Ministry of Justice

Evaluating Circles of Support and Accountability (CoSA) in England and Wales

Justice Data Lab impact evaluation of proven reoffending

Justice Data Lab

Ministry of Justice

July 2026

1. Summary

1.1 About Circles of Support and Accountability

Circles of Support and Accountability (CoSA) provides a community-based approach to the support and management of people convicted of sexual offences across England and Wales. CoSA targets individuals at medium, high or very high risk of reoffending and aims to reduce reoffending and encourage personal accountability.

1.2 Evaluation aims and methodology

Based on the methodology proposed in the feasibility study published in October 2024 (Evaluating Circles of Support and Accountability (CoSA) in England and Wales: Feasibility study for an impact evaluation of proven reoffending), this evaluation aims to assess the impact of CoSA on proven general and sexual reoffending within a five-year follow-up period.

To assess programme impact, the offences included in the reoffending analyses over a five-year period are as follows:

- General reoffending: This includes all proven reoffences, in line with standard Justice Data Lab (JDL) evaluations.
- Sexual reoffending: This only includes offence codes categorised as 'sexual' by the Sexual Reoffending Predictor [1] (SRP).

As CoSA aims to improve outcomes related to sexual reoffending, offences codes classed as sexual offences by the SRP were used to derive the sexual reoffending measure. This list of offence codes was used because it incorporates additional offences not included within the Home Office Sexual offences group such as 'Indecent images' and 'Sexual offending order breaches'.

To assess the effectiveness of completing CoSA, a sub-analysis was undertaken looking at those who completed the programme. Completion status was defined as: the individual's participation lasted for 9 months or longer, or less than 9 months but termination was recorded as 'planned'.

Unless otherwise stated, the headline analysis and quoted statistics in this report refer to the overall general reoffending measure.

[1] The Sexual Reoffending Predictor is an actuarial risk measure, which predicts the likelihood of sexual reoffending. It was known as the OASys Sexual Reoffending Predictor (OSP) until 23 February 2026. See the Glossary (Annex H) for more information on OSP and SRP.

This analysis involved a treatment group of 533 male adults who began receiving support from CoSA sometime between 2005 and 2020 and this was compared to a matched comparison group of 47,142 male adults who did not participate in the programme. Propensity score matching (PSM) was used to ensure comparable treatment and comparison groups.

Due to the small number of females who began receiving support from CoSA sometime between 2005 and 2020 (fewer than 20), it was not possible to produce an analysis of their outcomes.

The impact of CoSA was evaluated against three key general and sexual proven reoffending metrics over a five-year follow-up period from intervention start date (or release date for participants who began in custody):

1. Binary measure of reoffending (reoffending rate) – the percentage of individuals in the group who commit at least one proven reoffence.
2. Frequency of reoffences committed – the number of proven reoffences committed per person within the group.
3. Time to first reoffence – the average number of days from the start of the follow-up period to the date of the first proven reoffence, calculated only for those who reoffend.

Additional outcomes were also analysed, relating to reoffence severity by court outcomes and custodial sentences.

1.3 Summary of findings for the three key outcome measures

Over a five-year follow-up period, the overall results for the key outcome measures did not show a statistically significant effect on a person's general reoffending behaviour or sexual reoffending behaviour.

The key outcome measures of a sub-analysis for those who completed CoSA also did not show a statistically significant effect on a person's general or sexual reoffending behaviour.

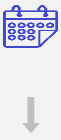
These results indicate that there is no clear evidence to confidently conclude that CoSA did or did not have an impact on reoffending. A non-significant result does not show that the intervention was unsuccessful in reducing reoffending. Instead, it indicates that there is no clear evidence to determine whether there is a difference in reoffending rates between the treatment group and the comparison group.

2. Headline results – general reoffending

The overall results do not show a statistically significant effect on a person's general reoffending behaviour.

The headline analysis in this report measured proven reoffences in a five-year period for a treatment group of 533 male offenders who began receiving support sometime between July 2005 and August 2020, and for a much larger comparison group (47,142) of similar offenders who did not receive it. The analysis estimates the impact of receiving support from Circles of Support and Accountability (CoSA) on general reoffending behaviour.

Overall measurements of the treatment and comparison groups: General reoffending

For **100** typical men in the **treatment** group, the equivalent of:		For **100** typical men in the **comparison** group, the equivalent of:
38 of the 100 men committed any proven reoffence within a five-year period (a rate of 38%), **1 man more** than in the comparison group.		**37** of the 100 men committed a proven reoffence within a five-year period (a rate of 37%).
168 proven reoffences were committed by these 100 men during the five year period (a frequency of 1.7 reoffences per person), **11 reoffences more** than in the comparison group.		**157** proven reoffences were committed by these 100 men during the five year period (a frequency of 1.6 reoffences per person).
546 days was the average time before a reoffender committed their first proven reoffence, **36 days earlier** than the comparison group.		**582** days was the average time before a reoffender committed their first proven reoffence

Overall estimates of the impact of the intervention: General reoffending

For **100** typical men who receive support, compared with **100** similar men who do not:

The number of men who commit a proven reoffence within five years after release could be **lower by as many as 3 men, or higher by as many as 5 men**. This is not a statistically significant result.

The number of proven reoffences committed during the five year period could be **lower by as many as 19 reoffences, or higher by as many as 40 reoffences**. This is not a statistically significant result.

On average, the time before an offender committed their first proven reoffence could be **shorter by as many as 98 days, or longer by as many as 26 days**. This is not a statistically significant result.

These estimates refer to the confidence intervals around the difference in means between the treatment and comparison groups. The measurements of the treatment and comparison groups are the best estimates for each group, because the groups are samples of larger populations, and so are subject to random error. Therefore, the observed difference in means between the two groups is also subject to random error. A 95% confidence interval provides the upper and lower bounds in which there can be 95% confidence that the real rate for each group lies. Wider confidence intervals mean that there is a wider range that the true value could be in, reflecting greater uncertainty in the estimate.

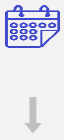
For example, the general reoffending rate for the post-matched treatment group was 38 reoffenders per 100 people (38%) and the comparison group was 37 reoffenders per 100 people (37%), a difference of 1 reoffender. The confidence intervals demonstrate 95% confidence that the true measurement of the treatment group falls within 3 reoffenders lower than, or 5 reoffenders higher than, the comparison group (per 100 people).

3. Headline results – sexual reoffending

The overall results do not show a statistically significant effect on a person's sexual reoffending behaviour.

The headline analysis in this report measured proven sexual reoffences in a five-year period for a treatment group of 533 male offenders who began receiving support sometime between July 2005 and August 2020, and for a much larger comparison group (51,875) of similar offenders who did not receive it. The analysis estimates the impact of receiving support from Circles of Support and Accountability (CoSA) on sexual reoffending behaviour.

Overall measurements of the treatment and comparison groups: Sexual reoffending

For **100** typical men in the **treatment** group, the equivalent of:		For **100** typical men in the **comparison** group, the equivalent of:
33 of the 100 men committed any proven sexual reoffence within a five-year period (a rate of 33%), **1 man more** than in the comparison group.		**32** of the 100 men committed a proven sexual reoffence within a five-year period (a rate of 32%).
127 proven sexual reoffences were committed by these 100 men during the five year period (a frequency of 1.3 sexual reoffences per person), **13 sexual reoffences more** than in the comparison group.		**114** proven sexual reoffences were committed by these 100 men during the five year period (a frequency of 1.1 sexual reoffences per person).
580 days was the average time before a reoffender committed their first proven sexual reoffence, **16 days earlier** than the comparison group.		**596** days was the average time before a reoffender committed their first proven sexual reoffence

Overall estimates of the impact of the intervention: Sexual reoffending

For **100** typical men who receive support, compared with **100** similar men who do not:

The number of men who commit a proven sexual reoffence within five years after release could be **lower by as many as 3 men, or higher by as many as 5 men.** This is not a statistically significant result.

The number of proven sexual reoffences committed during the five year period could be **lower by as many as 12 sexual reoffences, or higher by as many as 37 sexual reoffences.** This is not a statistically significant result.

On average, the time before an offender committed their first proven sexual reoffence could be **shorter by as many as 85 days, or longer by as many as 52 days.** This is not a statistically significant result.

4. Charts of key reoffending measures

The figures in this section present the key measures of reoffending for the treatment and comparison groups. Figures 1 and 2 show the five-year proven reoffending rate, figures 3 and 4 show the proven reoffending rate frequency, and figures 5 and 6 show the average days to first proven reoffence.

When interpreting the results, please refer to Section 7, Annex B and Annex G for details and key limitations of how this evaluation was conducted. A glossary of terms is also included in Annex H.

Figure 1: Five-year proven general reoffending rate for males after support from CoSA

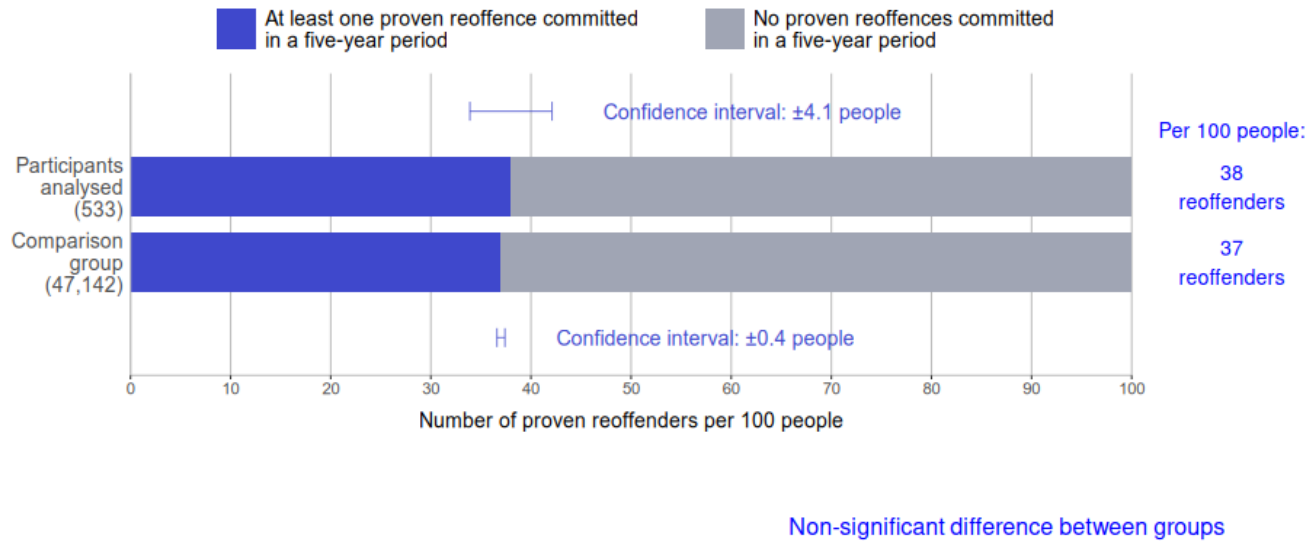


Figure 2: Five-year proven sexual reoffending rate for males after support from CoSA

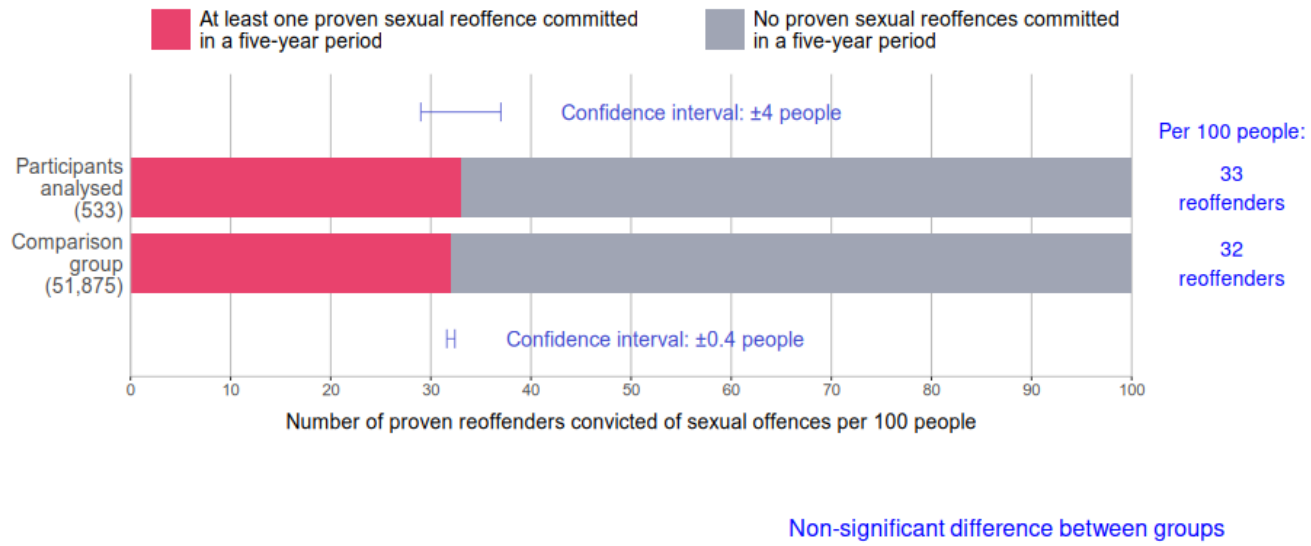


Figure 3: Five-year proven general reoffending frequency for males after support from CoSA

