



HM Prison &  
Probation Service

# Comparing two predictors of sexual recidivism: the Risk Matrix 2000 and the OASys Sexual Reoffending Predictor

**Philip Howard and Helen Wakeling**  
Ministry of Justice, and Her Majesty's Prison and Probation Service

Ministry of Justice Analytical Series  
2021

Preventing victims by changing lives



*Her Majesty's Prison and Probation Service is committed to evidence-based practice informed by high-quality social research and statistical analysis. We aim to contribute to the informed debate on effective practice with the people in our care in prisons, probation and youth custody.*

## **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 2021



© **Crown copyright 2021**

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](http://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 [National.Research@noms.gsi.gov.uk](mailto:National.Research@noms.gsi.gov.uk)

This publication is available for download at <http://www.justice.gov.uk/publications/research-and-analysis/moj>

ISBN 978-1-84099-866-5

# Contents

## List of tables

<b>1. Summary</b>	<b>1</b>
<b>2. Introduction</b>	<b>3</b>
<b>3. Method</b>	<b>5</b>
3.1 Data Sources	5
3.2 Measures	5
3.3 Procedure	6
3.4 Analysis Plan	7
3.5 Sample Characteristics	7
<b>4. Results</b>	<b>8</b>
4.1 Comparisons of predictive validity	8
4.2 The effect of reducing risk category after a five-year offence-free period	10
<b>5. Conclusions</b>	<b>12</b>
5.1 Implications	13
<b>References</b>	<b>15</b>
<b>Appendix 1: Further Analyses</b>	<b>17</b>
Method	17
Measures	17
Procedure	19
Results	19
Comparisons of Predictive Validity for a range of reoffending outcomes	19
OSP/C and RM2000/s Item Prediction	20

## List of tables

Table 1: Concordance Indices for Proven Sexual Reoffending	8
Table 2: Distributions and Failure Rates for Proven Contact Sexual and Indecent Image Reoffending, by Risk Scale and Category	9
Table A1: Concordance Indices for Proven Reoffending and other recidivistic outcomes	20
Table A2: Cox Regression Model: RM2000/s items as contact sexual reoffending predictors, in two steps	21
Table A3: Cox Regression Model: OSP/C items as contact sexual reoffending predictors	22

# 1. Summary

This study aimed to inform decisions about the risk assessment of men convicted of sexual offences. It did this by testing the predictive validity for sexual reoffending – the ability to separate lower- and higher-risk individuals – of two actuarial risk instruments, the OASys Sexual Reoffending Predictor (OSP) and the Risk Matrix 2000/s.

Previous research has found OSP to be a slightly better predictor of proven contact sexual reoffending than a simulated version of RM2000/s. This study re-examined this issue using full RM2000/s scores and minor revisions to OSP's scoring. It also compared a further OSP scale, developed for indecent image reoffending, with RM2000/s. Finally, the study examined whether length of time without offending in the community (on release from prison) affected contact sexual reoffending rates and whether this should be considered in determining individuals' risk ratings. It examines these issues using a sample of men released from custody in the 2000s.

## Approach

A prison sample of 2,728 men who had been convicted of at least one sexual offence and who had been risk assessed for participation in an offending behaviour programme, was utilised, and rates of proven reoffending were examined.

RM2000/s, OSP/Contact (OSP/C) and OSP/Indecent images (OSP/I) risk categories were produced for each individual in the sample, and the predictive validity of the tools were examined.

Caveats with the current study include the fact that the sample is historic, and that most had contact sexual offences as their index offence.

## Results

- OSP/C was a slightly better predictor of proven contact sexual reoffending than RM2000/s.
- By far the best predictor of proven indecent images reoffending was OSP/I.
- Analysis, based on a small sample, indicated that those contact sexual offenders who had been in the community offence-free for a five-year period had a lower risk of reoffending beyond the five-year point. These results, therefore, lend some support to the recent guidance proposing the reduction of risk by one category for

every five years an individual convicted of a sexual offence has been in the community offence-free.

## **Conclusions**

The results provide further support for the use of OSP over RM2000/s in assessing the risk of proven contact sexual reoffending in men with a sexual offending history. OSP has a number of advantages over RM2000/s: it focuses on specific types of sexual reoffending, it has a simpler scoring process, there is no age restriction on its use and it has better predictive validity for its two outcomes.

## 2. Introduction

Forensic actuarial risk assessment provides a means of predicting the likelihood of future offending. The prediction of the likelihood of recidivism is critical in making decisions regarding how convicted individuals should be managed and treated, and how resources should be prioritised. Actuarial risk assessment tools – which arrive at an overall risk level using empirically-supported risk factors and mathematical rules, rather than professional judgement – have been shown to be good predictors of proven reoffending (Hanson & Morton-Bourgon, 2009). The actuarial tool Risk Matrix 2000/s (RM2000/s) was developed using a large sample of adult males convicted of sexual offences in the United Kingdom (UK) and cross-validated with further UK samples (Thornton et al., 2003). Her Majesty's Prison and Probation Service (HMPPS) and the Police Service currently use it across England and Wales to assess the sexual recidivism risk of such men.

