Reoffending impact evaluation of the prison-based RESOLVE Offending Behaviour Programme

Ciara Robinson, Annie Sorbie, Johannes Huber, James Teasdale, Katy Scott, Mark Purver and Ian Elliott

Ministry of Justice

2021

Summary

RESOLVE is an accredited programme designed and delivered by Her Majesty's Prison and Probation Service (HMPPS). The programme is a cognitive-behavioural therapy-informed offending behaviour programme which aims to improve outcomes related to violence in adult males who are of a medium risk of reoffending. The aim of this evaluation was to assess the impact of prison-based RESOLVE on proven reoffending.

This analysis looked at the reoffending behaviour of 2,509 adult males who participated in the RESOLVE custody programme at some point between 2011 and 2018 and who were released from prison between 2011 and 2018.

The headline results for one-year proven general reoffending show that those who took part in the programme in England and Wales were less likely to reoffend, reoffended less frequently and took longer to reoffend than those who did not take part. The headline results for two-year proven general reoffending show that those who took part were less likely to reoffend, reoffended less frequently and took longer to reoffend than those who did not take part. These results were statistically significant.

For proven violent reoffences, the headline one and two-year results did not show that the programme had a statistically significant effect on a person's reoffending behaviour but this should not be taken to mean it fails to have an effect. Further analyses were also conducted to examine the specific effects of RESOLVE on relevant sub-groups, for proven general reoffending and violent reoffending. Among the one-year violent sub-analyses, the 'RESOLVE only' sub-group were significantly less likely to reoffend violently and reoffended violently less frequently than those who did not take part. There were no statistically significant sub-analyses for the two-year violent measures.

The headline analysis in this report measured proven reoffences in a one and two-year period for a 'treatment group' who received the intervention and for a much larger 'comparison group' of similar offenders who did not receive it. There may have been a different impact on participants who did not meet the criteria for analysis. Analyses considered one and two-year reoffending measures for both general reoffences and violent reoffences. General reoffending is comprised of all proven offences, including violence. This is a Justice Data Lab (JDL) study with supplementary analyses. For further information on the methodology and interpreting results, see the Summary of methodology below and Annex 1.

Key results

Headline general proven reoffending measure

One-year	26.9% of the treatment group reoffended in the year following release from prison	↓	This is significantly fewer than the comparison group (30.2%)
Two-year	44.7% of the treatment group reoffended in the two years following release from prison	↓	This is significantly fewer than the comparison group (47.4%)

One-year	An average of 0.73 proven reoffences were committed by each of the men in the treatment group	↓	This is significantly fewer than the comparison group (0.87)
Two-year	An average of 1.56 proven reoffences were committed by each of the men in the treatment group	↓	This is significantly fewer than the comparison group (1.85)

One-year	The average time before a reoffender committed their first proven reoffence was 166 days	↑	This is significantly later than the comparison group (151 days)
Two-year	The average time before a reoffender committed their first proven reoffence was 287 days	↑	This is significantly later than the comparison group (267 days)

Headline violent proven reoffending measure

One-year	5.7% of the treatment group reoffended with a violent reoffence in the year following release from prison	↓	This is not significantly fewer than the comparison group (6.3%)
Two-year	11.6% of the treatment group reoffended with a violent reoffence in the two years following release from prison	↓	This is not significantly fewer than the comparison group (12.0%)

One-year	An average of 0.08 proven violent reoffences were committed by each of the men in the treatment group	↓	This is not significantly fewer than the comparison group (0.09)
Two-year	An average of 0.16 proven violent reoffences were committed by each of the men in the treatment group	↓	This is not significantly fewer than the comparison group (0.18)

One-year	The average time before a reoffender committed their first proven violent reoffence was 169 days	↑	This is not significantly later than the comparison group (158 days)
Two-year	The average time before a reoffender committed their first proven violent reoffence was 318 days	↑	This is not significantly later than the comparison group (303 days)

**Green* *arrow for significant finding, grey arrow for non-significant*

Impact on general reoffences

Overall estimates and *what you can and can't say* statements

For any **100** typical men who receive the intervention, compared with any **100** similar men who do not receive it:

The number of men who commit a proven reoffence within ***one year*** could be **lower by between 2 and 5 men**. The number of men who commit a proven reoffence within ***two years*** after release could be **lower by between 0 and 5 men**. **These are statistically significant results**.

The number of proven reoffences committed within ***one year*** could be **lower by between 7 and 21 offences.** The number of proven reoffences committed within ***two years*** could be **lower by between 15 and 42 offences**. **These are statistically significant results.**

On average, the time before an offender committed their first proven reoffence within ***one year*** could be **longer by between 8 and 23 days**. On average, the time before an offender committed their first proven reoffence within ***two years*** could be **longer by between 7 and 33 days**. **These are statistically significant results**.

✔ **What you can say about the one and two-year general reoffending measures:**

"This analysis provides evidence that support from the RESOLVE intervention programme may decrease the **number of proven reoffenders** during a one [or two] year period."

"This analysis provides evidence that support from the RESOLVE intervention programme may decrease the **number of proven reoffences** committed by its participants during a one [or two] year period."

"This analysis provides evidence that support from the RESOLVE intervention programme may lengthen the **average time to first proven reoffence** for its participants."

✘ **What you cannot say about the one and two-year general reoffending measures:**

"This analysis provides evidence that support from the RESOLVE intervention programme increases/has no effect on the **reoffending rate** of its participants during a one [or two] year period."

"This analysis provides evidence that support from the RESOLVE intervention programme increases/has no effect on the **number of proven reoffences** committed by its participants during a one [or two] year period."

"This analysis provides evidence that support from the RESOLVE intervention programme shortens/has no effect on the **average time to first proven reoffence** for its participants."

Impact on violent reoffences

Overall estimates and *what you can and can't say* statements

For any **100** typical men who receive the intervention, compared with any **100** similar men who do not receive it:

The number of men who commit a proven violent reoffence within ***one year*** could be **lower by as many as 2 men and higher by as many as 0 men**. The number of men who commit a proven violent reoffence within ***two years*** after release could be **lower by as many as 2 men and higher by as many as 1 man**. **These are not statistically significant results**.

The number of proven violent reoffences committed within ***one year*** could be **lower by as many as 3 offences and higher by as many as 1 offence**. The number of proven violent reoffences committed within ***two years*** could be **lower by as many as 4 offences and**

On average, the time before an offender committed their first proven violent reoffence within ***one year*** could be **shorter by as much as 6 days or longer by as much as 27 days**. On average, the time before an offender committed their first proven violent reoffence within ***two years*** could be **shorter by as much as 12 days or longer by as**

✔ **What you can say about the one and two-year violent reoffending measures:**

"This analysis does not provide clear evidence on whether support from the RESOLVE intervention programme increases or decreases the **number of participants who commit a proven violent reoffence** in a one [or two] year period."

"This analysis does not provide clear evidence on whether support from the RESOLVE intervention programme increases or decreases the **number of proven violent reoffences** committed by its participants during a one [or two] year period."

"This analysis does not provide clear evidence on whether support from the RESOLVE intervention programme shortens or lengthens the **average time to first proven violent reoffence**."

✖ **What you cannot say about the one and two-year violent reoffending measures:**

"This analysis provides evidence that support from the RESOLVE intervention programme increases/decreases/has no effect on the violent **reoffending rate** of its participants during a one [or two] year period."

"This analysis provides evidence that support from the RESOLVE intervention programme increases/decreases/has no effect on the **number of proven violent reoffences** committed by its participants during a one [or two] year period."

"This analysis provides evidence that support from the RESOLVE intervention programme shortens/lengthens/has no effect on the **average time to first proven violent reoffence** for its participants."

Table of contents

HMPPS description of RESOLVE

RESOLVE, in custody, is a cognitive-behavioural[1] group work accredited programme (see Annex 4). The programme aims to reduce violence in medium risk men (aged 18 years and above)[2] and is designed and delivered by HMPPS (Her Majesty's Prison and Probation Service). RESOLVE defines violence as "actual, attempted, or threatened harm to a person or persons" (Webster, Douglas, Eaves & Hart, 1997) and therefore it is inclusive of those who inflict both physical and psychological harm on individuals. It is a programme underpinned by research that suggests those with violent convictions often display difficulties relating to self-control and emotional management and possess attitudes supportive of violence (for more information, see Annex 2). Participants range from those who have come to see the world and other people through a lens of fear or hostility, to those who appear to put their own needs above harm to others. In turn RESOLVE directly targets these need areas.

The overall aim of RESOLVE is to help individuals make a positive change in their lives and move away from the use of aggression and violence. In working toward this, individual goals are identified to help participants develop knowledge and learn skills to live a rewarding and purposeful life free of aggressive and violent behaviour. A future focused, strengths-based approach helps participants form meaningful therapeutic relationships with facilitators, supports open and genuine engagement and encourages a positive approach to the future.

The programme comprises 22 group sessions and 4 individual sessions. These are divided into six modules, offering knowledge and skills to manage impulsivity, improve emotional management, dispute attitudes supportive of violence and improve conflict management. Group sessions offer the opportunity to discuss learning, personal experiences and practice skills within a supportive and safe environment. There are also four individual sessions at key points within the programme. These are delivered by a 'named facilitator' who provides support for participants to personalise the work further in an open and focused environment. Throughout the programme, participants are expected to apply their learning and practice skills outside of sessions in order to take maximum benefit.

The final module of the programme allows participants to bring everything together into a plan for future violence-free living and they are given an opportunity to practice skills within their plan. Post programme objectives are set to encourage programme graduates to continue to apply and refine their plan for living non-aggressively beyond the programme.

---

[1] RESOLVE is informed by cognitive-behavioural therapeutic (CBT) approaches to behavioural change in offenders. It is one of several accredited programmes delivered by HMPPS across England and Wales in both prison and probation settings and more information can be found here https://www.gov.uk/guidance/offending-behaviour-programmes-and-interventions

[2] RESOLVE participants were required to meet each of the following criteria, guided by the risk/need/responsivity principal of rehabilitation:

1. Violent index offence or previous violent offence (convicted or non-convicted in prior two years)
2. Are identified as medium risk of reoffending and violence
3. Have needs addressed by the programme
4. Are able to engage in and benefit from the intervention

Summary of methodology

The aim of this evaluation is to assess the impact of RESOLVE on proven reoffending outcomes.

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

- Police National Computer (PNC) to provide reoffending outcomes and criminal history and employment information

- Offender Assessment System (OASys) to provide risks and needs information

This suite of data, comprised of 87 matching factors (Annex 6), was used as the basis for building propensity score matching (PSM) models. 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'[3]. This propensity score reflects the likelihood that an offender received the intervention, given the recorded characteristics. Using this propensity score, each offender was weighted by similarity to those in the treatment group. Offenders in the treatment group were matched to similar untreated offenders. Overall, the matching quality achieved for the headline and sub-analyses was very good[4].

The reoffending rates for the treatment and comparison groups were then compared. The rates are calculated using the weighted values for each person after matching. Three reoffending outcomes were used to estimate the impact of the intervention in a one and two-year period. The outcomes were applied for both the general (all reoffences) and violent measures. The outcomes are as follows:

1) A binary reoffending outcome: the number of people who commit a proven reoffence, expressed as a percentage of the group

2) A frequency reoffending outcome: the number of proven reoffences committed, expressed per person

3) Days to reoffence: the average number of days between a person's prison release date and the date on which they commit their first proven reoffence, including only those who reoffend

[3] A propensity score is a value between 0 and 1 which represents the likelihood of receiving treatment. More details on the matching methodology are included in Annex 1.

[4] Matching quality in JDL analyses uses a traffic light scale (see Annex 1). The mean absolute standardised differences for all sub-analyses was less than 5%. Therefore, the matching quality achieved based on recorded factors was 'green' or 'very good' (see standardised differences annex to this report).

The same general and violent reoffending headline measures were applied to a range of sub-analyses examining specific sub-groups. For a summary, see 'Explanation of sub-analyses' below and for detail, see Annex 1. Each analysis undergoes a different and unique PSM process.

Interpreting results

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. The direction of the difference in reoffending rates indicates whether the treatment effect is positive or negative. The estimated differences shown are the 95% confidence intervals for the differences between the relevant treatment and comparison group outcomes.

The effect size or the strength of the relationship between the two groups is measured by Cohen's *d*. Cohen's *d* values estimate the size of any effects and are provided as another measure of impact. To inform interpretation, these effect size values are typically categorised as small (0.2) medium (0.5) and large (0.8) (Cohen, 1998). Small, medium and large categories are however relative to the area of behavioural science or specific research method being employed (Cohen, 1988). In the field of criminal justice and offender interventions evaluations, effect sizes are on average small to medium (see for example, Barnes, TenEyck, Pratt & Cullen, 2020). For guidance on interpreting effect sizes, please see Annex 1.