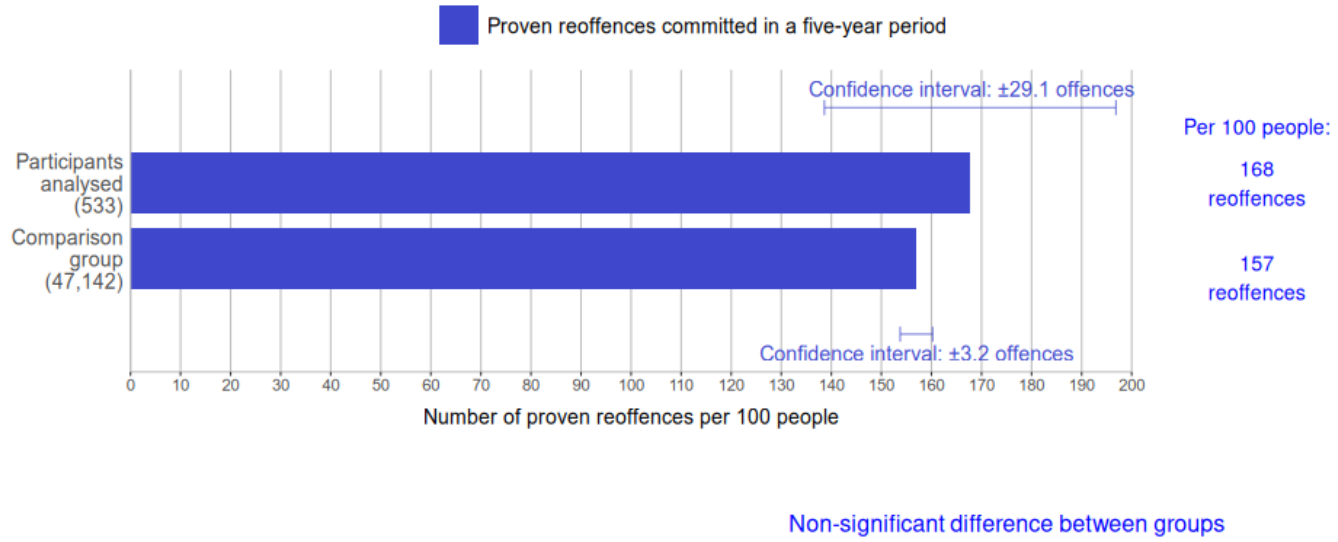


Figure 4: Five-year proven sexual reoffending frequency for males after support from CoSA

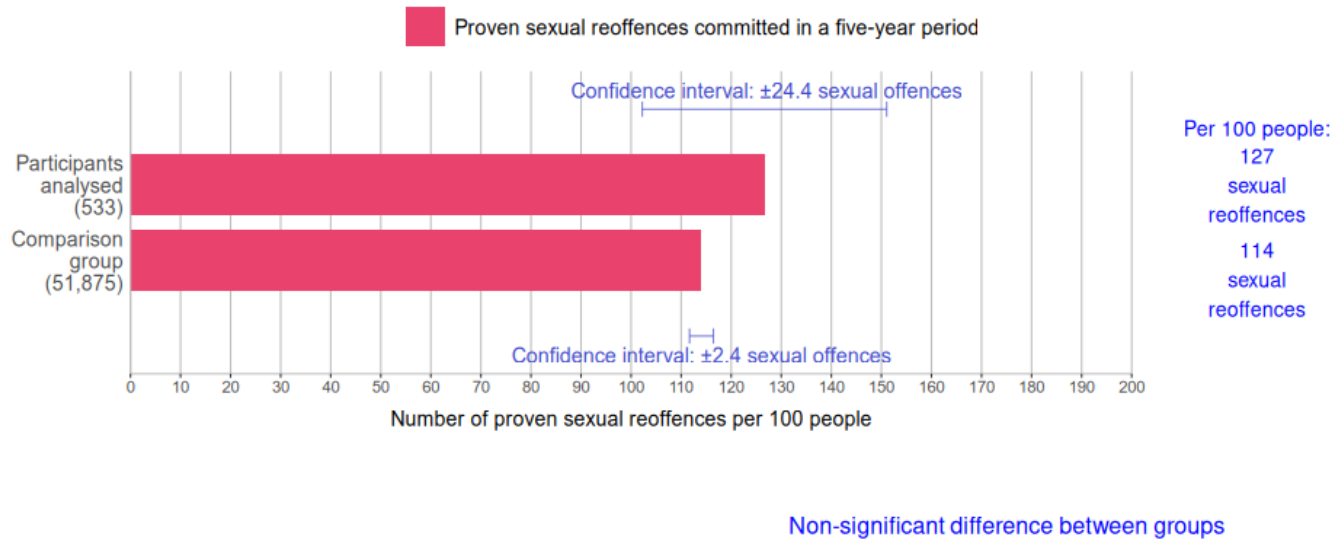
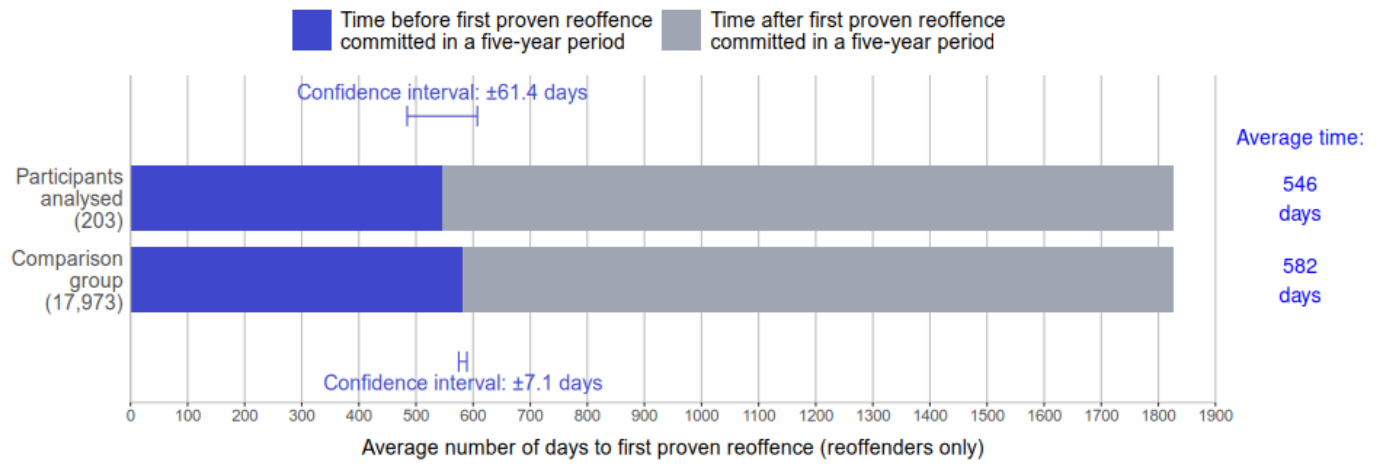
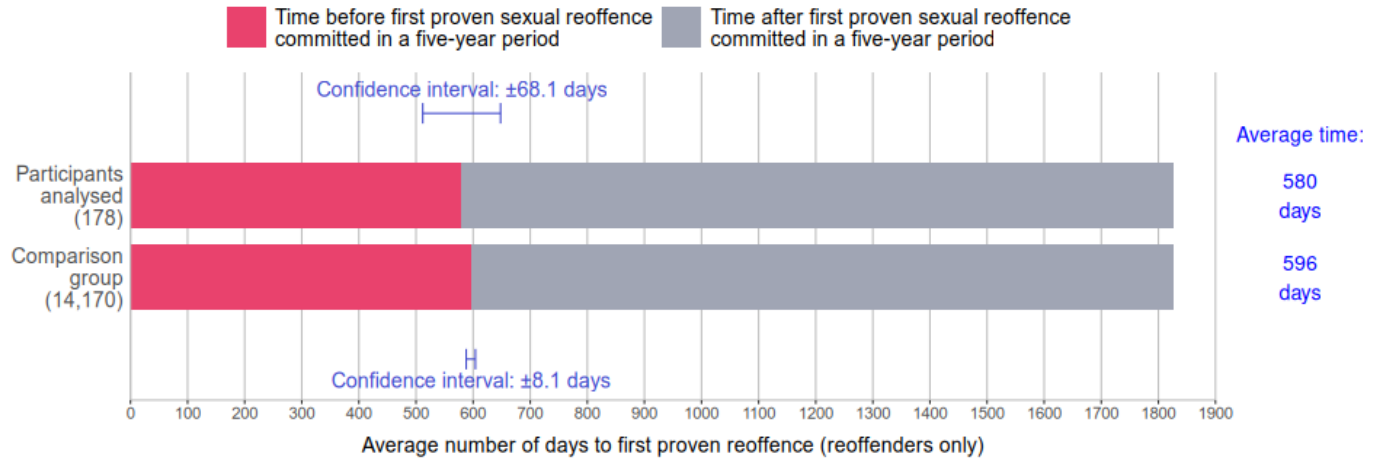


Figure 5: Average time (days) to first proven general reoffence for males after support from CoSA



Non-significant difference between groups

Figure 6: Average time (days) to first proven sexual reoffence for males after support from CoSA



Non-significant difference between groups

5. Circles of Support and Accountability intervention information

Circles of Support and Accountability (CoSA) provides a community-based approach to the support and management of people convicted of sexual offences. CoSA aims to reduce reoffending and help individuals convicted of sexual offences in taking accountability for their actions.

CoSA was established in Canada in 1994 to support the safe reintegration of a high-risk offender with a long history of sexually abusing children, who was leaving prison without supervision. CoSA expanded across Canada and into the USA before reaching the UK in 2002, when it was set up by the Religious Society of Friends (the Quakers). In 2008, Circles UK was launched as the national umbrella organisation for the development and delivery of CoSA provision.

Currently, CoSA is available across much of England and Wales. CoSA targets individuals who have committed sexual or sexually motivated offences against children and/or adults. CoSA is suitable for individuals with a medium, high or very high risk of reoffending, and people convicted of contact and non-contact sexual offences.

CoSA operates by creating a support network around the Core Member (an individual convicted of a sexual offence). A group of four to six trained volunteers from the local community form a ‘Circle’ around the Core Member to encourage them to take responsibility of their ongoing risk and behaviour, by offering support and practical guidance, such as developing social skills, securing suitable accommodation and finding appropriate hobbies, interests and work. These efforts aim to improve the Core Member’s self-esteem and confidence, strengthen their social connections and facilitate their safe reintegration into the community, ultimately reducing their risk of reoffending.

See the Circles UK website for more information on the CoSA programme.

6. Feasibility study (2024) overview and recommendations

In 2021, His Majesty's Prison and Probation Service (HMPPS) requested that the Ministry of Justice's (MoJ) Justice Data Lab (JDL) conduct an impact evaluation of Circles of Support and Accountability (CoSA) in England and Wales. Because of the complexities of evaluating the reoffending behaviour of those convicted of sexual offences, the JDL conducted a study to assess the feasibility of conducting that impact evaluation, ensuring that it is robust and that resources are used effectively.

The JDL published the feasibility study (Evaluating Circles of Support and Accountability (CoSA) in England and Wales: Feasibility study for an impact evaluation of proven reoffending) in 2024, concluding that an impact evaluation of CoSA was feasible. The study included a review of relevant literature, previous evaluations of CoSA, and exploratory analyses on a provisional CoSA treatment group dataset.

The feasibility study recommended propensity score matching (PSM), a statistical method that creates comparable treatment and comparison groups by accounting for differences in their characteristics, thereby allowing for an assessment of the intervention's effectiveness. PSM forms the basis of the standard JDL methodology, and the JDL therefore have extensive experience in using PSM.

The study explored several research questions and provided specific analytical recommendations for a CoSA impact evaluation. A summary of the analytical recommendations can be found in Annex F.

The methodology of the current evaluation is based on this 2024 feasibility study. Recommendations from the feasibility study were adhered to where possible, however further inspection of datasets resulted in some changes. Details of where the current evaluation deviates from the feasibility study can be found in Annex F.

7. Summary of methodology

This evaluation aims to assess the impact of Circles of Support and Accountability (CoSA) on proven general and sexual reoffending outcomes in a five-year follow-up period. The study includes individuals who began receiving support from CoSA sometime between 2005 and 2020. Men and women would be analysed separately due to known differences in reoffending behaviour. In this evaluation, only male participants were analysed because the number of female participants was insufficient for analysis.

7.1 Outcome measures

CoSA aims to improve outcomes in adults at medium or above risk of sexual reoffending. To assess programme impact, reoffences were counted over a five-year follow-up period from intervention start date (or release date for participants who began in custody). The analysis included:

- General reoffending: This includes all proven reoffences, as per the standard Justice Data Lab (JDL) method which uses general reoffending outcomes to estimate the impact of the intervention.
- Sexual reoffending: This only includes offence codes categorised as sexual by the Sexual Reoffending Predictor (SRP).

A five-year follow-up period to measure reoffending was chosen because sexual offenders typically exhibit lower reoffending rates, necessitating a longer follow-up period to increase the likelihood of detecting treatment effects, as outlined in the feasibility study.

As CoSA aims to improve outcomes related to sexual reoffending, offences codes classed as sexual offences by the SRP were used to derive the sexual reoffending measure. The SRP is an actuarial risk measure, which predicts the likelihood of sexual reoffending. This list of offence codes was used because it includes 284 separate offences, incorporating additional offences not included within the Home Office Sexual offences group such as ‘Indecent images’ and ‘Sexual offending order breaches’, which are categorised by the Home Office as ‘Miscellaneous crimes against society’ and ‘Public order offences’, respectively. The SRP does not include some offences that are included in the Home Office ‘Sexual offences’ category, such as prostitution offences.

In line with the Ministry of Justice’s proven reoffending statistics, the reoffending measures counted only new and separate offences as a reoffence. This includes breach offences that are

classed as new and separate offences, such as those included in the SRP category 'Sexual offending order breaches'. All other breaches, for example breaches of a court order and recalls to prison, were excluded (see Annex B.8 and the Variables and Offence Codes Excel annex accompanying this report for more information on offence codes used within the analysis).

Unlike standard proven reoffending, which measures from standard index date (see Glossary in Annex H), reoffending in this analysis was measured from intervention start date. As the comparison group did not have intervention start dates, pseudo-start dates were computed to enable their reoffending periods to be calculated (see Annex B.7 for more detail). For participants who began in custody, reoffending was measured from release date.

7.2 Completion of CoSA

It is important to account for differences in completion status because the effectiveness of CoSA may vary between individuals who complete the intervention as intended and those who do not. A sub-analysis consisting of participants who completed CoSA was therefore conducted, matching 'completers' to a 'no treatment' comparison group.

Circles UK provided data on programme termination for each participant. Completion status was defined as: the individual's participation lasted for 9 months or longer, or less than 9 months but termination was recorded as 'planned'.

The completion status definition was based on previous research that showed that the effects of CoSA on an individual's wellbeing can be seen at 9 months[2].

7.3 Applying eligibility criteria to the treatment and comparison groups

The following criteria were considered to select offenders for inclusion in the analysis. Offenders should:

- Be male and aged 18 or over when they started the intervention;
- Be convicted of an index offence categorised as a sexual offence by the Home Office or Sexual Reoffending Predictor (SRP), or an offence outside of this categorisation flagged as 'sexually motivated' by OASys[3];
- Have a previous conviction for an offence categorised as a sexual offence by the Home Office or an offence outside of this categorisation flagged as 'sexually motivated' by OASys;

[2] Winder, B., Blagden, N., Lievesley, R., Dwerryhouse, J., Kitson-Boyce, J., & Elliot, C. (2020). UK National Evaluation of Big Lottery Funded Circles of Support and Accountability: Evaluation Report March 2020, Nottingham Trent University.
[3] Offender Assessment System (OASys) provides offending-related risks and needs information

- Be deemed to be at medium or above risk of reoffending (assessed by SRP);
- Have acknowledged committing harmful sexual behaviour (assessed by OASys);
- Be under statutory supervision by probation and/or police;
- Be living within the community or within six months of release into the community.

