Ministry
of Justice

Impact evaluation of the prison-based Core Sex Offender Treatment Programme

Aidan Mews, Laura Di Bella and Mark Purver
Ministry of Justice

Ministry of Justice Analytical Series
2017

Analytical Services 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 2017

Any enquiries regarding this publication should be sent to us at mojanalyticalservices@justice.gsi.gov.uk

This publication is available for download at www.gov.uk/government/publications

ISBN 978-1-84099-783-5

Acknowledgements

From the Ministry of Justice the authors would like to thank Imran Ejaz, Kathryn Hopkins and Michael Cohen for their significant contributions to this project, and John Marais and Osama Rahman for their oversight.

The authors would also like to thank an external advisory panel comprising of Conor Duggan, Friedrich Lösel, Karl Hanson, Lorraine Dearden, Mike Daly, Paul Ainsworth and Richard Dorsett. Thanks are also due to Gerald Gaes and Martin Schmucker for their comments on the report.

Contents

List of tables

List of figures

Summary

Aims and background

The aim of the research was to extend the evidence base on the effectiveness of treatment for sexual offenders. This study measures the impact of the prison-based Core Sex Offender Treatment Programme (SOTP) on the re-offending outcomes of sex offenders in England and Wales, whilst controlling for the different observable characteristics, needs, and risk factors of offenders.

Core[1] SOTP is a cognitive-behavioural psychological intervention designed by the HM Prison and Probation Service (HMPPS) for imprisoned men who have committed sexual offences. The Programme is intended to reduce sexual reoffending amongst participants by identifying and addressing known criminogenic needs. It was accredited for use in prisons in 1992 by the then HM Prison and Probation Service Prison and Probation Services Joint Accreditation Panel, which later became the Correctional Services Accreditation and Advice Panel (CSAAP).[2] The CSAAP help the MOJ and HMPPS to develop and implement high quality offending behaviour programmes and promote excellence in programmes designed to reduce reoffending. Programmes are assessed against a set of criteria derived from the "what works" evidence base. These include having a clear model of change, effective risk management, targeting offending behaviour, employing effective methods, ensuring relevance to individual learning styles, and maintaining the quality and integrity of delivery. Changes have been made to the targets, the content, and the methods used in Core SOTP since its introduction in response to emerging research. As a result, during the course of this study (and in the period thereafter) the Programme has changed. However, it remains a cognitive behavioural group based treatment approach. It was, and remains, available in approximately one-sixth of male prison establishments in England and Wales and is intended for individuals sentenced to 12 months or more, who had either a current or previous (sentence) sex offence, were willing to engage in treatment, and were not in denial of their offending.

There were 2,562 convicted sex offenders who started treatment under the prison-based Core Sex Offender Treatment Programme between 2000 and 2012 in England and Wales. These were matched to 13,219 comparison sex offenders using 87 matching factors from

1 Other variants of the Programme also exist, including the Adapted Programme for lower IQ prisoners, and the Extended Programme for higher risk offenders.

2 CSAAP is a non-statutory advisory body for HMPPS. It provides independent expert advice on effective corrections for the Ministry of Justice and HMPPS.

Police National Computer (PNC) records, SOTP treatment records, and the Offender Assessment System (OASys) database (where available). Standardised mean differences between the matched treated and comparison groups for the matching factors showed that the matching quality achieved was excellent.

Propensity score matching (PSM) was used to match sexual offenders who participated in Core SOTP (treated sex offenders) to similar sexual offenders who did not.[3] PSM is a statistical matching technique which uses factors theoretically and empirically associated with both receiving the treatment and the outcome variable (i.e. reoffending) to predict a 'propensity score', which represents the likelihood of entering treatment. This propensity score is then used to match treated individuals to comparison offenders who are similar to them.

The matched treatment and comparison groups were then compared on an extensive range of proven reoffending outcomes (sexual and non-sexual). These outcome measures were calculated over a period of up to 13.9 years (average of 8.2 years) starting from each offender's release from prison between 2002 and 2012, with the follow-up period finishing in October 2015. For all individuals in this study (the treatment group plus the unmatched comparison group), the binary reoffending rate for all offences was 38.3% and the sexual reoffending rate excluding breaches,[4] was 7.5%. These are low when compared to international studies but are within the range of other UK-based studies on reconviction rates for sex offenders (Craig et al., 2008).

PSM can provide a robust quasi-experimental approach, although offenders can only be matched on observable variables. While extensive efforts were undertaken in identifying relevant factors, it is possible that unobserved factors could influence the findings that emerge from this research. Such factors include deviant sexual interest, general self-regulation problems and the degree of violence associated with the current sexual offence.

[3] A suitable comparison group of sex offenders who did not participate in Core SOTP could be formed because many were in prisons where Core SOTP was not run, while others were in prisons where the programme was run but could not participate due to a shortage of places.

[4] Breaches are non-compliance against the conditions of an offender's release from prison or the requirements placed on them whilst serving a sentence in the community.

Key findings

The main findings of the analysis were as follows:

- Some statistically significant differences were detected over an average 8.2 year follow up period. They were small in magnitude although they widened over the follow-up period. In particular:
  - More treated sex offenders committed at least one sexual reoffence (excluding breach) during the follow-up period when compared with the matched comparison offenders (10.0% compared with 8.0%).
  - More treated sex offenders committed at least one child image reoffence during the follow-up period when compared with the matched comparison offenders (4.4% compared with 2.9 %).

  Otherwise, the matched treated and comparison groups had similar reoffending rates across a variety of outcome measures.
- A variety of sensitivity analyses were performed, which mostly focused on the sexual reoffending measure. The sexual reoffending treatment effect was found to be reasonably stable across these.

As previously noted, it is possible that these results could be materially influenced by unobserved factors. However, such factors would need to increase both the odds of treatment and the odds of reoffending after controlling for the observable factors that were included within the matching process. In fact to conclude that the sexual re-offending treatment impact is not statistically significantly different from a reduction of 2 percentage points, the odds of treatment and re-offending would both need to increase by 122%. This increases to 219% for a 5 percentage point reduction. While the sensitivity analysis, involving both treatment and comparison groups, shows reoffending rates to be higher for individuals who have higher risk profiles, the matching process includes a range of factors that are used to determine risk.

Conclusions

The results suggest that while Core SOTP in prisons is generally associated with little or no changes in sexual and non-sexual reoffending, there were some statistically significant differences. The small changes in the sexual reoffending rate suggest that either Core SOTP does not reduce sexual reoffending as it intends to do, or that the true impact of the Programme was not detected.

This study draws on large treatment and comparison groups, long follow-ups, and many matching factors, thus addressing the most common shortcomings in the research field on sex offenders' reoffending behaviour. However it still has a number of limitations that could either bias the findings or the interpretation of them. In particular:

- It is impossible to conclusively rule out the absence of variables relating to deviant sexual interest,[5] general self-regulation problems and the degree of violence associated with the current sexual offence that could possibly influence the results. Moreover, it is possible that the available data do not fully account for issues such as motivation to address offending behaviour. However, these absences are at least partly accounted for by matching factors included in this study (e.g. sexual deviancy by matching factors covering previous offending). Furthermore as shown above, what remains unaccounted for would need to have strong relationships both with participation onto treatment and reoffending to conclude that Core SOTP is associated with a reduction in sexual reoffending.
- The estimated impact of Core SOTP was found to be similar when removing from the comparison group those who were identified as having done community SOTP. However, it will include some differences between the matched treatment and comparison groups that reflect changes occurring after the prison sentence has commenced and which are not associated with the provision of Core SOTP. Such factors include participation on other treatment programmes in prison and in the community, differences in offender management and in supervision, and regional demographics e.g. in employment rates.
- Availability of good quality data on all factors which determine an offenders' participation on core SOTP, was also a particular issue. It is possible that paucity of data on some key offender characteristics including denial of offending, and a degree of self-selection, could bias the results.

[5] This can be defined by grouping together factors relating to sexual preoccupation, sexual preference for prepubescent or pubescent children, sexualised violence and multiple paraphilias (Mann, Hanson & Thornton, 2010).

One of the main issues that will need to be addressed in any future studies on the effectiveness of SOTP in prisons, using some form of matching approach, is the collection of information on all potentially important variables. The lack of comprehensive empirical data on deviancy is a major issue that needs further investigation. Additional factors that need incorporating into any future study include other interventions received in prison and in the community, and the level of supervision once released from custody. Additionally, it is recommended that there be a focus on improving the quality of the data already collected on SOTP, e.g. a single unified record per offender.

This study does not reveal the extent to which Core SOTP reoffending outcomes are due to treatment design or poor implementation. However the treatment approach should be modified in line with the latest evidence base, of which this study is part. In particular, it could include individual sessions as well as group sessions. It could also focus more on factors that have been established to predict reoffending.

Whilst this research uses a recognised evaluation methodology, a randomised control trial could more robustly estimate the impact of any subsequent programme and this should be considered for the future.

Background

Research question

The aim of the research was to extend the evidence base on the effectiveness of treatment for sexual offenders. This study evaluates the impact of the prison-based Core Sex Offender Treatment Programme (SOTP) on the reoffending of sex offenders in England and Wales, whilst controlling for the different characteristics, needs, and risk factors of offenders by using propensity score matching.

The England and Wales Sex Offender Treatment Programme

The Core[6] SOTP in England and Wales is a cognitive-behavioural psychological group work programme delivered by the HM Prison and Probation Service (HMPPS) to imprisoned sex offenders, and has been accredited by the Correctional Services Advice and Accreditation Panel (CSAAP) since 1992. There have been a number of changes to the Core programme both during the evaluation period and thereafter. The main changes have related both to the assessment and selection of men into the programme, and the programme curriculum.

Assessment and selection changes to the Core Programme.

Before 2006, all men convicted of sexual offences (regardless of their risk of reoffending) were eligible for the Core SOTP. As validated risk tools were developed, and when research emerged indicating that intensive programmes for low risk men were ineffective or even harmful, the eligibility criteria were changed. In 2006 Core SOTP was designated as suitable for men who were medium or higher risk but not for lower risk offenders (unless they were convicted of an offence considered to be a sexual murder) who were instead redirected to the Rolling SOTP. The risk of reoffending (measured by the actuarial tool RM2000/S[7]) therefore became the main criteria for allocation to the Core SOTP from 2006 onwards. In addition, suitability guidance advised treatment managers to prioritise men with less time left in prison.

Content changes to the Core Programme

In the early years of the Core SOTP, the content of the programme was closely modelled on what other jurisdictions offered. Hence, a significant part of the Core programme delivered

---

6 Other programmes for men with convictions for sexual assault are also available including the Adapted Programme for lower IQ prisoners, and the Extended Programme which would be completed in addition to Core SOTP for higher risk offenders.

7 Measured by RM2000/S, a static risk tool widely used by practitioners and researchers as an indicator of risk of sexual reoffending amongst sex offenders.

during the evaluation period was the use of cognitive restructuring to challenge offence-related thoughts and beliefs, work to provide an account of the offending which was consistent with the victim's statement, and exercises for increasing victim empathy. These targets have in recent years been questioned and have been phased out of many programmes, including the Core Programme, based on a clearer understanding of the risk factors for sexual recidivism (Mann, Hanson and Thornton, 2010).

The Core programme was delivered to groups of 8 men who completed the course content over approximately 180 hours. At the time of this research, various additional treatment programmes were available to men who had completed the Core programme depending on need, including the Extended programme (for higher risk men) and the Booster programme (a pre-release course). Therefore, a maximum treatment dosage of 360 hours[8] was available to some men depending on their risk and needs. Participation in SOTP has always been conditional on offenders consenting to treatment. Core SOTP was, and remains, available in approximately one-sixth of male prison establishments in England and Wales, and it is intended for individuals with the following characteristics:

- sentenced to 12 months or more;
- with either a current or previous sex offence[9];
- willing to engage in treatment; and
- not in categorical denial of their offending.

In practice, participation also depends on other factors such as the availability of treatment, cognitive ability, mental and general health, their openness to discussing their sexual offending, and whether there were associated legal issues e.g. people who are appealing against sentence length or conviction. A study of treatment manager decision-making criteria (Barnett & Wakeling, unpublished) established that when treatment managers made prioritisation decisions they also took into account:

- Apparent motivation to change;
- Risk of serious recidivism – a clinical rather than actuarial judgement, mainly based on evidence of the present of offence-related sexual interests or preferences such as sadism or paedophilic preference;
- Ability to engage in treatment, in terms of cognitive ability and mental health;

8 SOTP was also delivered to low-risk offenders between 2000 and 2010.

9 Some recipients of SOTP do not have a sex offence conviction, but are referred onto the Programme as a result of a serious offence which was considered to have involved a sexual motive.

- Legal issues – the extent to which individuals were held up by not having engaged in treatment or were serving indeterminate sentences and past their tariff date.

Offenders in non-SOTP prisons who were considered suitable for Core SOTP may have been transferred to SOTP prisons to enable them to participate.

Efficacy of treatment for sex offenders – UK and international evidence

Evaluations of treatment programmes for sex offenders have so far produced mixed results on the effectiveness of such interventions, and no clear overarching conclusions can be drawn from any individual study. Indeed, while numerous studies into the effectiveness of treatment for sex offenders on sexual recidivism have been conducted, some have found positive results, some neutral or even negative results, and overall many involved small samples and/or low quality control groups (see for example: Pérez et al., 2012; Ruddijs and Timmerman, 2000; Olver et al., 2008; Abracen et al., 2011). There is widespread agreement that there is a need for high quality evaluations of treatment programmes for sex offenders (e.g. Dennis et al., 2012; Hanson et al., 2009; Långström et al., 2013; Schmucker and Lösel, 2015), and the lack of high quality evidence is so acute that if someone were to consider higher quality published studies only, there would be virtually no evidence that treatment programmes can reduce sexual recidivism (Hanson et al., 2009). Few studies have used stronger research designs (e.g. randomised control trials), with the evaluation of these treatment programmes rendered particularly challenging by the low base rate of proven sexual reoffending, and by the large heterogeneity between sex offenders (Schmucker and Lösel 2015), which then requires larger samples and longer follow-ups in order to reach sufficient statistical power.

A 2012 Cochrane Review of randomised control trials evaluating psychological interventions for adult males convicted or cautioned of sexual offences, or those seeking treatment for sexual offending (Dennis et al., 2012) concluded that there was no difference in sexual reoffending between treatment and control groups. Conversely, three meta-analyses that included both randomised and non-randomised designs have found a significant impact of treatment on recidivism (Hanson et al., 2002; Hanson et al., 2009; Lösel and Schmucker, 2005), providing promising evidence for the efficacy of cognitive-behavioural and pharmacological interventions with sexual offenders. Revised and improved versions of Lösel and Schmucker (2005) have corroborated the earlier findings that there is promising evidence on the impact of treatment for sex offenders. However, the authors highlight in their

conclusions that there is still room for improvement in establishing a robust evidence base on such treatment programmes (Schmucker and Lösel, 2008; Schmucker and Lösel, 2015). Importantly, there were only two studies that were included in both the Dennis et al. (2012) and Schmucker and Losel (2015) reviews so that these meta-analyses are really a synthesis of different studies. This disparity alone could easily explain the different findings. In addition, Losel and Schmucker included adolescent programmes which, generally, show a more positive outcome compared to those delivered to adults, these being the focus of the Dennis et al. review and of this report. Moreover, Losel and Schmucker found weaker effects for treatment in prisons and purely group treatments, while Hanson et al. (2009) found differing effects depending on the adherence to risk-need-responsivity principles.

While the effectiveness of the Core SOTP has not yet been assessed in its current format, the pre-2000 version of the Programme has been subject to evaluation. The evaluation examined reconviction[10] outcomes for offenders sentenced to four years or more in prison who participated in treatment between 1992 and 1994 (Friendship et al., 2003). Results showed no significant differences in sexual reconviction rates between treatment and control groups (2.6% of the treatment group were reconvicted for a sexual offence within two years, compared with 2.8% of the comparison group). However, there was a statistically significant difference in the overall combined sexual and violent reconviction rate (4.6% of the treatment group were reconvicted of a sexual or violent offence within two years, compared with 8.1% of the comparison group). This significant difference was most marked for medium risk offenders (though also significant for low risk offenders, but not high risk offenders). The study authors concluded that this suggested that SOTP may reduce sexual and violent reconviction for medium and low risk offenders but that higher risk offenders would benefit from additional treatment.

[10] Reconviction was defined as a conviction in court for a new offence during the follow-up period (cautions were excluded).

Approach

The study evaluates the reoffending patterns of a cohort of treated and comparison (untreated) offenders after their *first* release from a prison sentence (defined as the index sentence) within the period 2002 to 2012. The treatment group includes those who started Core SOTP during the index sentence, an intent-to-treat design. The comparison group includes those who did not attend Core SOTP during the index sentence (regardless of whether they had SOTP in a later sentence).