Despite efforts to include all observed factors known to be predictive of selection onto RESOLVE and of reoffending risk, the importance of information that is not recorded cannot be known and unobserved factors may influence these results. Other limitations include: small sample sizes for certain sub-analyses and non-proven reoffending not having been included in the analysis. For detail on methodology and a fuller list of limitations, see Annex 1.

Profile of the treatment group

The following descriptive statistics aim to provide an overview of the characteristics of 2,509 treatment group offenders included in the analysis. The treatment group was 100% male with an age range from 18 to 69 years old. The tables below contain information on demographics, offence history, offending-related risks/needs, and participation in other accredited programmes.

Demographic information for the treatment sample.

Variable	**Frequency** (or mean where stated)
Age	29 (IQR 23-33)
Ethnicity	
White	72%
Black	21%
Asian	5%
Other	1%
Unknown	1%
Nationality	
UK national	97%
Non-UK national	3%
Unknown	1%
Learning difficulties and/or learning challenges (LDC)	
No problems	84%
Some problems	13%
Significant problems	3%
Difficulties with either numeracy, reading or writing	
No problems	74%
Some problems	21%
Significant problems	5%
Participation in other accredited programmes (APs)	
No other APs	69%
1 other AP	22%
2 other APs	6%
3 other APs	2%
4 or more other APs	1%
Other accredited programmes attended	
Participated in Thinking Skills Programme (TSP)	23%
Participated in another violence-based accredited programme	5%

Offence-history information for the treatment sample.

Variable	Frequency (or mean/average where stated)
Sentence length	
Less than or equal to 6 months	2%
Between 6 and 12 months	1%
12 months to less than 4 years	35%
4 to 10 years	47%
More than 10 years	3%
Indeterminate or life sentence	13%
Index offences	
Violence against the person	45%
Robbery	29%
Theft offences	10%
Possession of weapons	6%
Drug offences	3%
Summary offences excluding motoring	2%
Public order offences	2%
Criminal damage and arson	2%
Miscellaneous crimes against society	1%
Time between treatment and release	
Less than or equal to 12 months	41%
Over 12 months to 5 years	59%
Prior criminal appearances	
Mean number of previous offences	29 (IQR 10-38)
Mean number of previous convictions	13 (IQR 5-18)
Mean number of previous violent offences (violence against the person)	5 (IQR 2-6)
Mean number of previous custodial sentences	4 (IQR 1-6)
Risk assessment	
Mean Offender Violence Predictor (OVP) score	47 (IQR 36-58)
Mean Offender Group Reconviction Scale (OGRS3) score	63 (IQR 52-76)

Explanation of sub-analyses (see Annex 1)

The headline analyses were performed on those in the RESOLVE treatment group[5]. Further analyses were also conducted to examine the specific effects of RESOLVE on relevant sub-groups. There is an evidence base outlined in the literature review in Annex 3 that suggests evaluations ideally should investigate how treatment can vary with different factors. The themes included treatment completion, programme integrity and the importance of targeting a theoretically ideal programme participant group. Each sub-group undergoes a separate PSM process and therefore results are not comparable across the sub-analyses:

Participation in RESOLVE only

This analysed those who solely participated in RESOLVE and did not take part in any other accredited programme, compared to a matched comparison group who had not taken part in any accredited programmes. This aims to measure the effect specifically associated with RESOLVE and not a combined effect with another programme.

Completion of the Programme

The treatment group was divided into RESOLVE completers and non-completers. Subsequently, each sub-group was matched to a 'no treatment' comparison group. This sub-analysis aims to determine whether the treatment effect was different for those who completed the programme and also those who started RESOLVE but did not complete it.

Ideal Suitability for RESOLVE

This analysis created a subgroup of 'ideal' RESOLVE candidates by removing those deemed *potentially less suitable* for receiving treatment. This compared a primary target group of RESOLVE to a matched comparison group who received no treatment.

Programme Integrity

This analysis examined the difference in the treatment effect in prisons where programme integrity[6] is broadly maintained and in prisons where programme integrity is compromised (as outlined in programme and management manuals). This was conducted for two different timeframes to mirror the two sets of quality assurance data collected by HMPPS: 2014-2016 and 2016-2019 (see Annex 5).

[5] This is inclusive of participants where there was intention-to-treat (ITT), regardless of whether they fully completed or complied with programme requirements (for more information on ITT, see Annex 3).
[6] Programme Integrity data collected by HMPPS Interventions Services used different methodologies across two different time periods (2014-2016 and 2016-2019). To note, these sample sizes are substantially smaller than the overall group. Approximately 75% of the overall sample was not included because the prison they were in was not classified in either of the programme integrity time categories, or they did not spend at least 50% of their time in the prison assessed. Further details can be found in Annexes 1 and 5.

Analyses undertaken

The total suite of analyses undertaken are listed below. Each analysis was conducted for one and two-year general and violent reoffending measures.

1. **Overall**: treatment group matched to offenders in England and Wales using demographics, criminal history and individual risks and needs.

2. **Participated in RESOLVE only and not in any other Accredited Programmes**: treatment group matched to offenders in England and Wales using demographics, criminal history and individual risks and needs.

3. **Completed RESOLVE**: treatment group matched to offenders in England and Wales using demographics, criminal history and individual risks and needs.

4. **Did not complete RESOLVE**: treatment group matched to offenders in England and Wales using demographics, criminal history and individual risks and needs.

5. **Participants who met the ideal suitability criteria**: treatment group matched to offenders in England and Wales using demographics, criminal history and individual risks and needs.

6. **Prisons where the programme integrity was broadly maintained (2014-2016 assessment)**: treatment group matched to offenders in England and Wales using demographics, criminal history and individual risks and needs.

7. **Prisons where the programme integrity was compromised (2014-2016 assessment)**: treatment group matched to offenders in England and Wales using demographics, criminal history and individual risks and needs.

8. **Prisons where the programme integrity was broadly maintained (2016-2019 assessment):** treatment group matched to offenders in England and Wales using demographics, criminal history and individual risks and needs.

9. **Prisons where the programme integrity was compromised (2016-2019 assessment):** treatment group matched to offenders in England and Wales using demographics, criminal history and individual risks and needs.

Results in detail

General reoffending

The standard JDL method uses general reoffending outcomes to estimate the impact of the intervention. This includes all proven reoffences and therefore all violent reoffences are part of this measure.

Table 1 presents the sample sizes for the analyses. This includes the unweighted and weighted number of reoffenders in the comparison group, of which the weighted are used to calculate the reoffending rate in Table 2. Where sample sizes are relatively small, they will be unlikely to produce a statistically significant result and thus have a lower likelihood of supporting conclusions with an acceptable level of confidence.

The one and two-year general reoffending outcomes (as described in the summary of the methodology) can be found in Tables 2 to 4.

Table 1: Sample sizes after matching for general reoffending analysis.

Analyses	Outcome measure	Treatment group size	Comparison group size	Reoffenders in treatment group	Reoffenders in comparison group (weighted number)
Overall	**1-year**	2,506	176,491	673	81,947 (53,293)
	2-year	1,916	81,343	857	47,398 (38,536)
Participated in RESOLVE only	**1-year**	1,731	122,948	477	55,176 (40,386)
	2-year	1,316	133,014	616	80,430 (66,967)
Completed RESOLVE	**1-year**	2,205	100,672	570	42,751 (29,262)
	2-year	1,692	145,709	731	87,099 (66,839)
Did not complete RESOLVE	**1-year**	290	78,622	103	36,683 (31,360)
	2-year	231	69,685	129	43,406 (39,772)
Participants who met ideal suitability criteria	**1-year**	799	13,531	219	5,915 (4,087)
	2-year	589	22,675	277	15,092 (10,576)
Programme integrity broadly maintained 2014-2016	**1-year**	239	4,606	66	1,848 (1,429)
	2-year	229	4,396	99	2,414 (1,995)
Programme integrity compromised 2014-2016	**1-year**	62	1,412	19	575 (498)
	2-year	57	898	32	530 (556)
Programme integrity broadly maintained 2016-2019	**1-year**	189	5,119	54	2,460 (1,825)
	2-year	97	1,309	45	803 (784)
Programme integrity compromised 2016-2019	**1-year**	148	2,849	55	1,376 (991)
	2-year	79	1,226	44	773 (714)

Results Summary

Statistically significant results of the one and two-year general reoffending measures

This table shows there were 27 statistically significant results among the analyses. These provide evidence that:

Overall

One-year

- **Participants are less likely to commit a general reoffence** than non-participants.
- **Participants commit fewer general reoffences** than non-participants.
- **Participants who reoffend within a one-year period commit their first proven reoffence later** than non-participants.

Two-year

- **Participants are less likely to commit a general reoffence** than non-participants.
- **Participants commit fewer general reoffences** than non-participants.
- **Participants who reoffend within a two-year period commit their first proven reoffence later** than non-participants.

Participated in RESOLVE only

One-year

- **Participants are less likely to commit a general reoffence** than non-participants.
- **Participants commit fewer general reoffences** than non-participants.
- **Participants who reoffend within a one-year period commit their first proven reoffence later** than non-participants.

Two-year

- **Participants are less likely to commit a general reoffence** than non-participants.
- **Participants commit fewer general reoffences** than non-participants.
- **Participants who reoffend within a two-year period commit their first proven reoffence later** than non-participants.

Completed RESOLVE

One-year

- **Participants are less likely to commit a general reoffence** than non-participants.
- **Participants commit fewer general reoffences** than non-participants.
- **Participants who reoffend within a one-year period commit their first proven reoffence later** than non-participants.

Two-year

- **Participants are less likely to commit a general reoffence** than non-participants.
- **Participants commit fewer general reoffences** than non-participants.
- **Participants who reoffend within a two-year period commit their first proven reoffence later** than non-participants.

Results Summary

Statistically significant results of the one and two-year general reoffending measures

Did not complete RESOLVE

Two-year

- **Participants commit fewer general reoffences** than non-participants.

Participants who met the ideal suitability criteria

One-year

- **Participants who reoffend within a one-year period commit their first proven reoffence later** than non-participants.

Two-year

- **Participants who reoffend within a two-year period commit their first proven reoffence later** than non-participants.

Prisons where the programme integrity was broadly maintained (2014-2016 cohort)

One-year

- **Participants commit fewer general reoffences** than non-participants.

Two-year

- **Participants commit fewer general reoffences** than non-participants.

Prisons where the programme integrity was compromised (2014-2016 cohort)

Two-year

- **Participants who reoffend within a two-year period commit their first proven reoffence later** than non-participants.

Prisons where the programme integrity was broadly maintained (2016-2019 cohort)

One-year

- **Participants are less likely to commit a general reoffence** than non-participants
- **Participants who reoffend within a one-year period commit their first proven reoffence later** than non-participants.

Two-year

- **Participants are less likely to commit a general reoffence** than non-participants.

Tables 2-4 show the measures of general reoffending. Rates are expressed as percentages and frequencies expressed per person. The sub-analysis is highlighted in green if it is statistically significant (at the 0.05 level). Effect sizes (expressed as Cohen's *d* statistic) are also included to indicate the strength of the relationship between the two groups. The estimated differences shown are the 95% confidence intervals for the differences between the relevant treatment and comparison group measures.

Table 2: Proportion of men who committed a proven general reoffence in a one and two-year period after support from RESOLVE, compared with matched comparison groups.