Although RM2000/s is well evidenced and validated, a new actuarial tool called the OASys Sexual Reoffending Predictor (OSP) was developed by Howard and Barnett (2015) in order to address some concerns with RM2000/s. In its initial form (now OSP/C), OSP focused on those offences which tend to cause the most serious harm to the victims (i.e., contact sexual offences), rather than all sexual recidivism. It now also has an additional scale, OSP/I, which predicts the risk of indecent images of children recidivism.

In addition to its focus on specific offences, OSP was designed to be less resource intensive than RM2000/s, and to be used with all men who have committed sexual offending (RM2000/s cannot be used on those whose sexual offending was confined to childhood, i.e., all offending took place before the age of 16).

For HMPPS to adopt this new predictor and set aside RM2000/s, which is well-established and based on large samples with long follow up periods, it would require strong evidence that OSP is a better or equivalent predictor *as well as being* easier to implement. The claim that OSP is easier to use has been checked recently in a separate study in which 22 HMPPS users each scored both tools using real materials for up to ten cases (Bell, 2018). The research confirmed that the two tools have similar inter-rater reliability but that OSP is scored more quickly. In the OSP development research (Howard & Barnett, 2015), OSP was found to be slightly better able to discriminate contact sexual reoffenders from non-reoffenders than a simulated version of RM2000/s. The present study aims to strengthen the evidence base on which tool is the better predictor, by directly comparing *real* (as opposed to simulated) RM2000/s and OSP scores on a large sample of men convicted of sexual offences. Due to

data availability, this sample consists of those receiving custodial sentences and assessed prior to treatment designed to address sexual offending.

This study also examines whether the length of time an individual spends in the community is associated with a reduction in the likelihood of them committing a further sexual offence in line with recent research (Hanson et al., 2014). The current guidance (NOMS, 2014), as advised by the developer of RM2000/s, is that the RM2000/s risk category should be reduced by one level for every five years an individual convicted of a sexual offence has been in the community offence-free. By testing this advice against the most recent data for England and Wales, this study can provide an evidence base to maximise the effectiveness of ongoing community supervision.

In summary, the present research aimed to:

- 1) Compare the predictive validity of the OSP/C, OSP/I and RM2000/s, for sexual reoffending outcomes, and
- 2) Examine the effects of reducing risk category for those who have been in the community offence-free.

This study's Appendix provides technical details related to these two primary aims. It also sets out the methods and results relating to two secondary aims, to:

- 3) Compare the predictive validity of the above tools and three tools designed to predict general or nonsexual violent reoffending, for a range of other outcomes. This analysis aimed to inform wider protocols around the risk assessment of men convicted of sexual offences, and
- 4) Examine the predictive contribution of each RM2000/s and OSP/C item, to inform future development of risk tools.



## 3. Method

### 3.1 Data Sources

The final sample was created by combining records from a number of databases, including the Offender Assessment System (OASys; Home Office, 2006) database, and a treatment database held by the HMPPS Interventions Services (IS) unit. The IS database is the only centrally-collated RM2000/s source, and is restricted to men being considered for treatment in custody. Police National Computer (PNC) data was obtained from the Ministry of Justice Police National Computer (MoJPNC) database. As detailed by Howard and Barnett (2015), these databases had earlier been accessed to create the datasets used to construct OSP. Although there was initially some overlap between the current sample used in this study and the OSP development sample, development sample cases were removed (see section 3.3), and the current sample included a number of additional cases (e.g., those with past sexual offences).

### 3.2 Measures

#### ***RM2000/s (Thornton et al., 2003)***

RM2000/s is a risk assessment tool for use with adult males who have ever been convicted of a sexual offence committed when the individual was age 16 or over. The RM2000/s predicts sexual recidivism and is made up of seven items, generally static in nature (i.e., only subject to change due to aging or new criminality). These relate to age, sexual and general criminal history, sexual offence/victim type and relationship history. The Appendix details these items and how they are combined to generate a rating of low, medium, high or very high risk. The RM2000/s has been shown to have moderate predictive validity in a recent large sample of individuals convicted of sexual offences in England and Wales (Barnett, Wakeling & Howard, 2010), as well as a variety of other populations of males convicted of sexual offences (Helmus, Babchishin & Hanson, 2013).

#### ***OSP (Howard & Barnett, 2015)***

The OSP is a risk assessment tool for use with adult males who have ever been convicted of a sexual offence, committed at any age. It was intended for use alongside HMPPS's Offender Assessment System (OASys: Home Office, 2006), the risk and need assessment tool used with adults across HMPPS, but can also be used when no OASys has been completed.