Offenders in the treatment group were matched to untreated offenders using propensity score matching (PSM). PSM is a statistical matching technique which uses factors theoretically and empirically associated with both receiving the treatment and the outcome variable (i.e. reoffending) to predict a 'propensity score', which represents the likelihood of entering treatment. This propensity score is then used to match treated individuals to comparison offenders who are similar to them. For more information on the methodology, see Sadlier (2010), Mews et al. (2013) and Ministry of Justice (2015).

Criteria for inclusion in the study

Offenders were selected using the following criteria:

- Being released from prison within the period 2002 to 2012;
- Receiving a custodial sentence of no less than 12 months;[11]
- Being 18 or over at release from custody;[12]
- Having no record of receiving any type of prison-based sex offender treatment commencing prior to the index sentence;
- Having a sexual primary index offence[13] (excluding breaches).

[11] If the offender was released more than once during the cohort period, the *first* release when the offender received a sentence of 12 months or more and was aged 18 or over was selected.

[12] See footnote 11.

[13] The index sentence may relate to more than one offence, the most severe of which is categorised as the primary offence. SOTP can also be given to prisoners who have sexually offended in the past, or if any of the non-primary index offences are sexual. However, in order to increase the homogeneity of the sample, eligible index offences were restricted to primary sexual offences. Including cases where any of the index offences were sexual increased the data set by 784 individuals (5%). As such, no secondary analysis was run using this enlarged data set.

Offence categorisation

The offences and reoffences have been categorised according to the impact of the crime and the type of offender, listed as follows:[14]

- Adult - Violent/Serious: Rape, sexual assault etc. of a victim who is over 16 years old.
- Adult Other: Possession and distribution of banned material (excluding child images), incest, offences against vulnerable people, exposure etc.
- Child Contact: Any sexual offence involving direct sexual contact with a victim who is under 16 years old (or under 18 years old in cases of incest, prostitution and abuse of a position of trust).
- Child Other: Abuse of a position of trust (not involving direct sexual contact) with a victim who is under 18 years old, facilitating child contact and child image offences, meeting a victim who is under 16 years old following grooming, and other child non-contact offences.
- Child Image: Creation, possession and distribution of indecent images of a victim who is under 18 years old.
- Breach: Breach (non-compliance) of conviction/release conditions related to a sexual offence.
- Prostitution/Soliciting: Prostitution or soliciting involving a victim who is over 18 years old, and related offences such as running or promoting a brothel etc.
- Non-sexual violent: Violent non-sexual offences, e.g. robbery, assault.
- Non-sexual non-violent: Non-violent and non-sexual offences.

Data set and analysis

The data set used for the analysis included a total of 2,562 treated and 13,219 untreated sex offenders. The vast majority of these either committed a primary index offence relating to child contact (58%) or adult serious (28%). The treated offenders tended to receive longer sentences and were of similar general risk of reoffending, but higher risk of sexual reoffending.

[14] Produced by Ministry of Justice and Home Office officials. Offenders' criminal records obtained from the PNC specify each individual offence according to a Home Office offence code, of which around 330 define sex offences. Each sexual offence has a corresponding description, which often provides further information about the victim (e.g. gender, age).

Using PSM, the treated and untreated offenders were then matched using 87 factors derived from Police National Computer[15] (PNC) records, SOTP treatment records, prison release records, and the Offender Assessment System[16] (OASys) database (where available). Standardised mean differences between the matched treated and comparison groups showed that the matching quality achieved based on the observed factors was excellent.[17] The two matched groups were then compared on an extensive range of proven reoffending outcomes (sexual and non-sexual). These outcome measures were calculated over a period of up to 13.9 years (average of 8.2 years) starting from each offender's release from prison between 2002 and 2012, and finishing in October 2015. The overall binary reoffending and sexual reoffending[18] rates for the combined unmatched treatment and comparison groups were 38.3% and 7.5% respectively. For more information about the data set, reoffending rates, and the implementation of PSM see Annexes A-D of the technical annex.

Limitations

There are caveats that should inform interpretation of findings.

- Whilst this research uses a recognised evaluation methodology, it is not as robust as a prospectively-matched evaluation or a randomised control trial. For a detailed discussion of the strengths and limitations of propensity score matching, see Mews et al. (2013) and Ministry of Justice (2015).
- While propensity score matching can provide a robust quasi-experimental approach, it can only match, and therefore reduce bias, on observed factors (information that is recorded). Sexual deviancy is possibly the most important characteristic that is lacking in coverage in the study, as it has been argued that sexual deviancy is the best predictor of sexual recidivism amongst sexual offenders (Hanson and Bussière, 1998). Sexual deviancy varies amongst sex offenders, and Hanson and Bussière suggest that sexually deviant interests were most prevalent amongst sexual offenders who victimize strangers (compared with victims who are known to them), use overt force, select boy victims (compared to girl victims), or select victims very different in age from themselves. Despite efforts to include all observed factors known to be predictive of selection onto Core SOTP and of reoffending risk, the importance of information that is not

[15] The PNC contains police and court records, including convictions for recordable offences from the first half of the 20th Century until the present.

[16] OASys assessments are practitioner-led offender assessments used for offender management and risk assessment purposes.

[17] The highest standardised mean difference was 3.8%, with an overall average of 1.2%.

[18] Unless otherwise stated, the sexual reoffending measure includes all sexual offences except breaches.

recorded cannot be known and it is possible that unobserved factors could influence these results.

- Core SOTP is intended for individuals who are not in denial of their offending and are willing to engage in treatment. While there is OASys evidence about whether the offender accepts responsibility for the current offence, valid OASys records were found for less than half of the treatment and comparison group individuals. Moreover, 25% of treatment individuals with valid OASys records were recorded as not accepting responsibility for the current offence. Reasons for this apparent discrepancy include individuals accepting responsibility after their OASys assessment but before doing Core SOTP (since the OASys data used was recorded around the time of conviction), and also OASys data quality issues. The comparison group therefore included individuals regardless of whether they were willing to engage in treatment and whether they were in denial of their offending. This could bias the results.
- It is possible that a large number of comparison individuals (those who did not attend Core SOTP during the index sentence) did some other cognitive behavioural programme either during the index sentence (e.g. Enhanced Thinking Skills[19]) or after being released from prison. In particular, at least 8% of comparison individuals attended Community SOTP run from 2008 onwards.[20] While sensitivity analyses were performed to isolate the influence of Community SOTP, it is possible that many of those who did neither Core SOTP nor Community SOTP did other cognitive behavioural programmes.
- Some comparison individuals attended Core SOTP in the follow-up period. However, this influence is small as less than 1% of comparison individuals were recorded as such.
- Sex offenders constitute a highly heterogeneous group of offenders (Robertiello and Terry, 2007), and as such this increases the challenges associated with drawing conclusions on the effectiveness of treatment for the sex offender population. However, matching offenders on the type of sexual index offence and on the types of previous sexual offences committed should help in minimising the influence of this limitation.

[19] Enhanced Thinking Skills is a cognitive-behavioural offending behaviour programme that addresses thinking and behaviour associated with offending with the objective of reducing general reconviction rates. It was replaced in 2009 by the Thinking Skills Programme (TSP), representing a refresh and update of the cognitive skills programme in line with advances in theory and practice.

[20] The Community SOTP data are not available from before 2008.

- There is a potential for geographical bias. For example, if an SOTP prison covers an area with a relatively large offender population and which has substantially different outcomes than for individuals in other areas, this could skew the results.
- While care was taken to link together a disparate selection of data sets, these were drawn from administrative IT systems which, as with any large scale recording system, are subject to possible errors with data entry and processing. For instance, not all SOTP treatment records could be assigned a PNC/CRO identifier[21] which may have resulted in around 1% of comparison individuals receiving Core SOTP during their index sentence.
- Information from OASys records was included in the matching process for individuals for whom an assessment was completed between 30 days before and 90 days after the conviction[22] for the index offence (the offence relating to the index prison sentence). As such, on occasions the OASys assessment may have been completed following the onset of Core SOTP and so have been influenced by the treatment. This could reduce the estimated impact of the treatment.
- This research is only able to consider proven reoffending (as recorded on the PNC), which is generally an underestimate of the true level of reoffending. This is particularly true of sexual reoffending, which is frequently unreported. If detection rates differ between treatment and comparison groups then the results will reflect this.

[21] PNC and CRO numbers are unique identifiers for individuals, one originating from the Police National Computer and the other from the Criminal Records Office. An individual may have one or both of these identifiers.

[22] If sentencing took place on a later date, the sentencing date was used instead.

Results

Overall findings

The **binary reoffending rates**[23] over an average 8.2 year follow up period were mostly similar between the matched treatment and comparison groups (see Table A.8, parsimonious model[24]). The only statistically significant differences emerged on the following outcomes:[25]

- Sexual reoffending:[26] the reoffending rate for the treatment group was 2.0pp higher than the matched comparison group (10.0% vs. 8.0%);
- Child image reoffending: the reoffending rate for the treatment group was 1.6pp higher (4.4% vs. 2.9%).

The **frequency of reoffending**[27] outcomes were also mostly equivalent between the matched groups (see Table A.9, parsimonious model), with statistically significant differences emerging only for:

- Sexual reoffending: the matched treatment group had 0.15 more reoffences per offender than the matched comparison group (0.59 vs. 0.45).
- Adult other: the matched treatment group had 0.02 more reoffences per offender than the matched comparison group (0.04 vs. 0.02).
- Non-sexual: the matched treatment group had 0.27 fewer reoffences per offender than the matched comparison group (0.98 vs. 1.25).
- Non-sexual non-violent: the matched treatment group had 0.27 fewer reoffences per offender than the matched comparison group (0.92 vs. 1.19). This seems to 'drive' the above difference in the overall non-sexual frequency of reoffending.

Given that the non-sexual and the non-sexual non-violent measures elicited significant differences only when looking at the frequencies of reoffending,[28] further tests were

---

[23] Proportion of offenders who committed at least one reoffence during the follow-up period.

[24] See Annex C for a description of the methods used to develop the parsimonious and less parsimonious models.

[25] Propensity score matching was performed in Stata with *psmatch2* (Leuven and Sianesi, 2003). Although the statistical significance calculations in *psmatch2* do not take account of the fact that the propensity scores themselves are estimated, very similar results were obtained using 95% CIs generated by bootstrapping, which does take account of propensity score estimation.

[26] Unless otherwise stated, the sexual reoffending measure used in this report includes all sexual offences except breaches.

[27] The number of reoffences per offender during the follow-up period.

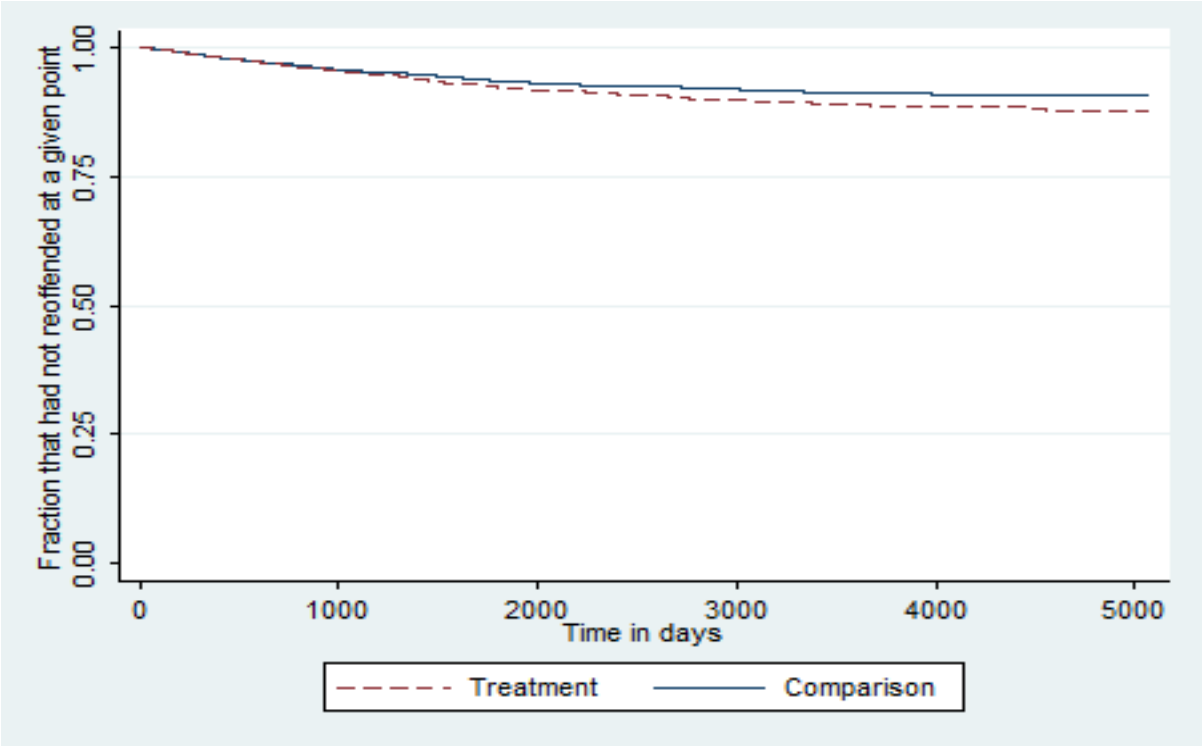
[28] The binary adult other measure was statistically significant at the 10% level even if not at the conventional 5% level.

performed to confirm that these differences were not due to a small minority of highly prolific reoffenders in the matched comparison group. For both measures around two per cent of the matched treatment group and three per cent of the matched comparison group committed ten or more reoffences. After producing winsorized estimates,[29] the matched treatment groups had 0.08 fewer reoffences per offender than the matched comparison groups, with these differences not being statistically significant. This is suggestive of the effect being driven by a minority of highly prolific reoffenders in the matched comparison group.

Unlike binary reoffending, the difference in the frequency of reoffending for the child image measure was not statistically significant. However, as a criminal reconviction can cover a large number of image reoffences, the results were also obtained in terms of reconvictions.[30] These showed that the matched treatment group had a statistically significant 0.02 more reconvictions/cautions per offender than the matched comparison group (0.05 vs. 0.03).

Reoffending rates over time

Figure 1: Sexual reoffending survival curves



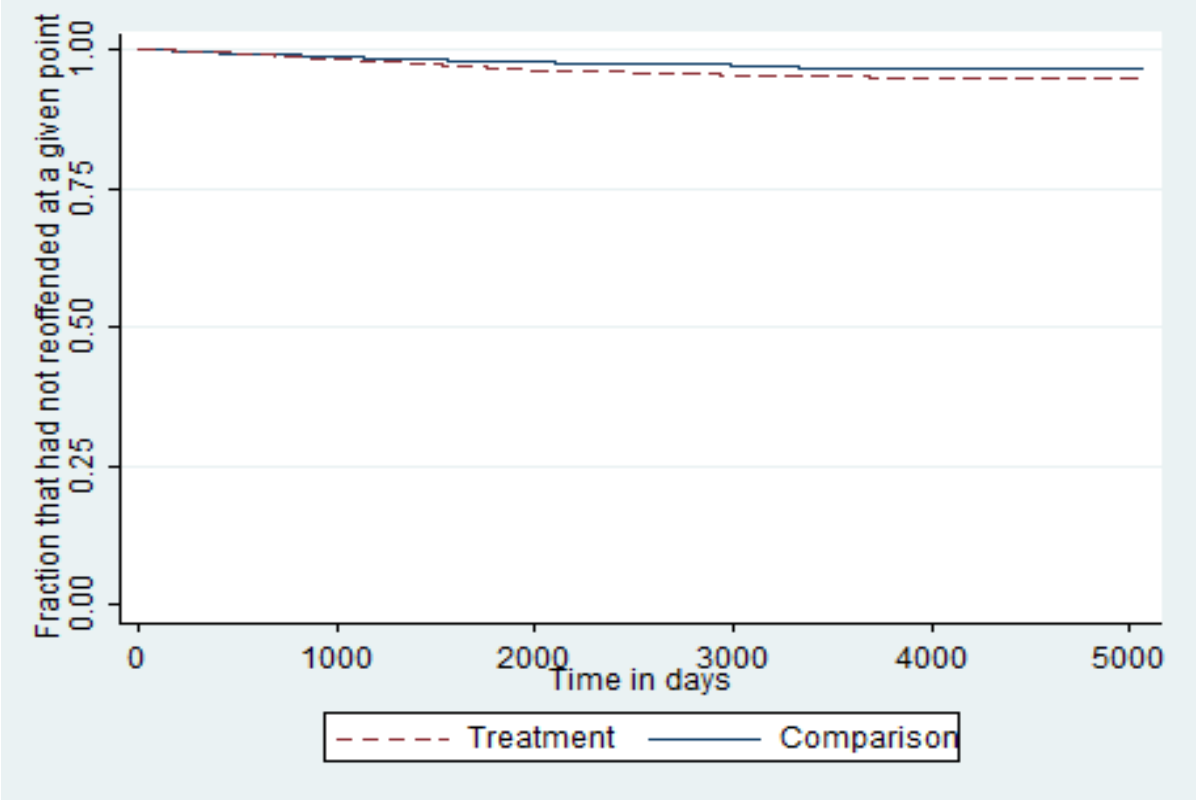
Figures 1 (sexual reoffending) and 2 (child image reoffending) show that the differences between the matched treatment and comparison group survival rates, which represent the

[29] The number of reoffences was set to ten for those individuals who committed ten or more reoffences.

