Other studies, however, produced results that did not replicate the CPORT's original predictive effectiveness. Using a truncated version of the CPORT scale[6] on 279 persons convicted of CSEM offenses in Canada and with a follow-up period of over

[4] It should be noted that about 1 out of 5 persons on federal supervision for CSEM offenses has a valid Static-99/R score (Cohen & Spidell, 2016). The Static-99/R is an actuarial risk prediction instrument that estimates the probability of sexual and/or violent reconviction for adult males who have already been charged with or convicted of at least one contact sexual offense against a child or non-consenting adult (Hanson et al., 2016; Helmus & Hanson, 2007). The Federal Bureau of Prisons attempts to score the Static-99/R on all sex offenders; however, valid scores are calculated for only those persons with current or prior arrest/conviction records for contact sex offending. Should the CSEM supervisee have a Static-99/R score, policy mandates that the officer default to the risk tool (i.e., PCRA or Static-99/R) that recommends the highest levels of supervision intensity.

[5] For a complete overview of the CPORT items, see Eke et al. (2018), as well as this paper's methods section.

[6] The truncated version omitted two items (questions #6 and #7) measuring the content of boy vs. girl material in the collections of persons convicted of CSEM offenses.

three years, Pilon (2016) generated results in the mediocre predictive range (AUC = .56). Another study conducted by Black (2018) using a shortened version of the CPORT scale[7] covering 547 persons with CSEM convictions in New Zealand and tracking their arrest activity for a period ranging from 2 to 19 years, found effect sizes ranging from the small (AUC = .60) to large (AUC >= .80), depending upon the arrest outcome examined. Last, Soldino et al. (2021) examined the CPORT's predictive efficacy on a sample of 304 men arrested for CSEM offenses in Spain and tracked for a duration of 5 years. The Soldino et al. (2021) study used the complete CPORT scale as well as the Correlation of Admission of Sexual Interest in Children (CASIC)[8]; overall, the results were mixed, with the CPORT total scores mostly producing AUC values of below .60 irrespective of the presence or absence of missing data. The authors, however, were able to generate AUC scores of .70 when applying the CASIC to a subset of the study population (Soldino et al., 2021).

Conducting a Pilot Test of the CPORT

Given the CPORT's potential effectiveness as a risk classification tool for CSEM supervisees, PPSO decided to ascertain whether this instrument could be integrated into the federal supervision system. The effort to integrate the CPORT occurred through two initiatives. Initially, PPSO attempted to conduct a pilot test of the CPORT by bringing in about 20 probation officers to manually code the CPORT on a random sample of 200 persons placed on federal post-conviction for CSEM offenses. For the pilot effort, PPSO contracted with one of the CPORT developers (Doctor Angela Eke) and she, along with Detective Sergeant Monica Denreyer, trained the federal probation officers on how to accurately code this risk tool. As a result of this training, CPORT and CASIC data were coded for 195 CSEM supervisees placed on federal post-conviction supervision between fiscal years 2011 and 2012.[9] The coding primarily involved examining presentence reports (PSRs) and other materials produced at supervision intake. While a great deal of information was learned from the pilot, unfortunately the officers had difficulty coding the CPORT items measuring boy to girl content in the child pornography material (CPORT item #6) and nude/other material (CPORT item #7). Moreover, officers were unable to code most of the CASIC items to determine a CSEM supervisee's sexual interest in children or teenagers. The combination of high rates of missing data for several CPORT and CASIC items, along with a relatively low rate of sexual reoffending for the pilot sample (only 9 of the 195 persons sexually reoffended), resulted in AUC scores in the mediocre to poor range (AUC = .54) for this instrument.

As a result of the pilot's poor performance, PPSO decided to rethink how to empirically test the CPORT's predictive performance for persons placed on federal supervision for CSEM offenses. Ultimately, PPSO decided to conduct a larger test of the CPORT using a population of 5,768 male supervisees placed on federal post-conviction supervision between fiscal years 2010 through mid-2016. Rather than have officers manually code the CPORT items, PPSO contracted with the MITRE Corporation (hereafter MITRE) to conduct a text mining endeavor aimed at collecting the CPORT elements. The efforts the MITRE project entailed, along with the data elements collected and analyzed, are further detailed in the methods section.

Method

CPORT and CASIC elements

Extracting both the CPORT and CASIC elements from PPSO's case management system (i.e., The Probation and Pretrial Services Automated Case Tracking System or PACTS) is problematic, because many of the risk factors scored in these instruments are not readily available for electronic data extraction. Persons attempting to score the CPORT, for example, are required to mark the following items as present or absent: (1) age at the time of index investigation, 35 or younger; (2) any prior criminal history; (3) any failure on conditional release, including charge at index; (4) any contact sexual offending, including a charge at index; (5) indication of pedophilic or hebephilic interests; (6) more boy than girl content in the child pornography material; (7) more boy than girl content in the nude/other child material (Eke et al., 2018). It's also important to note the instrument allows scorers to substitute the CASIC as a method for assessing CPORT item #5 (indication of pedophilic or hebephilic interests). The CPORT's developers suggested using the CASIC in lieu of attempting to directly ascertain the presence of sexual interest in children or teenagers because of concerns that many persons being scored on this instrument would not readily admit to these deviant forms of sexual behavior (Eke et al., 2018; Soldino et al., 2021).

The CASIC measures whether the CSEM supervisee manifests key characteristics associated with admission of pedophilic or hebephilic sexual interests (Eke et al., 2018; Seto & Eke, 2017). In the CASIC, six items are coded as being present or absent: (1) never married; (2) had child pornography videos; (3) had child pornography text stories; (4) child pornography material spanning two or more years; (5) volunteering in a role with high access to children; (6) engaging in online sexual communications with a minor or undercover officer posing as a minor (Eke et al., 2018; Seto & Eke, 2017). CASIC scores of 3 or higher are indicative that the person being scored is sexually interested in children or teenagers and hence should receive a score for CPORT item #5.

Several of the CPORT and CASIC items are stored in PPSO's case management system (i.e., PACTs) in a format that allows for further analysis. For example, the CPORT items measuring age and criminal history (CPORT items #1 through #4) are available in PACTS and can be readily extracted to generate a truncated CPORT score. The remaining CPORT items (items #5 through 7), however, are not entered into the PACTs system in a structured format that can be easily retrieved, assuming they are entered at all. For example, a CSEM supervisee's admission of sexual interest in children or preference in boys over girls might be manifested in the text embedded in a presentence report or psychosexual assessment uploaded into PACTs, but this information is typically stored in unstructured PDF files or images; none of these items are entered into specific numeric fields.