OSP's development was informed by evidence emerging after the creation of RM2000/s, showing that men convicted of sexual offences tend to specialise in particular types of sexual offence (this was demonstrated by Howard, Barnett & Mann, 2015, who summarized past evidence and analysed recent English and Welsh data). Its scoring involves dividing sexual offences into four groups: 1. contact adult, 2. contact child, 3. indecent images and 4. other noncontact, which are fully detailed in user guidance.<sup>1</sup> The initial version of OSP focused on offences causing the greatest harm, by predicting offences involving victim contact (contact sexual reoffending – offences in the contact adult and contact child groups). In the current version of OSP, this scale has been slightly rescored and renamed OSP/C, while an additional scale, OSP/I, has been developed in response to police and HMPPS requests for a predictor of indecent image reoffending. As originally designed, each scale rates individuals as low, medium, high or very high risk of the relevant type of proven reoffending, using scoring rules set out in the Appendix. To better meet operational needs, OSP/I has been recategorized so that individuals are scored as low, medium or high risk.

In response to HMPPS concerns that the switch from RM2000/s to a new predictor would cause operational disruption, the thresholds dividing OSP/C scores into low, medium, high and very high categories were set in a manner that minimised the difference in RM2000/s and OSP/C distributions (e.g., so the percentage of cases rated low on OSP/C was as close as possible to the percentage rated low on RM2000/s).

### 3.3 Procedure

Recidivism outcomes were coded blind to risk assessment scores. Proven reoffending was restricted to any conviction or caution for a new offence. To ensure that the periods at risk of each type of reoffending were identified correctly, separate recidivism follow up periods and outcomes were coded for each studied reoffence type. When studying any given type of reoffence, the follow up period lasted until that offence was committed, or the start of the waiting period, or the individual was imprisoned due to a new conviction for any other offence, whichever was the earliest. The waiting period comprised the last six months for which PNC data was available, to ensure there was time for all detected new offences to be converted into cautions and convictions. When an individual was recalled to prison prior to any of these dates, the days between recall date and rerelease or their follow up end date were subtracted from the follow up duration.

---

<sup>1</sup> Where an individual was simultaneously convicted of offences from more than one of these categories, only one category is scored, using the numbering above to determine preference (i.e., contact adult offences are selected when present, whereas indecent image offences are never preferred to other offence categories).

Data to score RM2000/s was available for 3,548 individuals who had been discharged from prison between 2003 and 2008 and were not included in the OSP development sample. 2,833 of these individuals were able to be matched to the MoJPNC database (an unusually low matching rate of 80%, due to poor recording of personal identifiers in the IS database). After a process of data cleaning (further individuals were removed for inadequate OASys or recall data, or being released after the start of the waiting period), complete and consistent records for a total of 2,728 individuals were included in the data analysis. As noted above, reoffending follow up periods varied according to outcome; in the study of contact sexual reoffending, the mean reoffending follow up length was 1,649 days (4 years, 6 months).

### 3.4 Analysis Plan

The two principal aims of the study were taken forward by:

- Examining the predictive validity of the three tools for sexual reoffending outcomes (Table 1) and how contact sexual reoffending rates vary by RM2000/s and OSP/C risk band (Table 2). Comparisons of predictive validity use the Concordance Index (C Index); its values have a theoretical range from 0 to 1, where .5 represents a prediction that is no better than chance, and 1 represents perfect prediction, obtained by rating all reoffenders as higher risk than all nonreoffenders.<sup>2</sup>
- Checking whether contact sexual reoffending risk reduces over time, by assessing whether the current guidance to reduce risk level every five years offence-free is appropriate for both RM2000/s and OSP/C.

### 3.5 Sample Characteristics

The mean age of the matched sample (n=2,728) was 45.0 years, which was markedly higher than in previous research (Barnett et al., 2010; Howard & Barnett, 2015). This reflects the fact that this is a prison sample. Where ethnicity was recorded, around 90% were White British. The majority (63%) had an index conviction for contact sexual offending with a child, followed by contact sexual offending with an adult (22%), and indecent images (9%). As this was a prison-only sample, the risk distributions of OSP/C and RM2000/s scores did not match closely, unlike the all-HMPPS distributions as tested on a previous HMPPS caseload (September 2016).

---

<sup>2</sup> The C Index statistic compares the trade-off between false negative and false positive predictions at varying risk thresholds. A high C Index indicates that the predictor manages to achieve low rates of both false positive and false negative prediction across most of the empirical range of scores; in other words, it is possible to set risk thresholds that most nonreoffenders score below while many reoffenders score above and/or vice versa. It is similar to the Area Under the ROC Curve (AUC) often found in reoffending research. The C Index has the advantage that it takes into account different lengths of follow-up and therefore avoids inappropriate comparisons between those not reoffending within short follow-ups and those reoffending after longer periods at risk.

## 4. Results

### 4.1 Comparisons of predictive validity

Table 1 shows the C Indices (or Concordance Index) of the three sexual risk scales for the two sexual reoffending outcomes. Because OSP/C can be used as an uncategorised (64-point) score or in its banded form (i.e., Low, Medium, High or Very High risk), both are presented.

