



Ministry
of Justice

The Actuarial Prediction of Sexual Reoffending

Responding to Changing Offending Patterns

**Chudi Emeagi, Liam Sullivan, Dr Julia Landsiedel,
Andrew Craik and Dr Philip Howard**

Ministry of Justice

Ministry of Justice Analytical Series

2024



Data and Analysis exists to improve policy making, decision taking and practice by the Ministry of Justice. It does this by providing robust, timely and relevant data and advice drawn from research and analysis undertaken by the department's analysts and by the wider research community.

Disclaimer

The views expressed are those of the authors and are not necessarily shared by the Ministry of Justice (nor do they represent Government policy).

First published 2024



© **Crown copyright 2024**

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit nationalarchives.gov.uk/doc/open-government-licence/version/3

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

Any enquiries regarding this publication should be sent to us at

Evidence_partnerships@justice.gov.uk

This publication is available for download at

<http://www.justice.gov.uk/publications/research-and-analysis/moj>

ISBN 978-1-911691-21-1

Acknowledgements

We acknowledge our external reviewers of both methodology and content.

Further, we acknowledge the subject matter advice and guidance provided by our colleagues in HMPPS, including Dr Mark Farmer, Louise Hubbard, Cindy Keehner and Helen Walton.

The authors

Chudi Emeagi, Dr Julia Landsiedel, Liam Sullivan, Dr Philip Howard

Contents

List of tables

1.	Executive Summary	1
2.	Introduction	3
2.1.	Overview of sexual reoffending prediction tools	3
2.2.	Reflection on risk of recidivism tools revalidation	4
2.3.	Brief typology of sexual offending behaviours	5
2.4.	OSP in the context of evolving sexual offending behaviours	6
2.5.	Research questions and study outline	8
3.	Methods	10
3.1.	Analytical dataset	10
3.2.	Statistical analysis and modelling	12
3.2.1.	Model discrimination - Concordance index	12
3.2.2.	Survival modelling explained	13
3.2.3.	Survival model specification	14
3.2.4.	Counting logic for sexual history related risk factors used in survival modelling	15
3.2.5.	Reoffending outcomes	17
3.2.6.	Risk predictor rescaling	18
3.2.7.	Limitations	19
4.	Results	21
4.1.	Performance of operational OSP risk scores by sexual reoffending behaviour	21
4.2.	Survival model results	22
4.2.1.	Contact sexual reoffending	23
4.2.2.	Contact adult reoffending	23
4.2.3.	Contact child reoffending	23
4.2.4.	Direct contact child reoffending	24
4.2.5.	Indirect contact child reoffending	24
4.2.6.	IIOC only reoffending	24
4.3.	Performance of revised OSP risk scores by sexual reoffending behaviour	25
4.3.1.	OSP/DC	25
4.3.2.	OSP/IIC	27

5. Conclusions	29
5.1. Performance of Operational OSP/C and Survival Models	29
5.2. Risk Predictor Rescoring and Policy Implications	30
5.3. Decision Framework	31
Glossary	32
References	37

List of tables

Table 3-1. Comparison of descriptive characteristics for men with sexual history in the 2018 vs 2022 prison and probation caseload: Proportion of the caseload by subgroup	11
Table 3-2. Comparison of details of sexual offending history by offending behaviour for men with sexual history in the 2018 vs 2022 caseload	12
Table 3-3. Risk factors used in OSP/C survival models	15
Table 4-1. Overview of concordance (model discrimination) results by reoffending outcome and operational OSP risk predictor	22
Table 4-2: Rates of proven reoffending split by offence type and OSP/C risk band	27
Table 4-3: Rates of proven reoffending split by offence type and OSP/DC risk band	27
Table 4-4: Rates of proven reoffending split by offence type and OSP/I risk band	28
Table 4-5: Rates of proven reoffending split by offence type and OSP/IIC risk band	28

1. Executive Summary

Within the suite of tools used by His Majesty's Prison and Probation Service (HMPPS) to estimate the risk of reoffending in its various forms, there are two relating to sexual offending:

- The Offender Assessment System (OASys) Sexual reoffending Predictor for contact sexual offending (OSP/C); and
- The OASys Sexual reoffending Predictor for indecent images of children (OSP/I).

These risk predictors are actuarial: each predictor assigns a risk level and/or probability to the offender, based on a statistical model that estimates how likely they are to reoffend in a certain way given their risk factors (e.g., age, previous offending).

Over time, as reoffending patterns change, it may be that some of these risk predictors become less effective at estimating risk of reoffending. Therefore it is important for the HMPPS to periodically check that the risk predictors in use continue to perform well. Craik et al. (2024) recently carried out one such check, a key finding of which was a decrease in OSP/C's performance. They found that OSP/C underpredicted contact sexual reoffending and demonstrated acceptable, rather than good, validity in differentiating between higher and lower risk offenders. As part of the report Craik et al. recommended further investigation to understand and address the noted drop in performance. This report arose as a result of those recommendations.

Using a dataset of individuals who were in the prison and probation system in 2018, distinct sexual offending behaviours were examined to investigate how OSP/C and OSP/I performance varied depending on the type of sexual reoffending. We then constructed a series of statistical models to understand which factors are indicators of each type of sexual reoffending.

These analyses revealed that model performance clearly varied based on the type of sexual reoffending studied – particularly those involving online contact between offender and victim (referred to in this paper as *indirect contact child offences*) – and that sexual reoffending in a specific category may be best explained by previous similar offending.

Based on those results several changes to the OSP/C and OSP/I models as remedies have been tested and proposed, identifying two potential modified predictors:

- OSP Direct Contact (OSP/DC), predicting reoffences involving direct sexual contact between offender and victim; and
- OSP Indecent images and Indirect Contact (OSP/IIC), predicting reoffending involving indecent images of children and/or indirect contact child offences.

As a result of this work, HMPPS will have to assess the policy implications of replacing OSP/C with OSP/DC and OSP/I with OSP/IIC and decide whether they should be introduced. A policy addendum outlining whether the two new sexual offence predictors are to be introduced will be published alongside this report.

2. Introduction

As part of the Ministry of Justice's wider offender management strategy, several predictive tools are used to estimate an offender's risk of reoffending after starting the community element of their sentence. These Actuarial Risk Assessment Instruments (ARAI)s are derived by statistically modelling reoffending patterns from large samples of offenders and are periodically evaluated on their performance and updated over time.

Operationally the ARAI)s inform practitioner decision-making regarding the supervision of a person in prison or on probation by providing an objective quantification of their probability of future proven reoffending. In this respect, the performance and maintenance of ARAI)s upholds the key departmental priorities of reducing reoffending and protecting the public.

Given the breadth of types of offending which fall under the broad category of sexual offences and the potential harm to the public in these cases, it is important that prison and probation practitioners have a clear understanding of the risk of sexual reoffending presented by any one offender. This introduction outlines the tools currently in use operationally by His Majesty's Prison and Probation Service (HMPPS) and the results of recent performance evaluations of these predictive tools, and signposts research available on behavioural typologies of those who commit sexual offences. Finally, the research questions at the heart of this paper are outlined.

2.1. Overview of sexual reoffending prediction tools

As part of the suite of tools for the prediction of sexual reoffending, two risk predictors focussing on two types of offending behaviours are currently in use:

1. The Offender Assessment System (OASys) Sexual reoffending Predictor for contact sexual offending (OSP/C); and
2. The OASys Sexual reoffending Predictor for indecent images of children (OSP/I).

The OSP/C and OSP/I algorithms (see Appendix A for how they are calculated) are based on static risk factors only. Static risk factors for reoffending are factors which cannot be changed over the course of a sentence (e.g. criminal history, or age). Due to the low

sexual reoffending rates of women with a sexual offending history, OSP risk scores are only calculated for men with a sexual offending history. Operationally, these tools provide risk of reoffending bands (low, medium, high, and very high; the latter only for OSP/C) that guide practitioners in their work.

Both risk predictors were originally developed and validated using data from the early 2000s and became operational in 2021 after research found they generated more accurate predictions (Howard & Wakeling, 2021) and were easier to use (Bell, 2018) compared to the operational tool preceding them (Risk Matrix 2000 / sexual , RM2000/S; Thornton et al., 2003).

2.2. Reflection on risk of recidivism tools revalidation

Tools developed to predict various types of recidivism usually undergo an initial validation phase to evaluate their performance for operational use. The predictive tools used by HMPPS were generated in the 2000s and 2010s, and therefore, re-evaluating their effectiveness is necessary to ensure that any changes in offending patterns that occurred since their development have not compromised their predictive performance.