Analyses	Outcome measure	Number in treatment group	Number in comparison group	Treatment group rate (%)	One and two-year proven general reoffending rates				
					Comparison group rate (%)	Estimated difference (% points)	Standardised effect size (Cohen's *d*)	Statistically significant difference?	p-value
Overall	**1-year**	2,506	176,491	26.9	30.2	-5.1 to -1.6	-0.074	Yes	<0.01
	2-year	1,916	81,343	44.7	47.4	-4.9 to -0.4	-0.053	Yes	0.02
Participated in RESOLVE only	**1-year**	1,731	122,948	27.6	32.8	-7.4 to -3.2	-0.115	Yes	<0.01
	2-year	1,316	133,014	46.8	50.3	-6.2 to -0.8	-0.071	Yes	0.01
Completed RESOLVE	**1-year**	2,205	100,672	25.9	29.1	-5.1 to -1.4	-0.072	Yes	<0.01
	2-year	1,692	145,709	43.2	45.9	-5.0 to -0.3	-0.054	Yes	0.03
Did not complete RESOLVE	**1-year**	290	78,622	35.5	39.9	-9.9 to 1.2	-0.090	No	0.12
	2-year	231	69,685	55.8	57.1	-7.7 to 5.2	-0.025	No	0.71
Participants who met ideal suitability criteria	**1-year**	799	13,531	27.4	30.2	-6.0 to 0.4	-0.062	No	0.09
	2-year	589	22,675	47.0	46.6	-3.7 to 4.5	0.008	No	0.85
Programme integrity broadly maintained 2014-2016	**1-year**	239	4,606	27.6	31.0	-9.3 to 2.5	-0.075	No	0.25
	2-year	229	4,396	43.2	45.4	-8.8 to 4.5	-0.043	No	0.52
Programme integrity compromised 2014-2016	**1-year**	62	1,412	30.6	35.3	-16.7 to 7.4	-0.098	No	0.45
	2-year	57	898	56.1	62.0	-19.5 to 7.8	-0.117	No	0.40
Programme integrity broadly maintained 2016-2019	**1-year**	189	5,119	28.6	35.7	-13.7 to -0.5	-0.152	Yes	0.04
	2-year	97	1,309	46.4	59.9	-24.0 to -3.1	-0.272	Yes	0.01
Programme integrity compromised 2016-2019	**1-year**	148	2,849	37.2	34.8	-5.7 to 10.5	0.050	No	0.56
	2-year	79	1,226	55.7	58.2	-14.1 to 9.0	-0.051	No	0.66

Table 3: Number of proven general reoffences committed in a one and two-year period by men who received support from RESOLVE, compared with matched comparison groups.

Analyses	Outcome measure	Number in treatment group	Number in comparison group	One and two-year proven general reoffending frequencies (offences per person)					
				Treatment group frequency	Comparison group frequency	Estimated difference	Standardised effect size (Cohen's *d*)	Statistically significant difference?	p-value
Overall	**1-year**	2,506	176,491	0.73	0.87	-0.21 to -0.07	-0.075	Yes	<0.01
	2-year	1,916	81,343	1.56	1.85	-0.42 to -0.15	-0.090	Yes	<0.01
Participated in RESOLVE only	**1-year**	1,731	122,948	0.75	0.97	-0.31 to -0.14	-0.114	Yes	<0.01
	2-year	1,316	133,014	1.63	2.05	-0.58 to -0.25	-0.126	Yes	<0.01
Completed RESOLVE	**1-year**	2,205	100,672	0.70	0.83	-0.20 to -0.06	-0.072	Yes	<0.01
	2-year	1,692	145,709	1.50	1.75	-0.39 to -0.12	-0.082	Yes	<0.01
Did not complete RESOLVE	**1-year**	290	78,622	1.06	1.28	-0.48 to 0.03	-0.096	No	0.08
	2-year	231	69,685	1.98	2.48	-0.97 to -0.04	-0.133	Yes	0.03
Participants who met ideal suitability criteria	**1-year**	799	13,531	0.76	0.88	-0.26 to 0.01	-0.064	No	0.08
	2-year	589	22,675	1.60	1.78	-0.43 to 0.07	-0.056	No	0.17
Programme integrity broadly maintained 2014-2016	**1-year**	239	4,606	0.59	0.85	-0.42 to -0.08	-0.157	Yes	<0.01
	2-year	229	4,396	1.15	1.68	-0.78 to -0.28	-0.209	Yes	<0.01
Programme integrity compromised 2014-2016	**1-year**	62	1,412	0.82	0.98	-0.70 to 0.39	-0.076	No	0.58
	2-year	57	898	1.96	2.55	-1.41 to 0.23	-0.182	No	0.16
Programme integrity broadly maintained 2016-2019	**1-year**	189	5,119	0.89	1.07	-0.46 to 0.10	-0.087	No	0.21
	2-year	97	1,309	1.84	2.42	-1.25 to 0.09	-0.168	No	0.09
Programme integrity compromised 2016-2019	**1-year**	148	2,849	1.13	1.09	-0.29 to 0.37	0.019	No	0.80
	2-year	79	1,226	2.29	2.68	-1.16 to 0.37	-0.097	No	0.31

Table 4: Average time to first proven general reoffence in a one and two-year period for men who received support from RESOLVE, compared with matched comparison groups.

Analyses	Outcome measure	Average time to first proven general reoffence in a one and two-year period, for reoffenders only (days)							
		Number in treatment group	Number in comparison group *(unweighted)*	Treatment group time	Comparison group time	Estimated difference	Standardised effect size (Cohen's *d*)	Statistically significant difference?	p-value
Overall	**1-year**	673	81,947	166	151	8 to 23	0.152	Yes	<0.01
	2-year	857	47,398	287	267	7 to 33	0.101	Yes	<0.01
Participated in RESOLVE only	**1-year**	477	55,176	166	148	9 to 27	0.177	Yes	<0.01
	2-year	616	80,430	291	261	15 to 46	0.154	Yes	<0.01
Completed RESOLVE	**1-year**	570	42,751	168	153	7 to 23	0.149	Yes	<0.01
	2-year	731	87,099	289	270	5 to 33	0.097	Yes	<0.01
Did not complete RESOLVE	**1-year**	103	36,683	159	143	-6 to 37	0.142	No	0.17
	2-year	129	43,406	276	249	-7 to 60	0.135	No	0.12
Participants who met ideal suitability criteria	**1-year**	219	5,915	176	151	11 to 39	0.245	Yes	<0.01
	2-year	277	15,092	303	266	14 to 60	0.188	Yes	<0.01
Programme integrity broadly maintained 2014-2016	**1-year**	66	1,848	158	160	-28 to 23	-0.025	No	0.84
	2-year	99	2,414	291	269	-21 to 65	0.108	No	0.31
Programme integrity compromised 2014-2016	**1-year**	-	-	-	-	-	-	-	-
	2-year	32	530	365	284	2 to 159	0.368	Yes	0.05
Programme integrity broadly maintained 2016-2019	**1-year**	54	2,460	173	147	1 to 51	0.275	Yes	0.04
	2-year	45	803	284	240	-12 to 99	0.238	No	0.12
Programme integrity compromised 2016-2019	**1-year**	55	1,376	157	132	-2 to 53	0.260	No	0.07
	2-year	44	773	252	249	-55 to 62	0.019	No	0.90

(Note: "-" identifies suppressed results where the number of reoffenders in either the treatment or comparison group is lower than 30. Where sample sizes are relatively small, they will be unlikely to produce a statistically significant result and thus have a lower likelihood of supporting conclusions with an acceptable level of confidence.)

Violent reoffending

A violent reoffending measure was created for this evaluation. This measure uses offence groups ‘Violence against the person’ and ‘Robbery’ to categorise a reoffence as violent. For further detail on offence categories, see Annex 1.

Table 5 presents the sample sizes for the analyses. This includes the unweighted and weighted number of reoffenders in the comparison group, of which the weighted are used to calculate the reoffending rate in Table 6. As a subset of general reoffending, the number of violent reoffenders will be smaller relative to the findings above. Where sample sizes are relatively small, they may be unlikely to produce a statistically significant result, particularly in combination with a relatively low reoffending rate. Thus, there may be a lower likelihood of supporting conclusions with an acceptable level of confidence.

The one and two-year violent reoffending outcomes (as described in the summary of the methodology) can be found in Tables 6 to 8.

Table 5: Sample sizes after matching for violent reoffending analysis.

Analyses	Outcome measure	Treatment group size	Comparison group size	Reoffenders in treatment group	Reoffenders in comparison group (weighted number)
Overall	**1-year**	2,497	108,914	143	9,979 (6,880)
	2-year	1,923	97,074	223	14,408 (11,677)
Participated in RESOLVE only	**1-year**	1,727	81,540	102	7,825 (5,841)
	2-year	1,315	105,123	161	16,108 (13,762)
Completed RESOLVE	**1-year**	2,189	82,212	114	7,187 (5,003)
	2-year	1,691	128,638	188	19,062 (14,775)
Did not complete RESOLVE	**1-year**	290	102,297	27	10,199 (8,773)
	2-year	231	91,849	35	14,740 (13,541)
Participants who met ideal suitability criteria	**1-year**	812	20,302	45	1,379 (1,146)
	2-year	589	23,301	66	2,684 (2,328)
Programme integrity broadly maintained 2014-2016	**1-year**	239	4,751	20	375 (296)
	2-year	228	4,758	29	630 (490)
Programme integrity compromised 2014-2016	**1-year**	62	1,762	4	145 (109)
	2-year	59	1,139	9	158 (186)
Programme integrity broadly maintained 2016-2019	**1-year**	189	4,337	10	596 (358)
	2-year	106	2,994	13	646 (474)
Programme integrity compromised 2016-2019	**1-year**	141	2,531	14	279 (186)
	2-year	72	770	9	122 (114)

Results Summary

Statistically significant results of the one and two-year violent reoffending measures

This table shows there were 2 statistically significant results among the one-year analyses. There were no statistically significant results among the two-year analyses. The one-year analyses provide evidence that:

Participated in RESOLVE only

One-year

- **Participants are less likely to commit a violent reoffence** than non-participants.
- **Participants commit fewer violent reoffences** than non-participants.

Tables 6-8 show the measures of violent reoffending. Rates are expressed as percentages and frequencies expressed per person. The sub-analysis is highlighted in green if it is statistically significant (at the 0.05 level). Effect sizes (expressed as Cohen's *d* statistic) are also included to indicate the strength of the relationship between the two groups. The estimated differences shown are the 95% confidence intervals for the differences between the relevant treatment and comparison group measures.

Table 6: Proportion of men who committed a proven violent reoffence in a one and two-year period after support from RESOLVE, compared with matched comparison groups.

Analyses	Outcome measure	Number in treatment group	Number in comparison group	One and two-year proven violent reoffending rates: Treatment group rate (%)	Comparison group rate (%)	Estimated difference (% points)	Standardised effect size (Cohen's *d*)	Statistically significant difference?	p-value
Overall	**1-year**	2,497	108,914	5.7	6.3	-1.5 to 0.3	-0.025	No	0.21
	2-year	1,923	97,074	11.6	12.0	-1.9 to 1.0	-0.013	No	0.56
Participated in RESOLVE only	**1-year**	1,727	81,540	5.9	7.2	-2.4 to -0.1	-0.051	Yes	0.03
	2-year	1,315	105,123	12.2	13.1	-2.6 to 0.9	-0.025	No	0.35
Completed RESOLVE	**1-year**	2,189	82,212	5.2	6.1	-1.8 to 0.1	-0.038	No	0.07
	2-year	1,691	128,638	11.1	11.5	-1.9 to 1.1	-0.012	No	0.63
Did not complete RESOLVE	**1-year**	290	102,927	9.3	8.5	-2.6 to 4.2	0.028	No	0.65
	2-year	231	91,849	15.2	14.7	-4.3 to 5.1	0.011	No	0.86
Participants who met ideal suitability criteria	**1-year**	812	20,302	5.5	5.6	-1.7 to 1.5	-0.004	No	0.90
	2-year	589	23,301	11.2	10.0	-1.4 to 3.8	0.039	No	0.36
Programme integrity broadly maintained 2014-2016	**1-year**	239	4,751	8.4	6.2	-1.5 to 5.7	0.082	No	0.24
	2-year	228	4,758	12.7	10.3	-2.0 to 6.9	0.076	No	0.28
Programme integrity compromised 2014-2016	**1-year**	62	1,762	6.5	6.2	-6.1 to 6.7	0.012	No	0.93
	2-year	59	1,139	15.3	16.3	-10.8 to 8.6	-0.030	No	0.82
Programme integrity broadly maintained 2016-2019	**1-year**	189	4,337	5.3	8.2	-6.3 to 0.4	-0.117	No	0.08
	2-year	106	2,994	12.3	15.8	-10.1 to 2.9	-0.103	No	0.28
Programme integrity compromised 2016-2019	**1-year**	141	2,531	9.9	7.4	-2.5 to 7.7	0.092	No	0.32
	2-year	72	770	12.5	14.8	-10.5 to 5.9	-0.067	No	0.57

Table 7: Number of proven violent reoffences committed in a one and two-year period by men who received support from RESOLVE, compared with matched comparison groups.