**Table 1: Concordance Indices for Proven Sexual Reoffending**

Risk scale	Outcome (% who reoffended)	
	Contact sexual (2.3%)	Indecent images (2.4%)
Risk Matrix 2000/s	0.707	0.625
OSP/C (4 bands)	0.742	0.538
OSP/C (64-point scale)	0.763	0.555
OSP/I	0.491	0.742 <sup>3</sup>

The unbanded version of the OSP/C score (64 point scale) is the best predictor of contact sexual reoffending, with a C Index of .763. OSP/C has a C Index of .742 in its banded form, compared to a C Index of .707 for RM2000/s. While traditional tests of statistical significance are not available for follow-ups of variable length, a simulation provides strong evidence that the superiority of OSP/C will generalize beyond the present sample.<sup>4</sup> While the results of the simulation are not statistically significant on their own (one-sided p value of 0.11), they concur with the original results of Howard & Barnett (2015); while the latter was the OSP development study, it used a different bootstrap process to confirm that OSP/C's superiority to RM2000/s was not due to the optimism bias inherent in tool development. In short, while neither study offers definitive proof of OSP/C's superiority, their results concur.

<sup>3</sup> OSP/I may be implemented in a three-group version, merging the Low and Medium groups (i.e., Very High becomes High, High becomes Medium, Low and Medium become Low). The C Index of this three-group version is about 0.72. Estimation was necessary because the implementation proposal was made after the full data for this study became unavailable: making calculations from the data in Table 2 as if it were a fixed follow-up yields an Area Under the ROC Curve of 0.759, while a three-group version would have an AUC of 0.740. Given the close similarities between AUC and C Index formulae, it is reasonable to infer that the C Index should also be reduced by 0.02.

<sup>4</sup> The generalizability of these C Index results was tested by generating 1,000 bootstrap replications of the sample – a statistical simulation of what might happen if 1,000 similar studies were conducted - and calculating each new sample's C Indices for contact sexual reoffending of RM2000/s and OSP/C in its four-band form. Their mean C Indices, with 95% confidence intervals were 0.708 (0.651, 0.764) and 0.742 (0.686, 0.792) respectively, for a mean estimated improvement of 0.038 (95% CI of (-0.012, 0.080)) for OSP/C. OSP/C's C Index was superior to that of RM2000/s in 887 of the 1000 comparisons, and it was superior by a non-negligible amount - at least 2 points (0.02) - in 690 comparisons.

Indecent image recidivism was very clearly best predicted by OSP/I, with a C Index of .74 compared with .63 for RM2000/s. The specialised nature of OSP/C and OSP/I is confirmed: each was the best predictor of the outcome for which it was designed, and a poor predictor of the other sexual reoffending outcome.

Table 2 examines the distributions and contact sexual and indecent image reoffending failure rates by risk category, at the five-year point using life table methods.<sup>5</sup>

**Table 2: Distributions and Failure Rates for Proven Contact Sexual and Indecent Image Reoffending, by Risk Scale and Category**

Risk Scale	N of offenders	Failure rate (1 – survival rate) for 5-year proven reoffending	
		Contact sexual	Indecent images
RM2000/s			
Low	1,185	1.2%	2.5%
Medium	980	2.1%	2.7%
High	440	6.0%	5.5%
Very High	123	10.4%	9.6%
OSP/C			
Low	993	0.7%	
Medium	1,030	1.9%	n/a
High	526	4.4%	
Very High	179	12.6%	
OSP/I			
Low	1,884		1.3%
Medium	251	n/a	3.4%
High	546		8.8%
Very High	47		32.4%
All offenders	2,728	2.6%	3.3%

Both RM2000/s and OSP/C showed contact sexual reoffending percentages approximately 10 times higher in their 'Very High' risk categories than their 'Low' categories, with risk approximately doubling at each increase in category, with this pattern being stronger for OSP/C. This is a concrete example of the very good predictive validity – especially for OSP/C – demonstrated by the C Indices in Table 1. The figures show that both tools –

<sup>5</sup> Life table methods are used to calculate these estimates: the failure rates are estimates of the proportion in each category who would have reoffended within five years if their follow-up was at least that long and they were not imprisoned for any other offence. If an individual was recalled to custody, their follow-ups were paused at the point of recall, and resumed if they were discharged again. Standardising the length of follow up attempts to overcome the fact that the mean length of follow-up tends to differ between risk category, thus enabling a better comparison between risk tools and categories, and using life table (survival analytic) rather than fixed five-year follow-ups makes more use of the available data.

but especially OSP/C – are able to identify groups of offenders with very low and (in relative terms) very high risk of sexual reoffending, enabling criminal justice practitioners to effectively concentrate offender management and public protection resources on higher-risk cases.

For indecent images reoffending, OSP/I classifies two-thirds of offenders into the Low risk category, whose failure rate is below half the overall mean. Men with Very High OSP/I (i.e., multiple prior indecent image sanctions) present an elevated risk, though, due to the small number in this category, this reoffending rate cannot be generalized with any precision.<sup>6</sup>

Caution should be used in interpreting these results, as the sample is not fully representative of the HMPPS caseload: the September 2016 caseload (mentioned above) has balanced OSP/C and RM2000/s distributions, while in this sample OSP/C scores skewed slightly higher.<sup>7</sup> This is likely to be because this subsample reflects those receiving prison treatment some years ago – these were mostly contact sexual offenders, and contact offending is the predictive focus of OSP/C, as reflected in the items used (see the Appendix). The distribution of OSP/I scores was not designed to bear any similarity to that of RM2000/s scores, and indeed is very different.