A recent Ministry of Justice publication (Craik et al. 2024) carried out this work studying the current operational suite of actuarial risk predictors, using a dataset based on the probation and prison caseload as of June 30, 2018, consisting of 22,231 offenders, with a 2-year follow-up. A key finding of this report was on the performance of OSP/C, the risk predictor for *contact sexual* reoffending.

Performance of a predictive tool can be assessed in various ways. In this case, Craik et al's study focussed on a) how accurately OSP/C was able to predict the actual proven contact sexual reoffending rate (that is, the model calibration rate, represented by the difference between actual and predicted proven reoffending rate) and b) the predictive validity of OSP/C (model discrimination, see 3.2.1 for a detailed explanation). In brief, effective model discrimination in the context of OSP/C means that individuals who reoffend with contact sexual offences sooner should receive higher risk scores compared to those who reoffend in this way later or not at all. Given the very low rate of such proven reoffending in absolute terms, the comparison of reoffenders and non-reoffenders has far

more weight in model discrimination metrics than the comparison of earlier and later reoffenders.

Craik et al. (2024) found a decrease in OSP/C's predictive validity compared to the original validation by Howard and Wakeling (2021). Specifically, proven sexual reoffending rates differed less between lower and higher risk bands. Furthermore, Craik et al. (2024) found that OSP/C underpredicted the 2-year contact sexual proven reoffending rate by 0.42 percentage points (actual rate 1.37%, predicted rate 0.95%) though OSP/C was still deemed well-calibrated overall. This metric was not reported by Howard and Wakeling (2021), thus no comparison can be drawn in this case.

Regardless, OSP/C's clear drop in performance raises concerns about its continued operational use and therefore, it is imperative to investigate a) the underlying cause(s) for this change given that it was implemented operationally based on strong evidence of its advantages over RM2000/S and b) the potential pathways for remediation.

2.3. Brief typology of sexual offending behaviours

This section will briefly discuss broad typologies of sexual offending which may be relevant in the context of risk prediction. Various classifications of sexual offending behaviours and the individuals who commit them exist. Thornton (2021) outlined how conventionally, prior to the wider use of the internet and other online media, the main distinction was between noncontact sexual offences (e.g. indecent exposure) and contact sexual offences, dividing contact sexual offences into those that involve adult victims compared to those involving child victims in intra- and extrafamilial settings. Thus, this conventional distinction does not include internet-mediated sexual offending as it did not exist at the time. Today, in most cases, children are the main victims of internet facilitated sexual offending (Thornton, 2021), though adults can equally become victims (Almond, McManus, & Chatterton, 2017). Notably, research has found clear differences in the characteristics and psychological profiles (personality traits, cognitive distortions, empathy, and impulsivity) of sexual offenders who offend online vs offline (see Babchishin, Hanson, & VanZuylen, 2015; Lim et al., 2021). These have clear practical implications when it comes to understanding similarities and differences in offenders' criminogenic needs as well as situational risk factors associated with different sexual offending types.

2.4. OSP in the context of evolving sexual offending behaviours

OSP/C was originally developed and validated using data from the early 2000s to predict *contact* sexual reoffending. (Howard & Wakeling, 2021). At the time, contact sexual offences involved direct contact between offender and victim, such as rape, sexual assault, and sexual activity with adults unable to give legal consent or children. Since then, legislative changes in response to the emergence of new, mostly online, offending behaviours were introduced. Importantly, legislative changes were put into practice by updating probation practitioner guidance on how to evaluate these new offences when calculating OSP/C risk scores. Thus, the impact of these changes on actuarial predictions were unclear until the recent revalidation study (Craik et al. 2024).

Firstly, the Sexual Offences Act 2003 represented a major overhaul of the law on sexual offences in England and Wales (Sjölin, 2015). “Its purpose was to strengthen and update the law on sexual offences, whilst improving the protection of individuals from sexual offenders.” (Crown Prosecution Service, 2022). The Sexual Offences Act 2003 considerably reformed the concept of consent in the context of existing sexual offences such as rape, sexual assault, incest etc that involve *direct* sexual contact with victims (Sjölin, 2015), and introduced new sexual offence categories. Examples of these new offending behaviours are causing or inciting children to engage in, or watch, sexual activity, grooming children, and arranging or facilitating the commission of a child sexual offence. Secondly, the Serious Crime Act 2015 introduced a further novel offence statute, engaging in sexual communication with a child (see Ministry of Justice, 2015) to reflect how online chatrooms previously, now obsolete and replaced by social media networking sites (Wagstaff, 2012) created new communication channels for offenders (even when there is no intention to meet up offline). Furthermore, many of the above new offences may take place both online and offline (Howard & Barnett, 2015) proving it challenging to “label” them.

To address this issue in the context of the current study, preparatory analysis investigated patterns of offending behaviours that co-occurred to derive meaningful subgroups of child sexual offending (see Appendix F). The rationale here was that similar offending behaviours would co-occur in offenders’ criminal history and thus can be grouped together

as different typologies of offending. This analysis revealed that offences that were introduced into legislation in more recent times frequently co-occurred in offenders' most recent conviction for sexual offending. In discussion with subject matter experts, it was agreed that these offences could be classified as the offender "exclusively" having *indirect* sexual contact with the victim; i.e., the offender not directly engaging in the sexual activity with the victim.

In contrast, current OSP guidance (HMPPS Public Protection Group, 2022). considers these indirect contact offences such as 'inciting a child to engage in sexual activity' as 'contact' offences. Specifically, it defines contact offences "by there being actual or attempted contact with a live human being" (HMPPS Public Protection Group, 2022, p.4). At the time, the assumption was that offences associated with intent or attempts to seek sexual contact could essentially be seen as precursors to future contact that if successful would result in substantial sexual harm.

Finally, unrelated to child-related sexual offending legislation, additional offences relating to the possession of extreme pornographic images (for example, possessing extreme pornographic images that depict an act which threatens a person's life, images portraying rape, images of a person performing an act of intercourse or oral sex with an animal) were introduced (Criminal Justice and Immigration Act 2008; Criminal Justice and Courts Act 2015). The operational OSP/C algorithm currently accommodates these offences as part of sanction counts for other noncontact offences¹ (other noncontact OSP group, see Appendix E).

Any data used to validate OSP/C prior to the Craik et al. (2024) study pre-dated the above developments; i.e., indirect contact offences were not contained in offenders' criminal histories or current offending behaviour, which were studied during its validation. Therefore, OSP/C's predictive capability may have been inadvertently biased towards direct contact sexual reoffending.

¹ Some past research publications related to OSP used the historic term paraphilia offences instead of other noncontact offences. The term 'other' is relative to the separate category of indecent images of children offences.

In contrast, OSP/I (OSP/C's counterpart for the prediction of indecent images of children - IIOC - reoffending) may not have been affected by the above legislative changes in the same way. Indeed, IIOC offences were first introduced into legislation in the Protection of Children Act 1978, and the Criminal Justice Act 1988 long before the widespread use of the internet by the public, and only experienced minor amendments throughout the subsequent years (including the Sexual Offences Act 2003). The only change of note was the introduction of offences relating to the possession of prohibited (pornographic) images of children in the Coroners and Criminal Justice Act 2009. Therefore, a move of this type of offence to the online world (even in its early days) was therefore largely captured by the law (see also Akdeniz, 2002), and thus sanctions for IIOC offending online could be deemed equivalent to sanctions for offline IIOC offending in terms of its effect on OSP/I. If any, assuming that increased internet IIOC offending would lead to increased number of sanctions, then an improvement of predictive validity for OSP/I might be expected. Indeed, Craik et al. (2024) reported an increased predictive validity for OSP/I compared to its original validation (Howard & Wakeling, 2021).

Nonetheless, at least some of the more recently introduced indirect contact sexual offences share a common denominator with IIOC offences; i.e. they are often facilitated via the internet. The preliminary analysis of their co-occurrence patterns (see Appendix F) revealed that IIOC strongly co-occurred with some but not all indirect contact child offences, as well as extreme pornography offences. These findings support further investigation to elucidate how these associations may be related to the prediction of reoffending; that is, could these associations inform how to improve the offence classifications underpinning OSP/C and OSP/I.

2.5. Research questions and study outline

In sum, this study aims to investigate how well evolving sexual offending patterns are captured in the currently operational risk predictors for contact sexual reoffending (OSP/C) and IIOC reoffending (OSP/I). Specifically, it is hypothesized that a distinction should be made between direct and indirect contact sexual offending.