See Annex B.5 and Annex C for more detail on how CoSA eligibility and suitability criteria were applied. For more information on proportions and matching quality of the variables used for eligibility and suitability criteria, see the Standardised Differences Excel annex accompanying this report.

7.4 Data used for the analysis

Person-level intervention data from CoSA was matched to the following datasets:

- Police National Computer (PNC) to provide criminal histories and to determine reoffending outcomes.
- Offender Assessment System (OASys) to provide offending-related risks and needs information. Two OASys records were chosen: one at the time of the index offence to measure characteristics of the offence, and one at the time of intervention start to measure offender characteristics.
- National Offender Management Information System (NOMIS) to identify whether individuals had also participated in other related accredited programmes.
- National Delius (nDelius) to identify participation in other related accredited programmes, as well as disposals, sentencing and release data.

The linking of these datasets produced a comprehensive suite of data for each CoSA participant. These individuals mainly served custodial or community sentences between 2005 and 2020. The majority (94%) started CoSA after leaving custody or during community sentences, while a small proportion (6%) started CoSA during their custodial sentence and continued upon release.

An additional comprehensive suite of data was extracted for a comparison group of offenders with similar characteristics, serving custodial or community sentences during the same period, who met the eligibility and suitability criteria for participating in the programme.

This suite of data, comprising 189 variables, was used as the basis for building a propensity score matching (PSM) model. Variables were included in the model to account for offender risks and needs, offending behaviour, participation route, completion status, social capital, and participation in additional programmes.

Variables were also included to account for offence-related sexual interests (also known as sexual deviancy and/or paraphilia). These variables were included to minimise bias due to offence-related sexual interests and ensure these characteristics were well balanced across the treatment and comparison groups. These variables covered both offending history and index offences, and characteristics such as victim age and gender, victim-perpetrator relationships, offence types, and extent of offending. See Annex E for more detail on the variables included in this analysis.

7.5 Propensity score matching (PSM)

PSM is a statistical matching technique that uses factors associated with both receiving the intervention and the outcome, to predict a 'propensity score' representing the likelihood of receiving the intervention conditional on these factors. Individuals in the 'treatment group' were matched to a 'comparison group' of similar individuals who did not receive treatment.

In addition, because this intervention is delivered to both those with a custody and community disposal, and given the differences between these sentence types, comparison group records with a pseudo-start date during a custodial sentence could only be matched to treatment group records who began CoSA during a custodial sentence. Similarly, comparison group records with a pseudo-start date that began during a community sentence or upon release from custody could only be matched to treatment group records who began CoSA during a community sentence or upon release from custody, respectively. This method helps ensure the groups are similar, reducing bias and allowing for a more accurate comparison of treatment effects.

Matching quality in JDL analyses is assessed using a traffic light scale (see Standardised Differences Excel annex accompanying the report). The standardised differences of the variables included in the models indicated that the matching quality was very good across all analyses (see Section B.10 and the Standardised Differences Excel annex accompanying this report for further detail).

The reoffending rates for the treatment and comparison groups were then compared, using statistical significance testing (t-tests). The rates are calculated using the weighted values for each person after matching. Three key reoffending outcomes were used to estimate the impact of the intervention in a five-year period, for both proven general and sexual reoffending:

1. Binary measure of reoffending (reoffending rate) – the percentage of individuals in the group who commit at least one proven reoffence.
2. Frequency of reoffences committed – the number of proven reoffences committed per person within the group.

3. Time to first reoffence – the average number of days from the start of the follow-up period to the date of the first proven reoffence, calculated only for those who reoffend.

Additional outcomes were also analysed, relating to reoffence severity by court outcomes and custodial sentences. See Section 8 for more information.

7.6 Interpreting the results

Both effect sizes (expressed in this report as a Cohen's d statistic) and whether the result is statistically significant (unlikely to be explained by chance alone) should be taken into consideration when interpreting the findings of this impact evaluation.

The difference in reoffending outcomes between the treatment and comparison groups is compared using statistical significance testing, which returns a 'p' value. In this report, the results are examined using the standard 0.05 significance level. If less than 0.05, the difference between the two groups is considered to be statistically significant and unlikely to be due to chance alone. The direction of the difference in reoffending rates indicates whether the treatment effect is positive or negative. The estimated differences are accompanied by associated 95% confidence intervals. If the 95% confidence interval includes zero then the result is not statistically significant.

While statistical significance indicates whether an observed difference is unlikely to be due to chance alone, it does not indicate the size or magnitude of the difference. Smaller sample sizes may have large observed differences but not be statistically significant due to a higher level of uncertainty which leads to wider confidence intervals. In contrast, large sample sizes are associated with a lower level of uncertainty and narrower confidence intervals, meaning a smaller observed difference can be found to be statistically significant.

This report includes the Cohen's d statistic to give an indication of the effect's magnitude and aid in the interpretation of the results.[4] These are typically categorised as:

- Small: a Cohen's d equal to or greater than 0.2 but below 0.5
- Medium: a Cohen's d equal to or greater than 0.5 but below 0.8
- Large: a Cohen's d equal to or greater than 0.8

The results set out in this report should be interpreted using a combination of: (a) whether the statistical tests meet the standard threshold for "statistical significance" by considering the p-value

[4] Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences. Routledge. ISBN 978-1-134-74270-7.

and (b) the “effect size” associated with that statistical test which, in these tables, is the Cohen’s d. Together, these describe whether there appear to be genuine differences between the CoSA treatment and comparison groups and the magnitude of that change.

Despite efforts to include all observed factors known to be predictive of selection onto CoSA and of reoffending risk into the PSM model, the importance of information that is not recorded cannot be known. As a result, there may be factors that are unobserved and unaccounted for which affect the results of this study. Other limitations include unknown/non-proven reoffending which is not included in the analysis. For further detail on methodology, see Annex B. A fuller list of limitations can be found in Annex G.

7.7 Survival analysis

Survival analysis, also known as time-to-event analysis, is a statistical method used to analyse the time until an event of interest occurs. During consultation for the feasibility study, it was recommended that the JDL explore conducting a supplementary survival analysis. Survival analysis using a Kaplan-Meier methodology was conducted in the evaluation of the Core Sex Offender Treatment Programme (SOTP),[5] which better accounted for the variable follow-up period used in that evaluation and allowed for adjustments to be made for restrictions on reoffending opportunities, such as being recalled to prison.

To enable a robust survival analysis to be conducted, further technical work would have been required which was outside the scope of this evaluation. It is therefore not included within this analysis. The Justice Data Lab, along with other teams within the Ministry of Justice, will continue to explore the applications of survival analysis in the evaluation of programmes to reduce reoffending.

Without the survival analysis, this evaluation does not directly consider a variable follow-up period, or the restriction on reoffending behaviour of recalls. Variables were derived to account for recall events between an individual’s index date and their start or pseudo-start date. These variables were well matched between the treatment and comparison groups. For more information, see Annex B.11 and Annex G.8, and the Standardised Differences Excel annex accompanying this report.

[5] Mews, A., Di Bella, L., and Purver,M. (2017). Impact evaluation of the prison-based Core Sex Offender Treatment Programme, Ministry of Justice.

8. Results in detail

The headline results in this report refer to the following:

1. **General male overall analysis:** treatment group matched to offenders across England and Wales using demographics, criminal history and individual risks and needs.
2. **Sexual male overall analysis:** treatment group matched to offenders across England and Wales using demographics, criminal history and individual risks and needs

These headline results controlled for offender demographics and criminal history and the following risks and needs: accommodation, employment history, financial history, relationships, mental health, thinking skills, drug and alcohol use, and attitudes towards offending. These were included to reflect the Circles of Support and Accountability (CoSA) eligibility and suitability criteria, and areas of need (see Annex B.5 and Annex C for more information)

The sizes of the treatment and comparison groups for reoffending rate and frequency analyses are provided below. To create a comparison group that is as similar as possible to the treatment group, each person within the comparison group is given a weighting proportionate to how closely they match the characteristics of individuals in the treatment group. The calculated reoffending rate uses the weighted values for each person and therefore does not necessarily correspond to the unweighted figures. The comparison group size differs across analyses because separate propensity score models were run, resulting in comparison group records being weighted differently.

In addition to the headline analyses, a sub-analysis consisting of participants who completed CoSA was also conducted. **Completion status was defined as: the individual's participation lasted for 9 months or longer, or less than 9 months but termination was recorded as 'planned'.**

Reoffence Type	Analysis	Treatment group size	Comparison group size	Reoffenders in treatment group	Reoffenders in comparison group (weighted number)
General	Overall	533	47,142	203	17,973 (17,440)
	Completers	315	39,851	103	14,390 (13,892)
Sexual	Overall	533	51,875	178	14,170 (16,573)
	Completers	314	37,991	89	10,084 (11,414)

Three key measures of five-year reoffending were analysed, as well as four additional measures, for general and sexual reoffending (see results in Tables 1-7):

1. Rate of reoffending
2. Frequency of reoffending
3. Time to first reoffence
4. Rate of first reoffence by court outcome
5. Frequency of reoffences by court outcome
6. Rate of custodial sentencing for first reoffence
7. Frequency of custodial sentencing

8.1 Significant results

3 measures show a statistically significant result. These provide significant evidence that for:

General Completers analysis

- Completers who reoffend with any offence within a five-year period **receive fewer custodial sentences** than non-participants who reoffend (Table 7).

Sexual Completers analysis

- Completers who reoffend within a five-year period are **less likely to receive a custodial sentence** for their first sexual proven reoffence than non-participants who reoffend (Table 6).

- Completers who reoffend with a sexual offence within a five-year period **receive fewer custodial sentences** than non-participants who reoffend (Table 7).

Although none of the headline general and sexual reoffending measures show a statistically significant result, a non-significant result does not show that the intervention was unsuccessful in reducing reoffending. Instead, it indicates that there is no clear evidence to determine whether there is a difference in reoffending rates between the treatment group and the comparison group. These results therefore indicate that there is no clear evidence to confidently conclude that CoSA did or did not have an impact on reoffending.

8.2 Tables of all reoffending measures

Tables 1 to 7 show the overall measures of reoffending. Rates are expressed as percentages and frequencies expressed per person. Tables 3 to 7 include reoffenders only, and are only shown where the total number of offenders in the treatment group is greater than 30. Tables 4 to 7 outline the additional outcome measures relating to reoffence severity by court outcomes and custodial sentences. Court and custodial outcomes are only shown if the number of offenders or offences in both the treatment and comparison groups is greater than 10 for that outcome.

In Tables 4 and 5, court outcomes of Indictable, Triable-either-way, and Summary are classifications of offences based on severity, with Indictable being the most severe and Summary the least. Due to the nature of the offending behaviour of this cohort, the most common reoffences for the treatment and comparison groups in both the general and sexual reoffending measures are sexual offences (as defined by the Sexual Reoffending Predictor, SRP). The majority of SRP sexual offences are classed as Triable-either-way or Indictable, therefore the majority of the reoffences here fall into those two categories. For more information on the severity classification of offences, see the Criminal Justice Statistics Quarterly: December 2025 (Offence Group Classification). See Annex D.2 for descriptive statistics on reoffence categories.

The standard acceptable level of statistical significance to demonstrate impact is 0.05. This means that for the difference between the treatment and comparison groups to be considered statistically significant (so unlikely to be due to chance alone), the p-values in the tables below must be 0.05 or lower.

The Cohen's d statistic gives an indication of the treatment effect's magnitude and aids in the interpretation of the results. These are typically categorised as: Small (a Cohen's d equal to or greater than 0.2 but below 0.5), Medium (a Cohen's d equal to or greater than 0.5 but below 0.8), or Large (a Cohen's d equal to or greater than 0.8). For more information, see Section 7.6 above. The standardised effect sizes across most of the analyses are categorised as below small.

Confidence intervals for the treatment group, comparison group and estimated difference are given in brackets. See the Glossary (Annex H) for more information.

Table 1: Proportion of men who committed a proven reoffence in a five-year period (reoffending rate) after support from Circles of Support and Accountability compared with a matched comparison group

Reoffence Type	Analysis	Number in treatment group	Number in comparison group	Treatment group rate (%)	Comparison group rate (%)	Estimated difference (% points)	Cohen's d	Significant difference? (p-value)
General	Overall	533	47,142	38 (34 to 42)	37 (37 to 37)	1 (-3 to 5)	0.02	No (0.61)
	Completers	315	39,851	33 (27 to 38)	35 (34 to 35)	-2 (-7 to 3)	-0.05	No (0.42)
Sexual	Overall	533	51,875	33 (29 to 37)	32 (32 to 32)	1 (-3 to 5)	0.03	No (0.48)
	Completers	314	37,991	28 (23 to 33)	30 (30 to 31)	-2 (-7 to 3)	-0.04	No (0.51)

Table 2: Number of proven reoffences committed in a five-year period (reoffending frequency - offences per person) by men who received support from Circles of Support and Accountability compared with a matched comparison group

Reoffence Type	Analysis	Number in treatment group	Number in comparison group	Treatment group frequency	Comparison group frequency	Estimated difference	Cohen's d	Significant difference? (p-value)
General	Overall	533	47,142	1.68 (1.39 to 1.97)	1.57 (1.54 to 1.6)	0.11 (-0.19 to 0.40)	0.03	No (0.47)
	Completers	315	39,851	1.24 (0.92 to 1.57)	1.49 (1.46 to 1.53)	-0.25 (-0.58 to 0.08)	-0.07	No (0.14)
Sexual	Overall	533	51,875	1.27 (1.02 to 1.51)	1.14 (1.12 to 1.16)	0.13 (-0.12 to 0.37)	0.05	No (0.31)
	Completers	314	37,991	0.90 (0.65 to 1.16)	1.07 (1.05 to 1.10)	-0.17 (-0.42 to 0.08)	-0.06	No (0.19)

Table 3: Average time (days) to first proven reoffence in a five-year period for men who received support from Circles of Support and Accountability, compared with a matched comparison group (reoffenders only)