## 4.2 The effect of reducing risk category after a five-year offence-free period

Contact sexual reoffending rates were examined for 974 individuals from the sample with more than five years follow-up for this offence type. Some men included in Table 2 were not included: those released from custody too late to have a five-year follow-up, those who committed contact sexual offences in the first five years and those in custody due to recall at the five-year point.<sup>8</sup>

---

<sup>6</sup> The standard error for this group is .1106, and therefore the 95% confidence interval of the failure rate is (10.7%, 54.1%). Of 47 individuals, 8 had indecent image reoffences within five years, 31 had censored observations (i.e., short follow-up or reimprisonment) and 8 were still at risk at the five-year point. Note the difference between 8/47 and the 32% quoted in Table 2: the life table methods used in Table 2 take into account the gradual loss to follow-up of these 31 cases. One way of visualising this is to say that the life table methods converted these partial follow-ups of 31 cases to the equivalent to complete five-year follow-ups of 9 more non-reoffenders, and so the 32% failure rate is similar to an 8/25 reoffending rate, the denominator of 25 comprising the 8 reoffenders, 8 still at risk at the five-year point and 9 extra synthetic cases.

<sup>7</sup> Another obstacle to interpretation is the higher reoffending rate for three of the four RM2000 categories than the equivalent OSP/C category. This apparent paradox is explained by more offenders being in the higher categories of OSP/C. The C Index is used to help overcome such comparison problems.

<sup>8</sup> The proposals on reducing risk category are not specific about what should happen if recall but not proven reoffending occurs in the first five years. A decision was made to include individuals who had been recalled and re-released within the first five years, but not those who were in custody under recall at the five-year point.

The results of this analysis provide only weak evidence, as there were just 11 contact sexual reoffenders beyond the five-year follow-up point. This limited evidence is consistent with what is already known about risk reduction as, for contact sexual offences, the rates associated with each risk category in the later period appear similar to the rates of the next lowest category in the first five years. (A similar study for indecent images reoffending could not be meaningfully conducted, as there were just two such reoffenders beyond the five-year follow-up point.)

While this evidence alone would be insufficient to develop recommendations, it does provide tentative support to Thornton's existing recommendation that individuals should be downgraded by one risk category if they survive five years of follow-up without sexual reoffending. That is, if an individual was assessed as medium risk, and then was offence-free in the community for five years, it would be reasonable to treat him as a low risk individual from then on. Conditional on progress in addressing his dynamic risk factors, it would then be assumed that his risk of further sexual reoffending had reduced.

## 5. Conclusions

This study examined and compared the predictive validity of OSP (/C and/I) and RM2000/s for proven contact sexual offending. The best predictor of proven contact sexual reoffending was OSP/C in its unbanded form, which is of limited utility to practitioners as it does not provide a straightforward means of allocating resources to higher-risk individuals. However, OSP/C in its banded form was still a slightly better predictor of contact sexual offending than RM2000/s (C Index of .74 compared with .71). To express this in absolute terms, contact sexual rates were around 10 times higher in the Very High risk categories than the Low risk categories, and the rates approximately doubled at each increase in risk category. Therefore both predictors can be used to identify the level of risk presented by offenders, with implications for offender management and public protection, though OSP/C is superior.

OSP/I was very clearly the best predictor of indecent image reoffending. Previous research on the specialization of offending by individuals convicted of sexual offences (e.g., Howard, Barnett & Mann, 2015) shows that a history of indecent image offending is associated with raised rates of such reoffending and low rates of all other reoffending.

Our results lend tentative support to the recent guidance proposing the reduction of risk by one category for every five years an individual convicted of a sexual offence has been in the community offence-free. Contact sexual reoffending rates associated with each risk category in the post-five-years period seemed similar to the rates of the next lowest category in the first five years, for both tools. Limited sample sizes beyond the five year point prevented reliable testing of whether risk reduction also occurs among those who have reoffended non-sexually rather than desisted completely. Further research (including a later follow-up of the present sample) must re-examine this issue before firm conclusions can be drawn about this revised guidance and its application with OSP.

One caveat of the current study is that it uses a historic prison treatment sample. The offenders featured mostly had contact sexual index offences, and there were very few whose sexual offending was not their most recent offence. The former point potentially reduces the measured predictive validity of both tools by creating a more homogeneous sample. The latter point probably disadvantages OSP/C, which dedicates 10 of its 64 points to distinguishing those whose only sexual offending sanctions were before age 18, although those for whom all sexual offending occurred before age 16 cannot be assessed using RM2000/s at all.



Like all studies of actuarial prediction of sexual reoffending, these findings are only applicable for men. Women convicted of sexual offending are few in number and have very low sexual reoffending rates, rendering actuarial prediction impractical.

## 5.1 Implications

The current study has found further support for OSP/C in assessing contact sexual reoffending risk in men with a sexual offending history; it has a slightly better C Index than RM2000/s in predicting this outcome. This accords with previous research comparing the two tools (Howard & Barnett, 2015), but solidifies this finding by using *real*, and not *simulated* RM2000/s scores, a limitation of the previous study. Statistical replication of the current study suggests that it is very likely that this difference is real, and probable that the difference is large enough to have practical utility.