Hence, any attempt to obtain these data would involve officers having to read through case files and manually code the CPORT items measuring sexual interest in children or teenagers or preference in boys over girls.[10] The

[7] This version omitted the CPORT's last three items, including question #5 (indication of pedophilic or hebephilic interests) and questions # 6 and #7 (measuring boy vs. girl content).

[8] The CASIC is used to assess a CSEM person's sexual interest in children or teenagers (see Seto & Eke, 2015).

[9] Five supervisees were removed from the sample because subsequent data obtained from PPSO's case management system showed they did not meet the criteria of persons who should be scored on the CPORT.

[10] Obtaining data for the CASIC is even more challenging; only 1 of the 6 items (never married) could be readily extracted through PACTs.

level of time, effort, and resources involved in obtaining this information through a review of PDFs or scanned documents uploaded into PACTs could be enormous given the number of CSEM cases officers would potentially have to code. Generally, the rates of sexual recidivism for CSEM supervisees are relatively low. Cohen and Spidell (2016) showed about 3 percent of CSEM supervisees being rearrested for sexual offenses within three years of their supervision start dates. Given the low base rates of sexual re-offending for this population, any effort to validate the CPORT on this population would involve collecting CPORT information on potentially thousands of CSEM supervisees. The challenge, therefore, was to devise a way to collect CPORT data through a mechanism that minimized having officers individually go through case files while simultaneously extracting the CPORT elements from as many cases as possible. Ultimately, the AO's Department of Technology Services contracted with MITRE to engage in a proof-of-concept project on the feasibility of applying natural language processing and machine learning techniques to retrieve CPORT data elements from thousands of CSEM supervisees. The MITRE project and its results are detailed below.

The MITRE Data Collection Effort

The MITRE project's primary goal was to construct an algorithm for extracting information from various documents to complete the CPORT risk tool from an initial list of 8,896 male CSEM supervisees placed on federal post-conviction supervision between fiscal years 2011 through 2018.[11] MITRE extracted unstructured text data from numerous sources, including PSRs, with a particular emphasis on the sections containing information on charges and convictions, mental and emotional health, personal and family data, offense conduct, and victim impact statements; polygraph reports; and psychosexual assessments and psychological evaluations. In total, MITRE processed an estimated 11,000 PSR documents,[12] 60,000 psychological and psychosexual assessments, and 55,000 polygraph reports. The process resulted in the analysis of about 126,000 PDF and scanned documents containing over 8 million sentences. To these 8 million sentences, MITRE applied a combination of content extraction, natural language processing, and artificial intelligence reasoning capacities to automatically produce responses that could be used to complete the CPORT risk instrument.[13] The entire automated process took about 12 days to complete. In comparison, if PPSO had opted for manual data collection, and if the amount of time required to complete the instrument were similar to the pilot effort (about one hour per CSEM supervisee), it is estimated that it would have taken four full-time staff about one year to manually code the CPORT for the same 8,896 CSEM supervisees.

It is important to note that while the MITRE effort produced results that mostly adhered to the CPORT data elements, there was some divergence between the MITRE-generated and CPORT fields. CPORT element #5 (indication of pedophilic or hebephilic interests), for example, was split into two elements measuring the presence of pedophilic or hebephilic interests separately (see Table 3). In addition, CPORT element #6 (more boy than girl content in the child pornography material) and CPORT element #7 (more boy than girl content in the nude/other child material) were combined into one field measuring whether the CSEM supervisee evidenced greater sexual interests in boys over girls. Moreover, the MITRE project attempted to gather several additional elements that could be associated with sexual recidivism. This effort involved measuring the presence or absence of the following elements: evidence of deviant sexual interests (a catchall category created by MITRE); lives with lover or partner for less than 2 years; engaged in online communications for illicit purposes; any prior non-contact sexual offenses; and any prior violent (non-sexual) offenses.

Though the MITRE project involved a novel initiative to transform unstructured text files into structured datasets for nearly 9,000 CSEM persons on federal supervision, some limitations about this project should be noted. First, it's important to acknowledge that MITRE relied on admissions, rather than on an examination of actual child pornography collections, to gauge preferences for boys over girls. Admissions, and not CASIC, were also used to ascertain the presence of pedophilic or hebephiliac interests. The use of admissions over these other forms of obtaining the CPORT data could explain some of the differences in the study's primary findings compared to previous CPORT research.

Inclusion of Elements From PCRA and Rap Sheets as Additional Predictors of Sexual Recidivism

In addition to the factors extracted through the MITRE project, this effort attempted to determine whether any other factors collected by the federal post-conviction risk assessment tool (i.e., PCRA), the generalized assessments conducted on all supervisees, or the criminal history data embedded in rap sheets were correlated with sexual recidivism. An effort was made to examine these other factors because of the concerns that the CPORT might not be predictive of sexual recidivism, given the low rates of reoffending activity among CSEM supervisees (Cohen & Spidell, 2016; USSC, 2021). The specific non-CPORT elements identified through this effort, the processes for selecting these elements, and their predictive efficacy are further detailed in the findings section.

Outcome Measure

The primary outcome of interest involves whether the CSEM supervisee was rearrested for any new sexual offenses. Rearrests for new criminal activity were obtained from the National Crime Information Center (NCIC) and Access to Law Enforcement System (ATLAS). ATLAS is a software program used by the AO that provides an interface for performing criminal record checks through a systematic search of official state and federal rap sheets (Baber, 2010). Sexual recidivism was defined to include arrests for any sexual offenses—either violent or non-violent but excluding prostitution offenses—within a fixed five-year time frame from the supervision start date. Similar to other CPORT validation studies, an attempt was made to distinguish contact from non-contact sexual recidivism events; however, there were so few CSEM supervisees arrested for contact sex crimes (less than 1 percent) that it was ultimately not practicable to separate out these arrest outcomes in the extant study.

The five-year follow-up period aligns with the tracking time used in the CPORT development study (Eke et al., 2019; Seto & Eke, 2015) as well as subsequent CPORT validation efforts (Soldino et al., 2021). The decision to use a five-year fixed follow-up period resulted

[11] The challenges inherent in obtaining CASIC items necessitated that we focus solely on the CPORT for this project. An effort was made, however, to collect the CASIC field measuring online communications with a minor.

[12] Some CSEM supervisees have multiple PSRs.

[13] For a more in-depth overview of the processes MITRE applied to data-mine the judicial system's text documents, see Megerdoomian et al. (2019), which discusses this effort for a related PPSO-sponsored project.

in 3,128 of the 8,896 CSEM supervisees whose CPORT data were collected by MITRE being removed from the analysis because their arrest outcomes could not be tracked for a minimum of five years. The remaining cohort of 5,768 male CSEM supervisees who were included in the current study, however, constitute one of the largest samples attempting to validate the CPORT ever conducted.