Based on the above, this study set out to investigate three major questions:

1. What is driving the recently observed reduction in performance of OSP/C?
2. What potential remedies are there to improve the performance of OSP/C?
3. How do these potential remedies affect the performance of OSP/I?

The next section sets out the methodological foundations of this report including a description of the analytical dataset and statistical analysis. In the following section results are presented, which show the explanatory factors behind the fall in predictive validity for OSP/C, the results of survival modelling which seek to explain which factors make for good predictors of proven reoffending, and the predictive validity of several different remedies for the prediction of sexual reoffending. Finally, the conclusions section sets out the connotations of these results, details the analytical preference for each of the remedies, and sets out the decision framework which will be used to select remedies by HMPPS.

3. Methods

3.1. Analytical dataset

The study cohort was selected from a snapshot of the England and Wales probation caseload as of 30 June 2018 (N = 189,097); that is, every individual serving a sentence supervised in the community under a court order (community orders or suspended sentences) or under supervision on licence following release from custody. The start of the study for every individual was set to the date of the snapshot (30 June 2018). As a result, individuals differed regarding their amount of time in the community during their current sentence, that is zero or more complete months, which will be referred to as offence-free months (OFM, or offence-free time). A focussed look at the effect of OFM on an individual's risk of reoffending is discussed in detail in Appendix J.

Data was extracted from the Police National Computer (PNC), from which full profiles of previous and subsequent offending were derived, as well as demographic data from the probation caseload system Delius, and from the Offender Assessment System (OASys) to evaluate individuals' risks and criminogenic needs. The caseload dataset was filtered to include only individuals who had a proven conviction for sexual offending (sexual history). Only male individuals (N = 22,231) were studied due to the very small number of women with sexual history (N = 261).

Sexual offending patterns (see 3.2.5 Reoffending outcomes) were followed up over the course of two years. Thus, the follow up window ended on 30 June 2020. The total follow-up time was 731 days (due to 2020 being a leap year). Any reoffence committed during this two-year period was considered even if the conviction or sanction for these reoffences were incurred at a later date (the PNC data extract used to assess convictions or sanctions was taken on 2 November 2022). This method was chosen to allow inclusion of offences that would otherwise not be included due to waiting times for offences to be brought to justice.

In addition to the work done to understand patterns of reoffending within the two-year follow up period, the effect that time spent offence-free in the community prior to the start

of the follow up period has been considered. The effect that this has on an individual's risk of reoffending is set out in Appendix K.

Table 3-1 shows characteristics of the study cohort, including the distributions of age group, disability and housing status. Table 3-2 additionally provides information on the proportions of individuals in the caseload whose proven criminal histories involved different types of sexual offences (for a detailed explanation refer to section 3.2.5). As a comparative reference, both tables also provide this information for the most recent version of the community caseload (N = 22,403, as of June 2022). Only demographics were compared between the 2018 and 2022 caseload. Reoffending rates cannot be examined yet for the 2022 cohort as data across the entire 2-year follow-up period at the time of writing is not available yet (even without a waiting period for new convictions to occur, its follow-up period would end on 30 June 2024). This gives an indication of how well the results observed in the 2018 cohort generalise to a more recent dataset. Finally, rates and absolute volumes of reoffending can be found in Appendix B.

Table 3-1. Comparison of descriptive characteristics for men with sexual history in the 2018 vs 2022 prison and probation caseload: Proportion of the caseload by subgroup

Subgroup	2018 (N = 22,731)	2022 (N = 22,403)
Age: 18 - 24	10.60%	7.97%
Age: 25 - 29	12.11%	10.11%
Age: 30 - 39	24.42%	24.84%
Age: 40 - 49	19.46%	19.20%
Age: 50 - 59	18.54%	19.05%
Age: 60 and Over	14.84%	18.81%
Age: No data	0.02%	0.02%
Disability: Does not have disability	62.35%	51.61%
Disability: Has disability	37.65%	48.39%
Fixed abode: No	13.63%	13.99%
Fixed abode: Yes	80.43%	80.88%
Fixed abode: No data	5.94%	5.13%

Table 3-2. Comparison of details of sexual offending history by offending behaviour for men with sexual history in the 2018 vs 2022 caseload

Type of previous sexual offence	% 2018 caseload (N = 22,731)	% 2022 caseload (N = 22,403)
Contact adult	29.91%	30.48%
Contact child	37.06%	44.92%
Direct contact child	32.29%	31.97%
Indirect contact child	13.20%	19.42%
IIOC	33.12%	30.86%
IIOC and Extreme pornography	8.61%	8.28%

3.2. Statistical analysis and modelling

This study used the method of survival analysis to examine the predictive validity of OSP/C for several sexual reoffending outcomes. Additionally, operational OSP/C and OSP/I risk predictor algorithms were rescored to understand their predictive validity if they were to be modified. This section outlines how the predictive validity of OSP models was assessed and how survival analysis was implemented.

3.2.1. Model discrimination - Concordance index

The predictive validity of currently operational OSP/C and OSP/I risk scores was evaluated using the concordance index (Harrell's C-index, Harrell et al., 1996). The concordance index is a measure of how well a model can discriminate between those with high and low risk of reoffending. It is increased when reoffenders have higher risk scores than non-reoffenders and when early reoffenders have higher scores than those who reoffend later in the follow up period. For example, an individual with several previous sanctions for contact sexual offences would have a higher risk of reoffending (higher OSP/C risk score) than someone with just one sanction. In practice, the OSP/C risk scores would be considered to discriminate well between high-risk and low-risk offenders if a person with a higher risk score reoffends earlier compared to a person with a lower risk score. In contrast, a model's discrimination would decrease if individuals with very different risk scores (low vs high) reoffended after the same amount of time, or if indeed a person with a lower risk score reoffended whereas the person with the higher risk score didn't. Importantly, the C-index allows censoring to be used in the study (see section 3.2.2).

The higher the value of the C-Index, the better discrimination between low and high risk offenders. A C-Index of 1 indicates perfect discrimination, whereas a C-Index of 0.5 indicates poor discrimination; that is the predictor is as good as random guessing at discriminating low vs high risk offenders. A C-Index of 0 suggests an opposite relationship between risk scores and reoffending, i.e. low risk individuals reoffend whereas high risk individuals don't reoffend at all (or, if any do, later than those with low risk).

However, discrimination for subgroups should be interpreted cautiously as C-Index can vary due to the distribution of risk scores (e.g., higher or lower standard deviation) as reflects the inherent ease of discrimination. Assuming that risk is estimated correctly by a given predictor, a cohort comprised of some high risk and some low risk people will have a higher C-Index than a cohort comprised mostly of medium risk people. If choosing a subgroup inherently reduces the range of variation in risk, this subgroup is likely to have a lower C-Index (Howard, 2017).

This study examined how the discrimination of OSP/C and OSP/I varied when considering different sexual offending behaviours. Specifically, to determine what may be driving the observed reduction in OSP/C's concordance, discrimination was assessed for any contact sexual reoffending (*contact sex*), contact sexual reoffending with adult victims (*contact adult*) and child victims (*contact child*) respectively. Contact child reoffending was investigated in more detail distinguishing between *direct* and *indirect* contact. Additionally, discrimination of OSP/I was analysed for IIOC and indirect contact child reoffending (both separately and in combination).

3.2.2. Survival modelling explained

Survival models, also called time-to event models, are a statistical method used to predict an event such as reoffending, accounting for the time until it happens. That is, within a given time window (the follow-up period), the model considers how many days from a specific starting date it took for a reoffending event to happen. Survival models allow for an individual to be censored. Censoring occurs when an individual is removed from the study partway (for example, being imprisoned for another offence, not of interest), and therefore ensures information about individuals that were 'removed' from the study before the end of the follow-up period is not lost. For example, when studying OSP/C, censoring would be noted for those imprisoned for violent non-sexual reoffending at the number of days from

start of follow-up to this imprisonment, if they did not sexually reoffend (or incur another type of censoring event) on a prior date. This means, it is known how long someone did not sexually reoffend but that there is no data to understand their sexual reoffending after the censoring event (which effectively prevents sexual reoffending from happening in the first place).

For this study specifically, the Cox proportional hazards survival model (Cox, 1972) was applied to study which risk factors drive the observed reduction in performance of OSP/C when examining sexual reoffending outcomes at different levels of granularity (see section 3.2.5 for details). Specifically, Cox regression models the extent to which each risk factor increases or reduces the hazard of reoffending (i.e. how likely reoffending is at a given point in time). It is assumed that a risk factor's effect on the hazard of reoffending is constant across the two-year follow-up period.