For indecent image reoffending, the predictive validity of OSP/I far exceeds that of all other tools. In addition to better predictive validity, OSP has the following advantages over RM2000/s.

- 1) OSP/C predicts contact sexual reoffending and therefore focuses on those offences most likely to cause direct and intense harm. OSP/I is also of interest to practitioners. The less harmful “other noncontact” offences are set aside.
- 2) Scoring OSP is a simpler process than scoring RM2000/s, which requires intensive training and a detailed understanding of the individual’s criminal and relationship history.
- 3) OSP enables a risk score to be generated for those individuals who are under the age of 18, and for those whose sexual offending all took place before the age of 16, which the RM2000/s does not.
- 4) Use of OSP, as part of OASys, provides consistency with the tools used by HMPPS to predict other forms of reoffending (e.g. the OGRS tools) and will allow a more structured programme of recalibration and ongoing improvement in the future.

Due to these advantages along with its slightly superior predictive ability, we suggest that RM2000/s is replaced with OSP as the preferred static risk assessment tool, which should be used alongside dynamic risk assessment to provide a global risk assessment for men convicted of sexual offences.

The practical considerations of moving to a new tool however must not be overlooked. In order to enable consistency across the Criminal Justice System, it would make sense for all organisations within England and Wales to use the same tool, which includes risk assessors

from the Prison Service, probation and police. Accordingly, the police have amended their system to manage violent and sexual offenders – ViSOR – to ensure that the management of men convicted of sexual offences will remain consistent across agencies.

Police forces are required to manage the risk presented by individuals convicted of sexual offences for the duration of their registration – far longer than most individuals are managed on probation. Our confirmation of previous results, showing that their risk levels appear to be lower after five years offence-free, is therefore encouraging. An evidence-based process to reduce the intensity of their supervision could enable the police to realise considerable resource savings without endangering public safety.

## References

- Barnett, G. D., Wakeling, H. C., & Howard, P. D. (2010). An examination of the predictive validity of the Risk Matrix 2000 in England and Wales. *Sexual Abuse: A Journal of Research and Treatment*, 22, 443–470. doi: 10.1177/1079063210384274
- 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.
- Hanson, R. K., Harris, A. J. R., Helmus, L. & Thornton, D. (2014). High-risk sex offenders may not be high risk forever. *Journal of Interpersonal Violence*, 29, 2792–2813. doi: 10.1177/0886260514526062
- Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment*, 21, 1–21. Doi: 10.1037/a0014421.
- Helmus, L., Babchishin, K. M., & Hanson, R. K. (2013). The predictive accuracy of the Risk Matrix 2000: A meta-analysis. *Sexual Offender Treatment*, 8, 1.24.
- Home Office (2006). *Offender Assessment System Manual version 2*. London: Home Office.
- Howard, P. (2015). OGRS4. The revised Offender Group Reconviction Scale. 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. Available at [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/449357/research-analysis-offender-assessment-system.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/449357/research-analysis-offender-assessment-system.pdf)
- Howard, P. D., & Barnett, G. D. (2015). The development and validation of the OASys Sexual reoffending Predictor (OSP). In Moore, R. (Ed.) (2015) (see above)
- Howard, P. D., Barnett, G. D., & Mann, R. E. (2015). Specialization in and within sexual offending in England and Wales. *Sexual Abuse: A Journal of Research and Treatment*, 26, 225–251. doi: 10.1177/1079063213486934

Howard, P., Francis, B., Soothill, K., & Humphreys, L. (2009). *OGRS 3: the revised Offender Group Reconviction Scale* (Research Summary 7/09). London: Ministry of Justice.

National Offender Management Service (2014). *Guidance given to RM2000 Administrators*. Available from HMPPS Interventions Services on request.

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: 10.1111/j.1749-6632.2003.tb07308.

Wakeling, H.C., Freemantle, N., Beech, A.R., & Elliott, I.A. (2011). Identifying predictors of recidivism in a large sample of United Kingdom sexual offenders: A prognostic model. *Psychological Services*, 8, 307–318. doi:10.1037/a0025516.

## Appendix 1: Further Analyses

In addition to the analyses included in the main report, the research team sought to establish:

- 1) whether actuarial risk assessment instruments not developed specifically for sexual reoffending outcomes were predictive of sexual reoffending for men convicted of sexual offences;
- 2) whether these tools, OSP/C, OSP/I and RM2000 were predictive of other recidivism outcomes for men convicted of sexual offences, and
- 3) which OSP/C and RM2000 items contributed most to the prediction of contact sexual reoffending.

The first two questions are relevant to the process of assessing and managing men convicted of sexual offences, while the latter question informs the design of future risk assessment instruments.

All three of these questions are addressed in this Appendix.