Analytical Approach

The statistical techniques applied to this analysis involved a combination of descriptive techniques, chi-square tests, and AUC-ROC scores. The AUC-ROC scores were primarily used to assess the predictive accuracy of the CPORT risk tool as well as the risk tool PPSO constructed, which combined elements from the CPORT, PCRA, and rap sheets. In addition to these techniques, an attempt was made to apply machine learning approaches to predict sexual recidivism. Specifically, random forest machine learning approaches were employed to assess whether these novel methods could substantially improve prediction compared to traditional risk assessment approaches. The random forest analyses are further discussed in the findings section.

Findings

Validating the CPORT

Descriptive information about the study sample is provided in Table 1. Of the 5,768 CSEM supervisees in the study population, nearly 90 percent were non-Hispanic whites, while the remainder were a combination of Hispanics, Blacks, Asian/Pacific Islanders, or American Indians or Alaska Natives. The average age was about 46 years, and almost the entire population (97 percent) were placed on post-conviction supervision through a term of supervised release, meaning that these supervisees had served time in federal prison before being released. Over 90 percent of the study population were convicted of CSEM offenses involving child pornography possession (66 percent) or the distribution, receipt, or transportation of child pornography (26 percent). About 1 percent were convicted of actually producing child pornography materials. The study population skewed low risk, with three-quarters receiving a low-risk classification from the PCRA; about 4 percent were assessed as moderate or high risk. By comparison, about one-fourth of the general federal supervision population are classified as moderate or high risk at initial assessment (Johnson & Baber, 2015). The rates at which CSEM supervisees recidivated for sexual offenses were also relatively low. Approximately 5 percent of the 5,768 CSEM supervisees were rearrested for any sexual offenses within 60 months of their supervision start date. In contrast, 43 percent of all federal supervisees were rearrested for any new offenses within 5 years from supervision commencement (Markman et al., 2016). The low-risk distribution skew for the CSEM study population, combined with their minimal rates of sexual recidivism, gives rise to various challenges for risk assessment construction, development, and validation that are subsequently detailed.

Information about the PCRA's capacity to predict sexual recidivism among CSEM supervisees is provided in Table 2. Overall, the PCRA's capacity to predict sexual recidivism for the 5,768 CSEM supervisees placed on federal supervision is in the weak range (AUC = .61. 95% CI [.58 - .64]). The PCRA's inability to differentiate CSEM supervisees by risk is especially apparent when examining the sexual recidivism rates for low/moderate and moderate CSEM supervisees, which are essentially the same (8.1 percent vs. 7.9 percent). These findings further illustrate the need to move beyond the PCRA and apply other tools (e.g., CPORT) in attempting to distinguish the risk of sexual recidivism for the federal CSEM population.

The presence of the CPORT and other risk factors generated by MITRE are provided in Table 3 (next page) in a sorted format. The MITRE data collection effort showed over a third of the CSEM population evidencing sexual interests in children or teenagers through admissions to officers, treatment providers, or polygraph administrators and nearly two-fifths were 35 years or younger at the time of index investigation. Approximately one-fifth manifested any criminal history, but only 2 percent were determined by MITRE to have a background of contact sexual offending. The rates of prior contact sex offending are lower than those reported in other studies of CSEM supervisees (see Cohen & Spidell, 2016) and ultimately resulted in an effort to supplement the criminal history backgrounds of these persons with FBI rap sheet data (see next section). Last, MITRE identified 6 percent of CSEM supervisees evidencing greater sexual interests in boys over girls.

Information about the presence of other (non-CPORT) risk factors generated by MITRE is also provided in Table 3. The most common other risk factors included evidence of deviant sexual interests (57 percent) and lives with lover or partner for less than two years (30 percent). About 12 percent of the study population engaged in online communication for illicit purposes and less than 5 percent had an arrest record for non-contact sexual or violent offenses.

Data on the bivariate associations between the MITRE-generated risk factors—both CPORT and other—and the five-year sexual recidivism rates are provided in Table 4 (next page). Several of the CPORT risk factors were shown to be significantly associated with sexual recidivism ($p <.05$), including age at index investigation, presence of pedophilic interests, presence of previous criminal history, any failure on conditional release, and presence of contact sexual reoffending. Of all the CPORT risk factors, any failure while on conditional

TABLE 2.
Association Between Post Conviction Risk Assessment (PCRA) Risk Levels and Any Sexual Recidivism for Online Sex Offenders

PCRA risk levels	n	Percent sexually recidivated
All supervisees	5,768	4.5 %
Low	4,323	3.3 %
Low/moderate	1,234	8.1
Moderate	190	7.9
High	21	28.6
AUC-ROC	0.61 [0.58 - 0.64]	

Note: Includes 5,768 male supervisees convicted of online sex offenses placed on federal supervision between fiscal years 2010 through mid-2016 whose rearrest activity could be tracked for 60 months.
PCRA = Post Conviction Risk Assessment

release and presence of contact sexual offending had the strongest associations with sexual recidivism; CSEM supervisees with these characteristics were about three times more likely to sexually recidivate compared to the overall baseline sexual recidivism rates. Interestingly, the CPORT factors measuring the presence of hebephiliac interests and greater sexual interests in boys over girls were not associated with sexual recidivism. Among the other risk factors produced by MITRE, only those measuring the presence of prior non-contact sex offenses and violent offenses manifested significant associations with sexual recidivism. Over 15 percent of CSEM supervisees with these characteristics were rearrested for sexual offenses.

The predictive effectiveness of the CPORT risk tool and various modified versions of this tool for CSEM supervisees are detailed in Table 5 (next page). Initially, an attempt was made to ascertain the CPORT's efficacy by assigning scores of 0 or 1 to each CPORT risk factor and summing the scores into a total score; the scores were included in the sum irrespective of whether they were significantly associated with sexual recidivism (see Table 4). Using this approach generated some differentiation in the sexual recidivism rates. For example, the percentage of CSEM supervisees who sexually recidivated increased somewhat incrementally from 2 percent of supervisees with no CPORT risk criteria (score = 0) to 8 percent of supervisees with at least three CPORT risk factors (score = 3). CSEM supervisees with five or more CPORT risk factors were 10 times more likely to be rearrested for sexual offenses (20 percent rearrested) compared to their counterparts with zero CPORT risk factors (2 percent rearrested). Despite these promising patterns, the overall AUC scores for the CPORT are in the mediocre predictive range (AUC = .62. 95% CI [.58 - .65]). The low-risk skew of the CSEM population—56 percent manifested CPORT scores of ranging from to 0 to 1—provides a partial explanation for these poor prediction metrics.