Reoffence Type	Analysis	Number in treatment group	Number in comparison group	Treatment group time (days)	Comparison group time (days)	Estimated difference (days)	Cohen's d	Significant difference? (p-value)
General	Overall	203	17,973	546 (485 to 607)	582 (575 to 589)	-36 (-98 to 26)	-0.07	No (0.25)
	Completers	103	14,390	663 (577 to 748)	591 (583 to 599)	72 (-14 to 158)	0.15	No (0.10)
Sexual	Overall	178	14,170	580 (512 to 648)	596 (588 to 604)	-16 (-85 to 52)	-0.03	No (0.64)
	Completers	89	10,084	671 (576 to 766)	603 (594 to 613)	68 (-27 to 163)	0.14	No (0.16)

Table 4: Proportion of men supported by Circles of Support and Accountability with first proven reoffence in a five-year period (reoffending rate) by court outcome, compared with similar non-participants (reoffenders only)

Reoffence Type	Analysis	Number in treatment group	Number in comparison group	Court outcome	Treatment group rate (%)	Comparison group rate (%)	Estimated difference (% points)	Cohen's d	Significant difference? (p-value)
General	Overall	203	17,973	Either way	85 (80 to 90)	84 (83 to 84)	1 (-4 to 6)	0.04	No (0.61)
				Summary	9 (5 to 13)	11 (10 to 11)	-1 (-5 to 3)	-0.04	No (0.51)
	Completers	103	14,390	Either way	85 (79 to 92)	85 (84 to 85)	1 (-6 to 8)	0.02	No (0.81)
Sexual	Overall	178	14,170	Either way	98 (96 to 100)	98 (98 to 98)	0 (-2 to 2)	-0.004	No (0.96)
	Completers	89	10,084	Either way	97 (93 to 100)	98 (98 to 98)	-1 (-5 to 3)	-0.08	No (0.53)

Table 5: Number of proven reoffences in a five-year period (reoffending frequency) by court outcome for men supported by Circles of Support and Accountability, compared with similar non-participants (reoffenders only)

Reoffence Type	Analysis	Number in treatment group	Number in comparison group	Court outcome	Treatment group frequency	Comparison group frequency	Estimated difference	Cohen's d	Significant difference? (p-value)
General	Overall	203	17,973	Indictable	0.12 (0.03 to 0.20)	0.11 (0.10 to 0.12)	0.01 (-0.07 to 0.09)	0.01	No (0.83)
				Either way	3.73 (3.19 to 4.28)	3.50 (3.44 to 3.56)	0.23 (-0.31 to 0.78)	0.06	No (0.40)
				Summary	0.40 (0.24 to 0.57)	0.44 (0.42 to 0.46)	-0.04 (-0.21 to 0.13)	-0.03	No (0.67)
	Completers	103	14,390	Indictable	0.13 (0.04 to 0.22)	0.11 (0.10 to 0.12)	0.01 (-0.08 to 0.10)	0.02	No (0.75)
				Either way	3.14 (2.47 to 3.8)	3.65 (3.58 to 3.73)	-0.52 (-1.18 to 0.15)	-0.11	No (0.13)
				Summary	0.30 (0.07 to 0.54)	0.38 (0.36 to 0.4)	-0.08 (-0.32 to 0.15)	-0.06	No (0.49)
Sexual	Overall	178	14,170	Indictable	0.11 (0.01 to 0.20)	0.08 (0.07 to 0.09)	0.03 (-0.06 to 0.12)	0.05	No (0.54)
				Either way	3.69 (3.12 to 4.25)	3.49 (3.43 to 3.55)	0.19 (-0.37 to 0.76)	0.05	No (0.50)
	Completers	89	10,084	Indictable	0.11 (0.01 to 0.21)	0.08 (0.07 to 0.09)	0.04 (-0.06 to 0.13)	0.07	No (0.46)
				Either way	3.08 (2.40 to 3.76)	3.50 (3.43 to 3.57)	-0.42 (-1.10 to 0.26)	-0.11	No (0.22)

Table 6: Proportion of men and women who received a custodial sentence for their first proven reoffence after support from Circles of Support and Accountability, compared with similar non-participants (reoffenders only)

Reoffence Type	Analysis	Number in treatment group	Number in comparison group	Treatment group rate (%)	Comparison group rate (%)	Estimated difference (% points)	Cohen's d	Significant difference? (p-value)
General	Overall	203	17,973	46 (39 to 53)	44 (44 to 45)	1 (-6 to 8)	0.03	No (0.70)
	Completers	103	14,390	35 (26 to 44)	44 (43 to 45)	-9 (-19 to 0)	-0.19	No (0.051)
Sexual	Overall	178	14,170	50 (43 to 57)	51 (50 to 52)	-1 (-8 to 7)	-0.02	No (0.84)
	Completers	89	10,084	**40 (30 to 51)**	**52 (51 to 53)**	**-12 (-22 to -1)**	**-0.24**	**Yes (0.03)**

Table 7: Number of custodial sentences received in a five-year period by men and women who received support from Circles of Support and Accountability, compared to similar non-participants (reoffenders only)

Reoffence Type	Analysis	Number in treatment group	Number in comparison group	Treatment group frequency	Comparison group frequency	Estimated difference	Cohen's d	Significant difference? (p-value)
General	Overall	203	17,973	2.61 (2.10 to 3.12)	2.64 (2.57 to 2.7)	-0.03 (-0.54 to 0.49)	-0.01	No (0.92)
	Completers	103	14,390	**1.85 (1.27 to 2.44)**	**2.7 (2.62 to 2.77)**	**-0.84 (-1.43 to -0.25)**	**-0.18**	**Yes (0.01)**
Sexual	Overall	178	14,170	2.52 (1.98 to 3.05)	2.48 (2.42 to 2.55)	0.03 (-0.51 to 0.57)	0.01	No (0.9)
	Completers	89	10,084	**1.7 (1.15 to 2.24)**	**2.49 (2.42 to 2.57)**	**-0.79 (-1.34 to -0.24)**	**-0.21**	**Yes (0.01)**

Acknowledgements

From the Ministry of Justice, the authors would like to thank Aidan Mews.

The authors would also like to thank those involved in the external peer review process.

Final thanks are to His Majesty's Prison and Probation Service (HMPPS) and Circles UK for their communication and support throughout the evaluation process.

Contact points

Press enquiries should be directed to the Ministry of Justice press office.

https://www.gov.uk/government/organisations/ministry-of-justice/about/media-enquiries

Other enquiries about the analysis should be directed to:

Justice Data Lab team

E-mail: justice.datalab@justice.gov.uk

Official Statistics

Our statistical practice is regulated by the Office for Statistics Regulation (OSR). OSR sets the standards of trustworthiness, quality and value in the Code of Practice for Statistics that all producers of official statistics should adhere to.

You are welcome to contact us directly with any comments about how we meet these standards.

Alternatively, you can contact OSR by emailing regulation@statistics.gov.uk or via the OSR website.

Produced by the Ministry of Justice

Alternative formats are available on request from: justice.datalab@justice.gov.uk

Annex A: What you can and cannot say about these results

A.1 General reoffending

✓ **What you can say about the five-year reoffending rate:**

“This analysis does not provide clear evidence on whether support from CoSA increases or decreases the number of participants who commit a proven reoffence in a five-year period.

✗ **What you cannot say about the five-year reoffending rate:**

“This analysis provides evidence that support from CoSA increases/decreases/has no effect on the reoffending rate of its participants.”

✓ **What you can say about the five-year reoffending frequency:**

“This analysis does not provide clear evidence on whether support from CoSA increases or decreases the number of proven reoffences during a five-year period”

✗ **What you cannot say about the five-year reoffending frequency:**

“This analysis provides evidence that support from CoSA increases/decreases/has no effect on the number of proven reoffences committed during a five-year period by its participants.”

✓ **What you can say about the time to first reoffence:**

“This analysis does not provide clear evidence on whether support from CoSA shortens or lengthens the average time to first proven reoffence.”

✗ **What you cannot say about the time to first reoffence:**

“This analysis provides evidence that support from CoSA shortens/lengthens/has no effect on the average time to first reoffence for its participants.”

A.2 Sexual reoffending

✓ **What you can say about the five-year sexual reoffending rate:**

“This analysis does not provide clear evidence on whether support from CoSA increases or decreases the number of participants who commit a proven sexual reoffence in a five-year period.”

✗ **What you cannot say about the five-year sexual reoffending rate:**

“This analysis provides evidence that support from CoSA increases/decreases/has no effect on the sexual reoffending rate of its participants.”

✓ **What you can say about the five-year sexual reoffending frequency:**

“This analysis does not provide clear evidence on whether support from CoSA increases or decreases the number of proven sexual reoffences during a five-year period”

✗ **What you cannot say about the five-year sexual reoffending frequency:**

“This analysis provides evidence that support from CoSA increases/decreases/has no effect on the number of proven sexual reoffences committed during a five-year period by its participants.”

✓ **What you can say about the time to first reoffence:**

“This analysis does not provide clear evidence on whether support from CoSA shortens or lengthens the average time to first proven sexual reoffence.”

✗ **What you cannot say about the time to first reoffence:**

“This analysis provides evidence that support from CoSA shortens/lengthens/has no effect on the average time to first sexual reoffence for its participants.”

Annex B: Methodological approaches

B.1 Datasets used in the evaluation

Multiple datasets were used to produce a comprehensive set of data for Circles of Support and Accountability (CoSA) participants for analysis:

- Police National Computer (PNC) to provide criminal histories and to determine reoffending outcomes.
- Offender Assessment System (OASys) to provide offending-related risks and needs information.
- National Offender Management Information System (NOMIS) to identify whether individuals had also participated in other related accredited programmes.
- National Delius (nDelius) to identify participation in other related accredited programmes, as well as disposals, sentencing and releases data.

B.2 OASys data

Two OASys records were selected for each treatment and comparison group record. The first record was at the time of the sexual index offence, to identify offence characteristics such as victim age and gender, victim-perpetrator relationship, and escalation in seriousness of offending. The second record was around intervention start, to identify personal risks and needs at the time of starting their circle. For the comparison group, the pseudo intervention start dates were used to select the second OASys record (see B.7 for more information on creation of pseudo start dates).

Based on the feasibility study, the recommended OASys record selection at intervention start/pseudo-start date was as follows:

- For those who started CoSA in custody, select the most complete OASys record within the 12 months before or 1 month after intervention start date.
- For those who started CoSA through the gate (following release from custody) or in the community (during a non-custodial sentence), select the most complete OASys record within the 18 months before or 1 month after intervention start date.

As recommended in the feasibility study, individuals in the treatment and comparison groups without an OASys record around intervention start/pseudo-start, in the specified time periods outlined above, were removed from analysis. However, individuals with incomplete OASys records remained in the analysis to minimise bias and ensure methodological consistency with similar evaluations.

B.3 Data cleaning and PNC matching

Intervention data from CoSA provided date of birth, gender and Police National Computer ID (PNCID) as identifiers for matching. PNCID is a person-level identifier used on the PNC and other related databases. The data was cleaned according to standard Justice Data Lab (JDL) procedures, such as formatting dates and PNCIDs.

Records in the CoSA dataset were matched to corresponding records on the Police National Computer using matching criteria based on date of birth, gender and PNCID. Records without a match were dropped at this stage, and records with a match proceeded to the next stage. See Annex G.7 for more information on the limitations of matching across datasets.

B.4 Dealing with duplicate records and multiple participations

The dataset from CoSA included some duplicate records and individuals who had participated in CoSA twice. Exact duplicates (two rows with the same PNCID, date of birth, intervention start date and intervention end date) were assessed as a data error and the duplicate record was removed.

Multiple participations (two rows with the same PNCID and date of birth, but different intervention start and end dates) were included and treated as separate participations, if a distinct index offence was identified between participations. If no distinct index offence was identified between participations, the second participation was removed from analysis.

The frequency outcomes included in this report are expressed as frequency per person. Of the individuals in the treatment group, 99% had a single CoSA participation over the period of analysis, and 1% had two participations. To be precise, the frequency statistics quoted are per participation, rather than per person, but given the tiny number of those with multiple participations, per person and per participation are effectively interchangeable.

B.5 Accounting for CoSA eligibility and suitability criteria

Offenders in the CoSA dataset and the comparison group who were successfully matched to a record on the PNC were assessed against the CoSA eligibility criteria, see Annex C.1 for the full list.

The first stage of eligibility was a requirement that the offenders ‘have committed a sexual offence or a sexually motivated offence’.

The measures used to assess the first stage of eligibility were:

- Must have been convicted of an index offence which falls under the Home Office offence category of 'Sexual offences'[6], or
- Must have been convicted of an offence not within the Home Office offence category of 'Sexual offences' but has been flagged to be 'sexually motivated' on the individuals' OASys record.

For each offender in the treatment and comparison groups, offence records up to the point of intervention start were searched for an offence that falls under the Home Office offence category of 'Sexual offences'; and OASys records up to the point of intervention start were searched for a record that indicated an offence was 'sexually motivated'. Any offenders who were included in the CoSA dataset and matched to a record on the PNC but did not meet these criteria were removed from analysis. For more information, see Annex C.

The second stage of eligibility criteria was applied to the treatment and comparison groups to ensure that all offenders had acknowledged committing harmful sexual behaviour. This was assessed through OASys records either at the time of the index offence or at the time of starting the intervention. Any offenders in the treatment and comparison groups who had not scored 'Yes' on at least one of the following OASys items was removed from analysis:

- S2Q6 – Does the offender recognise the impact and consequences of offending on victim/community/wider society?
- S2Q11 – Does the offender accept responsibility for the current offence(s)?
- S12Q6 – Does the offender understand their motivation for offending?

The final two stages of eligibility criteria required offenders to be 'at medium or above risk of reoffending' and 'have statutory supervision by probation and/or police'. These criteria were accounted for in the propensity score matching (PSM) model but were not applied as filters on the treatment and comparison groups (see Table C1 for more information).

Suitability criteria (including CoSA areas of need, such as social isolation) were also accounted for in the PSM model but not applied as filters because participants may meet all, some or none of these criteria (see Table C2 for more information).

The variables mentioned above were sourced (for example, from OASys) and derived to reflect CoSA eligibility and suitability criteria. Individuals in the treatment and comparison groups were

[6] Details of Home Office offence classification can be found in the Criminal Justice Statistics Quarterly: December 2025 document (Offence Group Classification).

well matched on these variables, which indicates that the comparison group was reflective of the characteristics of those who took part in CoSA. See the Standardised Differences Excel annex accompanying the report for the proportions and matching quality of these variables (which are highlighted with a footnote).