				One and two-year proven violent reoffending frequencies (offences per person)					
Analyses	**Outcome measure**	**Number in treatment group**	**Number in comparison group**	**Treatment group frequency**	**Comparison group frequency**	**Estimated difference**	**Standardised effect size (Cohen's *d*)**	**Statistically significant difference?**	**p-value**
Overall	**1-year**	2,497	108,914	0.08	0.09	-0.03 to 0.01	-0.025	No	0.21
	2-year	1,923	97,074	0.16	0.18	-0.04 to 0.004	-0.035	No	0.10
Participated in RESOLVE only	**1-year**	1,727	81,540	0.08	0.11	-0.04 to -0.001	-0.049	Yes	0.04
	2-year	1,315	105,123	0.18	0.20	-0.06 to 0.004	-0.043	No	0.09
Completed RESOLVE	**1-year**	2,189	82,212	0.07	0.09	-0.03 to 0.001	-0.038	No	0.07
	2-year	1,691	128,638	0.16	0.18	-0.04 to 0.005	-0.035	No	0.12
Did not complete RESOLVE	**1-year**	290	102,927	0.13	0.13	-0.05 to 0.06	0.018	No	0.76
	2-year	231	91,849	0.22	0.23	-0.09 to 0.07	-0.015	No	0.80
Participants who met ideal suitability criteria	**1-year**	812	20,302	0.08	0.09	-0.04 to 0.02	-0.015	No	0.67
	2-year	589	23,301	0.15	0.14	-0.03 to 0.05	0.020	No	0.61
Programme integrity broadly maintained 2014-2016	**1-year**	239	4,751	0.12	0.08	-0.02 to 0.09	0.095	No	0.19
	2-year	228	4,758	0.19	0.16	-0.04 to 0.11	0.065	No	0.34
Programme integrity compromised 2014-2016	**1-year**	62	1,762	0.06	0.08	-0.08 to 0.05	-0.042	No	0.68
	2-year	59	1,139	0.19	0.29	-0.23 to 0.03	-0.154	No	0.13
Programme integrity broadly maintained 2016-2019	**1-year**	189	4,337	0.11	0.12	-0.10 to 0.07	-0.033	No	0.68
	2-year	106	2,994	0.31	0.27	-0.17 to 0.24	0.039	No	0.72
Programme integrity compromised 2016-2019	**1-year**	141	2,531	0.13	0.10	-0.04 to 0.11	0.075	No	0.40
	2-year	72	770	0.15	0.20	-0.15 to 0.06	-0.090	No	0.41

Table 8: Average time to first proven violent reoffence in a one and two-year period for men who received support from RESOLVE, compared with matched comparison groups.

		Average time to first proven general reoffence in a one and two-year period, for reoffenders only (days)							
Analyses	**Outcome measure**	**Number in treatment group**	**Number in comparison group (unweighted)**	**Treatment group time**	**Comparison group time**	**Estimated difference**	**Standardised effect size (Cohen's *d*)**	**Statistically significant difference?**	**p-value**
Overall	**1-year**	143	9,979	169	158	-6 to 27	0.107	No	0.21
	2-year	223	14,408	318	303	-12 to 41	0.072	No	0.29
Participated in RESOLVE only	**1-year**	102	7,825	164	156	-12 to 28	0.081	No	0.41
	2-year	161	16,108	313	303	-22 to 41	0.048	No	0.54
Completed RESOLVE	**1-year**	114	7,187	166	163	-15 to 22	0.034	No	0.72
	2-year	188	19,062	323	307	-12 to 46	0.083	No	0.26
Did not complete RESOLVE	**1-year**	-	-	-	**-**	**-**	**-**	**-**	**-**
	2-year	35	14,740	288	297	-81 to 62	-0.044	No	0.79
Participants who met ideal suitability criteria	**1-year**	45	1,379	166	169	-32 to 24	-0.038	No	0.80
	2-year	66	2,684	313	304	-39 to 57	0.046	No	0.71
Programme integrity broadly maintained 2014-2016	**1-year**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
	2-year	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
Programme integrity compromised 2014-2016	**1-year**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
	2-year	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
Programme integrity broadly maintained 2016-2019	**1-year**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
	2-year	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
Programme integrity compromised 2016-2019	**1-year**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**
	2-year	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**

(Note: "-" identifies suppressed results where the number of reoffenders in either the treatment or comparison group is lower than 30. Where sample sizes are relatively small, they will be unlikely to produce a statistically significant result and thus have a lower likelihood of supporting conclusions with an acceptable level of confidence.)

Acknowledgements

From the Ministry of Justice, the authors would like to thank David Dawson and Alana Diamond for their oversight of the study. Thanks to Wuraola Gerasimov for her substantial contribution to the production of the report. We would also like to thank Jessie Brener, Adam Fair, Katie Bullock and Eleanor Martin for their individual contributions in the early scoping stages for this report. Thanks are also due to Aidan Mews for his comments on this report.

The authors would also like to thank members of CSAAP (Correctional Services Accreditation and Advice Panel) who provided advice for this evaluation and accredited RESOLVE, Erica Bowen, Don Grubin and Mike Maguire. Thanks are also due to Gerald Gaes and Emma Palmer who peer reviewed the report.

Final thanks are to HMPPS Intervention Services for their communication and support throughout the evaluation process.

Contact points

Press enquiries should be directed to the Ministry of Justice press office. Other enquiries about the analysis should be directed to:

Justice Data Lab and **Reducing Reoffending Flexible Pool** teams

Ministry of Justice

7th Floor

102 Petty France

London

SW1H 9AJ

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

General enquiries about the statistical work of the Ministry of Justice can be e-mailed to: statistics.enquiries@justice.gov.uk

General information about the official statistics system of the United Kingdom is available from: www.statisticsauthority.gov.uk/about-the-authority/uk-statistical-system

Produced by the Ministry of Justice

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

Annexes

Annex 1: Methodological approaches

This Justice Data Lab (JDL) study evaluates the reoffending patterns of a cohort of treated and comparison (untreated) offenders after their release from a prison sentence, to estimate the impact of the intervention on proven reoffending. The treatment group is comprised of those who started RESOLVE during a prison sentence[7]. This includes participants where there was intention-to-treat (ITT) but did not necessarily complete the full programme requirements (see Annex 3 for the theoretical context of ITT). The comparison group includes those who did not attend RESOLVE during their sentence.

Propensity score matching

Offenders in the treatment group were matched to untreated offenders using propensity score matching (PSM). PSM is a statistical matching technique which uses factors theoretically and empirically associated with both receiving the treatment and the outcome variable (reoffending) to predict a 'propensity score' (see Annex 6 for variables). This propensity score reflects the likelihood that an offender received the intervention, given the recorded characteristics. It is a value between 0 and 1. Treatment group members were matched to similar untreated offenders, where their propensity scores were within a certain tolerance level. Where several comparison group members had propensity scores within the required tolerance for a given treatment group member, the comparison group records all received the same weighting factor. For example, if 10 comparison records were matched to a single treatment group record, each comparison group record would have a weight of 1/10 applied, with the treatment group record having a weight of 1. Where treatment group records had no corresponding comparison group record within the tolerance level, they were excluded from the analysis (their weight was set to 0). Similarly, comparison group records were given a 0 weight if their propensity scores were not close enough to any treatment group records.[8] Using the post-matched groups, the weighted reoffending rates for the treatment and comparison groups were compared.

[7] Exact duplicates have been removed but where an individual started RESOLVE in different prison sentences, they have been included as separate entries.

[8] The term for the matching methodology applied is radius matching (with replacement), based on a uniform kernel. The tolerance level (sometimes referred to as caliper or bandwidth) was selected in order to minimise the mean absolute standardised difference of all variables included in the model, while retaining as high a proportion of the treatment group as possible. Those with OASys records in the treatment group (90.4%) are only matched to those with OASys records in the comparison group. Likewise, those without OASys records in the treatment group are only matched to comparison group members where this data is unavailable. This is an additional step in the matching process. See *Matching process and quality* section of this annex for more detail. Technical information on the standard JDL methodology and matching process can also be found on pages 13-19 at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/392929/justice-data-lab-methodology.pdf

PSM can provide a robust quasi-experimental approach, although offenders can only be matched on observable variables. While extensive efforts were undertaken in identifying relevant factors, it is possible that unobserved factors could influence the results that emerge from this research.

Criteria for selection onto RESOLVE

The RESOLVE programme is for medium-risk violent adult male offenders (aged 18 years and above). RESOLVE participants therefore needed to have met each of the following criteria, guided by the risk/need/responsivity principle of rehabilitation:

1. Violent index offence or previous violent offence (convicted or non-convicted in previous two years)
2. Were identified as of medium risk of reoffending and violence
3. Had needs addressed by the programme
4. Were able to engage in and benefit from the intervention

Final data sets

One-year outcome measures

The dataset consisted of 2,509 convicted violent offenders who started treatment under the prison-based RESOLVE programme between 2011 and 2018 in England and Wales. A comparison pool of 367,027 records was created from Police National Computer (PNC) records and the Offender Assessment System (OASys) database (where available).

Using PSM, the treated and untreated offenders were then matched using 87 factors derived from further PNC records, prison release records, and the OASys database (where available). These variables are listed in Annex 6.

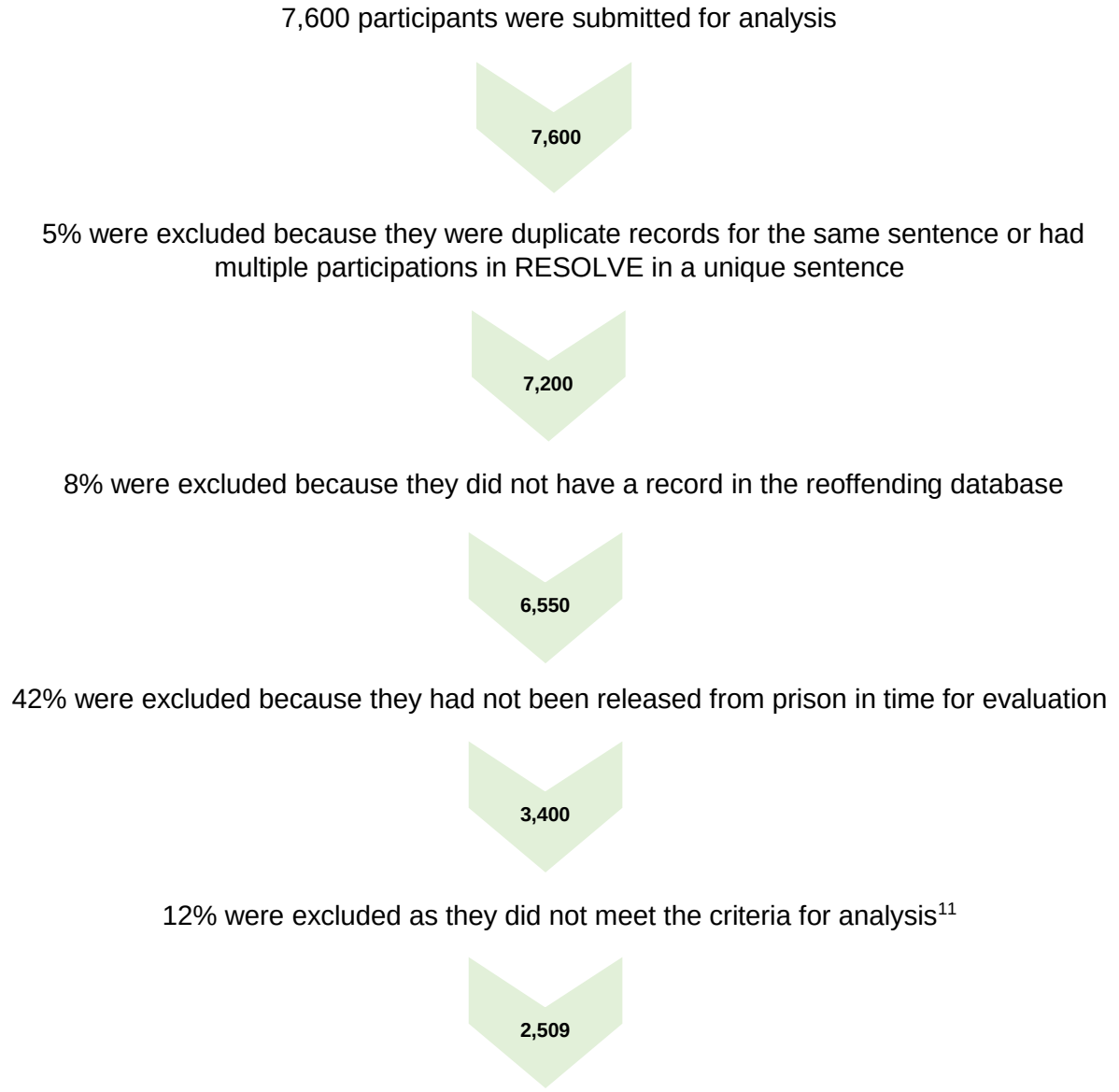
Two-year outcome measures

The two-year reoffending measures were created for general reoffences and violent reoffences. These were based on a methodology aligned with the one-year measures[9], applied to a dataset from the PNC that includes an additional 12-month follow-up period.