For each model, information on whether an individual reoffended, the time in days to reoffending and/ or censoring and their risk factors were analysed (see section 3.2.3 below for the details on the model specification). The model estimates each individual's hazard of reoffending which allows model concordance to be evaluated (see above section 3.2.1) as an indicator of model performance (goodness of fit, see Appendix G). Statistical assumptions for models were verified and assumption violations addressed where appropriate (see Appendix H).

3.2.3. Survival model specification

There were two types of models to investigate risk prediction for sexual reoffending currently covered by OSP/C. Table 3-3 gives an overview of each model's risk factors used for prediction (see also section 3.2.4 below for details on how sexual offence sanctions were aggregated).

The first type of model (hereafter referred to as *operational* model) used the same risk factors as currently implemented in the algorithms for OSP/C (see Appendix A). They were designed to understand which risk factors predict specific reoffending outcomes. The purpose of each operational model was to evaluate and compare its performance for the reoffending outcome it was designed for with a designated set of reoffending outcomes that diverged from the original specification, with a view to identify significant risk factors.

For OSP/C, this meant testing the operational algorithm for the prediction of specialised sexual offending behaviours.

The second type of model (hereafter referred to as the *revised* model) explored how the existing OSP/C algorithm might be improved upon. Specifically, revised models used modified risk factors to reflect specialisation in sexual offending behaviours in more detail than is currently the case.

Table 3-3. Risk factors used in OSP/C survival models

Operational OSP/C model	Revised OSP/C model
Age at last sexual sanction (under/over 18)	Age at last sexual sanction (under/over 18)
Number of sanctions for contact adult offences (as per original OSP definition)	Number of sanctions for contact adult offences (as per original OSP definition)
Number of sanctions for contact child offences (as per original OSP definition)	Number of sanctions for direct contact child offences (new counting rule)
	Number of sanctions for indirect contact child offences (new counting rule)
Number of sanctions for other noncontact offences (as per original OSP definition)	Number of sanctions for other noncontact offences (new counting rule, former paraphilia sanctions but now excluding extreme pornography offences in counts)
Age at discharge	Age at discharge
First time offender (yes/no)	First time offender (yes/no)
Victim type (stranger yes/no)	Victim type (stranger yes/no)

3.2.4. Counting logic for sexual history related risk factors used in survival modelling

An individual's sexual offending history is one of the key inputs into the analysis of what predicts sexual reoffending. When summarising someone's sexual history different levels of granularity are possible. Using PNC data, statutory sexual offences were identified, and categorised into the established OSP offence categories (HMPPS, 2021, contact adult, contact child, other noncontact, and indecent images). For the purposes of this report, sexual history was aggregated based on the number of sanction occasions. A sanction occasion represents a specific day (sanction date) that one or many sanctions (convictions, cautions or similar out-of-court criminal disposals) were given. Thus, if someone received two sanctions for contact adult sexual offences on the same day

(even if they represented offences committed on two separate occasions), they would count as one contact adult sanction occasion. This is in keeping with the existing OSP/C logic (HMPPS, 2022). Please note that from this point forward the terms 'sanction occasions' and 'sanctions' will be used interchangeably.

Additionally, the operational OSP/C model applies a prioritised counting method (Howard & Barnett, 2015) in cases where an offender committed multiple sexual offending types on the same sanction occasion. In this case, the primary offence was used which is coded in the PNC data extract on the basis of sentence severity. If this was unclear, a ranking was applied such that 'contact adult' was deemed the primary offence, if there were any such offences, otherwise 'contact child' if there were any such offences, otherwise 'other noncontact'. For example, if someone had a contact adult and contact child sanction on the same day, contact adult sanctions would be assumed as primary offences over contact child sanctions. Thus, only the contact adult offence but not the contact child offence would be counted as a sanction occasion. This ranking matches the strength of associations of the different types of sexual offending with contact sexual recidivism based on the reoffending rates found during the original development of OSP (Howard & Barnett, 2015).

When generating the revised survival models, a new counting logic was introduced which resulted in recalculation of certain sanction occasion counts. Specifically, abandoning any assumptions about which offence type should be prioritised when predicting a given reoffending outcome, the prioritisation method was eliminated to obtain the true sanction occasion count for each offence type. Thus, new sanction occasion counts were obtained for direct contact child, indirect contact child and other noncontact offending sanctions. Using the example above for comparison, in a situation where an offender had contact adult and contact child sanctions on the same day, it would here be counted as both a contact adult and contact child sanction occasion. Additionally, if the offender had direct contact child and indirect contact child sanctions on the same day it would be counted as a direct contact child and an indirect contact child sanction occasion due to the splitting of offence types. As a result, sanction occasion counts for direct contact child, indirect contact child and other noncontact used in the revised survival models were expected to be higher than those used in the operational models, whereas sanction occasion counts

for contact adult offences were unaffected. This logic was only used in the generation of the survival models and not the revised predictors.

3.2.5. Reoffending outcomes

Multiple reoffending outcomes were studied using operational and revised survival models (see section 3.2.3 above). These outcomes were selected based on the evidence of distinct (specialised) offending behaviours and sexual offending typologies as set out in the introduction (and Appendix F) of this report.

An offending outcome is shorthand for 'proven reoffending' (see Glossary) involving an offence of interest being committed within a specified time period (i.e. the study follow-up period, see above 3.1). For each reoffending outcome, every offender was categorised by the earliest of one of the following outcomes occurring within the follow-up period: a) they reoffended, b) they were sentenced to immediate custody (either for an offence not of interest or for any offence committed before the study start date; also called pseudo-reconviction), c) they were recalled to custody, or d) none of the above for the entire follow-up period. Individuals in groups b) and c) were censored from the analysis.

Investigation of OSP/C operational and revised models

For the prediction of contact sexual reoffending, a stepwise approach was taken, increasing the granularity of the reoffending outcome. Initially, the prediction of any contact sexual reoffending (*contact sex*) was investigated (akin to the revalidation study discussed earlier; Craik et al. 2024). Next, this analysis was repeated separating outcomes for any contact sexual reoffending by victim category (adults vs children; *contact adult* and *contact child*). These models still used the original OSP definition of 'contact' sexual offences; i.e. they made no distinction between direct and indirect sexual offending behaviours. Finally, models focussing on the prediction of specialised contact sexual offending behaviour were developed. This analysis focussed on three outcomes of specialised offending: 1) direct contact sexual reoffending regardless of victim category (adults and children combined, *direct contact*), 2) direct contact sexual reoffending with child victims only (*direct contact child*), and 3) indirect contact child sexual reoffending (*indirect contact child*). Additionally, prediction of IIOC only reoffending was investigated (IIOC only refers to IIOC reoffending without any other reoffences on the same sanction date).

3.2.6. Risk predictor rescoring

Based on the results from the survival modelling confirming differences in offending behaviour between indirect contact child compared with direct contact child and contact adult offending (see section 4.2), ways to refine and improve performance of OSP/C and OSP/I risk scores were explored. This involved revising their scoring methods, including sexual history risk factors and reoffending outcome. Importantly, the counting logic for sexual history risk factors mirrors the one currently operational, as outlined in section 3.2.4 above.

It should be noted that the purpose of the rescoring of risk predictors and testing of OSP/DC and OSP/IIC constitutes the recalibration of both OSP/C and OSP/I as opposed to the formulation of entirely new models. Therefore, the dataset used in this work essentially forms a test dataset used to score the performance of the new predictors. This dataset is entirely separate from the dataset used to train OSP/C in Howard and Barnett (2018).

For OSP/C, the revised risk score focused on *direct contact* sexual reoffending with either adult or child victims (OSP/DC, see Appendix A). OSP/DC narrowed OSP/C's sanction occasion count for any contact child offence to counting only direct contact child sanction occasions. Thus, no indirect contact child sanctions were included in the scoring of OSP/DC, which also affected sanction counts for other noncontact sanctions due to the prioritised counting rules. For example, a sanction occasion involving convictions for causing a child to watch a sexual act and exposure would now be scored as other noncontact based on the exposure offence. Previously (in OSP/C), the 'causing' offence would have taken priority over the exposure offence, however this is no longer the case in OSP/DC as it is not classed as direct contact child. All other risk factors remained the same.

For OSP/I, the revised risk score focussed on IIOC and *indirect contact child* sexual reoffending (OSP/IIC). It extended the existing simple algorithm for OSP/I to include *indirect contact child* offences in the sanction occasion counts (see Appendix A for details).