B.6 Identifying the index offence and offending history

The index offence for this evaluation was defined as the closest primary sexual offence identified before intervention start. This was determined in two stages:

- Identifying an offence code categorised as a sexual offence by the Home Office or the Sexual Reoffending Predictor (SRP).
- Where an offence in these categories was not identified, offence codes outside of these categories were allowed if a corresponding OASys record indicated the offence was sexually motivated.

This method was used to ensure that the identified index offence related to sexual offending and participation in CoSA. This was also important for offence-related sexual interests variables, which aim to reflect characteristics of sexual offending for each participant, and focus on the index offence as well as offending history. A dataset including each participant's previous offences was created to derive the offending history offence-related sexual interests variables.

B.7 Imputation of pseudo intervention start dates

As individuals in the comparison group did not participate in CoSA, they did not have intervention start dates. To address this, pseudo intervention start dates were calculated. Pseudo-start dates were used as the start point for measuring reoffending and to select appropriate OASys records for analysis.

Pseudo-start dates estimate when an individual would likely have begun treatment based on observed offence-related characteristics, measured prior to treatment to minimise the influence of any treatment effects. Multiple imputation was used for this estimation, which involved creating a predictive model for a variable with missing data and then filling the missing values based on this model. Imputation of pseudo-start dates has been used in previous JDL evaluations, such as An impact evaluation of the prison-based Thinking Skills Programme (TSP) on reoffending. See Annex E for more detail on the variables included in this analysis.

B.8 Deriving general and sexual five-year reoffending outcome measures

A five-year follow-up period to measure reoffending was chosen because sexual offenders typically exhibit lower reoffending rates, necessitating a longer follow-up period to increase the likelihood of detecting treatment effects, as outlined in the feasibility study. The five-year period was measured from each offender's intervention start date, rather than date of conviction or release from custody (as in usual JDL evaluations). As outlined in the feasibility study, this is due to lengthy waiting time between conviction or release and starting the intervention (average 1.3 years). However, for individuals who started CoSA in custody, reoffending was measured from release date.

To estimate the impact of the intervention over the five-year follow-up period, two reoffending outcome measures were created:

- General reoffending – a standard JDL outcome measure that includes all proven reoffences, and
- Sexual reoffending – a bespoke outcome measure that only includes offence codes categorised as sexual by the Sexual Reoffending Predictor (SRP).

These reoffending measures align with the department-wide approach to measuring reoffending which involves counting only new and separate offences as reoffences, excluding breaches[7] and recalls to prison. The 2024 feasibility study explored the possibility of including recalls in the reoffending measures but concluded that developing a new reoffending outcome measure that incorporates breaches and recalls would likely be complex and resource-intensive compared to the potential insights gained.

Participants from both the treatment and comparison groups were included in the analysis only if they had completed the full reoffending period. This period was defined as five years from the intervention start (or release date for those who began in custody), plus an additional six months to allow cases to progress through the courts. At the time of analysis, data availability extended through March 2026; therefore, individuals whose reoffending start date was before September 2020 (more than five years and six months before March 2026) were excluded from the analysis.

Offences codes classed as sexual offences by the SRP were used to derive the sexual reoffending measure. This list of offence codes was used because it includes 284 separate

[7] There are lots of different types of breaches. For the purposes of proven reoffending, the breaches that are excluded relate to non-criminal breaches (such as breaches of community orders or suspended sentence orders). Other types of breaches, and most importantly in the context of this evaluation (such as Breaches of Sexual Harm Prevention Orders), are included within proven reoffending statistics and within the reoffences analysed in this evaluation.

offences and covers a wide range of sexual offence types. It incorporates extra offences not included within the Home Office Sexual offences group such as ‘Indecent images’ and ‘Sexual offending order breaches’, which are categorised by the Home Office as ‘Miscellaneous crimes against society’ and ‘Public order offences’, respectively. It does not include some offences that are included in the Home Office ‘Sexual offences’ category, such as prostitution offences.

B.9 Measuring offence-related sexual interests

A wide range of variables associated with offence-related sexual interests (also known as sexual deviancy and/or paraphilia) were included in the propensity score matching (PSM) model, to minimise any bias due to offence-related sexual interests. These variables covered both offending history and index offences, and characteristics such as victim age and gender, victim-perpetrator relationships, offence types, and extent of offending. A full explanation of how these variables were chosen is set out in A.2 of the feasibility study.

The final suite of data, comprising 189 variables, was used as the basis for building PSM models. See Annex E for more detail on the variables included in this analysis

B.10 Propensity score matching

Propensity score matching (PSM) is a statistical matching technique which uses factors theoretically and empirically associated with both receiving treatment and the outcome variable (reoffending) to predict a ‘propensity score’. This propensity score reflects the likelihood that an individual in custody received the intervention, given the recorded characteristics. Individuals in the treatment group were matched to similar individuals who did not receive treatment on the basis of the propensity scores.

Matching quality in JDL analyses is assessed using a traffic light scale. The mean absolute standardised differences across all variables in the regression models, within each analysis, were all less than 1%. Within each analysis, all variables in the models used to calculate propensity scores had absolute standardised differences of less than 5%. Of all the variables considered as potential inputs to the models, none had an absolute standardised difference of more than 10%, and less than 2% of the variables had an absolute standardised difference of between 5% and 10%. This indicates that the matching quality was very good across all analyses (see the Standardised Differences Excel annex accompanying this report for further detail).

The outcome measures for each participant were used to create the main reoffending outcome metrics for general and sexual reoffending, for the treatment and comparison group. The rates were calculated using the weighted values for each person after matching. The reoffending rates

for the treatment and comparison groups were then compared. Three reoffending outcomes were used to estimate the impact of the intervention in a five-year period:

- Binary measure of reoffending (reoffending rate) – the percentage of individuals in the group who commit at least one proven reoffence.
- Frequency of reoffences committed – the number of proven reoffences committed per person within the group.
- Time to first reoffence – the average number of days from the start of the follow-up period to the date of the first proven reoffence, calculated only for those who reoffend.

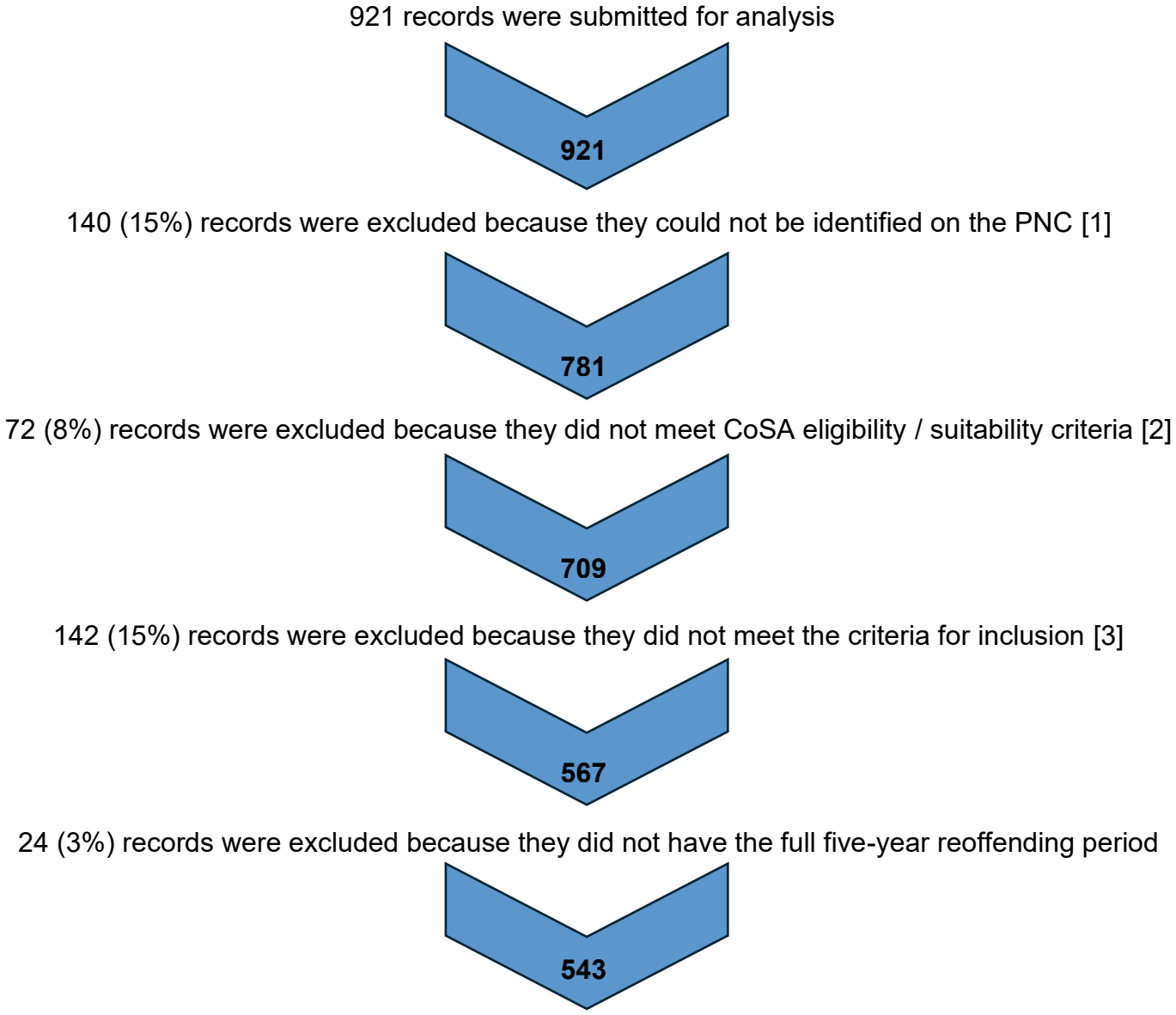
B.11 Survival analysis

In the feasibility study, the JDL committed to exploring the possibility of conducting a supplementary survival analysis for two reasons: firstly, to better accommodate variable follow-up periods for the treatment group, and secondly to allow for adjustments to restricted opportunities to reoffend when individuals were reimprisoned due to being recalled or receiving another custodial offence. Conducting the survival analysis was subject to resource and data quality constraints.

The Justice Data Lab scoped a proposed survival analysis based on the recommendations from the Working Group who fed into the development of the feasibility study. To enable a robust survival analysis to be conducted, further technical work would have been required which was outside the scope of this evaluation. It is therefore not included within this analysis.

B.12 Attrition from treatment group

Figure B1 Attrition from treatment group to create final pre-matched cohort



Final treatment group: 543 pre-matched records, 59% of records submitted

(Comparison group: 81,339 pre-matched records)

Notes

1. For data protection reasons, the names of CoSA participants were not submitted. Participants were identified on the PNC using PNC identifier (PNCID), date of birth and sex only.
2. The eligibility criteria required a previous sexual (Home Office) or sexually motivated (OASys) offence and acknowledgement of harmful sexual behaviour (see Annex B.5).
3. Criteria included being male, aged 18+ at intervention start date, relevant sexual index disposal, distinct disposal between multiple participations (see Annex B.4), OASys record within selection window (see Annex B.2).

The pre-matched treatment group refers to the group before propensity score matching (PSM) models were run, and treatment group records were matched to comparison group records. As such, the PSM model was run with 543 treatment group records, and 533 of those were matched to comparison group records to create the final post-matched treatment group (overall general reoffending analysis).

See Annex G.7 for more information about the implications and limitations of attrition from the treatment group.

B.13 Sensitivity analysis

A series of sensitivity analyses were run to measure possible effects of certain methodological decisions on the results. The principal aim of the sensitivity analyses was to test whether changing some of the assumptions or methods used would have a material impact on the key results.

Table B1 provides detail on each sensitivity analysis. In summary, the results of all sensitivity analyses were similar to the corresponding results presented in this report.

Table B1 Sensitivity analyses conducted and their findings

Sensitivity	Description/Explanation of alternative approach	Baseline: analysis chosen for comparison (including variables in final model and MASD)	Results/Findings/Comments
Matching approach: common support requirement	To explore the effect of having a restriction on common support. Treatment group members were automatically excluded where their propensity scores were outside the overall range of propensity scores of the comparison group, and comparison group members were automatically excluded where their propensity scores were outside the overall range of propensity scores of the treatment group.	Headline: general reoffending (final model has 150 variables and a mean absolute standardised difference (MASD) of 0.6%)	The results of the alternative approach were very similar to the baseline, showing any outliers have been appropriately excluded in the matching process of the headline analysis.
Matching approach: various matching rules	All analyses presented in this evaluation imposed a matching rule where treatment group records could only be matched to comparison group records if they shared the same participation route (custody, through-the-gate, or community). Various sensitivity analyses were run with different rules (e.g. remove all rules; apply the same rule to different variable(s)).	Headline: general reoffending (150 variables, MASD = 0.6%)	Whichever matching rule was tested, the results were very similar to the baseline, mainly because very good matching was observed across all variables that were considered (see the accompanying standardised differences Excel file). This shows that the use of these matching rules in the final model did not materially impact the results.

Sensitivity	Description/Explanation of alternative approach	Baseline: analysis chosen for comparison (including variables in final model and MASD)	Results/Findings/Comments
Matching approach: different caliper chosen	The PSM process matches individuals in the treatment and comparison groups where their propensity scores are within a given range of each other (known as the caliper). Where this caliper is quite wide (in order to improve the overall matching across all records), it is useful to check whether a narrower caliper would have resulted in materially different results. The analysis with the widest caliper was chosen as the baseline, and the caliper was then reduced significantly and results recalculated.	Completers: general reoffending (152 variables, MASD = 0.8%)	While the overall matching quality of this approach was worse, the results were very similar to the baseline, confirming that the final caliper selected for the baseline was appropriate.
Matching approach: different kernel (Epanechnikov)	This explored using an alternative type of kernel (a function used to determine weights assigned to comparison group records that are within the chosen caliper) matching sometimes used for PSM models.	Headline: general reoffending (150 variables, MASD = 0.6%)	The results were very similar to the baseline, which uses radius matching (with replacement) with a uniform kernel applied, confirming the baseline approach was appropriate.
Parsimonious model (fewer variables)	To reduce the risk of omitting confounding variables from regression modelling, JDL typically selects a large list of variables as potential inputs (189 variables for this evaluation). Such an approach can introduce 'noise' into the resulting model and lead to an 'overfitting' risk. This sensitivity analysis considered any differences to the findings that would have been observed had a model with fewer (104) potential inputs (less 'noise') been chosen instead.	Headline: sexual reoffending (151 variables, MASD = 0.6%)	The resulting parsimonious model comprised 87 variables (64 fewer than the baseline model). The matching quality of both the shared potential input variables and the final models were similar and the findings were very similar, confirming that any additional noise introduced in the baseline approach did not impact the results observed.