[30] A conviction may cover more than 1 offence.

proportions of offenders who have not reoffended within a fixed time period (as defined by the horizontal axis), widen as the follow-up time increases.

Figure 2: Child image reoffending survival curves



This widening can also be presented in terms of binary reoffending, in particular:

- The difference on the binary sexual reoffending measure increased from 0.2pp for (all) offenders with a 2-year follow-up period, to 2.1pp for those offenders who could be measured over a 7-year follow-up period.
- The difference on the binary child image reoffending measure increased from 0.3pp for (all) offenders with a 2-year follow-up period, to 1.7pp for those offenders who could be measured over a 7-year follow-up period.

As fewer offenders could be measured over a 7-year follow-up period than a 2-year follow-up period it is possible that the widening could be driven by differences in the composition of offenders or in the Core SOTP Programme over time (e.g. offenders commencing Core SOTP in December 2012 could only be measured over a 2-year follow-up period). However the same trend (with the differences between the matched and comparison groups becoming larger over time) was also detected when looking only at those offenders who could be measured over a 7-year follow-up period (N=9,434).

The widening effect was still present when the follow-up period excluded any prison spells (time from conviction date[31] to release date) during which it would be difficult for the individual to reoffend.[32] In particular:

- The difference on the binary sexual reoffending measure increased from 0.2pp for offenders with a 2-year follow-up period, to 1.3pp (although not statistically significant) for those offenders who could be measured over a 7-year follow-up period.
- The difference on the binary child image reoffending measure increased from 0.2pp for offenders with a 2-year follow-up period, to 1.0pp (statistically significant at the 10% level) for those offenders who could be measured over a 7-year follow-up period.

Further comparisons using survival analysis are shown in Table A.11 including associated significance testing and the censoring of individuals who were re-imprisoned during the follow-up period at the point they were sentenced to reimprisonment.

Sensitivity analyses

A number of sensitivity analyses were performed in order to ascertain whether the results (sexual reoffending unless otherwise stated) were stable across groups of offenders with different characteristics. In each case, the variables from the parsimonious model were used and the matching process was re-run[33] with the specific offenders under analysis, so that the matches between individuals in the treatment and comparison groups were able to change. The summary statistics, providing the main results together with information about the quality of matching for each comparison, are reported in Table A.10, with the main findings as follows.

Core SOTP received by comparison group. Less than 1% of comparison individuals commenced Core SOTP during their follow-up period so the amount of any associated bias was likely to have a negligible influence on the estimated impact of treatment. To verify this, the survival analysis for the sexual reoffending measure was rerun censoring comparison

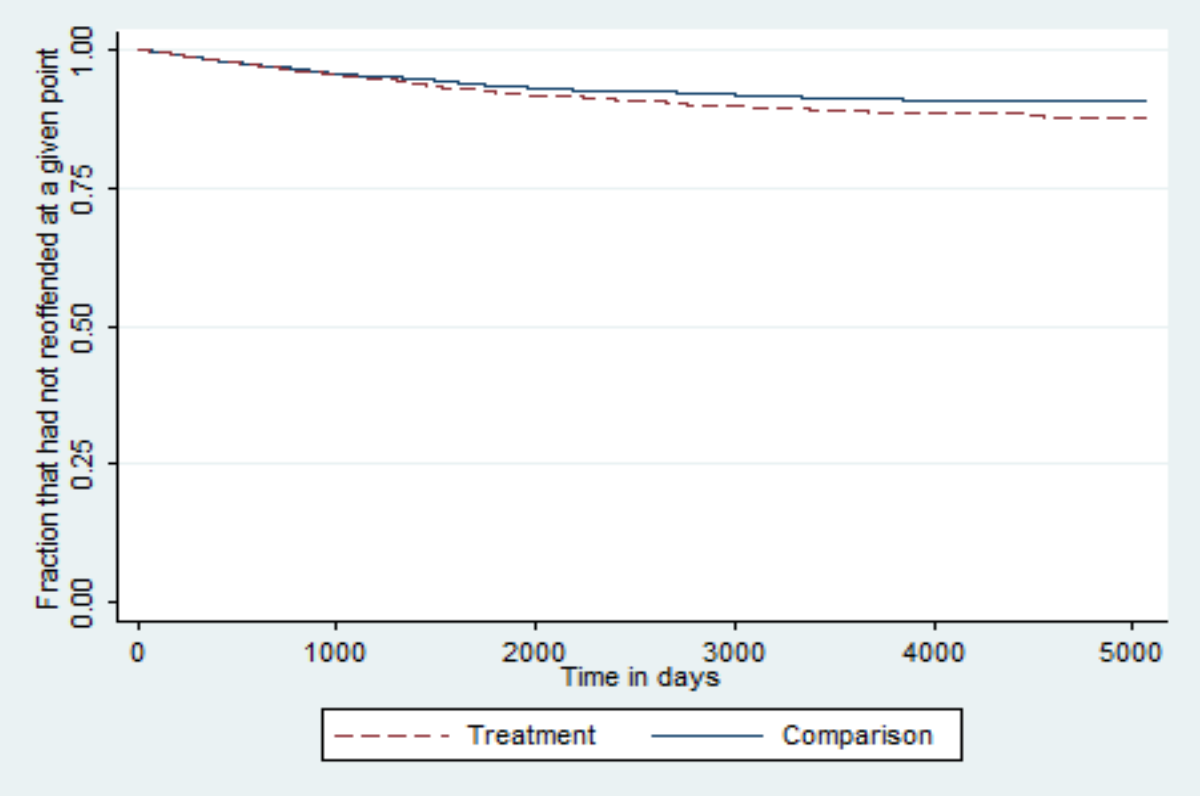
[31] If sentencing took place on a later date, the sentencing date was used instead.

[32] This approach is not unbiased due to a couple of drawbacks. Firstly, the length of time spent back in custody may itself be related to the treatment effect. Secondly, some who have been sentenced to a lengthy time in prison for sex reoffences will be discarded from the fixed follow-up period measures e.g. if someone is released from their index offence in 2003 having done core SOTP, reoffends in 2008 and then gets a 10 year sentence for a sex offence, he will be excluded from the 7-year follow-up period measure.

[33] The propensity scores were re-estimated and the treatment and comparison individuals then re-matched.

individuals who started Core SOTP during their follow-up period at the point when they commenced this treatment (so in other words, their follow-up periods ended when they started Core SOTP). Comparing Figure 3 to that without any censoring (see Figure 1) shows that it made little or no difference to the results.

Figure 3: Sexual reoffending survival curves, with comparison observations censored at the point of treatment



Core SOTP completers. The treatment effect was similar to that of the main analysis (1.4pp compared to 2.0pp)[34] when retaining in the treatment group only those who were recorded as having completed their treatment[35] (matched treatment N=1,918).

Non-Core SOTP treatment. Around one third of the treatment group also commenced some sort of non-Core SOTP (Adapted, Booster, Rolling, Extended or HSF) while serving their index sentence. The extent to which the overall treatment effect was influenced by this group was therefore examined. Firstly, only those in the treatment group who were registered as having also started a non-Core variant of SOTP during their index sentence were retained

[34] Although not statistically significant, due to the slightly smaller treatment effect size and the lower number of matched treatment individuals.

[35] Where there were multiple Core SOTP treatment records relating to a particular treatment, Core SOTP treatment was regarded as having been completed if flagged as such on one or more of these records.

(matched treatment N=816). The treatment effect was similar to the main analysis (2.2pp[36] compared to 2.0pp). A further analysis retaining those treated offenders who had started Core SOTP but not a non-Core SOTP variant during their index sentence (matched treatment N=1,735) obtained a similar effect (1.5pp) to the main analysis.

Extended SOTP. Around 9% of treatment individuals commenced Extended SOTP within their index sentence. The extent to which the overall treatment effect was influenced by this group was therefore also examined. Retaining in the treatment group only those who commenced Extended SOTP within their index sentence (matched treatment N=217), the reoffending rate for the matched treatment group was 7.5pp higher than the matched comparison group (17.5% vs. 10.0%). The outcome was the same whether including or excluding OASys assessment factors from the matching process. Possible explanations for the larger adverse treatment effect include:

- An adverse relationship between the amount of treatment sex offenders receive and the reoffending prognosis.
- The presence of substantial bias. The criteria for commencing Extended SOTP is different to that for commencing Core SOTP, in particular putting a higher emphasis on deviancy factors which may not be fully encapsulated by the propensity score matching process.

Community SOTP. One possibility is that the sexual reoffending rate was lower for the matched comparison group than the matched treatment group because many in the comparison group commenced community SOTP on release. This could potentially have a more positive impact in reducing reoffending than Core SOTP. To examine the influence of community SOTP, the results were found for offenders who were released from the year 2008 onwards (matched treatment N=1,084), with comparison individuals not having been identified as commencing community SOTP during the follow-up period. The effect was 1.0pp (7.7% reoffending rate for the matched treatment group vs. 6.6% for the matched comparison group), similar to the effect found when retaining all offenders who were released from the year 2008 onwards regardless of whether comparison individuals definitely commenced community SOTP (0.7pp; 7.7% reoffending rate for the matched treatment group vs. 7.0% for the matched comparison group). This suggests that community SOTP did not have a large influence on the main results.

[36] This was significant at the 10% rather than the 5% level, possibly due to the smaller number of treated individuals.

Sentence length. To check whether the full influence of sentence length was preserved using PSM, the main analysis was rerun performing PSM on all variables except sentence length. For this variable hard (or exact) matching was used (i.e. so treatment individuals with a certain sentence length could only be matched to comparison individuals with the same sentence length). The difference between the matched groups was found to be the same regardless of whether hard matching was performed (2.0pp). Further tests were then performed to see whether the estimated treatment effect varied for different lengths of sentence. It was found to be higher for longer sentences. In particular, the treatment effect for sentences of four years or more but less than ten years (matched treatment N=1,589) was 2.6pp compared to -0.5pp for sentences of one year or more but less than four years (matched treatment N=721). Rerunning the main analysis excluding offenders who had received a custodial sentence shorter than two years (matched treatment N=2,425) produced a similar effect to the main analysis (2.3pp). Further excluding offenders who had received a custodial sentence of less than four years (matched treatment N=1,823) resulted in a treatment effect of 3.0pp, similar to that for sentences of four years or more but less than ten years.

Lower vs. higher risk. The sample was also split into two groups based on the RM2000/S risk bands, one group representing low and medium (matched treatment N=1,694), and the other high and very high (matched treatment N=829). The Core SOTP impact was 3.2pp for the higher risk group and 3.2pp for the low and medium risk group. However the difference between these impacts was not statistically significant.

Early vs. late releases. The sample was split in two groups, those released between 2002 and 2005 (matched treatment N=950), and those released between 2006 and 2012 (matched treatment N=1,591). The Core SOTP impact in the first subsample was 2.8pp higher reoffending compared to 1.6pp in the second subsample, a statistically non-significant difference.

Time from treatment to release. To explore whether the length of the period between commencing Core SOTP and being released from prison affected the overall treatment effect, the treatment individuals were split into two groups. For individuals whose treatment started in the two years[37] before their index prison release (matched treatment N=1,148) the

[37] As the start date of treatment was often provided as a year rather than an exact date, the 'up to two years' category contained some people whose treatment began between two and three years before their release. Consequently, these people were absent from the 'more than two years' category.

treatment effect was 1.0pp. This compared to 2.5pp for those whose treatment started more than two years[38] before their index prison release (matched treatment N=1,378). However, as around half of individuals whose treatment started in the two years before their index prison release were sentenced to up to four years compared with less than 10% of the other treatment individuals, any differences may reflect those in sentence length, the treatment effect being higher for longer sentences (see above). Rerunning the analysis for treatment individuals whose treatment started in the two years before their index prison release but retaining only offenders who were sentenced to four years or more but less than ten years (matched treatment N=535) the treatment effect was 2.3pp. This was similar to the result for all offenders who were sentenced to four years or more but less than ten years (2.6pp).

OASys assessments. Retaining only offenders who had an OASys assessment (matched treatment N=704) resulted in a treatment effect of -1.1pp compared with 2.0pp overall. However, including the OASys assessment variables in the matching process made little or no difference. Rerunning the analysis retaining only offenders who had an OASys assessment but not including OASys variables in the matching process (matched treatment N=712) saw an effect of -0.9pp. Moreover, using all observations in the original sample but excluding OASys variables from the matching process resulted in an effect of 2.0pp. It is possible that receiving an OASys assessment might be a proxy of more effective and close management of offenders, and it might support greater involvement and 1-1 contact between the offender and the offender managers. In combination with treatment, this could potentially foster greater therapeutic alliance, which is thought to be crucial for treatment effectiveness (e.g. Cooper and Lesser, 2011).

SOTP prisons. To ascertain whether there was any evidence that suggests the treatment effect was really due to the treatment being undertaken at certain prisons rather than actually having any impact in itself, the comparison group was split into two groups. The first included those who had been discharged from prisons where SOTP was being delivered when they were serving their index sentence, and the second included the others. The treatment effects using the two comparison groups were found to be similar, being 1.5pp for the first and 2.0pp for the second.

Child contact offenders. The main analysis was rerun including only offenders who had committed a child contact index offence (matched treatment N=1,518). While the treatment effect on sexual reoffending (1.2pp) was not statistically significant, that for child image

[38] See footnote 37.

reoffending (1.7pp; matched treatment group 4.8% vs. matched comparison group 3.1%) was statistically significant.

Adult serious offenders. This analysis included only offenders who had committed an adult serious index offence (matched treatment N=776). The treatment effects on sexual reoffending (3.8pp; 10.8% for the matched treatment group vs. 7.0% for the matched comparison group) and adult other reoffending (2.1pp; 3.2% for the matched treatment group vs. 1.2% for the matched comparison group) were statistically significant. However, as expected the child image impact was not (1.0pp).

Quantifying bias: sensitivity of the results to unobserved factors

A further sensitivity analysis was run to quantify the magnitude of bias from unmeasured factors that would need to be present to change the key results of the study.[39] This indicated the amount of bias required for the sexual reoffending impact[40] to become not statistically significant (i.e. the p-value to rise above 0.05) when applying a two-sided hypothesis test.[41] The level of bias needed, after controlling for factors included in the model, is equivalent to:

- A 48% increase in the odds of treatment and a 48% rise in the odds of reoffending. To place this in perspective, it may be for example that we are missing one important binary factor from the PSM model, such as an overall measure of sexual deviancy. If, for instance, one third of the treatment group are highly deviant, then these odds would require their probability of commencing Core SOTP to be 20.3% (compared to 14.7% for those who are not highly deviant) and the sexual reoffending rate to be 9.6% (compared to 6.7% for those who are not highly deviant).
- A one quarter increase in the odds of treatment and a doubling of the odds of reoffending or vice versa.

To quantify the amount of unmeasured bias that would need to be present if the treatment genuinely has a beneficial impact on reducing the sexual reoffending rate, three scenarios were tested. Under the first, the real treatment effect was assumed to be -1pp which formed the null hypothesis of the two sided test, with the alternative hypothesis being that the treatment effect was greater or less than -1pp. The level of unobservable bias needed under

[39] Using *sensitivitymv* (Rosenbaum, 2015), an R package to run sensitivity analysis in observational studies with matched sets containing multiple controls.

[40] For simplicity in using *sensitivitymv*, the model was rerun using nearest neighbour 1-5 matching. Headline results were highly similar, with a 1.8 pp difference between the matched treatment and control groups instead of 2.0 pp obtained with our preferred kernel matching approach.