Finally, several new approaches to predicting indirect contact child reoffending using a new predictor were considered. These are discussed in more detail in Appendix J. Note

that while concordance is reported for various subgroups, the degree of uncertainty for these metrics would be wide for subgroups with low numbers of reoffenders.

The concordances of these revised risk scores evaluated against key reoffending outcomes can be found in section 4.3. As practitioners use risk bands (HMPPS 2022) (low, medium, high, very high risk of reoffending, the latter band only applies to OSP/C) rather than the numeric scores for OSP/C and OSP/I, the performance of the banded versions of OSP/DC and OSP/IIC were also assessed and compared to the operational banded risk predictors.

3.2.7. Limitations

There are also some acknowledged limitations to this analytical approach.

Firstly, using a fixed date to start the follow up period of all individuals in the dataset limits the ability to include offence-free time in survival models. Most individuals will have already spent some offence-free time in the community before their follow up period started. To mitigate this, work has been undertaken in Appendix J to understand the effect that offence-free time has on an individual's risk of reoffending.

In addition, by using data obtained from a snapshot of the Police National Computer on 2nd November 2022, any sexual convictions for offences committed during the follow up period but convicted after the snapshot date will not be included in the analysis. Statistics on the timeliness of sexual offence convictions are released quarterly by the Ministry of Justice (MoJ, 2024). These show that in Q3 2022, sexual offence cases had a median offence to conviction time of 769 days. Considering the offences committed in our July 2018 to June 2020 follow-up period that eventually resulted in a new conviction, this median time implies that most will have been counted as proven reoffences in this study (i.e., because the reconviction occurred by 2nd November 2022), but a certain proportion will have been excluded.

This study uses a follow-up period of two years, which only gives a partial picture of the reoffending habits of sexual offenders. Although previous studies (Howard and Wakeling, 2021, Thornton, 2003) show that individual risk of reoffending is highest at the start of the period after sentence / discharge into the community, some degree of sexual reoffending risk persists for many years. The way the cohort was drawn means that we did not study

only the first two years (for example, an individual who had been on the community caseload for three years prior to 30th June 2018 would be studied during their 'year four' and 'year five'). Nevertheless, as most offenders are only on the community probation caseload for a few years from the point of sentence/discharge, our reoffending data has less coverage of the risks presented many years later.

Finally, this report only takes into account proven reoffences. Any sexual reoffences which are either not reported or not convicted at court are not included in the study.

4. Results

4.1. Performance of operational OSP risk scores by sexual reoffending behaviour

Concordance indices for the operational OSP/C and OSP/I risk scores indicated a varied pattern of performance depending on the reoffending outcome. Table 4-1 results, while Appendix C, Table C-1 demographic and other characteristics. As Table B-2 demonstrated, the absolute number of reoffenders within a given subgroup for a given outcome was often very low, and the degree of statistical uncertainty will accordingly be considerable – while confidence intervals are not reported for reasons of space, they are often very wide and would overlap for many subgroups.

OSP/C

As previously reported in Craik et al. (2024), the concordance of the OSP/C score for any contact sexual reoffending was lower than had been found by Howard & Wakeling (2021) and gave reason for concern (difference of 0.064; 0.678 vs 0.763). Differentiating by victim category, concordance was good for adult but not child victims. Thus, comparatively poor concordance for contact child reoffending negatively affected overall concordance for contact sexual reoffending.

Further analysis of contact child reoffending based on the differential behavioural sexual offending typologies revealed that OSP/C's concordance was higher - though still only moderate - for direct contact child reoffending, and lower still for indirect contact child reoffending.

This revealed a potential cause of the reduced overall predictive validity of OSP/C; i.e. the data support the hypothesis that OSP/C in its current form may be biased towards direct contact reoffending.

OSP/I

As previously reported in Craik et al. (2024), OSP/I risk scores were highly concordant with indecent images reoffending, indicating effective performance of the predictor and

algorithm. OSP/I was a somewhat better predictor of indirect contact child reoffending than OSP/C, though concordance was below that of the best predictors of other outcomes. Combining the prediction of IIOC and indirect child contact reoffending led to a good concordance index in absolute terms, though lower than for IIOC alone.

Table 4-1. Overview of concordance (model discrimination) results by reoffending outcome and operational OSP risk predictor

Reoffending outcome	OSP/C	OSP/I
Contact sex	0.678	N/A
Contact adult	0.784	N/A
Contact child	0.614	N/A
Direct contact child	0.666	N/A
Indirect contact child	0.596	0.632
IIOC	0.452	0.808
IIOC and/or indirect child	N/A	0.754

Note. Only selected reoffending outcomes are relevant to both risk predictors in the context of this study.

In sum, the current performance of both OSP risk predictors suggest good performance for some but not all reoffending outcomes. Most critically, OSP/C underperforms when predicting indirect child contact offending. This highlights the need for further analysis into the relevance of each parameter used in the operational OSP/C and OSP/I algorithms for each of the reoffending outcomes discussed in this section. Additionally, the current results support the need for an investigation into how model performance could be improved by revising the underlying algorithms.

4.2. Survival model results

As detailed in sections 3.2.3 and 3.2.5, two types of models (operational and revised) were set up with several reoffending outcomes for OSP/C. The following sections will outline modelling results for relevant risk factors separately, contrasting how the two model types compared. The detailed statistical results for each model can be found in Appendix G.

4.2.1. Contact sexual reoffending

Across both model types, relevant risk factors affecting the hazard of reoffending were largely the same with one exception. While higher counts of any contact child sanctions raised the hazard in the operational model, it was increased by indirect contact child sanctions but not by direct contact child sanctions in the revised model. Furthermore, risk factors related to age (offenders that received their most recent sexual sanction aged younger than 18 and those that were older when they started the community element of their sentence) decreased the hazard of reoffending across both models. In contrast, offenders, who had a higher number of previous contact adult sexual sanctions and those who had any type of criminal history (sexual or non-sexual) had an increased hazard of reoffending.

4.2.2. Contact adult reoffending

Similar risk factors were identified by both model types when modelling contact sexual reoffending with adult victims only. Age-related risk factors were again associated with a decreased hazard of reoffending, as were higher counts of other noncontact sanctions (irrespective of the counting rules used in either model). In contrast, an increased hazard of reoffending was found for those who had a higher number of previous contact adult sexual sanctions as well as those who had any sort of previous criminal history.

4.2.3. Contact child reoffending

When modelling the hazard of contact sexual reoffending with child victims only, the pattern of results became more distinct between the operational and revised model though it did not diverge completely. Older age when starting the community-element of a sentence and a higher number of sexual sanctions involving adult victims were associated with a decreased hazard of contact child sexual reoffending. Additionally, the operational model found this association also for those who were aged younger than 18 years old at their last sexual sanction. Hazard of reoffending was associated with an increased number of previous contact child sanctions in the operational model. The revised model revealed that only previous indirect contact child sanctions as well as any previous criminal history increased hazard of reoffending.

4.2.4. Direct contact child reoffending

For each model, only two risk factors were found to influence the hazard of direct contact child reoffending. In both instances, older age when starting the community element of a sentence decreased the hazard. A higher number of contact child sexual sanctions increased the hazard in the operational model whereas, only direct but not indirect contact child sexual sanctions drove this effect in the revised model. Furthermore, sanctions for offences that were not child-specific (contact adult and other noncontact) did not have any associations with the direct contact child hazard. Thus, considering type of previous sanctions only, direct contact child sexual reoffending was best predicted by previous sanctions of the same nature.

4.2.5. Indirect contact child reoffending

For the operational model, the same set of risk factors were found to affect hazard of indirect contact child reoffending in the same way as contact child reoffending overall. Age-related predictors and number of contact adult sanctions were associated with decreased hazard, whereas contact child sanctions were associated with increased hazard. The revised model revealed that being older when starting the community element of a sentence and higher numbers of both contact adult and *direct* contact child sanctions were associated with decreased hazard of indirect contact child reoffending. In contrast, increased hazard of reoffending was associated with a higher number of indirect contact child sanctions as was having any prior criminal history. Indirect contact child reoffending was modelled using a different set of covariates when considering an indirect contact only risk predictor, the details of which can be found in Appendix J.