Sensitivity	Description/Explanation of alternative approach	Baseline: analysis chosen for comparison (including variables in final model and MASD)	Results/Findings/Comments
Model with squared variables	All analyses presented in this evaluation only included first order (linear) terms in relation to numeric variables as potential inputs to the regression models. There is the potential that the added inclusion of squared terms (quadratic effects) could better predict the likelihood of treatment assignment. A sensitivity analysis was run to test this variation in approach.	Parsimonious: sexual reoffending (87 variables, MASD = 0.7%)	The results from the resulting model (113 variables, MASD = 0.7%) were very similar to the baseline, confirming that the exclusion of squared terms from the regression models did not materially impact the results.
Sub-analysis: no previous sexual accredited programmes identified	Based on readily available data, the analyses presented in this evaluation were well-balanced in relation to the proportion who have participated in programmes aimed at sexual offenders, and how recently prior to their intervention start date these other participations happened. However, such an analysis cannot isolate the effect of CoSA alone, as there could be combined effects of participation in multiple programmes. This sensitivity analysis aimed to partially remove such combined effects, by only including those individuals in the treatment and comparison groups who were not observed as having participated in such programmes prior to their CoSA start (or pseudo start) date.	Headline: sexual reoffending (151 variables, MASD = 0.6%)	The results of this analysis were similar to the baseline. It is important to note that the cohort analysed in this sub-analysis excluded 37% of the pre-matched treatment group who had previously participated in a sexual offending accredited programme. Therefore, it may not be representative of the wider cohort used for this evaluation.

Sensitivity	Description/Explanation of alternative approach	Baseline: analysis chosen for comparison (including variables in final model and MASD)	Results/Findings/Comments
Model with different outcome: narrower measure of sexual reoffending	The outcome measure for sexual reoffending includes all reoffences described as sexual based on the SRP categorisation. This incorporates additional offences not included within the Home Office 'Sexual offences' group. The main differences, as described in this report, are that 'Indecent image' and 'Sexual offending order breaches' offences are excluded from the Home Office categorisation. This sensitivity recalculated the results of the headline sexual analysis using the Home Office definition of sexual offending, to illustrate how the results would differ had this version of the outcome measure been used instead.	Headline: sexual reoffending (151 variables, MASD = 0.6%)	While the sexual reoffending rates were much lower (owing to the much narrower definition of sexual offences), no significant results were observed across the three key outcome metrics, and the direction of treatment effects was the same as that observed in the wider definition of the sexual outcome measure presented in this evaluation.

Sensitivity	Description/Explanation of alternative approach	Baseline: analysis chosen for comparison (including variables in final model and MASD)	Results/Findings/Comments
Model with different outcome: no restriction on time to conviction	Sexual offences generally take longer to go through the courts (the period between the original offence and the conviction for that offence) than most other offence types. The analyses presented in this evaluation captured all reoffences in the five-year follow-up period, so long as they had been convicted within six months of the end of the follow-up period. This is in line with the follow-up methodology used for the published proven reoffending statistics. This sensitivity analysis investigated the impact of allowing for later convictions. It is worth noting that some convictions for reoffences may have happened after the available time period in the data, or are still going through the courts. Therefore, those with earlier participation dates will have had longer for convictions to take effect than those with later participations, potentially introducing bias into the results.	Headline: sexual reoffending (151 variables, MASD = 0.6%)	While the sexual reoffending rates were slightly higher (owing to some offences being included that took longer to go through the courts than was allowed in the headline analyses), the results observed were very similar to the baseline model.

Annex C: CoSA eligibility and suitability criteria

Individuals must meet specific eligibility criteria to be accepted onto Circles of Support and Accountability (CoSA). According to Circles UK guidelines, these criteria define personal characteristics which make an individual eligible for the intervention. CoSA guidelines also outline a set of suitability criteria to assess the potential benefits of the intervention for each individual.

This analysis used a set of measures to proxy eligibility and suitability criteria, which are outlined in Table C1 and Table C2. Some measures for eligibility and suitability criteria were used to filter the dataset, to ensure that only individuals in the treatment and comparison groups who met the criteria were included in analysis. Some measures were not used as filters on the dataset but were included in the propensity score matching (PSM) model.

C.1 Eligibility criteria

Table C1 outlines the eligibility criteria for all individuals accepted onto CoSA, along with the measures used to account for them.

Table C1: Eligibility criteria for the treatment and comparison groups

Eligibility criteria	Measure used in this analysis
Have committed a sexual offence or a sexually motivated offence.	Have a previous conviction for an offence categorised as a Sexual offence by the Home Office, or an offence outside of this category that is flagged as ‘sexually motivated’ on a corresponding Offender Assessment System (OASys) record. Have an index offence categorised as a Sexual offence by the Home Office or the Sexual Reoffending Predictor (SRP), or an offence outside of these categories that is flagged as ‘sexually motivated’ on a corresponding OASys record. This measure was applied as a filter on the dataset. Any individual who did not meet this criterion was removed from analysis.

Eligibility criteria	Measure used in this analysis
Acknowledge that he/she has committed harmful sexual behaviour.	Must score 'yes' on at least one of the following OASys items: • S2Q6 – Does the offender recognise the impact and consequences of offending on victim/community/wider society? (relating to the index offence) • S2Q11 – Does the offender accept responsibility for the current offence(s)? (relating to the index offence) • S12Q6 – Does the offender understand their motivation for offending? (relating to the point of intervention start) This measure was applied as a filter on the dataset. Any individual who did not meet this criterion was removed from analysis.
Are at medium or above risk of reoffending.	Must have a score of medium or above on the OASys Sexual Reoffending Predictor Contact (OSP-C) scale, or high on the OASys Sexual Reoffending Predictor Images (OSP-I) scale. This measure was not applied as a filter on the dataset. Although OSP scores were calculated for the current evaluation as recommended, individuals in the treatment or comparison groups were not removed from analysis based on these scores. The OSP was introduced to assess eligibility for CoSA in March 2021, however almost all (99%) of the cohort in this evaluation took part in CoSA before March 2021. However, scores were included in the PSM model to ensure treatment and comparison groups were well matched. See Annex F for more information on OSP scores.

Eligibility criteria	Measure used in this analysis
Have statutory supervision by probation and/or police.	Data from National Delius was used to identify whether individuals were still on licence/serving their community sentence in relation to their index offence, at their CoSA start date (or pseudo-start date). Where supervision had ended in relation to this disposal prior to the start date, the period from supervision end to start date was also calculated. These measures were not applied as filters on the dataset due to incomplete data being available to confidently determine the level of supervision at intervention start/pseudo-start date. However, the available data were included in the PSM model to help ensure balance between the groups.

C.2 Suitability criteria

Table C2 outlines the suitability criteria for all individuals accepted onto CoSA, along with the measures used to account for them. The suitability criteria are considered more flexible than the eligibility criteria – participants may meet all, some or none of these criteria. The following measures aim to account for this flexibility by balancing suitability criteria across both the treatment and comparison group.

The CoSA ‘areas of need’ measured by the OASys items in Table C2 are: Lack of employment/hobbies, Emotional loneliness, Lack of appropriate intimate relationships, Easily influenced by criminal associates, Social isolation, Low self-esteem, Impulsivity, and Poor problem-solving. These were all included as variables within the PSM model. There was no single OASys item to proxy Sexual preoccupation or Lack of pro social network for the time period considered for this evaluation, so these areas of need are not covered within the analysis.

For more information on proportions and matching quality of the variables used for eligibility and suitability criteria, see the Standardised Differences Excel annex accompanying this report.

Table C2: Suitability criteria for the treatment and comparison groups

Suitability criteria	Measure used in this analysis
Be living within the community or be within six months of release into the community.	The following variables were included in the PSM model to ensure balance between the post-matched groups, acting as proxies for this criteria: participation route (start in custody, through the gate, or on a community sentence), waiting times, whether the individual is (or was recently) a live probation case, or where data suggests that the individual is still in prison at intervention start.
Be motivated to stop their harmful sexual behaviour.	The OASys item 'S12Q8 Is the offender motivated to address the offending behaviour?' (relating to the point of intervention start), was included as a variable in the PSM model, rather than as a filter on the dataset.
Be prepared to engage with the intervention long-term.	The OASys item 'S13Q4 Understands the importance of completing programmes' (relating to the point of intervention start), was included as a variable in the PSM model, rather than as a filter on the dataset.
Display a CoSA 'area of need'.	The following OASys items (at the point of intervention start) were included as variables in the PSM model: • S4Q2 – Is the person employed? • S6Q1 – Current relationship with close family members • S6Q4 – Current relationship with partner • S6Q6 – Previous experience of close relationships • S7Q3 – Easily influenced by criminal associates • S10Q3 – Social isolation • S10Q4 – Offender's attitude to themselves • S11Q2 – Impulsivity • S11Q6 – Problem-solving skills

Annex D: Summary of descriptive statistics and the profile of the treatment group

D.1 Index offences

Table D1 shows the pre-matched treatment group's index offences by Sexual Reoffending Predictor (SRP) offence category, with the most common index offences being contact child (40%), indecent images (22%) and sexual offending order breaches (18%).

Table D1: Sexual Reoffending Predictor (SRP) offence categories for index offence, for pre-matched treatment group[8]

Sexual Reoffending Predictor (SRP) category description	Treatment group index offences	Proportion of index offences
Contact child	218	40%
Indecent images	118	22%
Sexual offending order breaches	97	18%
Contact adult	60	11%
Not sexual in statute	33	6%
Other noncontact	17	3%
Total	**543**	**100%**

Index offences categorised as 'Not sexual in statute' would be only included if they had an Offender Assessment System (OASys) flag which classified them as being sexually motivated.

Further descriptive statistics are available in the Descriptive Statistics Excel annex, including information on the most common index offences and a breakdown of index offence by Home Office offence group.

D.2 Reoffences

Tables D2 and D3 show all the reoffences of the matched treatment and comparison groups broken down by SRP offence category, from the general and sexual analysis, respectively. Different members of the treatment and comparison groups were matched across the general and sexual analyses, therefore the numbers across tables D2 and D3 may be inconsistent.

[8] This table is included in the Descriptive Statistics Excel annex published alongside this report, as Table A3. The figures relate to the pre-matched treatment group, so may not match those presented in D4 which contains the matched treatment group from the general analysis. The figures may also differ from those presented in the Standardised Difference Annex, which uses the corresponding matched treatment group for each analysis.

Table D2: Sexual Reoffending Predictor (SRP) offence categories of all general reoffences of the matched treatment and comparison group[9]

Sexual Reoffending Predictor (SRP) category description	Treatment group general reoffences	Proportion of general reoffences	Comparison group general reoffences (weighted)	Proportion of general reoffences (weighted)
Sexual offending order breaches	358	40%	338	40%
Not sexual in statute	220	25%	239	29%
Indecent images	217	24%	166	20%
Contact child	61	7%	50	6%
Other noncontact	21	2%	26	3%
Contact adult	15	2%	18	2%
Unknown	2	<1%	<1	<1%
Total	**894**	**100%**	**837**	**100%**

[9] This table is included in the Descriptive Statistics Excel annex published alongside this report, as Table A9. The figures relate to the matched treatment and comparison groups from the general analysis. Note that different members of the treatment and comparison groups were matched across the general and sexual analyses, therefore the numbers across tables D2 and D3 may be inconsistent. ‘c' denotes where numbers and proportions have been suppressed because there were 2 or fewer unweighted individuals in this category.

Table D3: Sexual Reoffending Predictor (SRP) offence categories of all sexual reoffences of the matched treatment and comparison group[10]

Sexual Reoffending Predictor (SRP) category description	Treatment group sexual reoffences	Proportion of sexual reoffences	Comparison group sexual reoffences (weighted)	Proportion of sexual reoffences (weighted)
Sexual offending order breaches	359	53%	341	56%
Indecent images	217	32%	177	29%
Contact child	61	9%	46	8%
Other noncontact	23	3%	25	4%
Contact adult	15	2%	18	3%
Total	**675**	**100%**	**608**	**100%**

[10] This table is included in the Descriptive Statistics Excel annex published alongside this report, as Table A14. The figures relate to the matched treatment and comparison groups from the sexual analysis. Note that different members of the treatment and comparison groups were matched across the general and sexual analyses, therefore the numbers across tables D2 and D3 may be inconsistent.

As shown in table D3, the SRP category of 'Sexual offending order breaches' represented the majority of all sexual reoffences in both the matched treatment group (53%) and matched comparison group (56%). This category includes offences such as 'Breaches of Sexual Harm Prevention Orders (SHPOs)', 'Sexual Offences Prevention Orders (SOPOs)', which SHPOs replaced, and breaches due to 'Failure to comply with notification requirements' as set out in the Sexual Offences Act 2003.

The majority of 'Sexual offending order breaches' and all 'Indecent image' offences are not included in the Home Office offence group classification of a sexual offence. This means that the sexual reoffending outcome measure in this report may be inconsistent with other measures of sexual reoffending which use the Home Office offence groups, such as Proven reoffending statistics and the evaluation of SOTP[11]. For a full breakdown of the classification of offences, see the accompanying Variables and Offence Codes Excel annex.

Further descriptive statistics are available in the accompanying Descriptive Statistics Excel annex, including the most common general and sexual reoffences, and further breakdowns of the number and proportion of general and sexual reoffences by Home Office offence group and SRP offence categories.

D.3 Profile of the treatment group

Table D4 provides an overview of the characteristics of the 533 male treatment group offenders included in the general headline analysis. The table contains information on demographics, offence history, and Circles of Support and Accountability (CoSA) participation route. For specific statistics on the make-up of each distinct analysis based on the matched treatment and comparison groups, please see the accompanying Standardised Differences Excel annex. Analysis of available demographic information from the PNC and Delius did not suggest that the post-matched treatment group was materially different to the original cohort of participants provided by Circles UK across age, ethnicity and nationality, for those who could be identified (see Annex G.7).

Further descriptive statistics are also available in the Descriptive Statistics Excel annex, including information on age, recalls, and participation in accredited programmes.

[11] Mews, A., Di Bella, L., and Purver,M. (2017). Impact evaluation of the prison-based Core Sex Offender Treatment Programme, Ministry of Justice.