[41] A permutational t-test was used.

this scenario for the statistically significant result to become statistically insignificant, after controlling for observable factors in the model, is equivalent to:

- An 87% increase in the odds of treatment and an 87% rise in the odds of reoffending. To place this into perspective using the example above (see the first bullet point of this quantifying bias section), these odds would require the probability of commencing Core SOTP for those who are highly deviant to be 23.4% (compared to 14.1% for those who are not highly deviant) and the sexual reoffending rate to be 11.3% (compared to 6.4% for those who are not highly deviant).
- A 50% increase in the odds of treatment and a 172% rise in the odds of reoffending or vice versa.

The second scenario was that the real treatment effect was -2pp, which formed the null hypothesis of the two sided test. The level of unobservable bias needed under this scenario for the result to become statistically insignificant, after controlling for observable factors in the model, is equivalent to:

- A 122% increase in the odds of treatment and a 122% rise in the odds of reoffending. Using the example above (see the first bullet point of this quantifying bias section), these odds would require the probability of commencing Core SOTP for those who are highly deviant to be 26.0% (compared to 13.7% for those who are not highly deviant) and the sexual reoffending rate to be 12.7% (compared to 6.1% for those who are not highly deviant).
- A doubling of the odds of treatment and a 150% rise in the odds of reoffending or vice versa.

The last scenario was that the real treatment effect was -5pp, which formed the null hypothesis of the two sided test. The level of unobservable bias needed under this scenario for the result to become statistically insignificant, after controlling for observable factors in the model, is equivalent to:

- A 219% increase in the odds of treatment and a 219% rise in the odds of reoffending. Using the example above (see the first bullet point of this quantifying bias section), these odds would require the probability of commencing Core SOTP for those who are highly deviant to be 32.3% (compared to 13.0% for those who are not highly deviant) and the sexual reoffending rate to be 16.3% (compared to 5.7% for those who are not highly deviant).

- A doubling of the odds of treatment and a 9 times rise in the odds of reoffending or vice versa.

As such, this further sensitivity analysis suggests that the detected adverse effect of treatment on the sexual reoffending binary rate is robust against some unmeasured bias. However, as the unmeasured bias is unknown, it is impossible to definitively rule out the impact of treatment being either not statistically significant or even favourable rather than adverse.

Conclusions

The results suggest that while Core SOTP in prisons is generally associated with little or no changes in reoffending, there were some statistically significant differences. In particular, there were small increases in the sexual and child image reoffending rates. The sexual reoffending treatment effect was -1.1pp rather than 2.0pp for those who received an OASys assessment, possibly reflecting these offenders having greater involvement with offender managers. This in combination with treatment could potentially foster greater therapeutic alliance, thereby enhancing treatment effectiveness (e.g. Cooper and Lesser, 2011). However, it was stable across most of the sensitivity analyses performed. Early evaluation of the Core SOTP found no treatment effect on the 2-year sexual reconviction rate for sexual offences (Friendship et al., 2003), and the international research on the effectiveness of treatment for sex offenders provides a mixed picture. While some quantitative reviews of available research have found that treatment reduces reoffending in sex offenders (Hanson et al., 2002; Hanson et al., 2009; Lösel and Schmucker, 2005; Schmucker and Lösel, 2008; Schmucker and Lösel, 2015), a review of randomised control trials only found no beneficial effect of psychological interventions (Dennis et al., 2012).

The present study draws on large treatment and comparison groups, long follow-ups, and many matching factors, thus addressing the most common shortcomings in the research field on sex offenders reoffending behaviour. However, it is possible that the results have been influenced by bias introduced by unobserved factors. Higher risk offenders (on measures of risk which have not been identified amongst the matching variables) may have been more likely to enter treatment. As discussed in the limitations section, the lack of comprehensive information on sexual deviancy or other appropriate dynamic risk factors in the data implies that treated offenders might have not been matched to comparison offenders with a similar likelihood of entering treatment. It is also possible that more manipulative offenders may have self-selected onto the Programme knowing that it may increase the likelihood of release and other benefits. However, the unobserved factors would need to have a substantial impact on both reoffending and entering treatment after controlling for the factors already included within the analysis for a magnitude of bias to be present that masks a favourable impact of treatment.

The estimated impact of Core SOTP in prisons was found to be similar when removing from the comparison group those who were identified as having done community SOTP. However, it is possible a much larger number of comparison individuals did some cognitive behavioural programme either in prison or on release. The estimated impact of prison based Core SOTP

will therefore also reflect changes between the matched treatment and comparison groups that occur after the prison sentence has commenced but which are not related to Core SOTP. Such factors include participation on other treatment programmes in prison and in the community, differences in offender management and in supervision upon release, and regional disparities such as in crime detection rates and employment opportunities. Moreover, good quality data were not available for all criteria determining entry onto Core SOTP, such as the individual not being in denial of their offending. This, along with participants needing to give consent prior to attending the programme, could bias the findings.

It is also possible that attendance on the Core prison-based SOTP may increase the propensity to sexually re-offend amongst sex offenders. This may have been as a result of the sole emphasis on group treatment: recent meta-analysis has indicated the importance of including more individualized modules in sex offender treatment (Schmucker & Lösel, 2015). A lack of tailoring has also been suggested as an explanatory factor for the failure of a randomised control trial by Marques et al. (2005) to show a treatment effect – Marshall and Marshall (2007) noted that the programme's detailed delivery manual constrained the ability of those delivering it to adapt to the particular needs of individual participants. Group treatment may also 'normalise' individuals' behaviour: when stories are shared, their behaviour may not be seen as wrong or different; or at worst, contacts and sources associated with sexual offending may be shared.

Further research into the effectiveness of SOTP using some sort of matching technique should focus on measuring potentially relevant confounding factors. These include exposure to other treatment delivered in prison and on release of offenders into the community, and the level of supervision received on release. General data issues would also be addressed, including ensuring that sex offender treatment names and dates[42] are accurately recorded in the HMPPS treatment database. This would enable more individuals to be linked to other data sources. More research could also be undertaken to explore in further detail the characteristics of certain subsamples of the treatment group, such as those offenders who are selected for Extended SOTP treatment, who were found to have the highest sexual reoffending rates. In addition, a process study could be undertaken to understand any systematic bias between being selected onto SOTP and not being selected. To undertake as robust as possible an evaluation of SOTP, a randomised control trial should be considered, as recommended by Ho and Ross (2012), Dennis et al., (2012), and Långström et al.,

[42] Year is currently recorded, but not month and day.

(2013). These trials can eliminate any systematic bias between entering the treatment or the control groups.

This study does not reveal the extent to which Core SOTP reoffending outcomes are due to treatment design or poor implementation. However the treatment approach should be modified in line with the latest evidence base, of which this study is part. In particular, it could include individual sessions as well as group sessions. It could also focus more on factors that have been established to predict reoffending.

References

Abracen, J., Looman, J., Ferguson, M., Harkins, L., & Mailloux, D. (2010). *Recidivism among treated sex offenders and comparison subjects: Recent outcome data from the Regional Treatment Centre (Ontario) high-intensity Sex Offender Treatment Programme*. Journal of Sexual Aggression: An international, interdisciplinary forum for research, theory and practice, 17(2), 142-152. doi:10.1080/13552600903511980.

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(4), 443–470. doi:10.1177/1079063210384274

Beech, A.R., Fisher, D., & Beckett, R.C. (1999). *STEP 3: An evaluation of the prison Sex Offender Treatment Programme*. London: HMSO. Available online: http://webarchive.nationalarchives.gov.uk/20110218135832/http:/rds.homeoffice.gov.uk/rds/pdfs/occ-step3.pdf (accessed 04/08/2016)

Cooper, M.G., & Lesser, J.G. (2011). Clinical social work practice an integrated *approach* (4th ed.) Boston, MA: Allyn & Bacon.

Craig, L.A., Browne, K.D., Stringer, I., & Hogue, T.E. (2008). *Sexual reconviction rates in the United Kingdom and actuarial risk estimates*. Child Abuse & Neglect, 32, 121-138. doi:10.1016/j.chiabu.2007.09.002

Dennis, J.A., Khan, O., Ferriter, M., Huband, N., Powney, M.J., & Duggan, C. (2012). *Psychological interventions for adults who have sexually offended or are at risk of offending* (review). Cochrane Library, 12. doi:10.1002/14651858.CD007507.pub2

Friendship, C., & Beech, A.R. (2005). *Reconviction of sexual offenders in England and Wales: an overview of research*. Journal of Sexual Aggression, 11(1), 1-15. doi:10.1080/13552600500063690

Friendship, C., Mann, R.M., and Beech, A.B. (2003). *Evaluation of a national prison-based treatment program for sexual offenders in England and Wales*. Journal of Interpersonal Violence, 18(7), 744-759. doi:10.1177/0886260503253236

Friendship, C., and Thornton, D. (2001). Sexual reconviction for sexual offenders discharged from prison in England and Wales: implications for evaluating treatment. British Journal of Criminology, 41, 285-292. doi:10.1093/bjc/41.2.285

Hanson, K. R., Bourgon, G., Helmus, L., & Hodgson, S. (2009). *The principles of effective correctional treatment also apply to sexual offenders: a meta-analysis*. Criminal Justice & Behavior, 36(9), 865-891. doi:10.1177/0093854809338545

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

Hanson, R.K., Gordon, A., Harris, A.J.R., Marques, J.K., Murphy, W., Quinsey, V L., and Seto, M.C. (2002). *First report of the collaborative outcome data project on the effectiveness of psychological treatment for sex offenders*. Sexual Abuse: A journal of research and treatment, 14(2), 169-194. doi:10.1023/A:1014624315814

Hanson, R.K. and Thornton, D., (2000). *Improving risk assessments for sex offenders: a comparison of three actuarial scales*. Law and Human Behaviour, 24(1), 119-136.

Ho, D.K., & Ross, C.C. (2012). *Cognitive behavioural therapy for sex offenders. Too good to be true?* Criminal Behaviour and Mental Health, 22, 1-6. doi:10.1002/cbm.1818

Howard, P.D., Barnett, G.D., & Mann, R.E. (2014). *Specialization in and within sexual offending in England and Wales*. Sexual Abuse: A Journal of Research and Treatment, 26(3) 225–251. doi:10.1177/1079063213486934

Howard, P.D., Barnett, G.D., & Wakeling, H.C. (2015). *Predicting nonsexual violent reoffending by sexual offenders: A comparison of four actuarial tools*. Legal and Criminological Psychology, 20, 267–287. doi:10.1111/lcrp.12027

Långström, N., Enebrink, P., Laurén, E., Lindblom, J., & Werkö, S (2013). *Preventing sexual abusers of children from reoffending: systematic review of medical and psychological interventions*. British Medical Journal, 347, f4630. doi:10.1136/bmj.f4630

Lösel, F., and Schmucker, M. (2005). *The effectiveness of treatment for sex offenders: a comprehensive meta-analysis*. Journal of Experimental Criminology*,* 1(1), 117-146. doi:10.1007/s11292-004-6466-7

Mann, R.E., Hanson, R.K., and Thornton, D. (2010). *Assessing risk for sexual recidivism: Some proposals on the nature of psychologically meaningful risk factors.* Sexual Abuse: A Journal of Research and Treatment, 22(2), 191-217. Doi:10.1177/1079063210366039

Marques, J.K., Wiederanders, M., Day, D M., Nelson, C., & van Ommeren, A. (2005). *Effects of a relapse prevention program on sexual recidivism: Final results from California's Sex Offender Treatment and Evaluation Project (SOTEP)*. Sexual Abuse: A Journal of Research & Treatment, 17(1), 79-107. doi:10.1007/s11194-005-1212-x

Marshall, W.L., & Marshall, L.E. (2007). *The utility of the random controlled trial for evaluating sexual offender treatment: The gold standard or an inappropriate strategy?* Sex Abuse, 19, 175-191. doi:10.1007/s11194-007-9046-3

Mews, A., Hillier, J., McHugh, M. and Coxon, C. (2013). *The impact of short custodial sentences, community orders and suspended sentence orders on re-offending*. London: Ministry of Justice. Available online: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/278133/compendium-reoffending-stats-2013.pdf (accessed 03/08/2016)

Ministry of Justice (2015). *2013 Compendium of re-offending statistics and analysis*. London: Ministry of Justice. Available online: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/399389/impact-of-short-custodial-sentences-on-reoffending.PDF (accessed 03/08/2016)

Olver, M.E., Wong, S.C.P., and Nicholaichuk, T.P. (2008). *Outcome evaluation of a high-intensity inpatient sex offender treatment programme*. Journal of Interpersonal Violence, 29(3), 522 – 536. doi:10.1177/0886260508317196

Pérez, D.M., & Jennings, W. G. (2012). *Treatment behind bars: the effectiveness of prison-based therapy for sex offenders*. Journal of Crime and Justice 2012, 35(3), 435-450. doi:10.1080/0735648X.2012.662060

Robertiello, G., & Terry, K. J. (2007). *Can we profile sex offenders? A review of sex offender typologies*. Aggression and Violent Behavior, 12, 508-518. doi:10.1016/j.avb.2007.02.010

Rosenbaum, P.R. (2015). *Two R Packages for Sensitivity Analysis in Observational Studies*. Observational Studies, 1, 1-17. Available online: http://obsstudies.org/files/packpaperos_published.pdf (accessed 04/08/2016)

Ruddijs, F., & Timmerman, H. (2000). *A comparative study of recidivism in first offenders in a Dutch outpatient setting*. International Journal of Offender Therapy and Comparative Criminology, 44(6), 725-739. doi:10.1177/0306624X00446008

Sadlier, G. (2010). *Evaluation of the impact of the HM Prison Service Enhanced Thinking Skills programme on reoffending: Outcomes of the Surveying Prisoner Crime Reduction (SPCR) sample*. London: Ministry of Justice. Available online: https://www.justice.gov.uk/downloads/publications/research-and-analysis/moj-research/eval-enhanced-thinking-skills-prog.pdf (accessed 04/08/2016)

Schmucker, M., and Lösel, F. (2008). *Does sex offender treatment work? A systematic review of outcome evaluations.* Psicotherma, 20(1), 10-19. PMID:18206060

Schmucker, M. & Lösel, F. (2015). *The effects of sexual offender treatment on recidivism: an international meta-analysis of sound quality evaluations*. Journal of Experimental Criminology, 11(4), 597-630. doi:10.1007/s11292-015-9241-z

Thornton, D. & Travers, R., (1991). *A longitudinal study of the criminal behaviour of convicted sexual offenders*. Paper presented at Her Majesty's Prison Service Psychology Conference, Scarborough, England, 16-18 October. Abstract available online: https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=143076 (accessed 04/08/2016)

Technical Annex
Annex A
The data set

Data sources

The following data sets from the Ministry of Justice (MoJ) and HM Prison and Probation Service (HMPPS) were used to create the data set for the analysis:

1. **Master Treatment Database.** This database was created by merging and cleaning seven HMPPS treatment databases.[43] The database holds information from 1996 to 2013 on treated sex offenders, including Programme dates, completion information, treatment type, and treatment establishment (prison).
2. **MoJ Reoffending Cohorts (combined).** These data sets are used by the MoJ to publish reoffending statistics, and comprise matched data from the MoJ prison release databases, the probation caseload databases, and the Police National Computer (PNC). Reliable reoffending cohorts are available from 2002 onwards. A combined 2002 to 2012 reoffending cohort was used in this analysis.
3. **MoJ Police National Computer (PNC) extract.** This contains police and court records, including convictions for recordable offences from the first half of the 20th Century until the present.[44] A bespoke extract was prepared for this analysis, including detailed information on sexual offending history and sexual reoffending.
4. **HMPPS Offender Assessment System (OASys) database**. This holds records of practitioner-led offender assessments used for offender management and risk assessment purposes (assessment is limited to prisoners sentenced to more than 12 months in prison). OASys assessments are available from 2004 onwards.

Creation of the treatment and comparison groups

The Master Treatment Database holds 30,398 records relating to the various types of SOTP treatment in English and Welsh prisons between 1996 and 2013.[45] Linking to the PNC database was undertaken in two main steps:

---

[43] As is common with large data sets, data quality in the HMPPS treatment databases varied and the final data set contains some incomplete information.

[44] The PNC was established in 1974, and offending history records are available for some offenders back to the 1920s. However, the reliability of early records is not known. The most reliable records date from the year 2000 onwards.

[45] Some of these records are duplicates.

- Linking records from the Master treatment database to the prison database using prison numbers, in order to obtain more information about the offenders.
- Linking records to the PNC using name, sex, date of birth and PNC/CRO[46] number. This process generated 10,854 automatic matches with a further 728 linked after manual inspection, thus a total of 11,582 individuals were identified.

The **treatment group** was then created selecting those offenders who met the inclusion criteria for the study, and had received Core SOTP during their index sentence and between 2000 and 2012. This generated a treatment group of 2,562 observations.

The **comparison group** was created by identifying sex offenders released from prison who met the inclusion criteria for the study, and who had not received any SOTP treatment during their index sentence (nor before). This generated a comparison group of 13,219 individuals.