4.2.6. IIOC only reoffending

Using the operational OSP/C model, several significant risk factors for IIOC only reoffending were identified. Most of these (age-related risk factors, contact adult and contact child sanction counts) decreased the hazard of reoffending except for one (having any previous criminal history), which increased the hazard. In contrast, for the revised model, there was an increased hazard of reoffending for higher counts of IIOC and extreme pornography sanctions (on the same sanction date) as well as IIOC sanction occasions without extreme pornography sanctions. Similarly to the operational model,

hazard of reoffending decreased with higher counts of contact adult or child sanctions as well as higher age at discharge.

4.3. Performance of revised OSP risk scores by sexual reoffending behaviour

In this section the discriminative performance of models currently in use are compared to each of the proposed models, across several different offence types. Bootstrapped confidence intervals have been used to outline the level of confidence with which one predictor can be said to outperform another.

This process involves selecting 1,000 bootstrapped samples, that is samples selected with replacement, meaning the same individual can appear in a sample more than once. These samples are then used to generate the comparison statistics: for each of these samples a c-index has been calculated for each of the predictors and the difference in c-index between the two of them computed. Where predictor 1 is the operational predictor and predictor 2 is the proposed new predictor, the proportion of times predictor 2 outperformed predictor 1 has been used as the p-value in a one tailed test, with the null hypothesis that predictor 2 did not outperform predictor 1.

Tables outlining the performance, reported as concordance index, on each specific reoffending outcome for the analysed cohort can be found in Appendix C. As noted above, while concordance is reported for various subgroups, the degree of uncertainty for these metrics would be wide for subgroups with low numbers of reoffenders.

4.3.1. OSP/DC

Table D-1 in Appendix D shows the comparison of OSP/C and OSP/DC both banded and unbanded for contact sexual, contact adult, direct contact child and direct contact child or contact adult offending.

For contact sexual offending, it is apparent that OSP/C outperforms OSP/DC for both the banded and unbanded algorithms, although the confidence intervals for their c-indices overlap. However, when indirect contact child offences are removed from the outcome reoffending, the evidence suggests that OSP/DC performs in a similar fashion to OSP/C,

with the banded versions of the predictors having an average difference in c-index of just 0.0006.

This is reflected in the c-index scores when predicting both contact adult offending and direct contact child offending, the unbanded versions of the two predictors show the OSP/DC algorithm outperforming the OSP/C algorithm 80% and 66% of the time respectively, whereas the banded versions of the two predictors are much more closely matched, with OSP/DC outperforming OSP/C for 48% and 40% of the samples respectively.

Comparing how both OSP/C and OSP/DC predict different outcomes, it can be seen that direct contact reoffending (i.e. the combination of contact adult and direct contact child) can be predicted far more successfully than all contact reoffending. Essentially, restricting the set of reoffences predicted is the important innovation, whereas changing algorithm from OSP/C to OSP/DC is conceptually logical and should provide clarity to assessors (i.e. how previous sexual sanctions are scored will line up with the type of sexual reoffending that is predicted) but does not in itself improve prediction significantly.

It is acknowledged that contact adult and direct contact child reoffending are themselves not predictively equal: the former seem easier to predict than the latter. As Table 4-2 below shows, the rate of proven direct contact child reoffending is very low – and there are too few such reoffenders for it to be feasible to create and validate a separate model for this offence type.

Practically, a move from OSP/C to OSP/DC would lead to little change in observed reoffending rates of those in each risk band, when it comes to direct contact offending (see Table 4-2 and Table 4-3). When considering direct contact offending against both adults and children, those labelled high or very high risk reoffend at similar rates. Counting only direct contact but not indirect contact sanctions in the OSP/DC scoring summation means that there is a net movement of individuals from very high and high risk bands on OSP/C into medium and low on OSP/DC. However, the removal of indirect contact child reoffending from the target of the risk predictor means that those in high or very high risk bands for OSP/DC reoffend at very similar rates to those with similar OSP/C risk bands.

Table 4-2: Rates of proven reoffending split by offence type and OSP/C risk band

Subgroup	Number of cases	Contact Sexual	Contact Adult	Direct Contact Child	Direct Contact Child or Contact Adult
All	22,231	1.37%	0.58%	0.18%	0.73%
OSP/C: Low	7,467	0.50%	0.05%	0.08%	0.13%
OSP/C: Medium	9,594	1.39%	0.45%	0.19%	0.59%
OSP/C: High	4,243	2.22%	1.27%	0.21%	1.44%
OSP/C: Very High	927	4.31%	3.02%	0.76%	3.78%

Table 4-3: Rates of proven reoffending split by offence type and OSP/DC risk band

Subgroup	Number of cases	Contact Sexual	Contact Adult	Direct Contact Child	Direct Contact Child or Contact Adult
All	22,231	1.37%	0.58%	0.18%	0.73%
OSP/DC: Low	7,774	0.62%	0.08%	0.09%	0.15%
OSP/DC: Medium	9,608	1.44%	0.44%	0.18%	0.58%
OSP/DC: High	4,009	2.07%	1.35%	0.27%	1.57%
OSP/DC: Very High	840	4.17%	3.21%	0.60%	3.81%

4.3.2. OSP/IIC

Similar methods were used to compare OSP/IIC to OSP/I and operational OSP/C for indirect contact child and IIOC offences. The results of these comparisons can be seen in Table D-2 in Appendix D.

There is strong evidence that OSP/IIC outperforms both OSP/C and OSP/I when it comes to the prediction of indirect contact offending. The one tailed test shows that in all of the bootstrapped samples OSP/IIC was able to score more highly than other algorithms, with no overlap in confidence intervals.

Similarly, there is evidence that in predicting IIOC offending, OSP/I and OSP/IIC give similar predictive performance. The banded version of OSP/IIC outperforms OSP/I for roughly 50% of the bootstrapped confidence intervals and the confidence intervals are almost completely the same.

Furthermore, when predicting either of the offence types, there is strong evidence, that both the banded and unbanded versions of the OSP/IIC algorithm outperform OSP/I. Indeed, the banded version of the algorithm outperforms OSP/I, 100% of the time. The evidence therefore suggests that moving indirect contact child offences from OSP/C to OSP/IIC does not negatively affect the performance of OSP/C and actually improves the performance of OSP/I. This uplift in performance is driven by OSP/IIC's ability to better predict indirect contact child offences, as observed in the first line of Table D-2.

In comparison to OSP/I, the reoffending rates for OSP/IIC are higher for indirect contact child offences, and slightly lower for indecent image-based offending, as can be seen in tables 4-4 and 4-5. Evidence of the improvement in predictive validity of OSP/IIC can be noted in the higher reoffending rate of those labelled high risk for indirect contact offences.

Table 4-4: Rates of proven reoffending split by offence type and OSP/I risk band

Subgroup	Number of cases	Indirect Contact Child	IIOC	IIOC or Indirect Contact Child
All	22,231	0.69%	1.55%	2.06%
OSP/I: Low	14,869	0.55%	0.33%	0.81%
OSP/I: Medium	6,504	0.86%	3.14%	3.64%
OSP/I: High	858	1.86%	10.61%	11.77%

Table 4-5: Rates of proven reoffending split by offence type and OSP/IIC risk band

Subgroup	Number of cases	Indirect Contact Child	IIOC	IIOC or Indirect Contact Child
All	22,231	0.69%	1.55%	2.06%
OSP/IIC: Low	12,783	0.27%	0.20%	0.43%
OSP/IIC: Medium	8,352	1.08%	2.63%	3.41%
OSP/IIC: High	1,096	2.65%	9.03%	10.77%

5. Conclusions

5.1. Performance of Operational OSP/C and Survival Models

Actuarial risk of reoffending prediction is an integral part of HMPPS's strategy for managing service provision to offenders in prison or on probation in support of its key objectives of protecting the public and reducing reoffending. Performance evaluation of Actuarial Risk Assessment Instruments (ARAI), the predictors of reoffending, is therefore central to ensuring that the decisions the ARAIs inform are well-founded in robust analytical evidence.

Recent revalidation of ARAIs in use within HMPPS saw a reduction in performance for the currently operational risk predictor for sexual reoffending involving contact with victims (OSP/C; Craik et al., 2024). The initial focus of the present study was to investigate drivers of this performance change compared with the initial OSP/C validation (Howard & Wakeling, 2021, but see Appendix I for further information on the potential impact of low reoffending rates in the current study). Specifically, the influence of specific types of sexual offending behaviours was examined. The analysis has revealed that OSP/C's performance did not reduce across all types of sexual offending behaviour. Instead, it was driven by poor prediction of indirect contact child sexual reoffending.