Attempts were made to evaluate whether the CPORT's predictive effectiveness could be enhanced by modifying this risk tool. Specifically, the modifications involved constructing an assessment score that included all the CPORT and other risk factors generated by MITRE regardless of their significant association with sexual recidivism as well as constructing a truncated assessment score that used only those CPORT and other risk factors significantly associated with sexual

TABLE 3.
Presence of CPORT or Other Risk Factors Associated with Sexual Recidivism for Online Sex Offenders

Risk items	Number	Percent
All supervisees	5,768	
CPORT risk factors		
35 years or younger at time of index investigation	2,168	37.6 %
Presence of indication of pedophilic interest	2,151	37.3
Presence of indication of hebephiliac interest	2,018	35.0
Presence of previous criminal history	1,276	22.1
Evidences greater sexual interest in boys over girls	350	6.1
Any failure on conditional release/a	236	4.1
Presence of contact sexual offending/a	130	2.3
Other risk factors		
Evidence of deviant sexual interests	3,265	56.6 %
Lives with lover or partner for less than 2 years	1,745	30.3
Engaged in online communication for illicit purpose	705	12.2
Presence of prior non-contact sex offenses	122	2.1
Presence of prior violent (non-sexual) offenses	101	1.8

Note: Includes 5,768 male supervisees convicted of online sex offenses placed on federal supervision between fiscal years 2010 through mid-2016 whose rearrest activity could be tracked for 60 months.
CPORT = Child Pornography Offender Risk Tool
The CPORT and other risk factors shown in table were generated by MITRE.
a/Includes previous or instant offenses

TABLE 4.
Association Between Individual CPORT or Other Risk Factors Associated with Sexual Recidivism for Online Sex Offenders

Risk items	n	Percent sexually recidivated
All supervisees	5,768	4.5 %
CPORT risk factors		
35 years or younger at time of index investigation	2,168	6.2 %***
Presence of indication of pedophilic interest	2,151	5.3 *
Presence of indication of hebephiliac interest	2,018	3.7
Presence of previous criminal history	1,276	7.9 ***
Evidences greater sexual interest in boys over girls	350	4.9
Any failure on conditional release/a	236	15.7 ***
Presence of contact sexual offending/a	130	14.6 ***
Other risk factors		
Evidence of deviant sexual interests	3,265	4.7 %
Lives with lover or partner for less than 2 years	1,745	4.9
Engaged in online communication for illicit purposes	705	4.1
Presence of prior non-contact sex offenses	122	18.9 ***
Presence of prior violent (non-sexual) offenses	101	15.8 ***

Note: Includes 5,768 male supervisees convicted of online sex offenses placed on federal supervision between fiscal years 2010 through mid-2016 whose rearrest activity could be tracked for 60 months.
CPORT = Child Pornography Offender Risk Tool
The CPORT and other risk factors shown in table were generated by MITRE.
Chi-square used to indicate statistical significance * $P < .05$; ** $P < .01$; *** $P < .001$
a/Includes previous or instant offenses

reoffending (see Table 4 for information about the types of risk factors significantly associated with sexual recidivism). The approach employing all the risk factors constructed by MITRE also produced sub-par predictive indices (AUC = .61. 95% CI [.57 - .64]). Conversely, employing a technique where only those risk factors significantly associated with sexual recidivism were included in the assessment calculations produced the highest AUC scores (AUC = .65. 95% CI [.62 - .69]) and patterns of sexual reoffending that increased somewhat *monotonic*ally by risk score. Though promising, even this method failed to generate predictive AUC scores in the high effect size range (e.g., AUC score > .70) (Rice & Harris, 2005).

TABLE 5.
Association Between CPORT Risk Scores and Risk Scores Using Other Factors with Sexual Recidivism for Online Sex Offenders

Risk scores	n	Percent sexual recidivated
All supervisees	5,768	4.5 %
Only CPORT risk factors		
0	1,314	2.4 %
1	1,907	4.0
2	1,526	4.2
3	765	7.7
4	210	10.0
5 plus	46	19.6
AUC-ROC	0.62 [0.58 - 0.65]	
CPORT plus other risk factors		
0	724	2.5 %
1	1,054	3.1
2	1,233	4.2
3	1,225	4.2
4	865	5.2
5	463	9.1
6	150	8.0
7	37	13.5
8 plus	17	23.5
AUC-ROC	0.61 [0.57 - 0.64]	
CPORT and other risk scores (reduced)/a		
0	1,748	2.4 %
1	2,407	3.7
2	1,187	6.0
3	322	10.6
4	84	22.6
5 plus	20	35.0
AUC-ROC	0.65 [0.62 - 0.69]	

Note: Includes 5,768 male supervisees convicted of online sex offenses placed on federal supervision between fiscal years 2010 through mid-2016 whose rearrest activity could be tracked for 60 months.
The scores calculated in table were based on factors generated by MITRE.
CPORT = Child Pornography Offender Risk Tool
95% confidence intervals shown.
a/ Reduced risk characteristics selected from CPORT and other factors significantly associated with any sexual recidivism at .05 level.

Building a CSEM Risk Instrument Based on CPORT, PCRA, and Criminal History Factors

Given the issues pertaining to sex offender prediction using the CPORT, PPSO decided to rethink its approach to developing a risk tool for CSEM supervisees. Specifically, an attempt was made to ascertain whether an in-house risk tool could be developed using data elements from a multitude of sources including the MITRE-generated CPORT and other risk factors, the risk elements collected by officers when conducting PCRA assessments, supervisee characteristics generated from officer assessments, and the FBI criminal history data. Elements were selected from these sources if they were associated with an increase of over three percentage points in the likelihood of sexual recidivism occurring within five years of the supervision start date. Though selecting elements through this approach might be viewed as less rigorous compared to selecting elements that are statistically significant, given the low base rates of sexual recidivism (4.5 percent), this method seemed to offer the best means for building a risk tool that could predict sexual recidivism among CSEM supervisees. In order to avoid the pitfall of generating a risk tool that overfits the data, and hence might not be useable when applied to a new group of CSEM supervisees, the database was randomly split into a training and testing data file. The variables associated with an increase of over three percentage points with sexual rearrest activity were selected from the training database and then applied to the testing file for the purpose of assessing this instrument's potential predictive efficacy.

The specific variables selected for CSEM risk construction and development are detailed in Table 6 (next page). In the training dataset, the following elements were selected from the MITRE-generated factors: presence of previous criminal history, evidences greater sexual interest in boys over girls, any failure on conditional release, presence of contact sexual offending, presence of prior non-contact sexual offenses, and presence of prior violent (non-sexual) offenses. All of these factors—with the exception of evidences greater sexual interest in boys over girls—were associated with significantly higher likelihoods of sexual recidivism in Table 4. Though the boy over girl content was not statistically associated with higher rearrest rates, in the training data this variable was correlated with a more than three percentage point increase in the likelihood of sexual recidivism and hence was included as a potential predictor variable.