All records were linked to the PNC extract and combined Reoffending Cohorts to retrieve individual-level criminal histories and reoffending data, and they were also linked to the OASys database to retrieve practitioner-led assessments on offenders. Valid OASys records were found for less than half of the sample (28% of the treatment group and 46% of the comparison group). A valid record was considered to be an assessment completed between 30 days before and 90 days after the conviction[47] for the index offence (the offence relating to the index prison sentence). Where more assessments were available, the one with the fewest missing values on the items of interest was selected. More information is provided about data quality and the extent of missing values in the notes of Table A.5.

[46] PNC and CRO numbers are unique identifiers for individuals, one originating from the Police National Computer and the other from the Criminal Records Office. An individual may have one or both of these identifiers.

[47] If sentencing took place on a later date, the sentencing date was used instead.

Annex B
Sample description

The majority of sex offenders included in this study committed a child contact (primary) index offence (58%). Most of the rest committed an adult serious index offence (28%) or a child image index offence (9%; see Table A.1 for the full breakdown).

Table A.1: Breakdown of index offences by category (unmatched groups)

	Adult Serious	Adult Other	Child Contact	Child Image	Child Other
Treatment number	778	122	1,519	116	27
Treatment proportion	30.4%	4.8%	59.3%	4.5%	1.1%
Control number	3,570	429	7,593	1,330	297
Control proportion	27.0%	3.2%	57.4%	10.1%	2.2%
Total number	4,348	551	9,112	1446	324
Total proportion	27.6%	3.5%	57.7%	9.2%	2.1%

Most treated offenders served a sentence of between 4 and 10 years (62%), with treated offenders generally serving longer sentences (see Table A.2 for the full breakdown). The mean age at time of release was 43.34 years (SD=14.88), and was similar for the two groups (M_{treat}=42.17, SD_{treat}=14.10; M_{ctrl}=43.57, SD_{ctrl}=15.02).

Table A.2: Breakdown by sentence length (unmatched groups)

	Less than 4 years	Between 4 and 10 years	More than 10 years	Life sentence
Treatment number	723	1,598	167	74
Treatment proportion	28.2%	62.4%	6.5%	2.9%
Comparison number	8,790	4,011	412	6
Comparison proportion	66.5%	30.3%	3.1%	0.0%
Total number	9,513	5,609	579	80
Total proportion	60.3%	35.5%	3.7%	0.5%

Reoffending risk was measured with the Offender Group Reconviction Scale (OGRS, Howard et al., 2009), which is a static[48] reoffending risk tool. The mean OGRS4 score was 0.28 (SD=0.19), and was similar across the two groups (M_{treat}=0.29, SD_{treat}=0.18; M_{ctrl}=0.28, SD_{ctrl}=0.18).

[48] Static factors are not changeable retrospectively, and include, for example, previous criminal history and age. In contrast, dynamic factors are considered changeable, and include factors such as attitudes and beliefs.

A simulated[49] Risk Matrix 2000/Sexual offending scale (RM2000/S, Thornton et al., 2003) was also computed for the study. The RM2000/S is a static risk measure widely used by practitioners and researchers as an indicator of risk of sexual reoffending amongst sex offenders. Overall, the treatment group was higher-risk than the comparison group (see Table A.3 for the full breakdown).

Table A.3: RM2000/S score distribution (unmatched groups)

	Low	Medium	High	Very high
Treatment number	306	1,393	610	227
Treatment proportion	12.1%	54.9%	24.1%	9.0%
Comparison number	3,782	6,423	2,239	579
Comparison proportion	29.0%	49.3%	17.2%	4.4%
Total number	4,088	7,816	2,849	806
Total proportion	26.3%	50.2%	18.3%	5.2%

Note: N=15,559. RM2000/S scores can only be computed for offenders who were at least 16 years old when they committed their most recent sexual offence.

SOTP dosage

The treatment group contained 2,562 offenders who commenced Core SOTP. Of these:

- 76% were recorded as having completed the Programme;[50] 13% were registered as not having completed; and 12% were registered as having an unknown completion status.[51]
- 9% also started Extended SOTP while serving their index sentence, and 1% commenced the SOTP Healthy Sexual Functioning (HSF) Programme. In total, 32% of offenders who commenced Core SOTP also commenced one or more of the non-Core SOTP variants available (Adapted, Booster, Rolling, Extended or HSF).

[49] The RM2000/S scores were simulated. Step 1 of the scale was computed as in Thorton (2007), and used information on the age of offenders at the time of release from prison, and the numbers of previous sexual and any criminal appearances. Step 2 of the scale was computed similarly to Barnett et al. (2010) and Howard et al. (2015), and used information on whether offenders had ever had male victims, and whether they had committed non-contact offences (excluding image offences). Real RM2000/S were known only for a subsample of the treatment group, with a comparison of the two scores revealing that the simulated score correctly classified 64.23% of offenders (X^2=663.00, p<0.001).

[50] Where 'duplicate' SOTP treatment records existed, the Programme was regarded as completed if flagged as such on one or more of the records. Otherwise if flagged as both uncompleted and completion unknown, the Programme was regarded as uncompleted.

[51] Completion rates were poorly recorded in earlier years. In 2004, for example, 85% of people who started Core SOTP were registered as having an unknown completion status.

- 2% were recorded as repeating Core SOTP during the follow-up period, while 2% were recorded as attending other prison-based SOTP variants during the follow-up period.[52]
- Around 22% were recorded as having started community SOTP during the follow-up period, with 82% of these offenders being released from 2008 onwards.[53]

The comparison group contained 13,219 offenders. Of these:

- Less than 1% were recorded as starting Core SOTP in the follow-up period, and less than 1% were recorded as doing any other prison-based SOTP treatment variants.[54]
- 8% were recorded as having started community SOTP during the follow-up period, with 85% of these being released from 2008 onwards.[55]

Reoffending rate

The average follow-up period was 8.2 years (8.5 for the treatment group, 8.2 for the comparison group), with a minimum follow-up period of 2.9 years, and a maximum of 13.9 (SD=3.2). The overall binary reoffending and sexual reoffending[56] rates for the combined unmatched treatment and comparison groups were 38.3% and 7.5% respectively, while the equivalent 2-year rates were 19.9% and 3.2% respectively (see the full breakdown in Table A.4; see Table A.10 for the 3-, 5-, and 7-year sexual reoffending rates).

[52] While the follow-up could extend up to October 2015, the SOTP treatment database only contained records up to 2013.

[53] Data on Community SOTP participation was only available from the year 2008 until 2013.

[54] See footnote 52.

[55] See footnote 53.

[56] Unless otherwise stated, the sexual reoffending measure includes all sexual offences except breaches. Soliciting/prostitution is not currently classified as a sexual offence and so is also excluded.

Table A.4: Overall and 2-year binary reoffending rates (unmatched groups)

	Treated overall (N=2,562)		Control overall (N=13,219)		Treated 2-year (N=2,562)		Control 2-year (N=13,219)	
	Rate	**Number**	**Rate**	**Number**	**Rate**	**Number**	**Rate**	**Number**
Total	39.3%	1,006	38.1%	5,038	17.9%	458	20.3%	2,679
Sexual	10.0%	256	7.0%	921	3.6%	91	3.2%	417
Adult serious	2.7%	68	2.1%	271	1.1%	27	0.9%	122
Adult other	2.1%	55	1.6%	205	0.7%	18	0.7%	91
Child contact	2.2%	56	1.7%	219	0.7%	18	0.7%	87
Child image	4.4%	113	2.6%	340	1.4%	35	1.0%	137
Child other	0.5%	14	0.6%	76	0.2%	6	0.3%	34
Breach	21.8%	558	22.6%	2,982	8.3%	212	9.8%	1,289
Soliciting	0.3%	7	0.1%	13	0.2%	4	0.0%	6
Non-sex violent	3.6%	93	4.1%	539	1.4%	37	1.5%	197
Non-sex non-violent	22.8%	584	23.1%	3,059	9.7%	249	12.3%	1,621

While these sexual reoffending rates are low when compared to international studies,[57] the present figures are within the range of other UK-based studies on reconviction rates for sex offenders, with an average of 6.0% with a follow-up period of up to 2 years, ranging from 1.2% to 10.3%, and an average of 19.5% for 6 or more years of follow-up, ranging from 8.5% to 25.0% (Craig et al., 2008). It has also been argued that, although sexual reconviction rates in the early 1980s were in line with international figures, these rates have been declining in England and Wales since the 1980s (Friendship and Beech, 2005), which suggests that a national difference may be the underlying cause of the seemingly low sexual reoffending rates detected in this study.

[57] For example, Hanson et al. 2010 detected a 15.2% sexual reoffending rate across 7 studies, with an average follow-up period of 7.5 years.

Annex C
Propensity score matching (PSM)

Matching process

Treatment and comparison group records were matched using PSM. A wide range of potential matching factors were identified following extensive review of the literature and in consultation with colleagues across the Ministry of Justice, the HM Prison and Probation Service and an expert advisory panel. Many of the factors are categorical and so were converted to 'dummy variables'. The final model, referred to as the 'parsimonious' model, included factors that were either deemed theoretically important (asterisked in Table A.5), and/or were empirically related to both selection onto the Programme and one of the main five outcome measures (overall, sexual, adult serious, child contact, and child image binary reoffending) at the 20% significance level. Interaction terms between the most important variables, as ascertained using machine learning techniques,[58] were also assessed. In total, this process generated 87 variables (listed in Table A.5) that were used to create propensity scores.

Table A.5: List of matching factors and matching quality

Variable	Sample[1]	Treated	Control	Bias[2]	T-val[3]	P-val[4]
Year of release 2003	U	0.08	0.09	-1.5%	-0.69	0.492
	M	0.08	0.08	0.3%	0.10	0.923
Year of release 2004	U	0.11	0.09	6.8%	3.26	0.001
	M	0.11	0.10	1.0%	0.33	0.741
Year of release 2005	U	0.10	0.09	4.0%	1.87	0.061
	M	0.10	0.10	0.2%	0.07	0.943
Year of release 2006	U	0.09	0.08	1.2%	0.57	0.567
	M	0.09	0.08	0.4%	0.15	0.885
Year of release 2007	U	0.11	0.09	8.1%	3.89	<0.001
	M	0.11	0.12	-1.9%	-0.63	0.531
Year of release 2008	U	0.11	0.09	4.9%	2.33	0.020
	M	0.11	0.11	0.7%	0.23	0.817
Year of release 2009	U	0.09	0.10	-1.7%	-0.80	0.425
	M	0.09	0.09	0.5%	0.19	0.850
Year of release 2010	U	0.06	0.08	-7.7%	-3.41	0.001
	M	0.06	0.06	-1.0%	-0.39	0.697

[58] Combinations of factors were considered for use in the model for the factors that were empirically most important. The criterion for each factor was an absolute importance of 0.001 or more in predicting treatment and in predicting at least one of the five main outcome measures. Absolute importance was evaluated using the R package RF-SRC (Ishwaran and Kogalur, 2015) and represents the change in the classification error we would expect as a result of randomly changing the value of the variable in the random forest. Due to lack of computing power, the 50 combinations thought most likely to have an impact were then included alongside the potential matching factors.

Variable	Sample[1]	Treated	Control	Bias[2]	T-val[3]	P-val[4]
Year of release 2011	U	0.08	0.10	-7.8%	-3.50	<0.001
	M	0.08	0.08	-0.4%	-0.14	0.891
Year of release 2012	U	0.09	0.12	-9.9%	-4.39	<0.001
	M	0.09	0.09	-0.4%	-0.15	0.878
Year of conviction 97-99	U	0.15	0.06	29.7%	15.96	<0.001
	M	0.15	0.15	1.3%	0.40	0.689
Year of conviction 00-02	U	0.30	0.20	24.8%	12.08	<0.001
	M	0.31	0.31	0.0%	-0.01	0.990
Year of conviction 03-05	U	0.28	0.26	4.2%	1.94	0.053
	M	0.28	0.29	-1.1%	-0.38	0.701
Year of conviction 06-08	U	0.16	0.23	-18.1%	-8.03	<0.001
	M	0.15	0.15	-0.1%	-0.05	0.962
Year of conviction 09-12	U	0.09	0.23	-41.2%	-16.99	<0.001
	M	0.09	0.08	0.9%	0.40	0.687
Asian ethnicity[7]	U	0.03	0.04	-7.7%	-3.36	0.001
	M	0.03	0.03	0.3%	0.12	0.905
Black ethnicity[7]	U	0.04	0.05	-3.8%	-1.7	0.09
	M	0.04	0.04	-0.1%	-0.04	0.971
Unknown ethnicity[7]	U	0.02	0.03	-5.9%	-2.58	0.01
	M	0.02	0.02	0.5%	0.20	0.840
Other ethnicity[7]	U	0.00	0.01	-5.6%	-2.35	0.019
	M	0.00	0.00	0.6%	0.26	0.791
Has a victim gender preference[8]	U	0.73	0.69	10.3%	4.69	<0.001
	M	0.73	0.75	-2.8%	-1.02	0.308
Age at release*	U	42.17	43.57	-9.6%	-4.36	<0.001
	M	42.17	42.47	-2.1%	-0.76	0.445
Age at index offence	U	33.46	36.34	-22.9%	-10.3	<0.001
	M	33.46	33.51	-0.4%	-0.13	0.895
Age at index offence squared	U	1,263.20	1,493.80	-23.3%	-10.26	<0.001
	M	1,264.10	1,265.60	-0.2%	-0.06	0.952
Age at first sex conviction[8]	U	33.43	38.64	-36.5%	-16.28	<0.001
	M	33.45	33.88	-3.0%	-1.14	0.256
Life index sentence	U	0.03	0.00	23.8%	18.75	<0.001
	M	0.02	0.03	-0.5%	-0.12	0.901
Index sentence of 4-10 years	U	0.62	0.30	67.8%	31.99	<0.001
	M	0.63	0.62	0.8%	0.28	0.783
Index sentence greater than 10 years	U	0.07	0.03	15.9%	8.4	<0.001
	M	0.07	0.07	-1.3%	-0.41	0.684
Prison of release was delivering SOTP at time of release[9]	U	0.83	0.62	49.2%	21.04	<0.001
	M	0.83	0.85	-3.5%	-1.46	0.144
Index prison sentence less than 1 year since previous prison sentence	U	0.04	0.05	-1.0%	-0.45	0.651
	M	0.04	0.05	-2.9%	-1.02	0.309
Adult serious index offence*,[8]	U	0.30	0.27	7.4%	3.49	<0.001
	M	0.30	0.30	1.9%	0.68	0.498
Child contact index offence*,[8]	U	0.59	0.57	3.8%	1.73	0.083
	M	0.59	0.60	-2.55	-0.88	0.377

Variable	Sample[1]	Treated	Control	Bias[2]	T-val[3]	P-val[4]
Child image index offence*,[8]	U	0.05	0.10	-21.4%	-8.91	<0.001
	M	0.05	0.04	1.1%	0.50	0.616
Child other index offence*,[8]	U	0.01	0.02	-9.4%	-3.90	<0.001
	M	0.01	0.01	1.0%	0.47	0.636
Victim less than 13 years old*,[8]	U	0.08	0.10	-5.1%	-2.30	0.021
	M	0.08	0.08	-1.8%	-0.67	0.502
Female victim*,[8]	U	0.67	0.71	-9.0%	-4.20	<0.001
	M	0.67	0.67	-0.4%	-0.14	0.891
Male victim*,[8]	U	0.15	0.10	14.2%	6.99	<0.001
	M	0.15	0.14	1.4%	0.46	0.646
Family member victim*,[8]	U	0.02	0.02	-1.6%	-0.73	0.466
	M	0.02	0.02	-0.4%	-0.15	0.881
Adult serious secondary index offence[8]	U	0.27	0.20	15.1%	7.21	<0.001
	M	0.27	0.27	-0.6%	-0.20	0.838
Child image secondary index offence[8]	U	0.15	0.14	0.4%	0.17	0.868
	M	0.14	0.13	3.6%	1.32	0.188
Child other secondary index offence[8]	U	0.03	0.03	0.2%	0.11	0.913
	M	0.03	0.02	2.3%	0.88	0.381
Victim family member secondary index offence[8]	U	0.04	0.03	5.6%	2.72	0.007
	M	0.04	0.04	-0.6%	-0.20	0.842
Number of child contact secondary index offences[8]	U	3.57	2.33	28.8%	14.96	<0.001
	M	3.57	3.64	-1.8%	-0.55	0.582
Number of secondary index offences involving male victims[8]	U	1.14	0.45	25.5%	15.02	<0.001
	M	1.14	1.14	0.0%	0.01	0.994
Ever previously committed adult other offence*,[8]	U	0.06	0.04	13.2%	6.75	<0.001
	M	0.06	0.06	2.2%	0.70	0.486
Ever previously committed child contact offence*,[8]	U	0.22	0.12	26.7%	13.54	<0.001
	M	0.22	0.22	-1.1%	-0.35	0.728
Ever previously committed child image offence*,[8]	U	0.03	0.02	6.1%	3.03	0.002
	M	0.03	0.03	2.9%	1.01	0.312
Ever previously committed child other offence*,[8]	U	0.00	0.00	4.0%	2.13	0.033
	M	0.00	0.00	1.4%	0.46	0.648
Ever previously had victim less than 13 years old*,[8]	U	0.01	0.01	3.9%	1.91	0.056
	M	0.01	0.01	0.2%	0.05	0.959
Ever previously had male victim*,[8]	U	0.11	0.05	22.8%	12.19	<0.001
	M	0.11	0.10	2.1%	0.64	0.519
Ever previously had female victim*,[8]	U	0.22	0.13	24.1%	12.05	<0.001
	M	0.22	0.22	-0.1%	-0.02	0.98
Ever previously had family member victim*,[8]	U	0.03	0.01	8.9%	4.63	<0.001
	M	0.03	0.03	-1.5%	-0.45	0.653
Ever (index or previously) committed exhibitionism offence[8]	U	0.06	0.04	11.1%	5.62	<0.001
	M	0.06	0.05	2.7%	0.89	0.376
Ever (index or previously) committed rape[8]	U	0.41	0.23	39.0%	19.13	<0.001
	M	0.41	0.41	-0.7%	-0.23	0.817
First offence was a non-sex non-violent offence[8]	U	0.34	0.35	-3.1%	-1.42	0.155
	M	0.34	0.35	-2.0%	-0.73	0.467