The dataset consisted of 1,926 convicted violent offenders who started treatment under the prison-based RESOLVE programme between 2011 and 2017 in England and Wales. A comparison pool of 334,628 records was created using the same process as the one-year measures. The smaller number of offenders in the treatment group are a result of the additional 12 months required for a participant to be at liberty to reoffend. In other words, participants will need to have been released from prison for a minimum of 24 months to commit an offence (plus an additional 6 months for this offence to have been processed in the system). As such, only those offenders who left prison between 2011 and August 2017 are considered for the two-year analyses.

[9] The methodology used to create the two-year proven reoffending rates is consistent with that used for the production of Ministry of Justice proven reoffending National Statistics (https://www.gov.uk/government/collections/proven-reoffending-statistics). Any minor variation in the results is a consequence of different PNC snapshots used to construct a two-year rate versus the one used at the earliest opportunity to create the one-year general reoffending rates, published in the proven reoffending National Statistics bulletin. All measures were validated and compared for consistency, and no issues were identified.

Figure A1.1: Treatment and comparison groups: exclusions and final numbers[10]



[10] Numbers are rounded to the nearest 50.

[11] The criteria included:

1. a prison sentence that matched to the time of treatment
2. exclusion of those with the offence group 'Sexual Offences' as their index offence and/or a previous history of sexual offending

Violent offence and reoffence categorisation

For the purposes of this analysis, offences were categorised using the type of crime (as represented by the offence code) and grouped using the standard offence groups reported in Ministry of Justice National Statistics on criminal justice outcomes and reoffending. In line with Ministry of Justice Proven Reoffending statistics, all proven reoffences (including violent) were considered for the general offending analyses[12].

The reoffences for the violent offending analyses were selected as follows:[13]

- Violence against the person
- Robbery

This selection consists of offence groups that represent acts of (non-sexual) violence, for consistency with other National Statistics. Only the most serious ('principal') offence has been considered for each case. It should be noted that, as with the two-year reoffending measures, the violent reoffending measures are based on methodology aligned with the one-year general reoffending measure.

Imputation

In statistics, imputation is the process of replacing missing data with substituted values. Imputation was used to deal with a small proportion of missing OASys data. The first consideration was where no OASys record was pulled from the OASys source data into the treatment and comparison group data sets, and all the OASys data was missing. This accounted for 10% (241) of people in the treatment group. For these records, blanks were left unchanged and treated as separate categorical variables.

Where an OASys assessment record was present in the data for a given person, but not all the OASys variables were populated, the following approach was taken:

- Where a missing entry could be logically inferred, mostly by reference to other OASys variables, then they were 'logically imputed'
- Where a missing entry related to an optional question in the OASys assessment, blanks were left unchanged and treated as separate categorical variables (essentially as 'no response')

[12] Proven reoffences are defined on page 8 of this document produced by the Ministry of Justice Proven Reoffending team:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/930457/Guide-to-proven-reoffending_Oct20.pdf

[13] Produced by Ministry of Justice and Home Office officials. Offenders' criminal records obtained from the PNC specify each individual offence according to a Home Office offence code. Each violent offence has a corresponding description, which often provides further information about the nature of the act. In the table provided by the link, the offence group can be filtered to only include 'Violence against the person' and 'Robbery':
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/886486/offence_group_classification2019.xlsx

- Where an OASys assessment record was not sufficiently complete (where less than three-quarters of the OASys variables were populated for a given record), blanks were left unchanged and treated as separate categorical variables
- Where more than three-quarters of the OASys assessment record was populated, statistical/multiple imputation was carried out to impute missing data[14]. This process models existing OASys data together with other demographic and offence variables, for all offenders in the study, to predict and populate entries for the missing OASys variables.

Matching process and quality[15]

An extensive collection of matching factors (see Annex 6) was identified following literature reviews and consultation with colleagues across the Ministry of Justice, HMPPS and CSAAP (Correctional Service Accreditation and Advice Panel, see Annex 4). The variables chosen were deemed of theoretical importance to programme selection and the relevant outcome measures. Additionally, factors were empirically investigated to determine whether they were related to programme selection or relevant outcome measures at the 20% significance level. Those that were theoretically and empirically significant were used to match participants to the comparison group. In this matching process, the main criteria for choosing technical specifications were: maximising the matching quality (via standardised mean differences of the matching variables); and minimising the number of treated individuals lost because they could not be matched. It should be noted that a unique model was run for each sub-analysis and therefore these criteria were independently considered each time.

For more information, standardised differences in means between the matched treated and comparison groups are presented in the Standardised Differences annex attached separately to this report. The standardised mean difference is expressed as a percentage; the smaller the percentage the more similar the groups are on that variable. The traffic light criteria used is as follows:

Green (very good): the two groups were well matched (less than or equal to 5%)

Amber (good): the two groups were reasonably matched (between 6% and 10%)

Red (poor): the two groups were poorly matched (greater than 10%)

The overall matching quality achieved based on the observed factors was very good. In other words, the mean absolute standardised differences for all models were less than 5%. However, standardised differences for each variable may vary. Though the majority are very well matched[16], where sample sizes are smaller in specific sub-

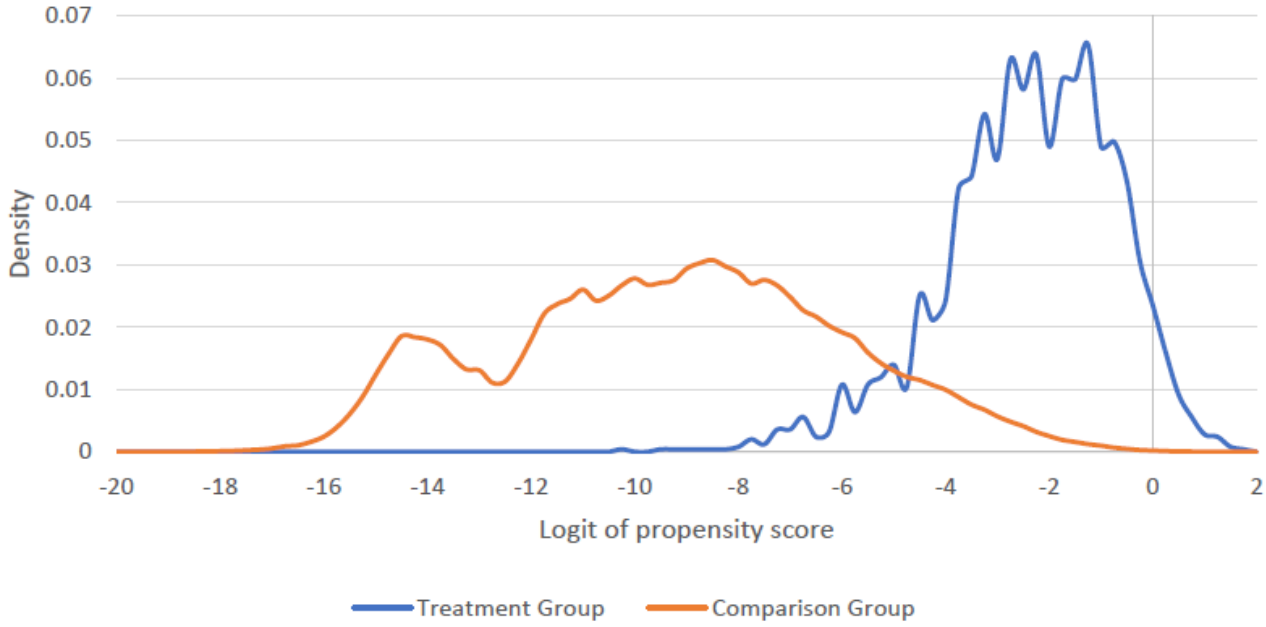
[14] The Multivariate Imputation by Chained Equations (MICE) R package was used for this condition. This is built from the algorithm described in Van Buuren and Groothuis-Oudshoorn (2011).
[15] Technical information on the standard JDL methodology and matching process can be found on pages 13-19 at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/392929/justice-data-lab-methodology.pdf
[16] For every analysis except programme integrity analyses, all variables in the final model were green (very good).

analyses, individual variables may not be as well matched[17]. Please see the standardised differences per sub-analysis for more detail on individual factors.

Common support

The propensity scores represent the probability of beginning RESOLVE treatment given an offender's recorded characteristics. There was a large region of common support (where the propensity scores for the treatment and comparison groups overlap), which implies they can be matched (see A1.2). After matching, the distributions of propensity scores in the two groups were closely aligned, showing that the likelihood of receiving treatment as predicted by the PSM model is well matched across the groups (see A1.3).

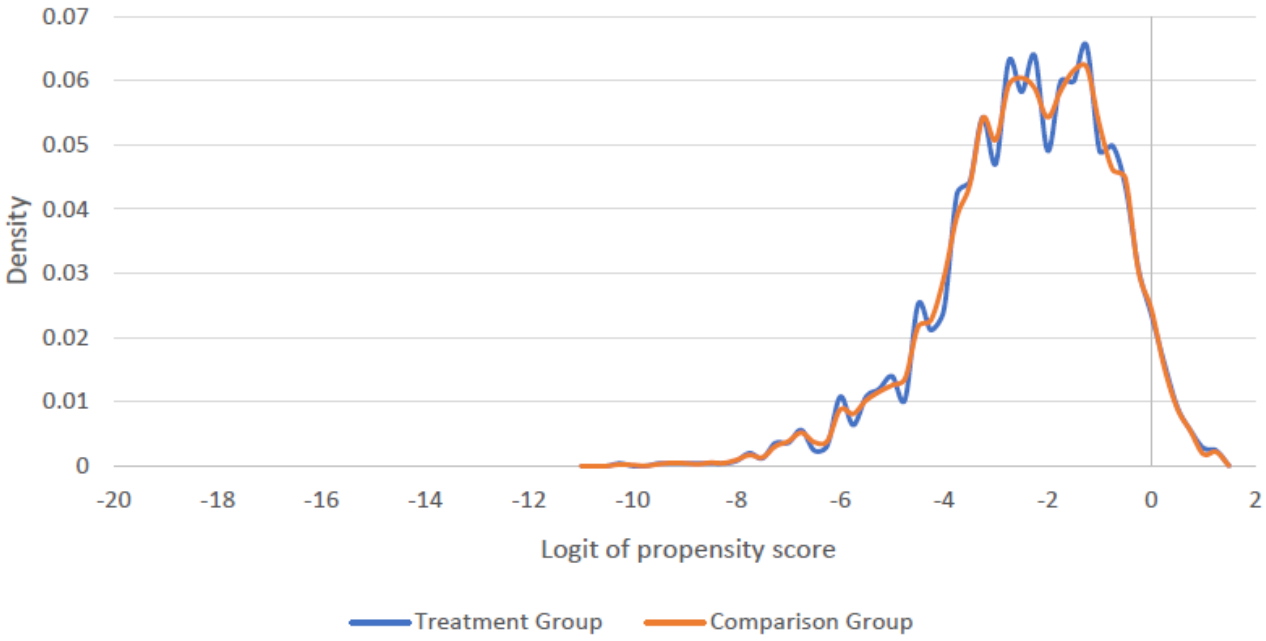
Chart A1.2: Distribution of the logit of the propensity scores across the treatment and comparison groups, before matching, for the one-year general reoffending headline analysis



In addition to the comparison above, a sensitivity analysis was run with a restriction on common support, so that treatment group members were automatically excluded where their propensity scores were outside the overall range of propensity scores of the comparison group. Similarly, comparison group members were automatically excluded where their propensity scores were outside the overall range of propensity scores of the treatment group. This run produced results that were very similar to the headline analysis, showing that any outliers have been appropriately excluded in our matching process.

[17] For example, the sub-analysis with the highest proportion of poorly matched variables was the two-year general reoffending analysis of programme integrity compromised (2016-2019), where in the final model 9% of variables had an absolute standardised difference above 10%.