Several non-MITRE risk factors also were associated with an increase of more than three percentage points in the likelihood of sexual recidivism. Many of these factors hailed from the PCRA and included officer scores measuring whether a supervisee manifested social problems associated with drug use or negative attitudes towards supervision or had a record of institutional adjustment. Another factor, denial of harm, hails from the Psychological Inventory of Criminal Thinking Styles section of the PCRA and essentially measures if the supervisee either rationalizes or minimizes the harm their criminal lifestyle might have done to others (Walters, 2013). In addition to these factors, an assessment indicating that the supervisee had a record of domestic violence was also associated with sexual recidivism. Last, the presence of an FBI record indicating that the supervisee had an arrest history for sex offenses (prostitution excluded) was shown to be associated with a more than three percentage point increase in sexual reoffending. The FBI criminal history records augmented the MITRE criminal history data, since MITRE recorded relatively few CSEM supervisees having any arrest histories for sexual offenses. The factors listed in Table 6 all received scores of 0 or 1 depending upon whether their presence was recorded for the CSEM supervisee, and their individual scores were summed to generate a total score. The predicative effectiveness of these total scores for both the training and testing data are shown in the next table and figure.

Results from the hybrid approach explicated above are provided in Table 7. Overall, the AUC-ROC scores approach acceptable levels for the training data (AUC = .68. 95 percent CI [.63 - .73]); however, there is a slight though not significant deterioration when moving to the testing data (AUC = .65. 95 percent CI [.60 - .70]). Among both the training and testing samples, CSEM supervisees with higher risk scores were more likely to sexually recidivate compared to their counterparts who scored lower on the assessment instrument. For example, the percentage of CSEM supervisees in the testing sample rearrested for sexual offenses manifested the following incremental increases: 3 percent (score = 0), 6 percent (score = 2), 19 percent (score = 4), and 41 percent (score = 5). The rearrest rates do fall off when moving to scores of 6 or above; however, that pattern is partially explained by the small number of CSEM supervisees (n = 5)

TABLE 6.
MITRE and non-MITRE Generated Factors Used to Predict Sexual Recidivism for Online Sex Offenders

	Training data		Testing data	
Risk items	Number	Percent	Number	Percent
MITRE generated factors/a				
Presence of previous criminal history	635	22.0 %	641	22.2 %
Evidences greater sexual interest in boys over girls	186	6.5	164	5.7
Any failure on conditional release	121	4.2	115	4.0
Presence of contact sexual offending	63	2.2	67	2.3
Presence of prior non-contact sex offenses	53	1.8	69	2.4
Presence of prior violent (non-sexual) offenses	48	1.7	53	1.8
Non-MITRE factors/b				
PCRA - Social problems associated with drug use	346	12.0 %	327	11.3 %
PCRA - Negative attitudes towards supervision	417	14.5	391	13.6
PCRA - Institutional adjustment	359	12.5	365	12.7
PCRA - Denial of harm	36	1.4	29	1.1
Assessment - Domestic violence	102	3.5	84	2.9
RAP Sheets -Prior arrests for sex offenses				
None	2,571	89.2	2,610	90.5
One or more	313	10.9	274	9.5
All supervisees	2,884		2,884	

Note: Includes 5,768 male supervisees convicted of online sex offenses placed on federal supervision between fiscal years 2010 through mid-2016 whose rearrest activity could be tracked for 60 months.
a/MITRE generated factors.
b/Information obtained from PCRA, PICTS, Assessments, and Rap sheets.
Unless otherwise noted, data available for 99% of supervisees with exception of denial of harm factor.
Data on denial of harm factor available for 93% of supervisees.

TABLE 7.
Association Between Calculated Risk Scores Using MITRE and Other Risk Factors with Any Sexual Recidivism Using Training and Testing Data

	Training data		Testing data	
Risk scores	n	Percent sexually recidivated	n	Percent sexually recidivated
All	2,652	4.2 %	2,643	4.9 %
0	1,241	2.1 %	1,266	2.8 %
1	781	3.7	769	4.7
2	370	7.0	359	6.4
3	156	5.8	164	8.5
4	68	17.7	63	19.1
5	27	25.9	17	41.2
6 plus	9	33.3	5	20.0
AUC-ROC	0.68 [0.63-0.73]		0.65 [0.60 - 0.70]	

Note: MITRE and other risk factors used to calculate risk scores selected from factors associated with a three percentage point increase in sexual recidivism.
Includes 5,295 male supervisees convicted of online sex offenses whose risk scores could be calculated and whose rearrest activity could be tracked for 60 months.

receiving these high scores.

Figure 1 highlights the predictive efficacy of the PPSO-generated risk tool with the combined training and testing data. Results show somewhat incremental increases in the sexual rearrest rates by risk score. In general, the sexual recidivism rates rise from 2.5 percent to 4.2 percent when moving from scores of 0 to 1; afterwards they plateau at about 7 percent between scores 2 and 3 and then move up again to 18 percent and then 32 percent for persons scoring 4 and 5, respectively. The combined data produces predictive metrics that approach (AUC = .67. 95% CI [.63 - .70]) but do not meet, nor exceed, the acceptable range for most risk instruments (AUC > .70).