Table D4: Demographic information for the PSM-matched male treatment group

Variable	Frequency (%)[12] (or mean where stated)
Age and sex	
Mean age at intervention start	44
Male	100%
Ethnicity	
White	95%
Asian	1%
Black	c%
Other	c%
Unknown	3%
Nationality	
UK nationality	94%
Foreign nationality	1%
Unknown nationality	5%
Index disposal	
Prison	69%
Youth Rehabilitation Order	18%
Suspended sentence	17%
Caution	10%
Community Rehabilitation Order	1%
Fine	c%
Community order	c%
Other	c%
Index offence - SRP category	
Contact child	40%
Indecent images	22%
Sexual offending order breaches	18%
Contact adult	11%
Not sexual in statute[13]	6%
Other noncontact	3%
Participation route	
Through the gate	62%
Community	32%
Custody	6%
Prior criminal appearances	
Mean number of previous offences	19
Mean number of previous convictions	5
Mean number of previous custodial sentences	2

[12] 'c' denotes where proportions have been suppressed because there were 1 or 2 individuals in a category or because secondary suppression was necessary.
[13] Index offences which are categorised as 'Not sexual in statute' would be only included in this evaluation if they have an OASys flag which classifies them as being sexually motivated.

Annex E: Variables included in analysis

The evaluation used propensity score matching (PSM), a statistical method that creates comparable treatment and comparison groups by accounting for differences in their characteristics, thereby allowing for an assessment of the intervention's effectiveness.

PSM models require factors associated with both receiving the intervention and the outcome to predict a 'propensity score' representing the likelihood of receiving the intervention conditional on these factors. Individuals in the treatment group are matched to similar individuals in the comparison group based on their propensity scores. This method helps ensure the groups are similar, reducing bias and allowing for a more accurate estimation of treatment effects. The full list of variables used in the matching is available in the Variables and Offence Codes Excel annex accompanying this report. Key derived variables bespoke to this evaluation, such as participation route and offence-related sexual interests, are explained below.

E.1 Participation route

When matching individuals across groups, it is important to consider the impact of different sentence types. A variable for participation route was therefore created to facilitate exact matching of similar individuals according to participation route following propensity score estimation. Pseudo intervention start dates were used to define participation route for the comparison group. Participation route was defined as follows:

- **Custody**: the individual received a custodial sentence for their latest offence before starting Circles of Support and Accountability (CoSA), their CoSA start date was during this sentence and before release from custody.
- **Through the gate**: the individual received a custodial sentence for their latest offence before starting CoSA, their CoSA start date was after release from custody.
- **Community**: the individual received a community-based sentence for their latest offence before starting CoSA, their CoSA start date was after their conviction for this offence.

E.2 Offence-related sexual interests

'Sexual deviancy' is an umbrella term which describes sexual interests or behaviours that are regarded as significantly different from the standards established by a culture or subculture. 'Paraphilia' is a more specific term used to describe such behaviour and can be defined as any intense and persistent sexual interest or arousal to atypical stimuli. Sexual deviancy is considered a legacy term, largely replaced by the more specific definition of paraphilia. 'Offence-related sexual interests' is the term used by His Majesty's Prison and Probation Service (HMPPS) to

describe sexual deviancy and paraphilia, as it is deemed less stigmatising. However, both sexual deviancy and paraphilia remain in use in academia. For more information, see A.2 of the feasibility study.

To proxy offence-related sexual interests, a wide range of variables associated with paraphilia from selected tools were derived and included in the PSM model, rather than a single tool. This includes variables from: SSPI-2 (Screening Scale for Paedophilic Interests), SGP (Scale for General Paraphilia), SRP (Sexual Reoffending Predictor), Offender Assessment System (OASys) items and offence-related variables considered to be associated with paraphilia. This decision was based on a detailed assessment of the strengths and limitations of each tool in the feasibility study and considered both resource and data constraints. For further detail, see section RQ1 of the feasibility study.

Annex F: Feasibility study recommendations

The 2024 feasibility study outlined research questions and recommendations for the Circles of Support and Accountability (CoSA) evaluation. Recommendations from the feasibility study were adhered to where possible, however further inspection of datasets resulted in some changes. Table F1 provides a summary of analytical recommendations from the feasibility study. Any deviations from the recommendations are detailed below and denoted by an asterisk within the table.

Table F1: Summary of analytical recommendations from the feasibility study

Section	Research question	Recommendation
Balancing offence-related sexual interests	**RQ1.** What are the most effective data sources for measuring offence-related sexual interests?	Instead of using a single tool to measure offence-related sexual interests, also known as sexual deviancy and/or paraphilia, it is recommended to include a wide range of variables associated with paraphilia in the propensity score matching (PSM) model. Controlling for these proposed variables should minimise bias due to offence-related sexual interests.
Defining an outcome measure	**RQ2.** Should breaches of licence conditions, particularly recalls to prison, be counted as a reoffence?	Developing a new reoffending outcome measure that incorporates breaches, including recalls to prison for failure to comply with licence conditions, would be complex and resource intensive. Therefore, it is recommended to remain aligned to the department-wide approach of counting only new and separate offences as a reoffence. Where possible, descriptive statistics on recall incidents will be provided to offer some insight into how recalls may affect results.

Section	Research question	Recommendation
	RQ3. From what starting point should reoffending be measured?	To avoid measuring any non-intervention effects during the waiting time between the standard index date and intervention start date, it is recommended that the follow-up period begin once an individual starts their CoSA participation (intervention start date). For the comparison group, pseudo intervention start dates will be calculated to ensure consistency in measuring outcomes.
	***RQ4.** What is the optimal follow-up period to measure reoffending?	To account for differences in sexual reoffending rates and patterns, as well as the extended time required to secure a conviction at court, a five-year fixed follow-up period is recommended. Complementary survival analysis to provide more detailed insights into reoffending patterns will also be conducted.
Using Offender Assessment System (OASys) records	**RQ5.** From what point should OASys records be selected?	OASys records that are more complete are more useful because they provide the greatest amount of data. However, it is important to balance the completeness of the OASys record with proximity to the analysis period. Therefore, it is recommended that: • For those who started CoSA in custody, the most complete OASys record within the 12 months before or 1 month after intervention start date be selected. • For those who started CoSA through the gate (following release from custody) or in the community (during a non-custodial sentence), the most complete OASys record within the 18 months before or 1 month after intervention start date be selected.

Section	Research question	Recommendation
	RQ6. How should missing and/or incomplete OASys records be handled?	Given the importance of OASys records for the impact evaluation, it is recommended that all individuals without an OASys record be excluded from the analysis. However, individuals with missing data in their OASys record should remain in the analysis to minimise bias and ensure methodological consistency with similar evaluations.
Creating the treatment and comparison groups	****RQ7.** How should CoSA eligibility and suitability criteria be applied?	To ensure balance across the treatment and comparison group, a number of measures have been recommended to account for CoSA eligibility and suitability criteria, including offence details and OASys items.
Accounting for observed differences	**RQ8.** How should differences in sex be accounted for?	Males and females exhibit different reoffending rates and patterns, so the standard Justice Data Lab (JDL) approach is to analyse them separately. However, due to the small number of females in the treatment group (fewer than 20), it is unlikely that statistically significant results will be reliably detected. Therefore, it is recommended that females be excluded from the analysis.
	RQ9. How should differences in participation route be accounted for?	When matching individuals across the treatment and comparison group, it is important to consider the impact of their CoSA participation route (whether in custody, through the gate or in the community). Different participation routes may introduce bias in the analysis, as individuals entering CoSA through different routes may have different characteristics and experiences. Therefore, it is recommended to create a flag for matching individuals according to participation route.
	RQ10. How should differences in participation in additional programmes be accounted for?	Participation in other reducing reoffending programmes could influence reoffending rates. To account for this, it is recommended to create a flag indicating whether an individual has participated in another accredited

Section	Research question	Recommendation
		programme during the analysis period. This flag should be included in the PSM model.
	RQ11. How should differences in completion status be accounted for?	Some individuals in the treatment group may have ended their participation before fully 'completing' the intervention. It is recommended that a sub-analysis be conducted to explore difference in CoSA effectiveness between those who completed the intervention as intended and those who did not.
	RQ12. How should differences in social capital be accounted for?	As CoSA aims to provide participants with a social network they might otherwise lack, it is recommended to include several OASys items in the PSM model to account for differences in social capital.
Sample sizes and power calculations	**RQ13.** What is the expected treatment group sample size after applying relevant filters?	From the 792 records submitted in the provisional treatment group dataset, 363 records remain after applying the relevant filters (prior to PSM). The filters include: • 252 records not identified on the Police National Computer (PNC) • 62 records without a corresponding entry in the reoffending database for their period of CoSA participation • 115 records that did not meet the inclusion criteria See Annex B.12 for the final treatment group sample size.
	RQ14. How should the impact of attrition be handled?	Attrition, or loss to follow-up, can occur for reasons such as leaving the country or death. This can result in a biased sample and an underestimation of reoffending. The JDL's preference would be to identify and exclude these individuals. However, since the data does not allow for their identification, they will be retained in the analysis, and their impact will be taken into consideration when interpreting the findings.

Section	Research question	Recommendation
	RQ15. Would an impact evaluation be sufficiently powered to detect a significant treatment effect?	Power calculations undertaken indicate that there is sufficient statistical power to detect a statistically significant treatment effect for the general reoffending measure, if such an effect exists. However, this is not the case for the sexual reoffending measure. Whilst low statistical power may indicate a limited ability to detect a statistically significant treatment effect for the sexual reoffending outcome, the possibility of finding significance cannot be completely ruled out.

*RQ4 of the feasibility study also recommended exploring the possibility of conducting a supplementary survival analysis. Survival analysis was not included as part of this evaluation. See Annex B.11 for more information.

**RQ7 of the feasibility study recommended using the OASys Sexual Reoffending Predictor (OSP) to apply the eligibility criteria ‘Are at medium or above risk of reoffending’, by removing any participants with a low-medium image offences score (OSP-I) or low contact offences score (OSP-C). Although OSP scores were calculated for the current evaluation as recommended, individuals in the treatment or comparison groups were not removed from analysis based on these scores. The OSP was introduced to assess eligibility for CoSA in March 2021, however almost all (99%) of the cohort in this evaluation took part in CoSA before March 2021. Removing individuals based on OSP scores would not reflect the eligibility criteria used at the time of their participation with CoSA. Instead, OSP scores were included in the PSM model to ensure treatment and comparison groups were well matched.

Scores were calculated using the methodology defined in the feasibility study. There have been revisions to the OSP since March 2021. The tool is now referred to as the Sexual Reoffending Predictor (SRP), see the Glossary for more information.

Annex G: Limitations and caveats to findings

G.1 Propensity score matching (PSM)

This evaluation uses propensity score matching (PSM), which is a quasi-experimental evaluation methodology. It is considered a level 4 on the SMS (Scientific Methods Scale). However, PSM is not as robust as randomised control trials or prospectively matched evaluations. For more information on the strengths and limitations of propensity score matching, see the following published reports: Mews, Hillier, McHugh, & Coxon (2013) and Eaton & Mews (2019).

A key limitation of PSM is that it can only reduce bias based on information about participants that is recorded and available (called 'observed factors'). Crucially, this information is required for both the treatment and comparison groups. There is a risk that unobserved or unavailable information could influence the results of a PSM evaluation. To mitigate this risk, the Justice Data Lab (JDL) strives to include all observed factors expected to be predictive of selection onto the intervention evaluated and of reoffending risk. For more information, see the Justice Data Lab Methodology Paper which outlines the methodology applied in more detail.

Of all the variables considered as potential inputs to the PSM models, which predict selection onto the intervention and reoffending risk, none had an absolute standardised difference of more than 10%, and less than 2% of the variables had an absolute standardised difference of between 5% and 10%. This indicates that the matching quality was very good across all analyses (see the Standardised Differences Excel annex accompanying this report for further detail). Therefore, although not all factors to predict selection onto the intervention and reoffending risk could be observed, observable factors were sufficiently accounted for within the analysis.

G.2 Completers sub-analysis

An individual was classified as a completer if their participation lasted for 9 months or longer, or less than 9 months but termination was recorded as 'planned'. For the completers sub-analysis, it is not possible to match on an observable "completion" counterfactual filter in the comparison group. As a result, participants who completed Circles of Support and Accountability (CoSA) were matched to the entire comparison group. It is therefore possible that the analysis did not fully control for unobserved factors relating to the likelihood that an individual would complete CoSA if they were assigned to the intervention. As such, the results of the completers sub-analyses should be interpreted with care, noting that there is a high positive correlation between completion and not having reoffended during the participation period.

Non-completers were classified as those who did not successfully complete a programme, intervention, or study, often due to dropout, non-compliance, or other reasons. It was not possible to report a sub-analysis on non-completers as the PSM model was not of sufficient quality.

G.3 Participation in accredited programmes

Individuals were included within the treatment and comparison groups, regardless of whether they had participated in another sexual offender programme. To account for this, participation in accredited programmes prior to treatment was included in the PSM model and their matching quality monitored. However, the effect of CoSA and any effect of other accredited programmes would be hard to distinguish and cannot be isolated within these results. Moreover, it is possible that individuals in the comparison group (those who did not participate in CoSA during the index sentence) instead undertook a non-accredited programme during the index sentence, which might have had an impact on reoffending rates.

Any treatment effects observed in this analysis are likely to reflect differences relative to other interventions, services, supervision or monitoring in place, as opposed to the absence of any treatment at all. It has not been possible to collect full details on other programmes or interventions that individuals in either the treatment or comparison group have participated in for this evaluation. This is an example of an unobserved factor, outlined above in Section G.1.

G.4 Reoffending outcomes and analysis

This evaluation uses the same definition of ‘reoffending’ as used in the Ministry of Justice’s Proven Reoffending National Statistics. More information on this methodology is available in the Proven reoffending statistics collection. Importantly, this definition does not include crimes that are committed but not recorded by the police or do not lead to a caution or conviction. It also does not adjust for any periods of time individuals are returned to custody, either due to recall or additional sentences, or time spent outside the UK.

There are other rehabilitation outcomes, in addition to reoffending, used in this evaluation, which may be important to consider for specific interventions. Examples could include employment, educational or skills attainment, physical or mental health, health of relationships, or attitudes.

G.5 Statistical significance and power

Statistical significance as defined in this report means that if no real differences exist, there is a 5% chance of each result nonetheless being found to be statistically significant. On the same basis though, the chance of at least one of the results being found to be statistically significant is much higher than 5%.

Analysis with low statistical power is less likely to detect a true difference between the treatment and comparison groups and indicate no significant difference between the two groups, even if there is a genuine treatment effect. The failure to detect a statistically significant effect in an impact evaluation may be due to the low statistical power of the study rather than the actual impact of the intervention. It may be beneficial to repeat the evaluation in the future when more participants have taken part in CoSA.