In the present study, follow-up analysis was conducted into the specific risk factors that predicted different types of sexual reoffending. Using survival models, which account for the length of time spent in the community before an offender reoffends, results indicated a pattern of like for like prediction. That is, a history of specific sexual offending behaviours was predictive of future offending of the same type. Specifically, previous sexual offending with adult victims predicted future such offences whereas a history of other noncontact sexual offences predicted an absence of such reoffences. Similarly, sexual reoffending with child victims for direct and indirect contact offences was predicted by a previous history of those specific offences, respectively. Furthermore, a history of adult sexual offences had no association with direct contact child reoffending and was associated with an absence of indirect contact child reoffending. These findings suggest that sexual

offending behaviours are becoming more distinct from each other. It seems that men who commit sexual offences may become “specialised”, meaning a history of one type of offending indicates that they are less likely to reoffend in another category of sexual offence.

5.2. Risk Predictor Rescoring and Policy Implications

The risk predictor rescoring has demonstrated that risk predictor performance changes when using different scoring algorithms and has provided several possibilities for the future of actuarial sexual risk prediction. Potential remedies, and their analytical consequences, are set out below.

There are three potential remedies for improving the performance of predicting sexual reoffending, each involving how indirect contact child offences factor into the prediction of future reoffending.

The first remedy would be to make no changes to operational risk predictors and instead change future guidance for practitioners. This could involve instructing them which offences are less likely to be predicted well by OSP/C and instructing them to use professional judgement alongside the compromised OSP/C score to assess the risk of reoffending for any one offender. In this case the performance of the predictors would, all other things being equal, remain the same.

Another remedy would be to remove the count of indirect child contact offences from the formulation of the OSP/C score and, in turn, change the scope of OSP/C from all contact sexual reoffending to just direct contact child and all contact adult reoffending. This would be the implementation of the proposed OSP Direct Contact (OSP/DC) method, which was shown no increase performance relative to OSP/C for all direct contact sexual reoffending. For example, the C-index, which measures the ability of a predictor to discriminate between low and high risk offenders, was 0.664 for the operational OSP/C predictor when predicting contact adult and both direct and indirect contact child reoffending. For OSP/DC, it was 0.743 when predicting just direct contact reoffending. This represents a substantial improvement in predictive performance. Although OSP/DC performed only slightly better than OSP/C at predicting direct contact sexual reoffending, removing indirect contact child offences from its scope arguably brings a substantial benefit, as the

most direct and effective way of addressing OSP/C's poor prediction of indirect contact reoffending.

In addition, this remedy involves the inclusion of indirect child contact offences into the count and target of the OSP indecent images of children (OSP/I) predictor. This would be the implementation of the proposed OSP/IIC method, with the count of previous indirect contact child offences included in the calculation of an offender's risk. This expanded solution does lead to better performance in predictive validity for indirect contact child reoffending. However, Table C-3 shows that the performance of OSP/I when predicting only indecent image offending is diluted by adjusting the sanction count method and the algorithm target to include indirect contact child offending.

The final remedy would involve both adjusting the scope of existing risk predictors by introducing OSP/DC, as illustrated above, and to target indirect contact child reoffending separately by developing an entirely new risk predictor. The survival modelling, formulation and scoring of a proposed indirect only risk predictor is set out in detail in Appendix J.

5.3. Decision Framework

The conclusion has summarised potential policies for revising the OSP predictors and their motivation. These options were presented to HMPPS who will be publishing an addendum to this report, in due course, advising which policy they will be taking forward.

Glossary

Type	Term or abbreviation	Description
Abbreviation	ARAI	Actuarial Risk Assessment Instrument (see actuarial)
Abbreviation	CI	Confidence Interval, a range of values so defined that there is a specified probability that the value of a parameter lies within it.
Abbreviation	coef	Cox regression coefficient for a given predictor, see hazard
Abbreviation	DV	Domestic Violence, where domestic violence perpetrator status is defined using two OASys items. Current DV perpetrator status is defined using an item that refers to physical violence against a partner as part of the current offence, former DV perpetrator status is defined using an item that refers to all previous or ongoing abuse against family members.
Abbreviation	HMPPS	His Majesty's Prison and Probation Service
Abbreviation	HR	Hazard Ratios describe relative hazard between two populations in survival analysis study, see hazard
Abbreviation	IIOC	Indecent images of children
Abbreviation	LDC	Learning Development Challenge, as defined by Wakeling (2018)
Abbreviation	MoJ	Ministry of Justice
Abbreviation	NCNI	Not Contact and Not Indecent Images, see ONC
Abbreviation	OASys	Offender Assessment System, a structured assessment instrument used to record the risks and needs of eligible offenders in prisons and probation trusts across England and Wales.
Abbreviation	OFM	Offence-free months, see also offence-free time
Abbreviation	ONC	Other Non-Contact, a type of offending categorised under OSP/C that includes offences such as voyeurism.

The Actuarial Prediction of Sexual Reoffending

Type	Term or abbreviation	Description
Abbreviation	OSP	OASys sexual predictor. Umbrella term used for sexual prediction tools (which form part of RSR)
Abbreviation	OSP/C	OASys sexual predictor for contact sexual reoffending
Abbreviation	OSP/DC	OASys sexual predictor for direct contact (child and adult) sexual reoffending
Abbreviation	OSP/I	OASys sexual predictor for reoffending involving indecent images
Abbreviation	OSP/IIC	OASys sexual predictor for reoffending involving indecent images and indirect contact child sexual reoffending
Abbreviation	OSP/Indirect	OASys sexual predictor for reoffending involving indirect contact child sexual reoffending only
Abbreviation	PNC	Police National Computer. The PNC is a national database of information available to all police forces, law enforcement agencies and other specified bodies throughout UK.
Abbreviation	p-val	Represents the p-value indicating statistical significance, i.e. how likely it is a given result could have occurred by random chance
Abbreviation	SE coef	Standard error of regression coefficient, where standard error is the approximate standard deviation of a sample population.
Abbreviation	SO	Sexual Offence
Conceptual term	Actuarial	Relating to actuarial science, the discipline that applies mathematical and statistical methods to assess risk
Conceptual term	AUC	Area under the curve. The AUC is an aggregated accuracy metric used to say how well a model has performed in its predictions analogous to Harrell's C-index used in this report. A value of 0.5 would indicate that a model has performed no better than 'random guessing'. The higher (and closer to 1.0) the value is, the better model's predictions. Conversely, the lower (and closer to 0.0) the value is, the poorer the model's predictions are.
Conceptual term	C-index	See Harrell's C-index

The Actuarial Prediction of Sexual Reoffending

Type	Term or abbreviation	Description
Conceptual term	Calibration	Also referred to as accuracy - a comparison of a predictor's (see Predictor) mean predicted risk score against the actual rates of reoffending observed. Also see residual.
Conceptual term	Caseload	Main cohort studied in the OSP study. A snapshot of the community (probation) population as at a single point in time, currently 30 June 2018. Have varied offence free time.
Conceptual term	Censored	Used in the context of survival analysis, and specifically relating to 'right censored' data. A right-censored data point is an individual who has been removed from the study (partway) or reasons not related to the event (offence) being studied. This can be a prison recall or a custodial sentence. A 'left censored' data point would be one where the study start date is unknown (not an issue within this study)
Conceptual term	Censoring	See censored
Conceptual term	Cohort	A group of offenders - used primarily in the context of 'caseload' cohort.
Conceptual term	Concordance Index	See discriminative validity
Conceptual term	Cox proportional hazards survival	The Cox proportional hazards is a statistical model used to study the relationship (or association) between a varying value (such as a person's predicted risk score) and the time for an 'event' of interest to happen. In the context of this study, we are looking to understand the relationship between the predicted risk score and the time till a reoffence.
Conceptual term	Discrimination	See discriminative validity

The Actuarial Prediction of Sexual Reoffending

Type	Term or abbreviation	Description
Conceptual term	Discriminative validity	Used in the context of the accuracy of a risk predictor. That is, how well a risk score (the prediction) was at discriminating between lower- and higher-risk offenders. In this study we use (for survival analysis) Harrel's C-index. The higher the value of the C-Index, the better discrimination. It can be interpreted as the 'probability that a randomly selected individual who reoffended had a higher risk score than another randomly selected individual who did not reoffend (or reoffended later)' See 'Model discrimination - Harrel's C-index' in report. As with AUC a value of 0.5 would indicate that a model has performed no better than 'random guessing'. The higher (and closer to 1.0) the value is, the better model's discrimination.
Conceptual term	Harrell's C-index	See discriminative validity
Conceptual term	Hazard	The instantaneous risk of an event occurring among those who are still at risk. Used to describe results in cox proportional hazards regression
Conceptual term	Odds ratio	An odds ratio (OR) is a statistic that quantifies the strength of the association between two events, A and B (here, pairs of offences). They provide a general sense of how often two offences A and B co-occur compared with only one of them occurring
Conceptual term	Offence-free time	The number of whole months that an individual has been in the community without a proven (conviction) reoffence
Conceptual term	Paedophile manual	Any item containing advice or guidance about abusing children sexually as defined by the Serious Crime Act 2015
Conceptual term	Predictive validity	See discriminative validity
Conceptual term	Predictor	An actuarial risk instrument used to estimate the likelihood of reoffending