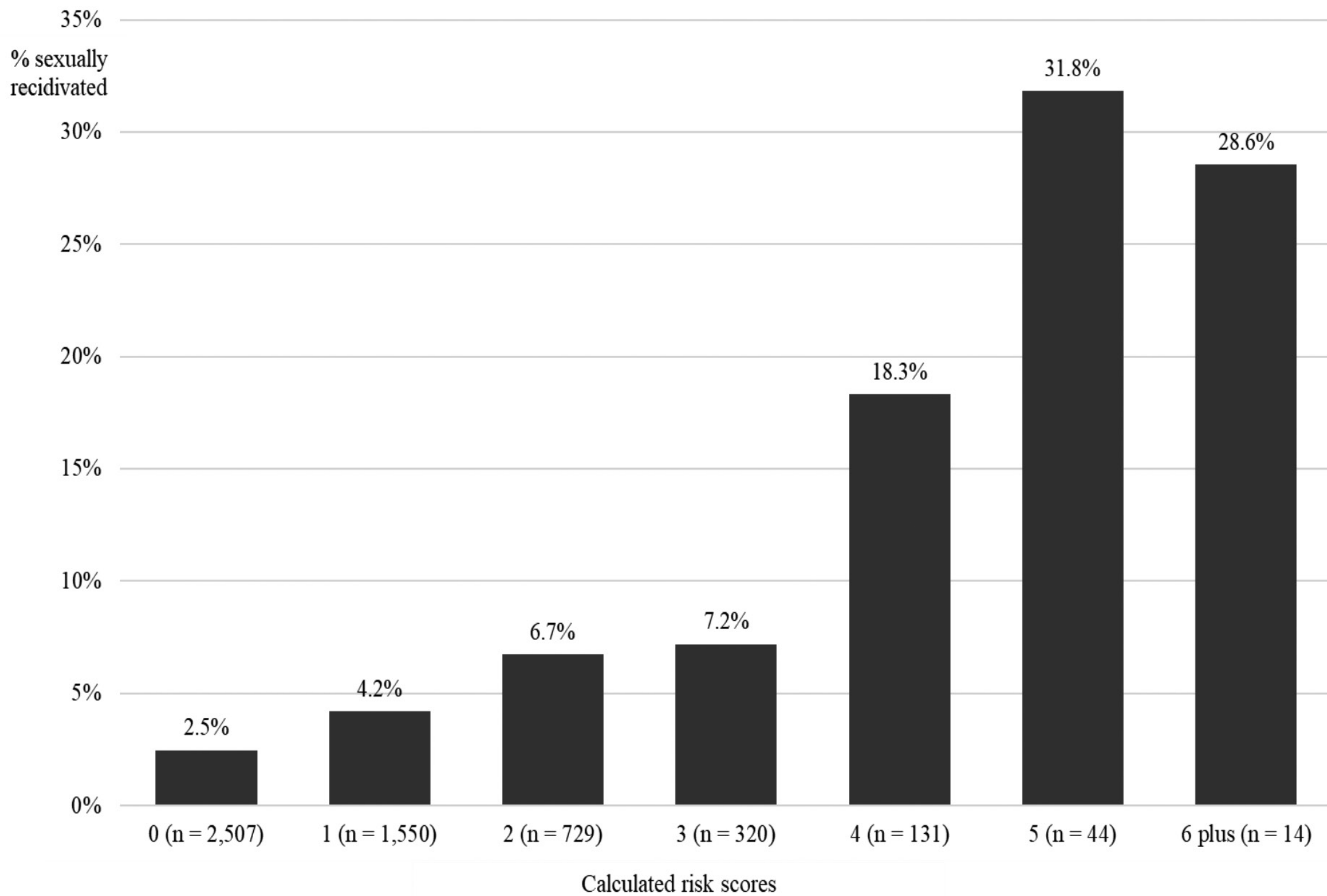
Using Machine Learning Approaches for CSEM Prediction

Though attempting to produce an in-house risk instrument geared to CSEM supervisees generated results that almost met the acceptable range of most criminal justice and sex offender risk assessments, this attempt fell short. Ultimately, PPSO was unable to construct a risk tool that could appreciably differentiate the risk of sexual recidivism among CSEM supervisees by using either the CPORT or a combination of CPORT and other risk factors embedded within the PCRA, officer assessments, and FBI criminal history records. In light of these results, PPSO made an additional effort to ascertain the feasibility of constructing an in-house risk assessment through the use of machine learning techniques. A brief description of PPSO's effort to apply machine learning applications to the problem of CSEM risk prediction is subsequently provided.

Machine learning is essentially an area of artificial intelligence that operates under the concept that a computer program can learn and adapt to data without the need for human intervention in the analytical process (Burkov, 2019). Over the past 20 years, machine learning has become increasingly used in the area of prediction, including investing, advertising, lending, fraud detection (Burkov, 2019) and, for purposes of this research, criminal justice risk assessment (Berk et al., 2019). An effort was made to apply random forests, which is a commonly used supervised machine learning approach. The random forest model works by growing algorithms called multiple decision trees,[14] which are then merged together for a more accurate prediction (Hartshorn, 2016). Specifically, multiple uncorrelated models

[14] The decision trees are basically algorithms used to classify data through a flowchart-type format. Each tree starts at a single point and then branches in two or more different directions, with each branch incorporating a variety of decisions until a final outcome is achieved (Hartshorn, 2016).

FIGURE 1.
Association Between Calculated Risk Scores Using MITRE and Other Risk Factors with Any Sexual Recidivism Combining Training and Testing Data



Note. Includes 5,295 male supervisees convicted of online sex offenses whose risk scores could be calculated and whose rearrest activity could be tracked for 60 months. AUC-ROC score for all supervisees is 0.67 with confidence interval of 0.63 - 0.70.

(e.g., decision trees) are applied to generate predictions superior to what would occur if only one decision tree was applied. Applying the random forests method results in each tree giving a classification or vote, and the forests picks the average of all outputs or trees (Hartshorn, 2016). In the current research, a total of 75 variables extracted from PPSO's case management system were used in the random forest models. The random forest models were configured to incorporate 1,400 trees with a maximum depth of 40 branches.

Results from the random forest models also failed to generate predictive indices that met the appropriate criminal justice risk assessment benchmarks (AUC > .70). When applied to the testing data, the random forest models generated AUC scores in the mediocre range (AUC = .54) (data not shown). Moreover, the true positive rate, or the percentage of CSEM supervisees arrested for sexual offenses who were predicted by the model to garner a new arrest, was 54 percent. The remaining 46 percent constituted false negatives, meaning that the model failed to accurately predict that these persons would be rearrested for sexual offenses. These suboptimal metrics of prediction remained constant even when differing random forest applications, including gradient boosting and other machine learning applications, were applied.

Discussion

This article documents PPSO's efforts to construct a risk assessment tool specifically geared towards predicting sexual recidivism among CSEM supervisees. Initially, the endeavor attempted to gauge whether the CPORT could be used for CSEM prediction. One of the challenges in using the CPORT involved the problem of coding several elements, including sexual interests in children and teenagers and preference for boys over girls, that are not readily extractable from PPSO's case management system. PPSO attempted to address this issue by employing MITRE, which used natural language processing for the purpose of text mining 126,000 PDF and scanned documents and, through this method, constructed a dataset composed of a modified version of the CPORT's elements as well as several additional factors believed to be associated with sexual recidivism. The construction of a structured dataset from a myriad of unstructured files embedded within PSRs, polygraph reports, and psychosexual assessments represented a novel effort to use many of the text files generated by federal probation officers during the course of supervision and is suggestive that many of the emerging data science techniques might be directed toward making PPSO's unstructured data more useful for research purposes.

Although MITRE was able to successfully transform unstructured files into structured data, regrettably this effort fell short of being able to construct and deploy a risk tool that could be used on CSEM supervisees. Overall, the modified version of the CPORT risk tool failed to adequately differentiate CSEM supervisees by their likelihood of sexual re-offending and produced AUC scores indicative of mediocre prediction (AUC = .62). While an effort to apply a truncated version of the CPORT performed somewhat better, it still resulted in predictive metrics (AUC = .65) that did not approach those reported by the CPORT's developers (AUC =. 78) (Seto & Eke, 2015).