Variable	Sample[1]	Treated	Control	Bias[2]	T-val[3]	P-val[4]
Number of previous adult serious offences[8]	U	0.20	0.11	12.1%	5.76	<0.001
	M	0.21	0.20	1.0%	0.35	0.723
Number of previous adult serious offences squared[8]	U	0.65	0.55	1.3%	0.5	0.619
	M	0.66	0.65	0.1%	0.08	0.934
Number of previous adult other (excluding images) offences[8]	U	0.22	0.11	6.2%	3.69	<0.001
	M	0.22	0.21	0.8%	0.23	0.818
Number of previous child contact offences[8]	U	0.81	0.40	19.9%	10.38	<0.001
	M	0.81	0.79	1.2%	0.39	0.698
Number of previous child contact offences squared[8]	U	6.43	3.15	9.4%	4.82	<0.001
	M	6.45	6.13	0.9%	0.30	0.768
Number of previous child victim offences[8]	U	0.99	0.62	13.4%	6.36	<0.001
	M	0.98	0.92	2.1%	0.76	0.446
Any previous offence is a serious offence (child contact / adult serious)[8]	U	0.27	0.15	28.6%	14.27	<0.001
	M	0.27	0.26	0.3%	0.10	0.924
Number of non-sex non-violent offences during last previous year[8]	U	0.36	0.36	-0.1%	-0.03	0.979
	M	0.36	0.36	0.3%	0.10	0.920
Number of non-sex non-violent offences during last previous year squared[8]	U	1.89	2.05	-0.9%	-0.41	0.684
	M	1.90	1.98	-0.5%	-0.20	0.841
Number of previous custodial sentences[8]	U	0.99	1.00	-0.4%	-0.19	0.847
	M	0.99	1.00	-0.2%	-0.09	0.928
OGRS 4 score*[,8]	U	0.29	0.28	5.9%	2.64	0.008
	M	0.29	0.29	0.7%	0.26	0.793
RM2000/S medium*[,8]	U	0.54	0.49	11.6%	5.36	<0.001
	M	0.54	0.56	-2.9%	-1.03	0.305
RM2000/S high score*[,8]	U	0.24	0.17	17.1%	8.29	<0.001
	M	0.24	0.23	1.9%	0.63	0.526
RM2000/S very high score*[,8]	U	0.09	0.04	18.1%	9.45	<0.001
	M	0.09	0.08	2.6%	0.83	0.408
Has an OASys assessment[10]	U	0.28	0.46	-36.8%	-16.52	<0.001
	M	0.28	0.28	0.2%	0.09	0.925
Is the victim of the index offence a stranger[11]	U	0.07	0.12	-18.9%	-8.08	<0.001
	M	0.07	0.06	1.4%	0.59	0.558
Suitability of location of accommodation on release: some problems[11]	U	0.03	0.04	-5.9%	-2.60	0.009
	M	0.03	0.03	1.4%	0.53	0.595
Suitability of location of accommodation on release: significant problems[11]	U	0.09	0.13	-13.7%	-5.99	<0.001
	M	0.09	0.09	-0.4%	-0.14	0.886
Learning difficulties, attended special school etc.: some problems[11]	U	0.01	0.03	-13.2%	-5.42	<0.001
	M	0.01	0.01	0.4%	0.21	0.837
Learning difficulties, attended special school etc.: significant problems[11]	U	0.01	0.02	-14.4%	-5.67	<0.001
	M	0.01	0.01	0.4%	0.20	0.843
Emotional wellbeing linked to offending behaviour[11]	U	0.12	0.16	-10.2%	-4.56	<0.001
	M	0.12	0.12	0.6%	0.21	0.832
Impulsivity: some problems*[,11]	U	0.10	0.15	-14.2%	-6.23	<0.001
	M	0.10	0.11	-3.8%	-1.45	0.146

Variable	Sample[1]	Treated	Control	Bias[2]	T-val[3]	P-val[4]
Impulsivity: significant problems*,[11]	U	0.05	0.08	-10.9%	-4.78	<0.001
	M	0.05	0.05	1.9%	0.76	0.45
Aggressiveness: some problems*,[11]	U	0.07	0.09	-8.1%	-3.62	<0.001
	M	0.07	0.06	2.2%	0.85	0.395
Aggressiveness: significant problems*,[11]	U	0.05	0.09	-15.9%	-6.78	<0.001
	M	0.05	0.04	1.8%	0.76	0.445
Temper control: significant problems*,[11]	U	0.03	0.04	-7.6%	-3.33	0.001
	M	0.03	0.03	1.1%	0.43	0.669
Pro-criminal attitudes: some problems[11]	U	0.03	0.06	-12.2%	-5.20	<0.001
	M	0.03	0.03	1.9%	0.83	0.405
Pro-criminal attitudes: significant problems[11]	U	0.03	0.04	-5.3%	-2.34	0.019
	M	0.03	0.03	2.2%	0.87	0.382
Successful completion of accredited programmes*[11,12]	U	0.02	0.03	-2.0%	-0.91	0.362
	M	0.02	0.02	2.1%	0.81	0.415
Has general health problems[11]	U	0.10	0.21	-28.9%	-12.29	<0.001
	M	0.10	0.10	-0.4%	-0.18	0.855
OASys Violent Predictor score*,[11]	U	3.05	4.14	-11.5%	-5.01	<0.001
	M	3.04	3.14	-1.0%	-0.38	0.703
Age at index release * ever committed rape[8]	U	16.03	9.57	31.9%	15.36	<0.001
	M	16.04	16.37	-1.6%	-0.55	0.584

Notes:

1. U = unmatched, M = matched
2. As indicated by the standardised (mean) difference which if <=5 indicates that the groups were closely matched, 5–10 that there was a reasonable match quality, and >10 indicates a poor quality of matching.
3. T-value
4. P-value
5. The following reference categories have been excluded; Year of release 2002; Year of conviction 78-96; White ethnicity; Has no known victim gender preference; Index sentence of 1-4 years; Prison of release wasn't delivering SOTP at time of release; Index prison sentence not less than 1 year after previous prison sentence; Adult other index offence; Has no recorded OASys assessment; The victim of the index offence is not a stranger; Suitability of location of accommodation on release: no problems; Learning difficulties: no problems; Emotional wellbeing not linked to offending behaviour; Impulsivity: no problems; Aggressiveness: no problems; Temper control: no significant problems; Pro-criminal attitudes: no problems; Hasn't successfully completed accredited programmes; No general health problems.
6. The variables asterisked were deemed theoretically important and were included in the PSM model regardless of whether they were found to be empirically related to both selection onto the Programme and one of the main five outcome measures (overall, sexual, adult serious, child contact, and child image binary reoffending) at the 20% significance level.
7. Ethnicity was self-reported.
8. While these PNC/prison system variables are complete, the reliability of early PNC offending history records before 2000 is not known. In addition, many of these variables will be undercounts due to many of the Home Office offence code descriptions not providing further information about the victim (e.g. gender, age). The terms 'previous' and 'previously' refer to prior to the index conviction date with 'previous' offences being those that occurred before the index conviction date and were not included within the index conviction.
9. Being in an SOTP prison raises the chance of entering treatment in the institution and also the general climate at an SOTP institution may be more supportive regardless of being in SOTP or not. As the start of the sentence prison data were unavailable, the prison of release data were used which may or may not have been the prison the offender stayed at for most of the sentence, and may or may not have been the prison where any SOTP treatment took place. SOTP treatment records indicated whether the release prison was a SOTP treatment one at the time of release.
10. Has an OASys assessment - Valid OASys records were found for less than half of the sample (28% of the treatment group and 46% of the comparison group).
11. The variables were from the OASys assessment system. Where valid OASys records were found, missing values were found to account for less than 10% of variables except Pro-criminal attitudes (32% missing),

Successful completion of accredited programmes (46% missing) and OASys Violent Predictor score (13% missing).

12. The 'Successful completion of accredited programmes' variable reflects whether or not the offender has a record of successfully completing accredited programmes at the time at which the OASys assessment was carried out. As to be included within this study, individuals had to have no record of receiving any type of prison-based sex offender treatment commencing prior to the index sentence, such accredited programmes should not include SOTP.

A less parsimonious model was also developed, where any factor *either* related to selection onto treatment or a main outcome measure was retained. The additional factors in the less parsimonious model (and so not listed in Table A.5 above) were as follows:

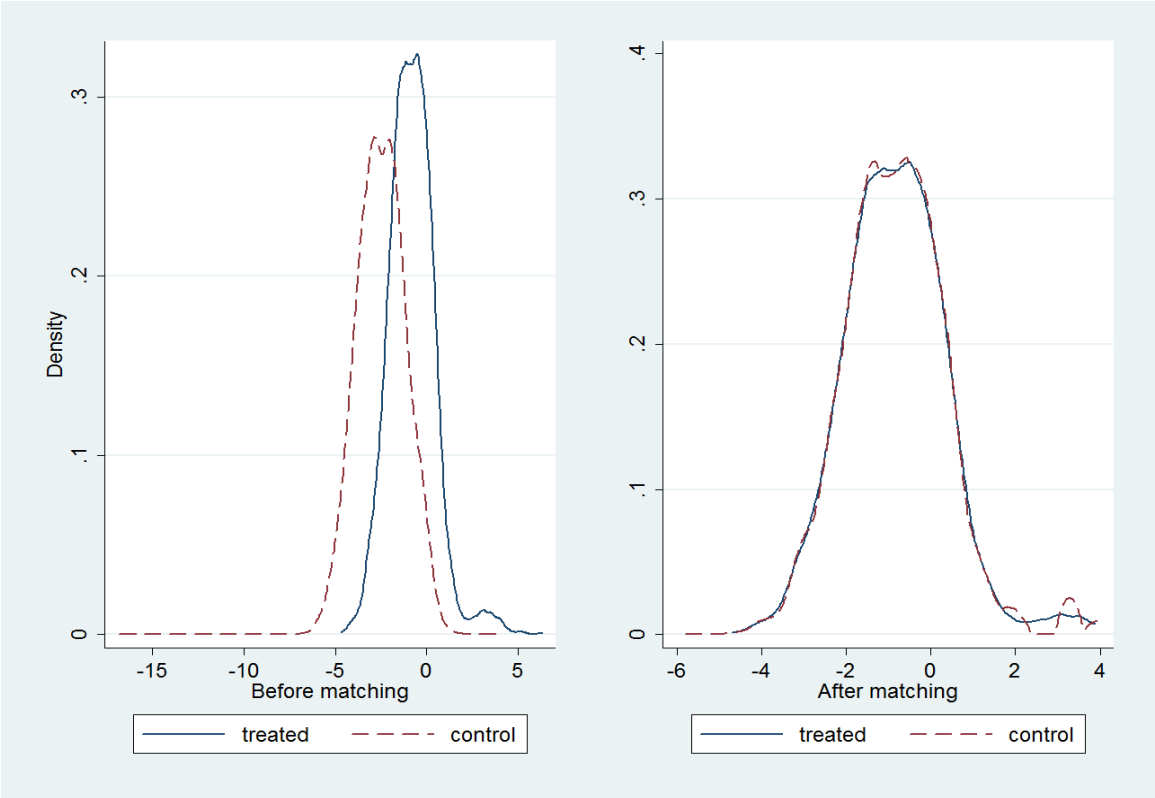
- age at first ever offence, at first conviction/caution, and at first ever sex offence.
- the offender had a secondary index offence (yes/no) in each of the following categories; non-sex violent, non sex non-violent, adult other, child contact, breach, child victim, victim under 13, male victim, and female victim.
- the numbers of secondary index offences in the following categories; non-sex non-violent, victim is child, and female victim.
- the offender had ever previously committed an offence (yes/no) in each of the following categories; sex, non-sex violent, non-sex non-violent, adult serious, soliciting, and victim is child.
- the offender's first ever offence related to each of the following categories (yes/no); child contact, male victim, victim is child, and female victim.
- the offender's first ever conviction/caution related to each of the following categories (yes/no); child contact, male victim, victim is child, female victim, non-sex non-violent.
- the numbers of previous offences in the following categories: non-sex violent, non-sex non-violent with conviction(s)/caution(s) 2 to 5 years before the index conviction, non-sex non-violent with conviction(s)/caution(s) more than 5 years before the index conviction, child images, breach, sexual (including prostitution/soliciting and breaches) committed under the age of 18, non-sex violent offences committed under the age of 18, male victim, female victim, and family member victim.
- OASys assessment variables as follows: the offender recognises the impact and consequences of his offending on the victim and the community (yes/no), the offender accepts responsibility for the current offence (yes/no), the offender usually lives alone (yes/no), the offender usually lives with partner (yes/no), the offender usually lives with children (yes/no), the person is unemployed or will be unemployed on release (yes/no), employment history (some problems, significant problems), childhood experience (significant problems), partner relationship or if

single the level of satisfaction with this (some problems, significant problems), previous relationships (some problems, significant problems), perpetrator of domestic violence (yes/no), the offender has social networks outside the family and friends that he interacts with on a regular basis (some problems, significant problems), psychiatric illnesses or symptoms have been diagnosed by a GP or a psychiatrist (significant problems), there is evidence of childhood behavioural problems (yes/no), there is some history of severe head injuries, fits, or period of unconsciousness (yes/no), the offender has a history of psychiatric treatment (yes/no), problem solving skills (some problems, significant problems), understands other people's views (some problems, significant problems), knows why offending (some problems, significant problems), motivation to address offending (quite motivated, very motivated), issues linked to offending behaviour (education, training and employability (yes/no), financial (yes/no), personal relationships (yes/no), lifestyle & associates (yes/no), drug misuse (yes/no), alcohol misuse (yes/no), and attitudes (yes/no)), issues linked to the risk of serious harm, risks to the individual and other risks (financial (yes/no), drug misuse (yes/no), alcohol misuse (yes/no), thinking & behaviour (yes/no), and attitudes (yes/no)), and a count of OASys sections where issues are linked to the risk of serious harm, risks to the individual and other risks.

- interaction terms involving firstly the age at release and the age at first conviction/caution for a sex offence, and secondly the age at first conviction/caution for a sex offence and the number of previous offences involving a female victim.

The propensity scores represent the probability of entering Core SOTP treatment given an offender's observed characteristics. There was a large region of common support (where the propensity scores for the treatment comparison groups overlap), which implies they can be matched. After matching, the distributions of propensity scores in the two groups was very similar. Figure A.1 shows the overlap in the logit of the propensity scores between the two groups before and matching.

Figure A.1: Distribution of the logit of the propensity scores across the treatment and comparison groups, before and after matching



The matching was conducted using the *psmatch2* programme (Leuven and Sianesi, 2003) in the Stata statistical analysis software package. A variety of PSM models were tested, including: 1-1, 1-3, and 1-5 nearest neighbour matching; radius matching; kernel matching using the Epanechnikov distribution; kernel matching using the normal distribution. Different specifications were tried for each of these models, such as matching either using the propensity score or the logit function of the propensity score, and varying the size of the caliper[59] or bandwidth[60] used for matching. A comparison of the various models and specifications is provided in Tables A.6 and A.7. The two main criteria for choosing these specifications were: maximising the matching quality, as determined by the standardised mean differences of the matching variables; and minimising the number of treated individuals who were lost from the analysis because they could not be matched. On balance, it was

[59] The caliper was varied from 0.1 to 0.2 standard deviations of either the propensity score or logit of the propensity score.