Chart A1.3: Distribution of the logit of the propensity scores across the treatment and comparison groups, after matching, for the one-year general reoffending headline analysis



Explanation of sub-analyses

The headline analyses were performed on all those who received treatment from RESOLVE in custody and for whom we could access reoffending data. Further analyses were also conducted to examine the specific effects of RESOLVE on relevant sub-groups. For the sub-analyses outlined below, there is an evidence base outlined in the literature review in Annex 3. This references the varying impact of treatment in different circumstances, such as completers vs non-completers or programme integrity. Analyses were conducted on the following characteristics.

1. Participation in RESOLVE only

This sub-analysis measures the isolated treatment effect of the RESOLVE accredited programme for those who have been on no other accredited programme. If offenders have participated in other intervention programmes, there could be combined effects of receiving treatment from multiple programmes[18]. This sub-analysis was conducted to partially control for any such effects.

This analysis looked at those who solely participated in RESOLVE and did not take part in any other accredited programme, compared to a matched comparison group who had not taken part in any accredited programmes during their sentence. Each reoffending record refers to the year/two years following release from a specific prison sentence; this sub-analysis only considered other accredited programmes delivered in custody during that specific prison sentence (for the treatment group, the prison sentence during which RESOLVE was delivered). It did not control for any other

[18] Cortoni, Nunes & Latendresse (2006) controlled for attendance on other treatment programmes as a potential factor for examining the effects of treatment. For more information, see Annex 3.

interventions or rehabilitative activity delivered before or after this specific sentence, either in custody or in the community.

Table A1.4: Proportion of treatment group who participated on other accredited programmes delivered in custody during the relevant sentence

Treatment Group (n = 2,509)		
Programme	**Record**	**Percentage of records**
Any other accredited programme	771	30.7%
Other violence accredited programme[19]	121	4.8%
Thinking Skills Programme (TSP)	580	23.1%
Controlling Anger and Learning to Manage it (CALM)	76	3.0%
Building Skills for Recovery (BSR)	67	2.7%
Prisons Addressing Substance Related Offending (P-ASRO)	60	2.4%
Twelve Step Programme	36	1.4%
Enhanced Thinking Skills (ETS)	27	1.1%
Healthy Relationships Programme (HRP)	27	1.1%
Kainos Challenge to Change (Kainos CTC)	25	1.0%
Alcohol Related Violence (ARV)	24	1.0%
Cognitive Skills Booster (CSB)	16	0.6%
Control of Violence for Angry Impulsive Drinkers (COVAID)	15	0.6%
Building Better Relationships (BBR)	14	0.6%
Short Duration Programme (SDP)	13	0.5%
Democratic Therapeutic Community Model (DTC)	12	0.5%
Alcohol Dependence Treatment Programme (ADTP)	9	0.4%
Bridges (short version of RAPT)	6	0.2%
Prison Partnership Twelve Step Programme (PPTSP)	5	0.2%
Self-Change Programme (SCP)	5	0.2%
CP	4	0.2%
FOCUS	4	0.2%
Juvenile Enhance Thinking Skills (JETS)	4	0.2%
Focus on Resettlement (FOR)	3	0.1%
RAPt 12-Step Alcohol Dependency Treatment Programme	2	0.1%
Adapted Better Lives Booster (ABLB)	1	<0.1%
Becoming New Me (BNM)	1	<0.1%
Cognitive Self Change Programme (CSCP)	1	<0.1%

Note: one individual can appear in the count for several different accredited programmes. Some of the programmes listed in Table A1.4 are no longer being delivered. For a list of the current suite of accredited offender behaviour programmes see the following website: https://www.gov.uk/guidance/offending-behaviour-programmes-and-interventions

[19] Other violence APs includes: CALM, ARV, COVAID, SCP, CP, CSCP.

2. Completion of the Programme

The effect of those participating in RESOLVE may have been different if participants did not complete the programme. This sub-analysis aimed to determine whether the treatment effect was different for those who completed the programme and also those who started RESOLVE but did not complete it.

The treatment group was divided into RESOLVE completers and non-completers[20]. Subsequently, each sub-group was matched to the wider 'no treatment' comparison group pool used for the headline analysis. This measured the treatment effect for completers separately from the treatment effect for non-completers.

This analysis looked at two groups:

1) Those who completed RESOLVE compared to a matched comparison group who did not receive the programme
2) Those who did not complete RESOLVE compared to a matched comparison who did not receive the programme

3. Ideal Suitability for RESOLVE

This analysis looked at those participants who fulfilled the most practically stringent RESOLVE criteria for treatment selection. In other words, it examined a subgroup of 'ideal' RESOLVE candidates by removing those deemed **potentially less suitable** to receive treatment. The remaining treatment group were those who were a guaranteed target group of the programme (compared to a matched comparison group who received no treatment).

Although RESOLVE is suitable for many people who may not fit this profile, this sub-analysis specifically concentrated on those whose risks and needs were measurably best suited to the RESOLVE selection criteria. A participant must have met each of the five criteria to be considered 'ideally suitable'. The comparison group was filtered by the same criteria and matched to the treatment group. Please note: there was no clear indicator for domestic violence and therefore a proxy indicator has been used[21].

The criteria included those with:

- An OGRS (see glossary) between 50 and 74 **AND/OR** an OVP score (see Glossary) between 30 and 59
- At least three risks or needs from the OASys variables specified for selection procedures in the RESOLVE handbook

[20] Both Cortoni et al. (2006) and a follow-up of that study by Higgs et al (2019) both found a statistically significant effect for a sub-group of completers but not for those where there was intention-to-treat. For more information, see Annex 3.

[21] The proxy indicator used was from Section 6 of the OASys assessment (Perpetrator of domestic violence or abuse towards a partner at any time) and is taken from matched OASys assessments.

The criteria excluded those recorded:

- With a history of domestic violence (to exclude all those whose violence could be predominantly within the context of intimate relationships and for whom the programme is ideally not intended).[22]
- With an index offence or history of sexual offending (for whom the programme is ideally not intended)[23]
- As having been on another accredited programme (to isolate the effect of RESOLVE only on this ideal sub group and to remove any potentially negative or positive effects of receiving an additional programme within the same sentence)

The numbers of participants remaining after filtering on each individual criterion is outlined below. Those with an index offence or history of sexual offending were excluded in the initial creation of the treatment group and are therefore not included in the counts below. It should be noted that these criteria have been considered separately and an individual may be excluded for more than one reason:

- 2248 of 2509: met OASys risk and needs (261 participants excluded)
- 1831 of 2509: met OVP or OGRS scores (678 participants excluded)
- 1738 of 2509: RESOLVE only (771 participants excluded)
- 1347 of 2509: No Domestic Violence flag (1,162 participants excluded)

After combining these filters, the cumulative number of individuals remaining in the **pre-matched** ideal suitability treatment group for the one-year analysis was 815 (out of 2,509).

4. Programme Integrity (see Annex 5)

This sub-analysis sought to evaluate the extent to which the programme integrity[24] of RESOLVE may have an impact on outcome[25]. Quality assurance of programme delivery was completed by HMPPS through the Interventions Integrity Framework (IIF). This was conducted for two different timeframes: 2014-2016 and 2016-2019. Given the differences in assessment between the timeframes, two cohorts of treatment and comparison groups were created. Using the quality assurance framework, the two

[22] The assessment manual specified that, "If violence history is predominantly within the context of intimate relationships, refer to a domestic violence programme as an alternative."
[23] The assessment manual specified that, "If violence history is sexually motivated or linked to sexual offending, refer to a sex offender treatment as an alternative."
[24]Data measuring the delivery quality of RESOLVE was collected by HMPPS Interventions Services using different methodologies across different time scales (2014-2016 and 2016-2019). Assessments only measured information which was recorded and therefore may not be fully representative of delivery settings.
[25] Lipsey, Landenberger, & Wilson (2007) found that quality of delivery had an impact on the success of treatment. Gannon et al. (2019) also suggested that certain delivery factors were likely to influence the efficacy of treatment. For more information, see Annex 3.

cohorts were sub-divided by two types of prison establishments: prisons where the programme integrity was broadly maintained and prisons where the programme integrity was compromised. In the rare cases that programme integrity could not clearly be categorised, those establishments were omitted from the analysis.

As this sub-analysis focused specifically on prison establishments, the treatment effect being measured is closely tied to prison-specific circumstances. To be included, participants had to spend at least 50% of their sentence in the prison where they participated in RESOLVE and be released from that same prison. This ensured the sample was composed of individuals who spent most of their sentence in the same prison where they received RESOLVE and therefore minimised prison-based fixed effects[26].

As a result, approximately 75% of participants were not included because the prison they were in was not classified in either of the programme integrity time categories, or they did not spend at least 50% of their time in the prison assessed. The sample size was therefore substantially smaller relative to other analyses. Smaller sample sizes are less likely to produce a statistically significant result and thus have a lower likelihood of supporting conclusions with an acceptable level of confidence.

This analysis examines the differences in the treatment effect when:

1) Programme integrity was broadly maintained when delivered in prison settings that met the guidelines outlined in programme and management manuals, compared to a matched comparison group. This category of prison setting was given either a Green or Green/Amber in the quality assurance ratings (see Annex 5)
2) Programme integrity was compromised when delivered in prison settings that did not meet the guidelines outlined in programme and management manuals, compared to a matched comparison group. This category of prison setting was given either an Amber/Red or Red in the quality assurance ratings (see Annex 5).

The treatment and comparison groups were divided as follows:

Programme integrity broadly maintained vs comparison (2014-2016)	Programme integrity compromised vs comparison (2014-2016)
Programme integrity broadly maintained vs comparison (2016-2019)	Programme integrity compromised vs comparison (2016-2019)

Standardised effect sizes

Effect sizes can be considered useful tools when evaluating the strength of a statistical relationship between two groups, or any differences between them. A standardised effect size calibrates the difference between outcomes of the treatment and comparison groups, in terms of the standard deviation. By standardising the measure,

[26] This involved creating an estimate for an equivalent intervention start date for comparison group members, using imputation techniques informed by the relationships between key variables in the treatment group data.

the units of measurement are removed, with the aim of making them easier to evaluate and compare (including comparisons with other similar interventions).

Cohen's *d* has been selected as the measure of standardised effect size. This is calculated, for each outcome measure, as the difference between the averages for the treatment and comparison groups, divided by the pooled standard deviation for both groups. It should be noted that averages for each group are calculated as weighted averages, based on the weights assigned to each record within each group as part of the propensity score matching process.

The formula is given below, where M_1 is the weighted average for the treatment group, M_2 is the weighted average for the comparison group, and S_1^2 and S_2^2 are the weighted variances of the treatment and comparison groups, respectively.

$$d = (M_1 - M_2) / \sqrt{\frac{S_1^2 + S_2^2}{2}}$$

To aid the interpretation of effect sizes, the Cohen's *d* statistic is typically categorised as follows (Cohen, 1988):

- **Small**: denoting an effect size greater than or equal to 0.2 but below 0.5
- **Medium**: denoting an effect size greater than or equal to 0.5 but below 0.8
- **Large**: denoting an effect size greater than or equal to 0.8

Small, medium and large categories are however relative to the area of behavioural science or specific research method being employed (Cohen, 1988). Moreover, effect sizes taken from 81 criminology meta-analyses were on average small to medium (Barnes et al., 2020). See Annex 3 for a review of selected meta-analyses specifically for offender behaviour programmes and details on effect sizes are included.

The point in estimating effect sizes is to provide some indication about the strength of an effect. This effect is separate from statistical significance which indicates whether or not the effect is likely due to chance. In this evaluation, where there are 'small' effect sizes (above 0.2 but below the 0.5 medium effect), the confidence intervals tend to be wider around the results. In other words, there is a higher degree of uncertainty in these estimates. Therefore, effect sizes should be considered carefully alongside the full suite of information and results.

Sensitivity analyses

A series of sensitivity analyses were run on both the one-year general and two-year violent headline analyses, to measure the possible effect of certain methodological decisions on the results. It should be noted that the chosen method was selected on the basis that the model should include variables both theoretically and empirically associated with selection and outcome. Given other theoretical considerations, having a lower mean absolute standardised difference does not necessarily mean that the matching is better. The following table provides an explanation of selected sensitivity analyses and their results, with reference to the one-year general reoffending headline analysis.