In light of these results, PPSO attempted to build its own CSEM risk tool that was based on a combination of MITRE-generated factors and elements obtained from the PCRA and FBI criminal history records. This approach performed somewhat better at distinguishing a supervisee's risk of sexual recidivism and produced AUC values approaching the acceptable range for the training data (AUC =. 68), but there was some fall-off in prediction when moving to the testing data (AUC = .65). While PPSO's efforts geared toward building a CSEM risk tool from a combination of factors was somewhat more favorable, this approach produced predictive indices that did not meet the standard benchmarks of many criminal justice risk assessment instruments (AUC > .70). Finally, PPSO attempted to employ machine learning techniques (i.e., random forests) in order to evaluate whether these approaches might assist with CSEM risk prediction. In findings mirroring other analyses discussed in this report, the machine learning approach failed to provide an effective method for ascertaining a CSEM supervisee's likelihood of sexual recidivism.

In general, these findings were disappointing, given the level of effort PPSO expended in attempting to use the CPORT or build its own risk tools for CSEM risk prediction. The results should not be taken, however, as a denigration of the CPORT, which has been shown to be predictive in several studies assessing this risk instrument (Eke et al., 2019; Savoie et al., 2022; Seto & Eke, 2015). A variety of reasons could explain why the current research failed to replicate prior efforts highlighting the CPORT's predictive efficacy. First, MITRE's use of text mining and natural language processing precluded the generation of CPORT factors in a manner similar to that used by Seto and Eke (2015). Specifically, Seto and Eke (2015) combed through the collections of CSEM supervisees to assess the extent to which these collections indicated preferences of boys over girls. Moreover, Seto and Eke (2015) recommended using the CASIC to gauge a CSEM supervisee's sexual interests in children and teenagers. Unlike the approach taken by the CPORT's developers, the limited information available on the types or characteristics of the child pornography collections, the length of time engaged in child pornography activity, or the extent to which CSEM supervisees volunteered in roles with high access to children precluded the CASIC from being used to gauge pedophilic or hebephilic interests or the child pornography collections from being employed to ascertain sexual interests in boys over girls. Ultimately, MITRE relied on admissions to officers, treatment providers, and polygraph administrators to address the CPORT items related to boy over girl preferences or sexual interests in children or teenagers, and this reliance on admissions could have resulted in a diminishment in the predictive efficacy of the CPORT tool.

Other potential explanations for the study's results include the lower base rates for sexual recidivism for the federal CSEM sample (4.5 percent sexually recidivated) compared to study sample used by Seto and Eke (2015) to construct the CPORT (16 percent sexually recidivated). The low-risk skew of the federal CSEM population was also problematic. Over half the population had CPORT risk scores of 0 or 1, and about half received a score of 0 using the risk tool constructed by PPSO. The fact that so many CSEM supervisees garner few if any points using the various risk tools employed in this study and that relatively few sexually recidivated produces various challenges when it comes to developing and deploying an effective risk tool. In addition to these issues, differences between the U.S. and Canadian CSEM populations and the typical degradation in effect sizes when moving from the development to validation samples could explain the study's results (Copas, 1983; Soldino et al., 2021). Last, similar to other studies (see Soldino et al., 2021), the divergence in data quality between the Seto and Eke's (2015) CPORT development study and PPSO's data collection efforts might also explicate these findings. Basically, text mining

126,000 PDF and scanned documents cannot approximate in quality the work conducted by the CPORT's developers to manually code the instrument through a careful review of the case files. While text mining may have potential future applications in PPSO's research, it is possible that some types of information are better obtained through manual (i.e., non-machine) methods.

Future Directions for CSEM Research

While this initial attempt to develop a CSEM-based risk tool failed to generate an instrument that officers could use to supervise this key subpopulation of sex offenders, the research suggests several directions for future risk assessment development. First, several factors embedded within PPSO's risk tool (e.g., PCRA) were identified as being correlated with sexual recidivism for the CSEM population, including social problems associated with drug use, negative attitudes towards supervision, institutional adjustment, presence of criminal thinking style indicating denial of harm, and an assessment for domestic violence. Moreover, the presence of prior criminal behavior and in particular an arrest history for sex offenses were associated with sexual recidivism. At the very least, CSEM supervisees possessing one or more of these characteristics should be subjected to higher levels of supervision intensity compared to their CSEM counterparts without any of these attributes. In addition to these factors, PPSO has begun collecting data that might prove valuable for future efforts aimed at CSEM prediction. The fields currently include prior arrests for any type of sexual assault or production of child pornography, stranger victimization during any type of violent or sex offense, sexual assault of an unrelated male under the age of 17, and presence of valid Static-99 scores. Moreover, officers are being asked to collect information on whether the CSEM supervisee admitted to any hands-on sexual behavior irrespective of any arrests associated with this conduct. Information about the number of victims associated with this behavior is also being collated. The endeavor currently underway to obtain information on admissions of contact sex behavior represents a first-time national level effort to measure the extent to which CSEM supervisees have a history of contacting sexual offending that did not result in an official arrest. Future research efforts conducted by PPSO will attempt to ascertain whether these newly collected risk factors in conjunction with factors already scored by the PCRA might be combined to generate a new risk tool centered on CSEM supervisees.

Regarding the CPORT and CASIC, the viability of any future efforts aimed at using this risk tool depend upon the availability of information that is currently not being systematically collected during the supervision terms for persons convicted of CSEM offenses. Specifically, greater resources would be required at the sentencing stage to obtain information on the details of the child pornography collections gathered by CSEM supervisees. This information could then be used to address the CPORT and CASIC questions pertaining to the nature of the child pornography collections. Additionally, more methodical approaches would be required to address CASIC questions about volunteering in a role with high access to children and engaging in online sexual communications with minors. Purposefully attempting to extract the CASIC elements would enhance the feasibility of accurately addressing the CPORT question concerning sexual interests in children and teenagers. PPSO is exploring the viability of making changes to its case management system in order more uniformly and comprehensively to obtain data measuring the CPORT and CASIC elements.

Last, relying on FBI criminal history files to track the sexual recidivism behavior of CSEM supervisees has serious limitations. Essentially, the literature shows more than half of persons convicted of CSEM offenses engaging in contact sex behavior that never resulted in an actual arrest via admissions (Seto et al., 2011). Given the potential of many CSEM supervisees to engage in behavior that remains unknown to law enforcement officials, it might be advisable to move away from relying on official criminal history records and instead use polygraphs to track any self-reported behavior involving new sex crimes committed while on federal supervision. The practicability of using self-reporting methods should be more fully explored by federal probation.