[60] The bandwidth was varied from 0.03 to 0.06.

decided to use kernel matching[61] on the logit of the propensity score with a normal distribution, a bandwidth of 0.03 and a requirement of common support (this drops treatment individuals with propensity scores that are outside the range of propensity scores of the comparison group).

The model was then run using the more and less parsimonious set of matching factors. As the results were very similar across the two models (see Tables A.8 and A.9), the parsimonious set of factors was selected as the main model for the analyses.

Matching quality

Treatment and comparison offenders were matched using the main model described above in order to control for the observable differences in characteristics between groups. Based on the observed variables, the quality of the match was excellent (see Table A.5 for matched averages in each group), with no statistically significant differences[62] observed between the matched treatment and comparison groups across any of the 87 matching variables. The standardised mean differences for the 87 variables that were included were all less than 5%, with the highest being 3.8%, and an overall average of 1.2%.

A total of 11 offenders in the treatment group could not be matched. They had the following characteristics:

- They all had high propensity scores.
- They all received index life sentences (69 matched treatment group offenders also received life sentences).
- Two of them reoffended.

[61] Kernel matching is a weighted matching method, with each member of the matched treatment group receiving a weight of 1 and each member of the matched control group receiving a weight that increases with the closeness of their matches to treatment group members as based on the propensity score. Significance levels were computed using *psmatch2*. This uses Welch's t-test, and takes weights into account in its standard error calculation by multiplying the control group standard deviation by the sum of the squares of the weights divided by the square of sum of the weights (instead of dividing the standard deviation by the number of control group members, as in an unweighted calculation). The treatment group standard deviation is calculated using the matched treatment group members.

[62] Unless otherwise stated, the level of significance used in this study is 5%.

Table A.6: Parsimonious model, model specification matching quality comparison

Type	Bandwidth[1]/ caliper	Treat no. dropped[2]	Mean st diff[3]	No. st diffs[4] > 5
Kernel (epan[5],	0.015	101	1.10	0 (largest 4.18)
common[6], logit of ps[7])	0.03	75	1.15	0 (largest 4.05)
	0.06	62	1.04	0 (largest 3.15)
	0.09	53	1.07	0 (largest 3.13)
Kernel (epan[5],	0.015	30	1.20	0 (largest 3.53)
common[6], ps[7])	0.03	17	1.26	0 (largest 3.76)
	0.06	11	1.35	0 (largest 4.80)
	0.09	11	1.51	1 (largest 5.74)
Kernel (normal,	0.03	11	1.22	0 (largest 3.81)
common[6], logit of ps[7])	0.06	11	1.26	0 (largest 3.79)
Kernel (normal,	0.03	11	1.37	0 (largest 4.99)
common[6], ps[7])	0.06	11	1.95	3 (largest 8.41)
Radius (logit of ps[7])	0.1 SD[8]	45	1.12	0 (largest 3.12)
	0.2 SD[8]	26	1.23	0 (largest 3.70)
Radius (ps[7])	0.1 SD[8]	16	1.27	0 (largest 3.97)
	0.2 SD[8]	2	1.44	0 (largest 4.92)
NN1-5 (logit of ps[7])	0.1 SD[8]	45	1.39	0 (largest 4.04)
	0.2 SD[8]	26	1.42	0 (largest 4.16)
NN1-5 (ps[7])	0.1 SD[8]	16	1.52	0 (largest 4.48)
	0.2 SD[8]	2	1.55	1 (largest 5.02)
NN1-3 (logit of ps[7])	0.1 SD[8]	45	1.60	1 (largest 5.20)
	0.2 SD[8]	26	1.63	1 (largest 5.64)
NN1-3 (ps[7])	0.1 SD[8]	16	1.73	1 (largest 5.88)
	0.2 SD[8]	2	1.82	1 (largest 6.08)
NN1-1 (logit of ps[7])	0.1 SD[8]	45	2.15	More than 5%
	0.2 SD[8]	26	2.21	More than 5%
NN1-1 (ps[7])	0.1 SD[8]	16	2.31	More than 5%
	0.2 SD[8]	2	2.35	More than 5%

Notes:

1. For more explanation of the bandwidth/caliper please see Caliendo and Kopeinig (May 2015).
2. Number of treatment individuals that could not be matched.
3. The mean of the standardised (mean) differences that were produced for each matching variable.
4. Number of standardised differences that were more than 5 (one of <=5 indicates that the groups were closely matched, 5–10 that there was a reasonable match quality, >10% a poor quality of matching).
5. Epanechnikov distribution.
6. Common support (this drops treatment individuals with propensity scores that are outside the range of propensity scores of the comparison group).
7. Propensity score.
8. Standard deviation of either the propensity score or logit of the propensity score.

Table A.7: Less parsimonious model, model specification matching quality comparison

Type	Bandwidth[1]/ caliper	Treat obs dropped[2]	Mean st diff[3]	No. st diffs[4] > 5
Kernel (epan[5],	0.015	128	0.94	0 (largest 3.1)
common[6], logit of ps[7])	0.03	107	1.08	0 (largest 3.0)
	0.06	75	1.20	0 (largest 3.3)
	0.09	60	1.22	0 (largest 3.2)
Kernel (epan[5],	0.015	22	1.34	0 (largest 3.9)
common[6], ps[7])	0.03	10	1.44	0 (largest 4.1)
	0.06	9	1.39	0 (largest 3.6)
	0.09	9	1.44	0 (largest 3.4)
Kernel (normal,	0.03	9	1.55	0 (largest 4.5)
common[6], logit of ps[7])	0.06	9	1.52	0 (largest 4.4)
Kernel (normal,	0.03	9	1.39	0 (largest 3.6)
common[6], ps[7])	0.06	9	1.64	6 (largest 6.4)
Radius (logit of ps[7])	0.1 SD[4]	49	1.32	0 (largest 3.2)
	0.2 SD[4]	31	1.26	0 (largest 3.7)
Radius (ps[7])	0.1 SD[4]	4	1.46	0 (largest 4.5)
	0.2 SD[4]	0	1.48	1 (largest 5.1)
NN1-5 (logit of ps[7])	0.1 SD[4]	49	1.30	0 (largest 4.0)
	0.2 SD[4]	31	1.35	0 (largest 4.4)
NN1-5 (ps[7])	0.1 SD[4]	4	1.49	2 (largest 5.2)
	0.2 SD[4]	0	1.53	1 (largest 5.3)
NN1-3 (logit of ps[7])	0.1 SD[4]	49	1.44	0 (largest 4.5)
	0.2 SD[4]	31	1.51	0 (largest 4.6)
NN1-3 (ps[7])	0.1 SD[4]	4	1.62	3 (largest 5.4)
	0.2 SD[4]	0	1.65	3 (largest 5.3)
NN1-1 (logit of ps[7])	0.1 SD[4]	49	1.94	More than 5%
	0.2 SD[4]	31	1.99	More than 5%
NN1-1 (ps[7])	0.1 SD[4]	4	2.11	More than 5%
	0.2 SD[4]	0	2.13	More than 5%

Notes:

1. For more explanation of the bandwidth/caliper please see Caliendo and Kopeinig (May 2015).
2. Number of treatment individuals that could not be matched.
3. The mean of the standardised (mean) differences that were produced for each matching variable.
4. Number of standardised differences that were more than 5 (one of <=5 indicates that the groups were closely matched, 5–10 that there was a reasonable match quality, >10% a poor quality of matching).
5. Epanechnikov distribution.
6. Common support (this drops treatment individuals with propensity scores that are outside the range of propensity scores of the comparison group).
7. Propensity score.
8. Standard deviation of either the propensity score or logit of the propensity score.

Table A.8: Overall binary reoffending rates for the matched treatment and comparison groups using the parsimonious model and the less parsimonious model, with numbers of reoffenders[63]

	Parsimonious model (treated N=2,551, control N=13,219)						Less parsimonious model (treated N=2,553, control N=13,219)					
	Treated rate	Treated no.[1]	Control rate	Control no.[1]	Difference	T-value	Treated rate	Treated no.[1]	Control rate	Control no.[1]	Difference	T-value
Total	39.4%	1,004	38.9%	5,038	0.5%pts	0.35	39.3%	1,004	38.9%	5,038	0.4%pts	0.30
Non-sex	24.0%	612	23.7%	3,147	0.3%pts	0.27	24.0%	613	24.0%	3,147	0.0%pts	0.02
All sexual[64]	27.8%	710	27.1%	3,477	0.7%pts	0.55	27.8%	709	27.1%	3,477	0.7%pts	0.52
Sexual	10.0%	255	8.0%	921	2.0%pts	2.45*	9.9%	254	7.9%	921	2.1%pts	2.48*
Adult serious	2.7%	68	2.4%	271	0.3%pts	0.59	2.6%	67	2.4%	271	0.3%pts	0.56
Adult other	2.1%	54	1.4%	205	0.7%pts	1.86	2.1%	54	1.4%	205	0.7%pts	1.82
Child contact	2.2%	56	2.1%	219	0.1%pts	0.18	2.2%	56	2.2%	219	0.0%pts	0.05
Child image	4.4%	113	2.9%	340	1.6%pts	2.96**	4.4%	113	2.7%	340	1.7%pts	3.17**
Child other	0.5%	14	1.0%	76	-0.4%pts	-1.86	0.5%	14	1.0%	76	-0.4%pts	-1.96*
Breach	21.9%	558	22.4%	2,982	-0.5%pts	-0.41	21.9%	558	22.5%	2,982	-0.7%pts	-0.52
Soliciting	0.3%	7	0.1%	13	0.2%pts	1.36	0.3%	7	0.1%	13	0.1%pts	1.14
Non-sex violent	3.6%	93	4.4%	539	-0.8%pts	-1.43	3.6%	93	4.5%	539	-0.9%pts	-1.47
Non-sex non-violent	22.9%	583	22.9%	3,059	-0.0%pts	-0.03	22.9%	584	23.3%	3,059	-0.4%pts	-0.30

Notes:

* $p < 0.05$; ** $p < 0.01$.

1. Number of reoffenders (unweighted for the matched control group).

[63] The control group rates are weighted, but the control group numbers are not (see footnote 61 for more information on weights).

[64] This includes breaches and also soliciting/prostitution offences (which are not currently classified as sexual offences).

Table A.9: Overall frequency of reoffending for the matched treatment and comparison groups using the parsimonious model and the less parsimonious model, with standard deviations[65]

	Parsimonious model (treated N=2,551, control N=13,219)						Less parsimonious model (treated N=2,553, control N=13,219)					
	Treated freq	Treated SD	Control freq	Control SD	Difference	T-value	Treated freq	Treated SD	Control freq	Control SD	Difference	T-value
Total	2.03	4.93	2.19	6.21	-0.16	-0.99	2.03	4.92	2.17	6.21	-0.14	-0.85
Non-sex	0.98	3.25	1.25	5.24	-0.27	-2.13*	0.98	3.24	1.25	5.24	-0.27	-2.00*
All sexual[66]	1.05	3.45	0.94	2.62	0.11	1.25	1.04	3.45	0.92	2.62	0.12	1.36
Sexual	0.59	2.88	0.45	2.16	0.15	1.99*	0.59	2.88	0.44	2.16	0.16	2.09*
Adult serious	0.06	0.43	0.04	0.29	0.02	1.91	0.06	0.42	0.04	0.29	0.02	1.81
Adult other	0.04	0.40	0.02	0.31	0.02	2.15*	0.04	0.40	0.02	0.31	0.02	2.18*
Child contact	0.06	0.62	0.06	0.46	0.00	0.26	0.06	0.62	0.06	0.46	0.00	0.17
Child image	0.42	2.67	0.32	1.96	0.10	1.54	0.42	2.67	0.31	1.96	0.12	1.69
Child other	0.01	0.16	0.01	0.14	0.00	-0.88	0.01	0.16	0.01	0.14	0.00	-1.01
Breach	0.45	1.40	0.49	1.26	-0.04	-0.96	0.45	1.40	0.48	1.26	-0.04	-0.88
Soliciting	0.00	0.06	0.00	0.03	0.00	1.43	0.00	0.06	0.00	0.03	0.00	1.24
Non-sex violent	0.06	0.41	0.06	0.36	0.00	-0.45	0.06	0.41	0.07	0.36	-0.01	-0.57
Non-sex non-violent	0.92	3.10	1.19	5.11	-0.27	-2.16*	0.92	3.09	1.18	5.11	-0.26	-2.01*

Note:

* $p < 0.05$; ** $p < 0.01$.

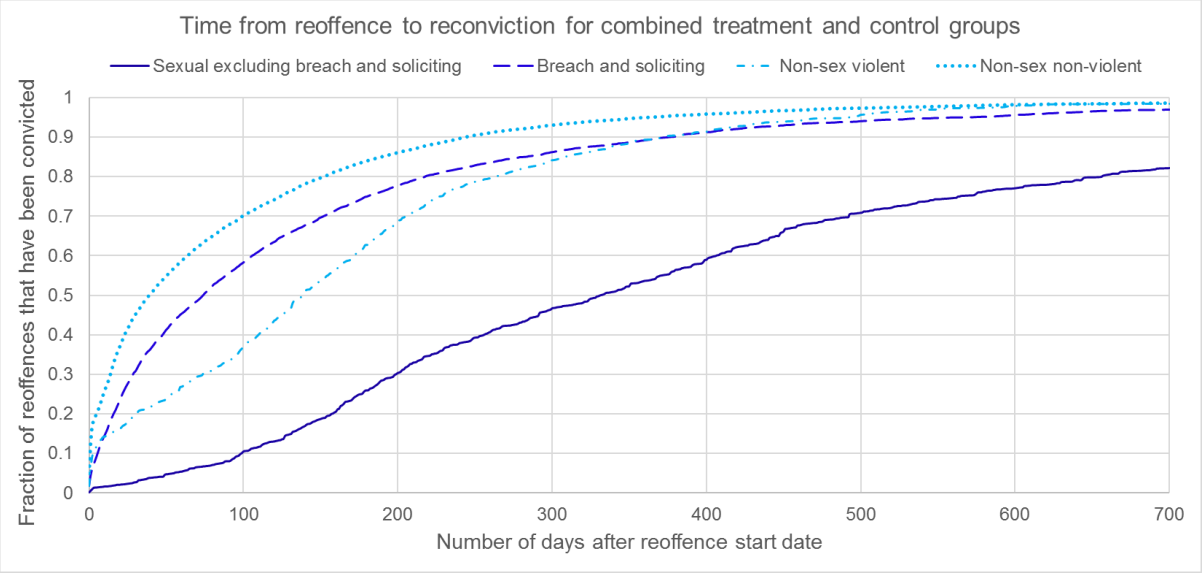
[65] The control group frequencies are weighted, but the control group standard deviations are not (see footnote 61 for more information on weights).

[66] See footnote 64.

Annex D
Outcome Measures

Proven reoffending outcomes (sexual and non-sexual) of the treatment and comparison groups after release from prison were compared and tested to identify statistically significant differences. A proven reoffence is defined as any offence committed during a follow-up period and receiving a court conviction, caution, reprimand or warning within the follow-up period or a further six-month waiting period. Given that the time from offence to conviction for sex offences seems to be longer than for other types of offences (see Figure A.2), in this study no waiting period was set for the court conviction (or caution) to occur.

Figure A.2: Time from reoffence to reconviction for different types of offences



The outcome measures include the binary reoffending rate and the frequency of reoffending per offender, both overall and by reoffence type. The follow-up period for each offender started from their first prison release and ended in October 2015.