G.6 Sentence selection

When selecting sentences relevant to CoSA participation, the method 'all participations in the treatment group, all non-participations for the comparison group' (APAN)[14] balances methodological robustness and sufficient sample size. However, there are still limitations to the method. A small proportion (1%) of individuals in the treatment group participated in CoSA twice. It is possible that participating in CoSA multiple times has a different effect compared to a single participation. Despite this, all participations in CoSA were treated as equal in this evaluation.

Individuals in the comparison group may also appear across multiple records. However, the weighted proportion that any given individual makes up in the matched comparison group is less than 0.5% for all individuals across all analyses, indicating that the impact of allowing for multiple records is immaterial in the context of the evaluation results. Furthermore, when selecting records for each group, all individuals that are included in the treatment group will not appear in the comparison group, and vice versa.

G.7 Attrition from the treatment group

Treatment group records are dropped from analysis at different stages for various reasons. In this evaluation, the pre-matched treatment group consisted of around half of the participant dataset provided by Circles UK. It is common in JDL evaluations for records to be dropped from analysis at the two stages outlined below. However, as the names of CoSA participants could not be submitted by CoSA for data protection reasons, more participants were dropped at the first stage of data matching due to the limited identifiable information available to match on. Demographic data from the PNC and Delius were used to compare the post-matched treatment groups to the original cohort of participants provided by Circles UK. Those who could be identified but were not included in the analysis had similar distributions across age, ethnicity and nationality to the post-matched treatment groups. However, the post-matched treatment groups may not be representative of the original cohort across other factors where data was not available.

[14] For more detailed information on APAN methodology, see Annex 13 of the Justice Data Lab Thinking Skills Programme evaluation report (July 2023).

Data matching

Matching individuals' records across datasets is complex and often results in partial or incomplete matches. A high level of confidence is required to match across datasets to ensure that the data attributed to an individual in the analysis is accurate and reliable. Consequently, a large number of individuals may be removed where there is not a high confidence in the match. At the first stage of data matching, participants could only be identified on the Police National Computer (PNC) using PNC identifier (PNCID), date of birth and sex. Without additional identifiable information, disparities in these three identifiers resulted in 15% of records being excluded from the analysis.

Inclusion criteria

Individuals in the cohort also had to meet various inclusion criteria to be included in the analysis. These criteria are separate to the CoSA eligibility criteria and cover demographic characteristics and offence data. As CoSA is aimed at adult offenders, individuals were removed from analysis if aged under 18 at the start of their intervention (and thus the start of the reoffending period, unless participation began in custody). Female participants were also removed from analysis because male and female offenders exhibit different offending behaviours and must be analysed separately, and there were not enough female participants in the cohort for a separate analysis. CoSA eligibility criteria had been applied to identify a past sexual or sexually motivated offence prior to starting CoSA. However, an index offence relating to CoSA participation must also be identified for inclusion in analysis. This required individuals to be identified in the Proven reoffending dataset, and their potential conviction and standard index dates to align with CoSA participation dates.

Any individuals who did not meet the criteria outlined above were removed, this resulted in an additional 15% of records being excluded from analysis.

G.8 Recalls and Survival Analysis

Without the survival analysis, this evaluation does not directly consider a variable follow-up period, or the restriction of recalls on reoffending behaviour. An analysis of recall events for individuals in the matched treatment and comparison groups for the general reoffending analysis showed that 20% of the treatment group had at least one recall event within the five-year reoffending period, compared to 16% of the comparison group. Further analytical work would be required to fully understand this observed difference in recalls between the treatment and comparison groups, including the relationship between receiving treatment from CoSA and additional monitoring, and subsequent impacts on reoffending behaviour.

G.9 Supervision effects

Participation in CoSA may have introduced a different level and method of supervision compared to that experienced by the comparison group. This could have impacted the incidence or timing of reoffending for the treatment group due to differing levels of supervision compared to the comparison group. It was not possible to fully account for differing levels of supervision across the treatment and comparison groups.

G.10 Quality of programme delivery

Quality of programme delivery could be an important factor for interventions aimed at reducing reoffending. Analyses of how the intervention was delivered could have helped to evaluate the extent to which the quality of delivery may have an impact on the outcome of CoSA. However, the quality of programme delivery was out of scope, it was not possible to include this within the analysis.

G.11 Impact of COVID-19

A small part of the cohort within this publication overlaps with the COVID-19 pandemic including lockdowns and operational restrictions. Delivery of CoSA may have been impacted for individuals in the treatment group who participated during the COVID-19 lockdowns.

Results may also be affected by the continued recovery of the courts system. Particularly, continued delays in the processing of cases mean that reoffence convictions for participants whose five-year follow-up period overlapped with the COVID-19 pandemic may fall outside of the six-month waiting period and therefore not be counted in these statistics.

Annex H: Glossary

Accredited programmes: Programmes that are accredited for use in the community and custody. The Correctional Services Advice & Accreditation Panel (CSAAP) helps HMPPS to accredit programmes by reviewing programme design, quality assurance procedures and findings, and programme evaluations. They make recommendations about whether to accredit to the HMPPS Rehabilitation Board. HMPPS is accountable for decisions to accredit programmes.

Average time to first reoffence: The average number of days between a person's index date and the date on which they commit their first proven reoffence, including only those who reoffend.

Circles of Support and Accountability (CoSA): A community-based intervention designed to support and monitor individuals who have committed sexual offences upon their reintegration into society.

Circles UK: The national body supporting the development, quality and coordination of CoSA

Comparison group: A group of offenders who did not receive the intervention being analysed. The comparison group is made up of offenders with similar characteristics to those in the treatment group.

Completers: Individuals who successfully complete a programme, intervention, or study according to predefined criteria or requirements. For CoSA, individuals will be classed as completers if participation duration were more than or equal to 9 months or participation duration was less than 9 months but termination was planned.

Core member: An individual with convictions for harmful sexual behaviour who is actively participating in a CoSA intervention.

Descriptive statistics: Statistical methods used to summarise and describe the characteristics of a dataset.

Eligibility criteria: Specific requirements or qualifications that individuals must meet to be considered eligible for participation in a study, programme or intervention.

Five-year proven reoffending rate – general: The proportion of offenders in a cohort who committed an offence during a five-year period starting on their intervention start date (or release date for participants who began in custody) and that resulted in a court conviction, caution,

reprimand or warning in England or Wales during the same period or a further six-month waiting period.

Five-year proven reoffending rate – sexual: The proportion of offenders in a cohort who committed a sexual offence (as categorised by the Sexual Reoffending Predictor) [see Annex B.8] during a five-year period starting on their intervention start date (or release date for participants who began in custody) and that resulted in a court conviction, caution, reprimand or warning in England or Wales during the same period or a further six-month waiting period.

Five-year follow-up period: The length of time proven reoffending is measured, from intervention start date or release date for participants who began in custody (index date). In this analysis, a fixed five-year follow up period was used (where all participants were monitored for five years).

Impact evaluation: An assessment method used to determine the effects or consequences of an intervention, programme or policy on certain outcomes of interest, often involving comparison groups or counterfactual scenarios.

Index date: The date from which the follow-up period for measuring reoffending begins. For participants who began in the community, this is the start date of their participation in CoSA. For participants who began in custody, this is the date of release from custody. This is different to the Standard index date, see later in the Glossary.

Index offence: The proven offence that leads to an offender being included in the cohort. This is the primary sexual offence for which the offender was convicted and received a custodial or community sentence (specifically, the index sentence). The index offence for this analysis was the offence closest to the start of the offender's participation in CoSA, categorised as a sexual offence by the Home Office or Sexual Reoffending Predictor (SRP). Where an offence in these categories was not identified, offence codes outside of these categories were allowed if a corresponding Offender Assessment System (OASys) record indicated the offence was sexually motivated.

Intervention end date: The date on which an individual ends participation with CoSA.

Intervention start date: The date on which an individual begins participation with CoSA.

Justice Data Lab (JDL): A team of analysts at the Ministry of Justice that provide a service to organisations working to reduce reoffending so that they can better understand the impact of interventions.

Level of confidence: A range of values within an upper and lower bound. A 95% level of confidence would mean you could be 95% confident that the real value for a population of interest lies within the upper and lower bound. Levels of confidence (otherwise known as confidence intervals) are a key output for Justice Data Lab analyses as the reoffending rates for the treatment and comparison groups are essentially samples of larger populations.

Mean: This is a measure of the average in the dataset. It is calculated by adding all the values of a dataset and dividing it by the number of values in the set.

Multiple imputation: A statistical technique used to address missing data by imputing multiple plausible values for missing observations, allowing for more robust analysis and inference.

Non-completers: Individuals who do not successfully complete a programme, intervention, or study, often due to dropout, non-compliance, or other reasons. For CoSA, individuals will be classed as completers if participation duration was less than 9 months but termination was planned or participation duration was more than 9 months.

No significant difference: This means that, based on the analysis, there is insufficient evidence to conclude that the intervention had an effect (positive or negative) on the outcome. The observed differences between the groups could plausibly be due to chance rather than a real effect.

Offender Assessment System (OASys): A system introduced in 2001 and built on the existing 'What Works' evidence base. It combines actuarial methods of prediction with structured professional judgement to provide standardised assessments of offenders' risks and needs, helping to link these risks and needs to individualised sentence plans and risk management plans.

OASys Sexual Reoffending Predictor (OSP): See Sexual Reoffending Predictor (SRP).

Offence-related sexual interests: A term used by HMPPS that describes paraphilia and sexual deviancy.

Offender Group Reconviction Scale (OGRS3): Percentage likelihood of committing any offence within 2 years leading to reconviction (proven reoffending). This is based on static factors such as age, gender and criminal history. An OGRS3 score of 50% or more means that an offender is more likely than not to commit a proven reoffence within 2 years.

Outcome measures: Variables used to assess the effects or outcomes of an intervention, programme or policy.

Police National Computer (PNC): The PNC is the administrative IT system used by all police forces in England and Wales and managed by the National Policing Improvement Agency. As with any large-scale recording system, the PNC is subject to possible errors with data entry and processing. The MoJ maintains a database based on weekly extracts of selected data from the PNC to compile statistics and conduct research on reoffending and criminal histories.

The PNC largely covers recordable offences. These are all indictable and triable-either-way offences, as well as many of the more serious summary offences. All figures derived from the MoJ's PNC database, and particularly those for the most recent months, are likely to be revised as more information is recorded by the police.

Propensity score matching (PSM): The methodology used for constructing a matched comparison group in Justice Data Lab analyses. It uses logistic regression to predict the likelihood of each offender receiving the treatment based on their recorded characteristics, the predicted probabilities being called propensity scores. Treated and non-treated offenders are matched based on the closeness of their propensity scores.

Propensity scores: Scores representing the likelihood of individuals receiving a treatment or intervention based on their recorded characteristics.

Proven reoffence: Where an offender in the cohort commits an offence within the follow-up period, and is convicted at court within either the follow-up period, or the waiting period.

Pseudo intervention start date: A hypothetical or simulated start date used for analytical purposes, often in studies where actual start dates are unavailable or inconsistent. As the comparison group does not have CoSA start dates, a CoSA pseudo start date was imputed.

P-value: The p-value is the probability of obtaining results as extreme as, or more extreme than, those observed, assuming that there is no true difference.

Recall: Where an offender is taken back to prison after being released on licence or parole for breaking the rules of their probation.

Regression model: A statistical model used to investigate the relationship between one or more independent variables and a dependent variable.

Reoffending frequency: The number of proven reoffences committed, expressed per person.

Scale for General Paraphilia (SGP): A measurement tool used to assess the presence and severity of paraphilic interests or behaviours across different domains.

Screening Scale for Paedophilic Interests (SSPI): A screening tool used to assess the presence and severity of paedophilic interests or tendencies.

Sexual Reoffending Predictor (SRP): A risk assessment tool used by HMPPS to predict sexual reoffending among men convicted of current or past sexual or sexually motivated offences. The SRP generates two scores, calculated as part of the Risk of Serious Recidivism (RSR) tool within section of OASys: DC-SRP, which predicts the risk of direct contact sexual reoffending, and IIC-SRP, which predicts the risk of offending related to indirect contact and indecent images of children. Level of risk is predicted by various factors such as age, victim characteristics and number of offences classed as Contact adult, Contact child, Indecent images and Other non-contact Previous iterations of this tool were known as the OASys Sexual Reoffending Predictor (OSP) and did not distinguish between direct and indirect contact (see Actuarial prediction of sexual reoffending: responding to changing offending patterns).

Sexual Reoffending Predictor (SRP) sexual offences: The SRP categorises Home Office offence codes into six categories – Contact adult, Contact child, Indecent images, Not sexual in statute, Other noncontact, and Sexual offending order breach. These, alongside the Home Office 'Sexual offences' group, were used to identify the index offence. Index offences categorised as 'Not sexual in statute' would be only included if they had an OASys flag which classified them as being sexually motivated. The sexual reoffending measure used in this evaluation includes all offence codes listed in these categories other than 'Not sexual in statute'.

Significant difference: This means there is statistical evidence that the difference between groups is unlikely to be due to chance alone. The significance level used in this analysis is 5%, meaning that such a difference would be expected to occur rarely if there were no true difference.

Social capital: The collective value of social networks, relationships, and interactions within a community or society, including trust, reciprocity, and cooperation, which can contribute to individual and collective well-being.

Standard index date: The standard starting point for measuring reoffending in Justice Data Lab evaluation. This is the earliest point an individual is at risk of reoffending: either the release date for those serving custodial sentences or the conviction date for those serving non-custodial sentences. This was not used in this evaluation, see Index date earlier in the Glossary.

Standardised effect size: A measure of the magnitude of the difference between groups, expressed on a standardised scale. This report uses Cohen's *d* as the effect size statistic, with positive and negative values indicating the direction of the effect.

Standardised mean difference: The standardised difference in means between the treatment and comparison groups, for an individual variable. The standardised mean difference is expressed as a percentage in this report; the smaller the percentage the more similar the groups are on that variable.

Suitability criteria: Criteria used to determine whether individuals are appropriate candidates for participation in a study, programme, or intervention, based on their characteristics, needs, or circumstances.

Survival analysis: A set of statistical techniques used to analyse the time until an event of interest occurs, such as reoffending or relapse.

Treatment effect: The impact or outcome resulting from an intervention or treatment.

Treatment group: The group of offenders who received the intervention delivered by the provider. In other words, the offenders who received ‘the treatment’.

Waiting period: The additional time beyond the follow-up period to allow for offences which are committed towards the end of this time frame to be proven by a court, resulting in a conviction, caution, reprimand or final warning.

Waiting time: The length of time between an individual’s standard index date and starting the intervention.