Conclusion

This report sought to document PPSO's efforts to develop an actuarial tool that could be used to gauge the risk of sexual recidivism for persons convicted of CSEM offenses placed on federal supervision. The report delved into PPSO's attempts to employ the CPORT, including an explication of the challenges inherent in extracting the CPORT data elements and the efforts to overcome these challenges by contracting the data collection process to MITRE. While MITRE was able to successfully extract the CPORT factors for nearly 5,800 CSEM supervisees using text mining and natural language extraction methods, the instrument produced through this process failed to generate predictive indices similar to those reported by its developers (Eke et al., 2019; Seto & Eke, 2015). Given these findings, PPSO then detailed its efforts to construct its own in-house CSEM risk tool using various elements from the CPORT, PCRA, assessment fields, and criminal history files as well as applying machine learning to CSEM risk prediction. These in-house efforts, while somewhat successful, ultimately fell short of PPSO's goal of constructing a risk tool that could effectively differentiate CSEM supervisees by their levels of risk. In light of these findings, at this time PPSO cannot recommend using an actuarial tool outside the PCRA and policy guidelines related to supervising CSEM supervisees. PPSO will continue to engage in the problem of CSEM risk prediction, with particular emphasis on assessing whether some of the new risk factors currently being collected by officers can be combined with the PCRA elements to construct a risk tool that officers could apply to CSEM supervisees. Finally, PPSO will explore the feasibility of more uniformly and systematically collecting information that can be used to re-examine the CPORT's predictive effectiveness. We hope that these approaches will result in a risk tool that officers can use to effectively and judiciously supervise persons convicted of CSEM offenses on federal supervision.

References

Administrative Office of the U.S. Courts (2018). *An overview of the federal Post Conviction Risk Assessment*. Administrative Office of the U.S. Courts. https://www.uscourts.gov/sites/default/files/overview_of_the_post_conviction_risk_assessment_0.pdf.

Baber, L. (2010). Results-based framework for post-conviction supervision recidivism analysis. *Federal Probation*, *74*(3), 5.

Berk, R., Berk, D., & Drougas D. (2019). *Machine learning risk assessments in criminal justice settings* (pp. 1-178). Springer.

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*. [Unpublished master's thesis]. University of Canterbury.

Burkov, A. (2019). *The hundred-page machine learning book* (Vol. 1, p. 32). Quebec City, Canada: Andriy Burkov.

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.

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), 456-48.

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.

Copas, J. B. (1983). Regression, prediction and shrinkage. *Journal of the Royal Statistical Society: Series B (Methodological)*, *45*(3), 311-335.

DeLisi, M., Caropreso, D. E., Drury, A. J., Elbert, M. J., Evans, J. L., Heinrichs, T., & Tahja, K. M. (2016). The dark figure of sexual offending: New evidence from federal sex offenders. *Journal of Criminal Psychology, 6*(1), 3-15.

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

Eke, A. W., Helmus, L. M., & Seto, M. C. (2019). A validation study of the Child Pornography Offender Risk Tool (CPORT). *Sexual Abuse, 31*(4), 456–476.

Faust, E., & Motivans, M. (2015). Sex offenders in the federal correctional system: The consequence of heightened attention on increased certainty and severity of punishment. *Justice Research and Policy, 16*(1), 81–98.

Hanson, R. K., Thornton, D., Helmus, L.-M., & Babchishin, K. M. (2016). What sexual recidivism rates are associated with Static-99R and Static-2002R Scores? *Sexual Abuse, 28*(3), 218–252.

Hartshorn, S. (2016). *Machine learning with random forests and decision trees: A mostly intuitive guide, but also some python*.

Helmus, L. M., & Hanson, R. K. (2007). Predictive validity of The Static-99 and Static-2002 for sex offenders on community supervision. *Sexual Offender Treatment*, *2*(2), 1-14.

Johnson, J. L., Lowenkamp, C. T., VanBenschoten, S. W., & Robinson, C. R. (2011). The construction and validation of the federal Post-Conviction Risk Assessment (PCRA). *Federal Probation*, *75*(2), 16–29.

Johnson, J. L., & Baber, L. M. (2015). State of the system: Federal probation and pretrial services. *Federal Probation*, *79*(2), 34–40.

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.

Lowenkamp, C. T., Holsinger, A. M., & Cohen, T. H. (2015). PCRA revisited: Testing the validity of the federal Post Conviction Risk Assessment (PCRA). *Psychological Services, 12*(2), 149–157

Luallen, J., Radakrishnan, S., & Rhodes, W. (2016). The predictive validity of the Post-Conviction Risk Assessment among federal offenders. *Criminal Justice and Behavior, 43*(9), 1173–1189.

Markman, J. A., Durose, M. R., Rantala, R. R., & Tiedt, A. D. (2016). Recidivism of offenders placed on federal community supervision in 2005: Patterns from 2005 to 2010. *Bureau of Justice Statistics*, *2*(1), 1-16.

Megerdoomian, K., Branting, K., Horowitz, C., Marsh, A., Petersen, S., & Scott, E. (2019). Automated narrative extraction from administrative records. In *AIAS@ ICAIL* (pp. 38-48).

Pilon, A. J. M. (2016). *The predictive validity of general and offence-specific risk assessment tools for child pornography offenders' reoffending*. [Unpublished master's thesis]. University of Saskatchewan.

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

Savoie, V., Quayle, E., Flynn, E., & O'Rourke, S. (2022). 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.

Serin, R. C., Lowenkamp, C. T., Johnson, J. L., & Trevino, P. (2016). Using a multi-level risk assessment to inform case planning and risk management: Implications for officers. *Federal Probation, 80*(2), 10–15.

Seto, M. C., Karl Hanson, R., & Babchishin, K. M. (2011). Contact sexual offending by men with online sexual offenses. *Sexual Abuse, 23*(1), 124–145.

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.

Seto, M. C., & Eke, A. W. (2017). Correlates of admitted sexual interest in children among individuals convicted of child pornography offenses. *Law and Human Behavior*, *41*(3), 305.

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.

United States Sentencing Commission. (2012). *2012 Report to the Congress: Federal child pornography offenses*. United States Sentencing Commission.

United States Sentencing Commission. (2021). *Federal sentencing of child pornography non-production offenses*. United States Sentencing Commission.

Walters, G. D. (2013). The psychological inventory of criminal thinking styles (PICTS) professional manual. *Center for Lifestyle Studies*.

Wolak, J., Finkelhor, D., & Mitchell, K. J. (2005). *Child pornography possessors and the internet: A national study*. National Center for Missing & Exploited Children.

Wolak, J., Finkelhor, D., & Mitchell, K. J. (2009). *Trends in arrests of "online predators."* Crimes Against Children Research Center.