### Method

#### Measures

##### **RM2000**

RM2000's seven items are combined into a final risk rating across two 'steps'. The first three items make up step one: Age of the individual on release, number of sentencing occasions for a sexual offence, and number of sentencing occasions for any criminal offence. Scores assigned to each of these items are summed and translated into one of four initial risk categories (low, medium, high or very high). The second scoring step has four items: any or all of the victims of sexual offending have been male, any or all of the victims of sexual offending have been strangers, the individual has never had a stable live in relationship for over two years (termed the 'single' item), and any or all of their sexual offences have been noncontact (excluding internet offences for those whose only sexual offences relate to the internet). If two or three of these items are present the initial risk category is raised by one level. If all four are present, the risk category is raised by two levels.

##### **OSP**

When designed by Howard and Barnett (2015), the total OSP/C score ran from 0 to 32, and it has now been revised to run from 0 to 64. (An analysis of the September 2016 HMPPS caseload, matched with RM2000/s scores from the multiagency ViSOR database, shows that a 64-point score generates a more similar distribution of cases across the risk bands than the

32-point version.) Unlike RM2000, the OSP/C score is created by adding points for all of its seven items in a single stage. The revised scoring rubric is as follows:

- Sanctions for contact adult sexual offences: zero (0 points); one (5); two (10); three or more (15).
- Sanctions for contact child sexual offences: zero (0 points); one (3); two (6); three or more (9).
- Sanctions for noncontact sexual offences other than indecent images: zero (0 points); one (2); two (4); three or more (6).
- Age at start of reoffending follow-up: 18 to 20 (14 points); 21 to 23 (13); 24 to 26 (12)... 57 to 59 (1); 60 and over (0).
- Age at last sanction for a sexual offence: 10 to 15 (0 points), 16 or 17 (5); 18 and over (10).
- Any previous criminal history: no (0 points); yes (6).
- Index sanction includes a contact sexual offence with a stranger victim: no (0 points); yes (4).

(In research settings, information on 'stranger victim' may not be available; if so, 2 points should be scored for a contact sexual index sanction.)

The total risk score is categorised as: Low (0–21 points); Medium (22–29); High (30–35); Very High (36–64).

OSP/I estimates indecent images risk based on sexual offending history alone, as other offending history is not associated with this outcome and age has a nonlinear relationship (Howard & Barnett, 2015) which does not improve predictive validity. Offenders are placed in four categories: the greatest risk is presented by those with multiple indecent images sanctions, followed by those with one indecent image sanction, then those with multiple contact child sanctions, and finally all other men with sexual offending sanctions. As mentioned in footnote 3, OSP/I can also be treated as a three-category predictor, by merging the last two groups (i.e., all those without indecent image sanctions are Low risk).

### ***OGRS3, OGRS4/G and OGRS4/V***

The OGRS3 (Howard, Francis, Soothill & Humphreys, 2009) is a predictor of all proven reoffending, in use throughout HMPPS. The OGRS4/G and OGRS4/V (Howard, 2015) are, respectively, an update to OGRS3 and a predictor of nonsexual violent proven reoffending. All three versions of OGRS use logistic regression to produce percentage estimates of the probability of an offence being committed within two years that later leads to caution or

conviction, based on static risk factors – gender, current offence type, age at first and current offence and at discharge from custody, number of previous sanctions and (for OGRS4/V only) violent sanctions.

## Procedure

The three secondary aims of the study are taken forward by:

- Examining the predictive validity of the three scores specific to sexual offending, and the three OGRS scores, for outcomes other than sexual reoffending outcomes (Table A1)
- Examining the component items of RM2000/s and OSP/C as predictors of contact sexual reoffending, using the Cox regression analysis method for data with varying follow up durations (Tables A2 and A3)

## Results

### Comparisons of Predictive Validity for a range of reoffending outcomes

Table A1 shows the C Indices of the three sexual risk scales and other risk assessment tools, with a range of reoffending outcomes. For completeness, it repeats those in Table 1.

The three OGRS predictors are stronger predictors of contact sexual reoffending than might be expected (C Indices range between .733 and .748). The continuous nature of OGRS predictors means that a statistical artefact partially explains their high C Indices (Howard, 2016); when scores on these OGRS predictors were grouped into bands of equal sizes to OSP bands, their C Indices typically fell by .02, and became comparable to those of RM2000/s. Nonsexual violence was very highly predictable, with the nonsexual violence instrument (OGRS4/V) having the best C Index (.882). The outcomes of all reoffending, breach and recall were all predicted best by OGRS4/G but only slightly less well by OGRS3 and OGRS4/V. All instruments achieved only fair predictive validity for noncompliance recidivism, even though the OGRS measures were strong predictors of breach recidivism and good predictors of recall. This suggests static risk factors for nonsexual recidivism, such as age and general criminal history, are closely associated with nonconformity with community supervision and restrictions, but noncompliance may be complicated by other factors (e.g., not all individuals are serving a Risk of Sexual Harm Order) which may demand more in-depth study.