The Actuarial Prediction of Sexual Reoffending

Type	Term or abbreviation	Description
Conceptual term	Proven reoffending	<p>A proven reoffence is defined as any offence committed in a follow-up period that leads to a court conviction, caution, reprimand, or warning in the follow-up period or within a further waiting period to allow the offence to be proven in court. Used to make the distinction between offending which is not detected by police or proven at court.</p> <p>In most Ministry of Justice reoffending publications the follow up period will begin at an offender's entry into the community; however, in this publication the follow up period for all offenders begins on the 30th June 2018</p>
Conceptual term	Residual	The difference (in percentage points) between the actual and the predicted rate of reoffending
Conceptual term	Risk band	Risk scores can be grouped into bands such as 'low', 'medium', 'high', as well as 'very high' for some predictors.
Conceptual term	Sexual history	A current or historic proven conviction for a sexual offence
Conceptual term	Static risk factors	Any input to a risk predictor which cannot change (i.e. is static) over the course of a sentence, such as 'age at the commencement of risk' and criminal history.
Conceptual term	Survival analysis	Survival analysis is a branch of statistics for analysing the expected duration of time until one event occurs, such as a proven reoffence

References

- Almond, L., McManus, M. A., & Chatterton, H. (2017). Internet facilitated rape: A multivariate model of offense behavior. *Journal of Interpersonal Violence*, 35(21-22), 4979-5004. <https://doi.org/10.1177/0886260517718187>
- 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>
- Bell, K. (2018). *A quantitative study of the inter-rater reliability and completion timings of the OASys Sexual reoffending Predictor (OSP) compared to the Risk Matrix 2000 (RM2000/s)*. Report prepared for HMPPS. Unpublished.
- Briggs, P., Simon, W. T., & Simonsen, S. (2011). An Exploratory Study of Internet-Initiated Sexual Offenses and the Chat Room Sex Offender: Has the Internet Enabled a New Typology of Sex Offender? *Sexual Abuse*, 23(1), 72–91. <https://doi.org/10.1177/1079063210384275>
- Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2), 187–220. <http://www.jstor.org/stable/2985181>
- Craik, A., Han, L., Sullivan, L., Landsiedel, J., Travers, T., Spaul, C., & Howard, P. (2024). *Revalidation: Risk of recidivism tools. An evaluation of the actuarial instruments developed to assess recidivism risk in England and Wales*. (Ministry of Justice Analytical Series, unnumbered). London: Ministry of Justice.
- Crown Prosecution Service. (2022). Chapter 7: Key legislation and Offences. In *Rape and Sexual Offences – Overview and index of 2021 updated guidance*. London: Crown Prosecution Service. <https://www.cps.gov.uk/legal-guidance/rape-and-sexual-offences-chapter-7-key-legislation-and-offences>

Harrell, F. E., Lee, K. L., & Mark, D. B. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine* 15, 361-387.

Harrell, Frank E, Jr. 2015. *Regression Modeling Strategies*. 2nd ed. Switzerland: Springer International Publishing.

Howard, P. D. (2017). The effect of sample heterogeneity and risk categorization on Area Under the Curve predictive validity metrics. *Criminal Justice and Behavior*, 44 (1), 103-120.

Howard, P. D., & Barnett, G. D. (2015). The development and validation of the OASys Sexual reoffending Predictor (OSP). In Moore, R. (Ed.) (2015) *A compendium of research and analysis on the Offender Assessment System (OASys) 2009 – 2013*. (Ministry of Justice Analytical Series, unnumbered). London: Ministry of Justice.

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/449357/research-analysis-offender-assessment-system.pdf

Howard, P. D., Craik, A., Han, L. & Spaul, C. (2023). *Escalation in the severity of offending behaviour*. (Ministry of Justice Analytical Series, unnumbered). London: Ministry of Justice.

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1152995/escalation-severity-offending-behaviour.pdf

Howard, P. D., & Wakeling, H. (2021). *Comparing two predictors of sexual recidivism: the Risk Matrix 2000 and the OASys Sexual Reoffending Predictor*. (Ministry of Justice Analytical Series, unnumbered). London: Ministry of Justice.

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/955345/comparing-2-predictors-sexual-recidivism.pdf

HMPPS (2021). *Implementation and use of OASys Sexual reoffending Predictor (OSP) Policy Framework*.

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1037749/implementation-osp-pf.pdf

HMPPS Public Protection Group (2022) *OASys Sexual reoffending Predictor (OSP) Guidance for Practitioners*.

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1060166/osp-guidance-practitioners.pdf

Lim, Y. Y., Wahab, S., Kumar, J., Ibrahim, F., & Kamaluddin, M. R. (2021). Typologies and psychological profiles of child sexual abusers: An extensive review. *Children*, 8(5), 333.

<https://doi.org/10.3390/children8050333>

Ministry of Justice (2015). *Serious Crime Act 2015 Fact sheet: Offence of sexual communication with a child*.

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/416003/Fact_sheet_-_Offence_of_sexual_communication_with_a_child.pdf

Ministry of Justice (2023). *Criminal Court Statistics Quarterly: July to September 2023*.

<https://www.gov.uk/government/statistics/criminal-court-statistics-quarterly-july-to-september-2023>

Naing N. N. (2000). Easy way to learn standardization : direct and indirect methods.

The Malaysian journal of medical sciences : MJMS, 7(1), 10–15. Available at:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3406211/>

Therneau, T. M., & Watson, D. A. (2017). The concordance statistic and the Cox model.

Technical Report, 85, Department of Health Science Research, Mayo Clinic. Rochester, MN, USA. <https://www.mayo.edu/research/documents/bsi-techreport-85/doc-20433003>

Thornton, D., Mann, R., Webster, S., Blud, L., Travers, R., Friendship, C., & Erikson, M. (2003). Distinguishing between and combining risks for sexual and violent recidivism. In R.

Prentky, E. Janus, M. Seto, & A. W. Burgess (Eds). *Understanding and managing sexually coercive behavior*. *Annals of New York Academy of Sciences*, 989, 225–235. doi:

<https://doi.org/10.1111/j.1749-6632.2003.tb07308.x>

Thornton, D., Hanson, R. K., Kelley, S. M., & Mundt, J. C. (2021). Estimating Lifetime and Residual Risk for Individuals Who Remain Sexual Offense Free in the Community: Practical Applications. *Sexual Abuse*, 33(1), 3-33.

<https://doi.org/10.1177/1079063219871573>

Thornton, D. (2021). Sexual offending and classification. *Aggression and Violent Behavior*, 59, 101436. <https://doi.org/10.1016/j.avb.2020.101436>

Uno, H., Cai, T., Pencina, M. J., D'Agostino, R. B., & Wei, L. J. (2011). On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in medicine*, 30(10), 1105–1117. <https://doi.org/10.1002/sim.4154>

Wagstaff, K. *The Good Ol' Days of AOL Chat Rooms*; CNN: Atlanta, GA, USA, 2012.

<https://edition.cnn.com/2012/07/06/tech/web/time-aol-chat-room/index.html>

Wakeling, H. (2018). *The development of a screen to identify individuals who may need support with their learning*. HMPPS Analytical Summary 2018.

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/740303/development-screen-identify-individuals-oasys-report.pdf

Hanson, R. & Harris, Andrew & Letourneau, Elizabeth & Helmus, Leslie Maaiké & Thornton, David. (2017). *Reductions in Risk Based on Time Offense Free in the Community: Once a Sexual Offender, Not Always a Sexual Offender*. Psychology, Public Policy, and Law. <https://doi.org/10.1037/law0000135>

Grambsch, P. M., & Therneau, T. M. (1994). *Proportional Hazards Tests and Diagnostics Based on Weighted Residuals*. *Biometrika*, 81(3), 515–526.

<https://doi.org/10.2307/2337123>