Table A.10: Sensitivity analyses' results and matching[67] quality on sexual reoffending binary and frequency outcomes

Analysis	Treated number (matched, dropped)	Control number (total)	Binary & frequency outcome (treated)	Binary & frequency outcome (control)	Difference	T-value	Matching quality (standardised differences[1])
Overall	2,551	13,219	10.0%	8.0%	2.0%pts	2.45	Mean st diff 1.22, largest 3.8
	11	-	0.594	0.449	0.145	1.99	All st diffs <5
Overall (less parsimonious model)	2,553	13,219	9.9%	7.9%	2.1%pts	2.48	Mean st diff 1.55, largest 4.5
	9	-	0.593	0.435	0.157	2.09	All st diffs <5
2-yr follow-up	2,551	13,219	3.5%	3.3%	0.2%pts	0.35	Mean st diff 1.22, largest 3.8
	11	-	0.188	0.142	0.047	1.24	All st diffs <5
3-yr follow-up	2,520	12,923	5.0%	4.7%	0.3%pts	0.52	Mean st diff 1.23, largest 3.8
	12	-	0.272	0.208	0.064	1.34	All st diffs <5
5-yr follow-up	2,099	10,133	8.4%	6.8%	1.6%pts	2.08	Mean st diff 1.08, largest 3.3
	10	-	0.489	0.357	0.133	1.93	All st diffs <5
7-yr follow-up	1,689	7,739	9.7%	7.6%	2.1%pts	2.31	Mean st diff 1.27, largest 4.6
	6	-	0.589	0.426	0.163	1.93	All st diffs <5
2-yr follow-up (excl prison time from conviction to release[2])	2,540	13,131	3.4%	3.2%	0.2%pts	0.38	Mean st diff 1.25, largest 3.8
	12	-	0.174	0.167	0.007	0.19	All st diffs <5
3-yr follow-up (excl prison time from conviction to release[2])	2,479	12,676	4.5%	4.1%	0.4%pts	0.70	Mean st diff 1.37, largest 3.9
	14	-	0.242	0.181	0.060	1.35	All st diffs <5
5-yr follow-up (excl prison time from conviction to release[2])	2,028	9,832	6.6%	5.4%	1.2%pts	1.70	Mean st diff 1.28, largest 5.1
	8	-	0.386	0.282	0.104	1.66	1 st diff >5
7-yr follow-up (excl prison time from conviction to release[2])	1,586	7,372	6.9%	5.6%	1.3%pts	1.64	Mean st diff 1.13, largest 3.5
	6	-	0.391	0.303	0.088	1.23	All st diffs <5
Overall (Core SOTP completers)	1,918	13,219	9.2%	7.8%	1.4%pts	1.59	Mean st diff 1.40, largest 5.2
	20	-	0.546	0.425	0.122	1.54	1 st diff >5

[67] Some treatment group members are dropped (not matched) during each analysis and make no contribution to the results. No control group members are discarded during the matching process, but different weights are assigned to each of them. Control group members with very small weights make a negligible contribution to the results. See Annex C for more information on the matching process.

Analysis	Treated number (matched, dropped)	Control number (total)	Binary & frequency outcome (treated)	Binary & frequency outcome (control)	Difference	T-value	Matching quality (standardised differences[1])
Overall (sentences of >=12months & <4yrs)	721	8,790	10.5%	11.0%	-0.5%pts	-0.39	Mean st diff 0.75, largest 3.4
	2	-	0.735	0.619	0.116	0.82	All st diffs <5
Overall (sentences of >=4yrs & <10yrs)	1,589	4,011	9.9%	7.3%	2.6%pts	2.80	Mean st diff 1.09, largest 4.5
	9	-	0.543	0.412	0.131	1.64	All st diffs <5
Overall (sentences of 2yrs or more)	2,425	8,621	9.6%	7.3%	2.3%pts	2.86	Mean st diff 1.47, largest 3.9
	15	-	0.563	0.395	0.168	2.38	All st diffs <5
Overall (sentences of 4yrs or more)	1,823	4,429	9.8%	6.9%	3.0%pts	3.09	Mean st diff 1.84, largest 9.2
	16	-	0.541	0.361	0.180	2.22	3 st diffs >5, 1 st diff > 6
Overall (hard matching on sentence length)	2,530	13,219	10.0%	8.1%	2.0%pts	2.63	Mean st diff 2.82, largest 12.0
	32	-	0.598	0.451	0.147	2.12	14 st diffs >5, 3 st diff > 7
Overall (index child contact offence)	1,518	7,593	9.1%	7.9%	1.2%pts	1.38	Mean st diff 2.01, largest 9.3
	1	-	0.594	0.537	0.056	0.66	4 st diffs >5, 1 st diff > 6
Overall (index adult serious offence)	776	3,570	10.8%	7.0%	3.8%pts	2.75	Mean st diff 1.45, largest 5.9
	2	-	0.365	0.174	0.191	2.85	2 st diffs >5
Overall (2002-2005 index releases)	950	4,500	11.8%	9.0%	2.8%pts	2.10	Mean st diff 1.47, largest 13.2
	12	-	0.839	0.537	0.302	2.12	1 st diff >5
Overall (2006-2012 index releases)	1,591	8,719	8.7%	7.1%	1.6%pts	1.49	Mean st diff 1.94, largest 6.2
	9	-	0.435	0.377	0.058	0.71	3 st diffs >5, 1 st diff > 6
Overall (only those with OASys assessments)[3]	704	5,940	5.1%	6.2%	-1.1%pts	-0.93	Mean st diff 1.78, largest 5.8
	21	-	0.264	0.256	0.008	0.09	2 st diffs >5
Overall (only those with OASys assessments, excl OASys vars)[3]	712	5,940	5.3%	6.2%	-0.9%pts	-0.71	Mean st diff 1.87, largest 5.6
	13	-	0.296	0.254	0.042	0.46	2 st diffs >5
Overall (without OASys variables)	2,547	13,219	10.0%	8.0%	2.0%pts	2.58	Mean st diff 1.30, largest 4.7
	15	-	0.594	0.440	0.154	2.17	All st diffs <5
Overall (RM2000/S low or medium)	1,694	10,205	6.7%	5.4%	1.2%pts	1.73	Mean st diff 0.87, largest 3.3
	5	-	0.347	0.280	0.067	1.11	All st diffs <5
Overall (RM2000/S high or very high)	829	2,818	16.9%	13.7%	3.2%pts	1.72	Mean st diff 2.12, largest 8.1
	8	-	1.107	0.766	0.341	1.82	6 st diffs >5, 3 st diffs >6

Analysis	Treated number (matched, dropped)	Control number (total)	Binary & frequency outcome (treated)	Binary & frequency outcome (control)	Difference	T-value	Matching quality (standardised differences[1])
Overall (Comparison discharged from SOTP prison)	2,536	8,201	9.9%	8.4%	1.5%pts	2.02	Mean st diff 1.17, largest 5.9
	26	-	0.584	0.433	0.151	2.16	2 st diffs >5
Overall (Comparison not discharged from SOTP prison)	2,551	5,018	9.9%	7.9%	2.0%pts	2.13	Mean st diff 2.16, largest 8.8
	11	-	0.588	0.458	0.130	1.58	9 st diffs >5, 2 st diffs >7
Overall (2008-2012 index releases)	1,087	6,487	7.7%	7.0%	0.7%pts	0.70	Mean st diff 1.65, largest 5.7
	10	-	0.386	0.295	0.091	1.19	1 st diff >5
Overall (2008-2012 releases, control not on community SOTP)[4]	1,084	5,545	7.7%	6.6%	1.0%pts	0.99	Mean st diff 2.03, largest 6.6
	13	-	0.387	0.287	0.100	1.30	3 st diffs >5
Overall (Core & Extended SOTP)[5]	217	13,219	17.5%	10.0%	7.5%pts	2.71	Mean st diff 1.93, largest 5.1
	8	-	0.797	0.493	0.304	1.55	1 st diff >5
Overall (Core & Extended SOTP, no matching on OASys variables)[5]	217	13,219	17.5%	10.0%	7.5%pts	2.73	Mean st diff 1.78, largest 5.6
	8	-	0.797	0.495	0.302	1.55	1 st diff >5
Overall (Core & any additional SOTP)	816	13,219	10.0%	7.9%	2.2%pts	1.80	Mean st diff 1.43, largest 4.2
	6	-	0.529	0.423	0.107	1.05	All st diffs <5
Overall (Core without Extended SOTP)	2,325	13,219	9.3%	8.0%	1.3%pts	1.76	Mean st diff 1.22, largest 4.4
	12	-	0.574	0.439	0.135	1.91	All st diffs <5
Overall (Core without any additional SOTP)	1,735	13,219	9.9%	8.5%	1.5%pts	1.80	Mean st diff 0.87, largest 3.9
	5	-	0.618	0.450	0.169	2.12	All st diffs <5
Overall (treatment started up to 2 years before index release)	1,148	13,219	9.1%	8.1%	1.0%pts	1.04	Mean st diff 0.53, largest 1.7
	7	-	0.636	0.460	0.176	1.76	All st diffs <5
Overall (as above but only for those sentenced >=4yrs & <10yrs)	535	4,011	8.6%	6.3%	2.3%pts	1.64	Mean st diff 1.68, largest 6.7
	1	-	0.520	0.287	0.232	1.80	2 st diffs >5, 1 st diff >6
Overall (treatment started more than 2 years before index release)	1,378	13,219	10.8%	8.3%	2.5%pts	2.26	Mean st diff 2.03, largest 7.7
	29	-	0.552	0.451	0.101	1.09	7 st diffs >5, 2 st diffs >6

Notes:

1. Standardised (mean) differences were produced for each matching variable. A standardised difference <=5 indicates that the groups were closely matched, one of 5–10 indicates that there was a reasonable match quality, while one of >10% indicates a poor quality of matching.
2. The follow-up period excludes any prison spells (time from conviction (or sentencing) date to release date). So if an offender was released on 31 January 2004 and sent back to prison on 31 July 2005 and released again on 31 January 2006, the 2 year follow-up period will begin on 31 January 2004 and end on 31 July 2006.
3. To improve the matching quality, this comparison used radius matching (caliper = 0.2 * SD(logit of propensity score)).
4. To improve the matching quality, this comparison involved matching on the propensity score rather than on the logit of the propensity score.
5. An enlarged bandwidth of 0.36 was used for these comparisons.

Annex E
Additional survival analysis

In addition to the binary reoffending rate and the frequency of reoffending per offender, survival analysis was used to better assess reoffending rates over time. Survival analysis is well equipped to cater for individuals having variable lengths of time during which they are at risk of reoffending.[68] The survival charts shown by Figures 1-3 represent the matched treatment and comparison groups using our preferred kernel matching approach (see Annex C). While significance tests could be run using the Cox proportional hazards model, an assumption of this model is that the curves for the matched treatment and comparison groups are proportional. This assumption was often found to be incorrect, so a non-parametric Kaplan-Meier approach to statistical significance testing was used instead. In order to run the non-parametric significance tests using Stata, the treatment and comparison groups were matched using the nearest neighbour one treatment to exactly five comparison individuals' algorithm,[69] which was used to quantify the sensitivity of the results to unobserved factors.

Although the different matching approaches yielded similar results where comparisons could be made, there were a few differences that should be taken into account when interpreting the survival analysis results found using the nearest neighbour algorithm. In particular, the estimated 'child other' and 'soliciting' binary reoffending 'treatment effects' were statistically significant using the nearest neighbour algorithm, but not using our preferred Kernel approach.

The survival analysis performed after matching the treatment and comparison groups using the nearest neighbour algorithm was run with the following variants:

1) Censoring at the end of the follow-up period of reoffending risk for each individual.
2) Censoring at the earliest of either the end of the follow-up period or, for a comparison individual, any commencement of core SOTP during the follow-up period.

[68] This is through the use of censoring at the point the risk period ends.

[69] The non-parametric tests in Stata assume the weights of comparison individuals are equal, which is not so using the kernel matching algorithm. With the nearest neighbour 1-5 algorithm the unweighted results are the same as the weighted ones.

3) Censoring at the earliest of either the end of the follow-up period, any reimprisonment during the follow-up period[70], or, for a comparison individual, any commencement of core SOTP during the follow-up period.

Overall the findings (see Table A.11) reaffirm the conclusions of the binary reoffending outcomes, with statistically significant differences in sexual reoffending and child image reoffending between the matched treatment and control groups. Differences in child other reoffending remain on the cusp of statistical significance at the 5% level.

[70] At the point of receiving a custodial sentence.

Table A.11: Binary reoffending rates and survival analysis for the matched treatment and comparison groups using 1-5 nearest neighbour matching (treated N=2,493)

	Censoring at the end of the follow-up period										
	Treated rate	**Treated number**	**Control rate**	**Control number**	**Difference in binary rates**	**t-test p-value**	**Treated survival rate***	**Control survival rate***	**Difference in survival rates**	**Log-rank p-value****	**Peto p-value****
Total	40.0%	997	39.5%	4,926	0.5%pts	0.660	54.4%	54.0%	0.4%pts	0.909	0.695
Non-sex	24.5%	611	24.4%	3,046	0.1%pts	0.939	71.9%	71.0%	0.9%pts	0.781	0.639
All sexual[71]	28.2%	703	27.4%	3,419	0.8%pts	0.435	66.2%	67.6%	-1.4%pts	0.606	0.727
Sexual	10.1%	252	8.3%	1,031	1.8%pts	0.005	87.6%	90.6%	-3.0%pts	0.003	0.004
Adult serious	2.7%	67	2.7%	332	0%pts	0.946	96.4%	97.0%	-0.6%pts	0.965	0.976
Adult other	2.1%	53	1.6%	195	0.6%pts	0.070	97.4%	98.2%	-0.7%pts	0.047	0.048
Child contact	2.2%	56	2.0%	253	0.2%pts	0.502	97.5%	97.7%	-0.2%pts	0.492	0.490
Child image	4.5%	111	2.8%	355	1.6%pts	0.000	94.7%	96.7%	-2.0%pts	0.000	0.000
Child other	0.6%	14	1.0%	127	-0.5%pts	0.009	99.3%	98.9%	0.4%pts	0.031	0.030
Breach	22.2%	553	22.5%	2,801	-0.3%pts	0.752	73.5%	72.7%	0.8%pts	0.596	0.564
Soliciting	0.3%	7	0.1%	12	0.2%pts	0.092	99.5%	99.9%	-0.4%pts	0.018	0.018
Non-sex violent	3.7%	93	4.5%	560	-0.8%pts	0.071	95.5%	94.5%	1.0%pts	0.085	0.088
Non-sex non-violent	23.3%	582	23.7%	2,951	-0.3%pts	0.723	73.0%	71.8%	1.2%pts	0.493	0.389

Notes:

* At 13 years and 11 months.

** p-values are shown for both the log-rank test and the Peto test. The log-rank test is equally sensitive to events over time, while the Peto test is more sensitive to earlier events.

[71] This includes breaches and also soliciting/prostitution offences (which are not currently classified as sexual offences).

Table A.11: continued

	Censoring at the earliest of either the end of the follow-up period or, for a comparison individual, commencement of core SOTP					Censoring at the earliest of either the end of the follow-up period, reimprisonment, or, for a comparison individual, commencement of core SOTP				
	Treated survival rate*	Control survival rate*	Difference in survival rates	Log-rank p-value**	Peto p-value**	Treated survival rate*	Control survival rate*	Difference in survival rates	Log-rank p-value**	Peto p-value**
Total	54.4%	53.9%	0.4%pts	0.879	0.669	54.2%	53.9%	0.3%pts	0.898	0.690
Non-sex	71.9%	71.0%	0.9%pts	0.757	0.616	72.6%	71.9%	0.7%pts	0.488	0.395
All sexual[72]	66.2%	67.5%	-1.3%pts	0.629	0.752	67.5%	68.9%	-1.5%pts	0.433	0.505
Sexual	87.6%	90.6%	-3.0%pts	0.003	0.004	88.3%	91.5%	-3.2%pts	0.001	0.002
Adult serious	96.4%	97.1%	-0.7%pts	0.909	0.921	96.4%	97.7%	-1.3%pts	0.334	0.344
Adult other	97.4%	98.2%	-0.8%pts	0.039	0.040	98.1%	98.4%	-0.3%pts	0.347	0.349
Child contact	97.5%	97.7%	-0.2%pts	0.522	0.520	97.4%	97.8%	-0.3%pts	0.236	0.235
Child image	94.7%	96.7%	-2.0%pts	0.000	0.000	94.7%	96.6%	-2.0%pts	0.000	0.000
Child other	99.3%	98.9%	0.4%pts	0.039	0.039	99.5%	99.0%	0.5%pts	0.018	0.018
Breach	73.5%	72.7%	1.2%pts	0.589	0.555	74.5%	73.9%	0.6%pts	0.675	0.662
Soliciting	99.5%	99.9%	-0.4%pts	0.020	0.020	99.5%	99.8%	-0.3%pts	0.064	0.064
Non-sex violent	95.5%	94.7%	0.7%pts	0.145	0.149	96.2%	95.9%	0.2%pts	0.593	0.602
Non-sex non-violent	73.0%	71.7%	1.4%pts	0.447	0.351	73.7%	72.7%	1.0%pts	0.306	0.245

Notes:

* At 13 years and 11 months.

** p-values are shown for both the log-rank test and the Peto test. The log-rank test is equally sensitive to events over time, while the Peto test is more sensitive to earlier events.

[72] See footnote 71