**Table A1: Concordance Indices for Proven Reoffending and other recidivistic outcomes**

Risk scale	Outcome (% who reoffended)						
	Contact sexual (2.3%)	Indecent images (0.6%)	Noncompliance (10.5%)	Nonsexual violence (5.7%)	All reoffending (21.2%)	Breach (1.3%)	Recall (8.0%)
Risk Matrix 2000/s	0.707	0.625	0.600	0.776	0.654	0.732	0.653
OSP/C (4 bands)	0.743	0.538	0.616	0.808	0.679	0.804	0.669
OSP/C (64-point scale)	0.763	0.555	0.630	0.833	0.694	0.835	0.691
OSP/I	0.491	0.742	0.508	0.421	0.495	0.443	0.526
OGRS3	0.743	0.567	0.652	0.873	0.726	0.856	0.731
OGRS4/G	0.748	0.582	0.654	0.870	0.727	0.861	0.737
OGRS4/V	0.733	0.551	0.647	0.882	0.725	0.854	0.725

*Note: "Noncompliance" comprised criminal offences such as failure to notify police of address changes (as a registered sex offender) or breach of civil orders (e.g., Risk of Sexual Harm Order). "Breach" represents the noncriminal behaviour, for which offenders can be returned to court, of breaching licence conditions. 'Recall' censors follow-up at the date of first reoffending, thereby excluding recalls which might have occurred in response to offences subsequently resulting in new caution/conviction. All recall, without censoring for reoffending, had a 10.5% rate and C Indices within  $\pm 0.01$  of the censored recall outcome.*

## OSP/C and RM2000/s Item Prediction

To examine further why the OSP/C was a slightly better predictor of contact sexual reoffending than RM2000/s, Cox regression was used to examine the relative predictive contribution of the RM2000/s items across the two steps of score calculation (see Table A2). As these Cox regressions are multivariate analyses, the results at both steps describe the contributions that each item makes when taking the other items into account. For example, while those with more criminal convictions tend to be older; the Cox model allows us to control for this in estimating the association between age and reoffending.

At Step One, the age and sexual sentencing items were strongly significant predictors of contact sexual reoffending, while the criminal sentencing appearance item was positively associated but not statistically significant. At step two, the Step One total score was highly predictive, but the stranger victim item was the only significant predictor of the four Step Two aggravating factors, though the non-significant 'single' item was associated with a 41% increase in reoffending hazard. The strong correlation between the latter item and age (and therefore the Step One total score) – may be linked to its lack of statistical significance. These results indicate that some RM2000/s items dilute its predictive validity, perhaps making it inferior to that of OSP/C. The finding that the age and sexual sentencing items were the best predictors of this outcome are in line with previous research (Wakeling,



Freemantle, Beech & Elliot, 2011). Similarly, the weak results for the male victim and noncontact offence items may in part be related to this choice of outcome measure; the noncontact item may be a better predictor of noncontact sexual reoffending, as supported by past specialization research (Howard et al., 2015).

**Table A2: Cox Regression Model: RM2000/s items as contact sexual reoffending predictors, in two steps**

Step One			
Risk factor (score range)	Beta	SE (Beta)	Hazard ratio
Age at commencement of risk (2)	0.90***	0.17	2.46
Sexual sentencing appearances (3)	0.67***	0.15	1.96
Criminal sentencing appearances (1)	0.29	0.30	1.33
Step Two			
Risk factor (score range)	Beta	SE (Beta)	Hazard ratio
Step One total score (6)	0.68***	0.15	1.98
Any conviction for a sexual offence against a male? (1)	0.16	0.30	1.18
Any conviction for a sexual offence against a stranger? (1)	0.82**	0.28	2.28
Single? (1)	0.38	0.27	1.47
Any conviction for a noncontact sexual offence (1)	-0.33	0.35	0.72

*Note.* \*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

A further Cox regression was conducted using scores on the seven OSP/C items to predict contact sexual reoffending (see Table A3). The first sanction, contact child and contact adult sanctions and age at start of follow-up items had significant positive associations with reoffending. Age at last sexual sanction and sanctions for other noncontact offences were not significantly associated with reoffending. The age at last sexual sanction result was likely to do with the fact that (unlike previous samples examined) almost all of this sample had index sexual offences, and so had the same score (“aged 18+”). Similarly, just 147 of the 2,728 men had ever been sanctioned for other noncontact offences. The limited size and somewhat unrepresentative nature of this sample means that precise conclusions about the weights of individual items should not be drawn: while ideally the beta estimates for all seven risk factors would be identical, the estimated weights do not vary greatly, allowing for the large standard errors of the other noncontact and age at last sexual sanction items.

**Table A3: Cox Regression Model: OSP/C items as contact sexual reoffending predictors**

Risk factor (score range)	Beta	SE (Beta)	Hazard ratio
First sanction for any offence? (6)	0.12*	0.06	1.13
Current contact offence with stranger victim (4)	0.17	0.09	1.18
Age at last sexual sanction (10)	0.12	0.24	1.13
Sanctions for other noncontact sexual offences (6)	-0.03	0.17	0.97
Sanctions for contact child offences (9)	0.18***	0.05	1.19
Sanctions for contact adult offences (15)	0.11***	0.03	1.12
Age at start of follow-up (14)	0.15***	0.04	1.16

*Note.* \*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ . 'Other noncontact sexual offences' are those not involving indecent images of children, of which the most frequent are indecent exposure and voyeurism.