Sensitivity	Explanation	Findings
Standard approach	The chosen approach is displayed here for comparison against other sensitivities. Radius matching (with replacement) with a uniform kernel was applied.	143 variables were included in the final model with a mean absolute standardised difference of 0.78%
Parsimonious	To explore the effect of having fewer variables in the model (tougher constraints imposed when determining model variables).	79 variables were included in the final model with a mean absolute standardised difference of 0.79%. The results and matching quality were very similar to the standard approach.
Non-parsimonious	To explore the effect of having more variables in the model (looser constraints imposed when determining model variables).	176 variables were included in the final model with a mean absolute standardised difference of 0.74%. The results and matching quality were very similar to the standard approach taken.
Epanechnikov kernel	This explores using an alternative type of kernel matching sometimes used for PSM models.	143 variables were included in the final model with a mean absolute standardised difference of 0.77%. The matching quality and also the results were very similar to the standard approach taken.
Exclusion of OASys (including OVP and OGRS) variables from the model	To explore the effect of including OASys data in the model.	64 variables were included in the final model with a mean absolute standardised difference of 0.56%. The results were very similar to the standard approach taken though the matching quality was marginally better than the standard approach.

The results of equivalent sensitivity analyses for the two-year violent reoffending headline analysis were also similar to the standard approach.

Limitations and caveats

- Whilst this study uses a recognised evaluation methodology (propensity score matching), which is considered level 4 on the SMS (Scientific Methods Scale)[27] it is not as robust as a randomised control trial or a prospectively-matched evaluation. For a detailed discussion of the strengths and limitations of propensity score matching, see Mews, Hillier, McHugh, & Coxon (2013), and Ministry of Justice (2015).

- While propensity score matching can provide a robust quasi-experimental approach, it can only match, and therefore reduce bias, on observed factors (information that is recorded). Despite efforts to include all observed factors known to be predictive of selection onto RESOLVE and of reoffending risk, the importance of information that is not recorded cannot be known and it is possible that unobserved factors could influence these results. This is a particular risk for the completers and non-completers sub-analyses. All efforts have been made to match on characteristics which may be proxies for completion. However, it is possible that the characteristics did not fully control for the lack of an observable "completion" counterfactual filter in the comparison group.

- These analyses only concern reoffending outcome data. Although outside the scope of Justice Data Lab analyses, there may be other important outcomes to consider for rehabilitation interventions. Examples of this might include increased employability or positive attitudes.

- A sub-analysis was performed to isolate the effect of RESOLVE from other accredited programmes. It is possible that comparison individuals (those who did not attend RESOLVE during the index sentence) instead undertook a non-accredited cognitive behavioural programme after being released from prison or engaged in other rehabilitative activity that might have had an impact on reoffending rates.

- This evaluation measures a treatment effect using proven reoffending outcomes in accordance with the standard Ministry of Justice definition as used in Proven Reoffending National Statistics.

[27] This is a five-point scale ranging from 1, for evaluations based on simple cross-sectional correlations, to 5 for randomised control trials. Systematic reviews and meta-analyses typically include impact evaluations scored 3 or above on the SMS to attempt to understand what works. (See Sherman et al. 1998).

- As such, the study only accounts for proven reoffending. This does not measure treatment effects on crimes that are committed but not recorded by the police or do not lead to a caution or conviction.

- This evaluation does not adjust for any restriction on the time individuals are at liberty to reoffend in the community. Such restrictions include recall, additional sentences and time spent outside the UK.

- A subset of programme participants (many of whom will have completed the programme) remain in prison. These will not be included in the treatment group, since there will be no data on reoffending in the community available. Therefore, the effect of the programme for this cohort will not be reflected in these analyses.

- As a subset of general reoffending, the number of violent reoffenders will be relatively smaller. Where there is a relatively low reoffending rate (particularly in combination with smaller sample sizes), they may be less likely to produce a statistically significant result. Thus, for the violent analyses, there may be a lower likelihood of supporting conclusions with an acceptable level of confidence.

- Following PSM, results will only be presented if they consist of at least 30 offenders. From a statistical perspective, and therefore considered in JDL analyses, any sample of fewer than 30 offenders will be unlikely to produce a statistically significant result. More generally, lower sample sizes lead to a reduced likelihood of concluding with an acceptable level of confidence that any difference in reoffending between the treatment and control groups was real rather than the result of chance. This is most prevalent in the programme integrity sub-analyses where the sub-groups are smaller.

- 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 many results being found to be statistically significant is much higher than 5%. Given the number of analyses, sub-analyses and outcome measures involved in this evaluation, care should therefore be taken when interpreting the findings. While multiple correction methods can be applied to reduce the risk of concluding that the treatment effect has a positive impact due to one or more measures being found to be statistically significant, these also have the undesirable effect of increasing the likelihood that real differences will not be detected. The results presented in this report have therefore not been adjusted in this way.

Annex 2: Description of RESOLVE

Overview

RESOLVE is a moderate intensity cognitive behavioural intervention designed by HMPPS which aims to reduce violent reoffending in medium risk men. HMPPS has a history of delivering interventions to this group. RESOLVE represents an evolution of learning in relation to programming for men with violence in their offending histories. RESOLVE was accredited by the Correctional Services Accreditation and Advice Panel (CSAAP) in October 2013 (see Annex 4).

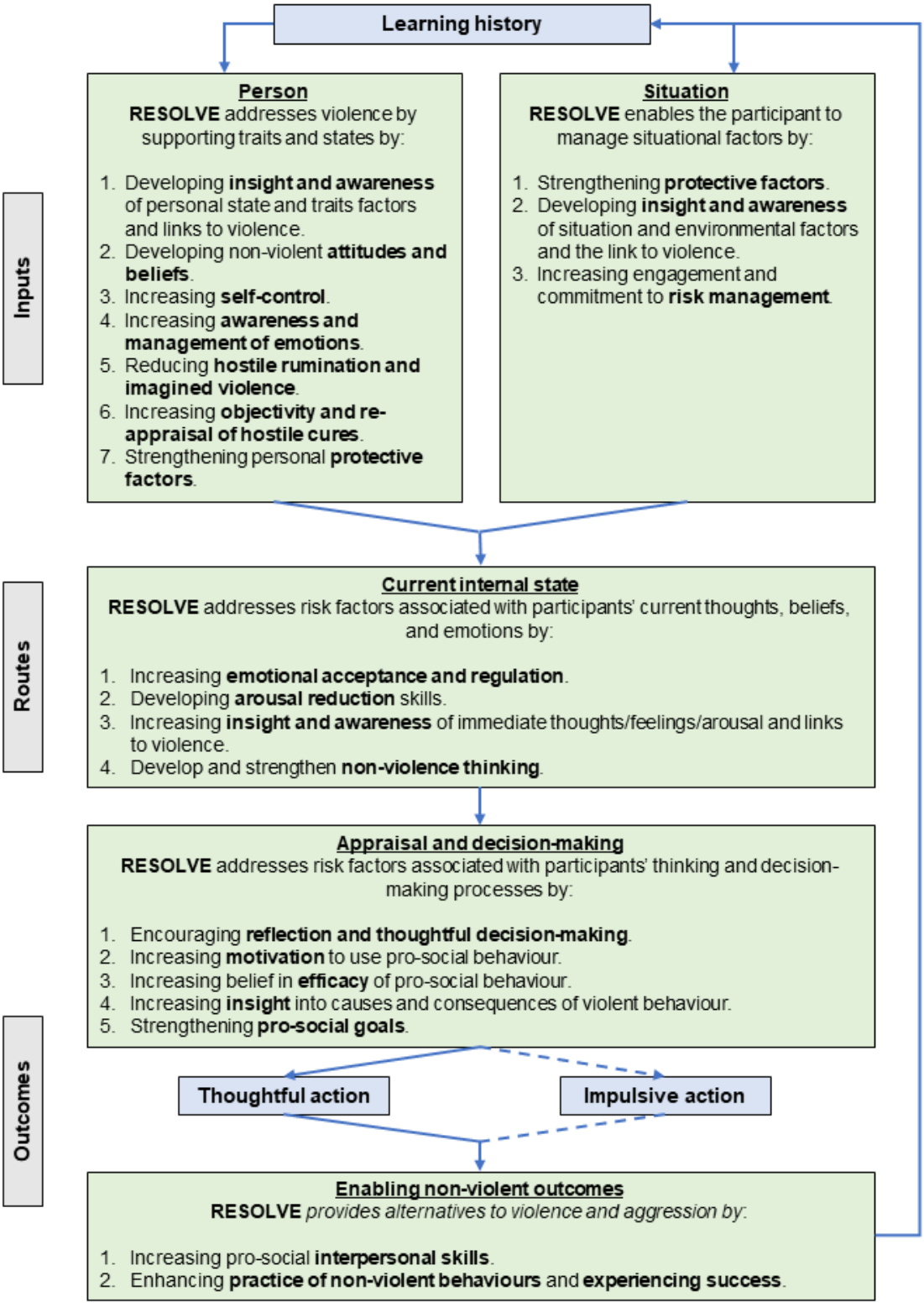
RESOLVE is predominantly a group work programme but incorporates several individual sessions and offers opportunities to tailor programme content to meet the individual needs of participants. The multi-modal design of the programme offers psycho-education alongside the development of cognitive and behavioural skills. These are designed to help participants address and modify unhelpful behaviours and build on their strengths so that they can develop a positive, pro-social identity for the future.

Theories of violence

The literature base offers multiple theories of violence, including social perspectives, biological, psychodynamic, behavioural and others. However, individual factors alone are unlikely to account for the complexities of human behaviour, or account for individual differences in these behaviours. Current thinking asserts that aggressive behaviours occur due to a combination of biological, psychological and social (environmental) factors (Huesmann & Guerra, 1997; Huesmann, 1998 Anderson & Bushman, 2002; Dodge, Coie & Lynam, 2006).

RESOLVE was developed to address a range of criminogenic needs associated with the use of aggression and violence. To support this approach, the programme integrates the General Aggression Model (GAM) (Anderson & Bushman, 2002) within the programme enabling the biological, psychological and social theories to be brought together into a broader multi-factorial theory of aggression. This allows individuals to explore and address their personal relationship with aggression and to consider where they are best placed to develop strengths and protective factors.

Figure A2.1 RESOLVE Model of offending based on the General Aggression Model.



Annex 3: Summary of findings from selected meta-analytic reviews of the efficacy of treatment for violent offenders

In the fields of criminology, criminal psychology, and forensic mental health, there has been a broad enterprise, spanning several decades, to examine "what works" in reducing criminal behaviour for individuals with criminal convictions. However, there is still sparse empirical evidence that intensive psychological interventions for individuals with violent convictions leads to reduced reconviction rates or changes in factors considered reliably predictive of recidivism (Polaschek, 2017).

Typically, evidence for criminal justice interventions is presented in studies known as meta-analyses, which examine aggregate data from multiple independent studies of the same subject to determine overall trends. This annex provides a non-systematic review of four recent meta-analyses of psychological interventions for individuals with violent convictions. The aim is not to provide an exhaustive review of the current literature but to provide a reasonable context for the study that was conducted.

Evidence for cognitive skills-based offender behaviour interventions: general reoffending

Lipsey et al. (2007)

Lipsey et al. (2007) conducted a widely-cited meta-analysis examining the general impact of cognitive skills-based interventions in criminal justice internationally. Although the interventions included were not specifically targeted at individuals with violent convictions the results are of interest given the nature of the interventions. Among the main conclusions, Lipsey et al. found:

- The odds of no-recidivism after participation in an intervention were more than one and a half times as great as those for individuals in the control group (an average "mean odds ratio" of 1.53). In other words, during the 12 months post-intervention, the odds of recidivism were higher for the control group.
- Lipsey et al. reported a 40% mean recidivism rate in the control group compared to a 30% mean recidivism rate in the treated group, which they reported as a relative reduction of 25% in recidivism.
- Lipsey et al. also reported larger odds of recidivism for the control group for the "most effective configurations of CBT" with a recidivism rate of 19% for treated offenders in those programmes compared to 40% in the control groups, where control group members were 2.86 times more likely to have reoffended.

The review concludes overall that there were three key factors that influenced the effect of CBT programmes for offenders: how well the treatment was implemented, the absence or presence of certain treatment elements, and the level of risk of reoffending

of the participants selected. The authors reported that the most critical factor was implementation quality. As such, the most effective treatment programmes were those with low drop-out rates, where activity was implemented in accordance with programme manuals, and where the providers have received appropriate training in CBT. There was nothing to suggest that the form of CBT used was important, rather it was the general therapeutic approach that had a positive effect on recidivism. When a comparison was made of the different courses of treatment, the presence of two treatment elements was shown to increase the effect: training in anger management and training in interpersonal problem solving. There were, however, two treatment elements that had the opposite intention and reduced the effect: behaviour modification and the inclusion of victim impact. The results also showed that CBT had a greater effect on offenders with a high risk of recidivism, than those with lower risk.

Lipsey et al. also found, however, that there was significant variation across the individual studies in effect sizes. In other words, aggregate findings are affected by the varying quality and representativeness of the individual studies. The limitations in the cited studies included (but were not limited to): (1) the increasing age of the studies; (2) the lack of randomised studies or quasi-experimental studies (those that use statistical techniques like propensity score matching or regression discontinuity to simulate randomisation); (3) a lack of studies using proven criminal sanctions, like reconvictions, as an outcome; (4) a lack of studies using at least a one-year follow-up period; (5) a lack of studies with a sample size greater than 200; (6) a lack of studies examining programmes with an adequate level of delivery quality; (7) high levels of attrition (drop-out) in the programmes evaluated; and (8) a large number of studies (one-half of the total) examining programmes for "low" or "low-medium" offenders.

Evidence for violent offender behaviour interventions

In terms of violent offending specifically, four recent meta-analyses that explicitly focus on correctional programmes targeting individuals with violent convictions were identified. The four meta-analyses evaluate different combinations from a pool of international outcome studies of varying methodological quality. One meta-analysis found no statistically significant overall effect of treatment on reconvictions (Henwood, Chou & Browne, 2015), but was focused on anger-management rather than violent behaviour. The other three meta-analyses found positive statistically-significant overall effects of treatment on reconvictions (Gannon, Olver, Mallion & James, 2019; Jolliffe & Farrington, 2007; Papalia, Spivak, Daffern & Ogloff, 2019).

Each of these meta-analyses included and excluded studies depending on their aims and objectives. All reported notable variation in effect sizes and methodological quality across those studies. All of the meta-analyses included at least some individual studies that are known to have methodological limitations, limiting the ability to draw credible and reliable conclusions from them. The following sections outline the findings of the meta-analyses followed by an examination of the individual studies included.

Jolliffe and Farrington (2007; 2009)

The earliest, by Jolliffe and Farrington (2009), was commissioned by Swedish National Council for Crime Prevention and was an update to a 2007 report commissioned by the Ministry of Justice. Jolliffe and Farrington identified 12 relevant impact evaluations. It is worth noting that Jolliffe and Farrington's meta-analysis also included one study that examined a supervision intervention (electronic monitoring) rather than a cognitive-skills programme[28]. Jolliffe and Farrington concluded that the combined effect sizes of those studies indicated that the interventions reduced both general and violent recidivism to a statistically-significant level[29]. Overall, this represented approximately 7-9% fewer general reconvictions in treated groups and 6-7% fewer violent reconvictions. However, the size of the statistical effects for those reductions were relatively small for both general and violent reoffending[30].

Jolliffe and Farrington also examined the extent to which study features influenced their findings. They found that low-quality studies generated higher effect sizes than the high-quality studies for violent recidivism (but not general recidivism). A moderate negative correlation between the method of analysis and effect size was found. They also found that intention-to-treat (ITT) analyses (where participants are allocated based on the assignment to treatment and not on the treatment they eventually received) were associated with lower effect sizes than analyses of completers only. A moderate negative correlation between date of publication and violent recidivism effect size was speculated to 'reflect the lower methodological quality of the studies that were undertaken earlier' (p. 31).

After controlling for methodological quality, Jolliffe and Farrington (2009) found that for general reoffending, lower reoffending rates were associated with the use of cognitive skills, role-playing and relapse prevention. For violent reoffending, lower rates were associated with role-play. Although it was not possible to determine which features were most influential, their analyses suggested that interventions employing two or three of the more successful features were associated with statistically significantly higher effects for general reoffending. There were no independent programme features that were related to higher effect sizes for violent recidivism. Jolliffe and Farrington concluded that interventions for violent offenders are successful dependent on the duration and content of those interventions, but that 'evaluative research of higher methodological quality is needed before firm conclusions can be drawn' (p. 41).

[28] A study by Finn and Muirhead-Steves (2002).

[29] Statistically-significant findings were reported in both their random-effects and fixed-effects meta-analytical models. A fixed effect meta-analysis assumes all studies are estimating the same (fixed) treatment effect, whereas a random effects meta-analysis allows for differences in the treatment effect from study to study (Riley, Higgins, & Deeks, 2011).

[30] Effect sizes (reported as Cohen's *d*) were *d* = .14 (fixed model) and *d* = .18 (random model) for general reoffending. Effect sizes were *d* = .12 (fixed) and *d* = .14 (random) for violent reoffending.

Henwood et al. (2015)

In 2015, Henwood et al. conducted a meta-analysis on the effectiveness of cognitive-behavioural therapy-informed anger management interventions on reducing reoffending amongst adult male offenders. They identified 14 relevant studies, including the nine in Jolliffe and Farrington (2007). However, five were subsequently excluded for not using proven outcomes (such as reconvictions), a lack of information about sample sizes, and the allocation of drop-outs to control groups. This left nine for meta-analysis, including 6 that measured violent reoffending. They found a statistically significant overall effect indicating that the likelihoods of both general and violent reoffending were lower in treatment groups. Like Jolliffe and Farrington, however, they also found significant variance in effect sizes across the studies, but only for general reoffending and not violent reoffending.

Henwood et al. also tested for and found sub-group differences between "anger-management", "correctional programmes", and "other CBT based" forms of treatment. For those programmes they labelled as "correctional programmes" they found a statistically-significant aggregate effect of treatment on violent reoffending but no aggregate effect of treatment on general reoffending. They concluded that 'the correctional programme subgroup showed the smallest [overall] effect' (p. 289) compared to the "anger-management" and "other" programmes, and that the effect sizes for the correctional programmes did not vary to a statistically significant extent. They suggested that the difference in effects between the anger-management and correctional programmes may be due to higher risk offenders being allocated to correctional programmes, as 'violence reduction programmes typically target offenders with an extensive history of violence and criminal convictions' (p. 290).

Furthermore, Henwood et al. conducted a separate meta-analysis examining treatment completion versus non-completion. These analyses found an overall positive effect of treatment completion for both general (6 studies) and violent reoffending (6 studies). For general reoffending, a 42% reduction in reconviction was found when treatment was completed but with significant variance in effect sizes across studies. For violent reoffending, a 56% reduction in reconviction was found when treatment was completed, also with significant variance in effect sizes across studies. Again, however, the smallest effects were found for the sub-group of "correctional programmes" in which there were no significant effects of treatment completion for general recidivism (4 studies) or violent reoffending (5 studies).

Gannon et al. (2019)

The third meta-analysis was conducted by Gannon et al. (2019) and examined the impacts of specialised psychological offence treatments for individuals with sexual, violent, and domestic violence convictions. After pooling samples from all included studies across the three programme types, they found rates of offence-specific recidivism (i.e., sexual reoffending for those on sexual violence programmes, violent

reoffending for those on violence programmes) of 13.4% for treated individuals and 19.4% for untreated comparisons over an average follow-up of approximately 5.5 years. They reported relative reductions in offence-specific recidivism of 32.6% for sexual offence programmes, 36.0% for domestic violence programmes, and 24.3% for general violence programmes.

Across all programme types, they reported a statistically significant reduction in both any form of violent recidivism over an average 65.4-month follow-up (14.4% for treated and 21.6% for untreated individuals) and any form of recidivism over an average 62.4 month follow-up (30.0% for treated and 37.7% for untreated individuals). For both any violent and any reoffending, however, Gannon et al. report significant variance in effect sizes across the included studies.

For violence-focused programmes specifically (excluding domestic violence) only four studies were included. This substantially limited the conclusions that could be drawn regarding interventions designed to reduce violent reoffending. Two of these studies had been included in Jolliffe and Farrington (2007) and three had been included in Henwood et al. (2015)[31]. Gannon et al. reported a statistically significant treatment effect for general violence[32], indicating a 9.3% absolute decrease in reoffending and a 24.3% relative decrease in reoffending for the treatment group. They also report that there was no statistically significant variance in effect sizes across the included studies. No separate analysis for the impact of violence-focused treatment on violent reoffending was conducted.

Gannon et al. also examined the effect of a variety of staff and programme factors on impacts. There were not enough studies available targeting general violence to conduct moderator analyses for staff and programme variables. However, there were features considered promising for programmes focused on individuals with sexual and domestic violence convictions (when all available studies were included[33]). These features, which might also be relevant and worthy of examination for violence-focused programmes, include:

- a licenced psychologist being consistently present in treatment;
- staff supervision being provided by psychologists;
- where service quality is rated as stronger.

[31] The additional study is a 2016 study of a new iteration of a previous programme by Polaschek and colleagues. Gannon et al. did not use the Motiuk, Smiley and Blanchette (1996) study used in the two previous meta-analyses.

[32] For both their random effects model and fixed effects model, statistically-significant findings were reported in both their random-effects (odds ratio = 0.60) and fixed-effects (odds ratio = 0.60) meta-analytical models. These odds ratios were converted into Cohen's d by multiplying the log odds ratio by the square root of 3 divided by pi (see Borenstein, Hedges, Higgins, and Rothstein, 2009: equation 7.1). An odds ratio of .60 is equivalent to d = 0.28)

[33] In some analyses, Gannon et al. provided findings with a study by Mews, DiBella, and Purver (2017) removed because the large sample size used in that study relative to other studies meant it had a large influence on the direction of the findings.

Other moderators, such as programme intensity and treatment across institutions (custody or community) resulted either mixed or non-significant findings. However, these should be interpreted with care given the small number of studies included.

Papalia et al. (2019)

The fourth meta-analysis was conducted by Papalia et al. (2019). This examined the impact of psychological treatments with adult violent offenders in both correctional and forensic mental health contexts.

Papalia et al. identified 27 studies and found that, overall, there was a statistically significant difference in treated and control groups on violent and general/nonviolent reoffending. These findings indicated that the combined likelihood of a new violent reoffence was 31% lower in the treated groups compared to comparison groups representing difference of 10.2% points. For general reoffending, the likelihood of a reoffence was 35% lower in the treated groups representing a difference of 11.2% points.

Papalia et al. also examined moderators of treatment effects. They found larger effects on violent reoffending in studies that did not use an ITT analysis, had larger sample sizes, and were assessed as having lower methodological quality. They also found larger treatment effects were associated with programmes based on multimodal cognitive behavioural or anger management techniques and that are longer in duration, have more sessions per week, and are delivered by clinical and correctional staff. Group-based rather than individual delivery and delivery in dedicated units were also associated with greater differences in violent, but not general, reoffending. Additionally, they found associations between role-play and relapse prevention as well as homework, interpersonal skills, and anger management, led to larger effects on violent reoffending.

It should be noted that Papalia et al. (2019) examined the impact of psychological treatments with adult violent offenders in both correctional and forensic mental health contexts. The inclusion of 7 samples (out of 27) from forensic mental health settings (as indicated by Table 1 in Papalia et al.), such as psychiatric inpatient or outpatient facilities either in custody or in the community, distinguishes the findings of this meta-analysis from those described above and makes the findings less generalisable to treatments targeting non-psychiatric criminal justice populations (like RESOLVE).

Why would meta-analyses of the same topic produce different results?

Each of the meta-analyses used different criteria for including or excluding individual studies, based on their aims and objectives, leading to different outcomes in their statistical analyses. For example, Henning and Frueh (1996) included interventions focusing on anger management where others did not, and Papalia et al. included some studies using forensic psychiatric populations where others did not. The individual studies themselves also vary in both their outcomes and their quality. For example, Papalia et al. indicate that risk of selection bias (the presence of systematic differences between groups being compared) was high/probably high in 63% of the included studies reflecting 'the use of unmatched or poorly matched treatment and comparison groups and/or the selection of treatment completers [only]' (p. 9). Gannon et al. was the only study meta-analysis to include studies below the recognised score of "3" and above on the Maryland Scientific Methods Scale (MSMS) for meta-analyses. Studies that fall short of a score of 3 are those with no comparable control or comparison condition. As each meta-analysis used a combination from the same set of studies it is worth examining them in more detail.

Randomised controlled trials

These 6 studies use the randomised controlled trial (RCT) design, whereby participants are prospectively allocated to either an experimental group or control group on a random basis. Like any research, RCTs can be limited by the quality of their design and implementation, but they are considered the most robust method by which to evaluate interventions. This is because they create groups that are comparable on both known and unknown factors.

Robinson (1995, *n* = 2,125) used a waiting-list RCT design (where participants were allocated to a cognitive-skills training programme or placed on a waiting list for the duration of the study) using federally-incarcerated Canadian males with an approximate 2-year follow-up. They reported no statistically-significant differences in reconvictions between the control group and all programme participants for violent, robbery, and drug offences, but positive statistically-significant differences for sexual and non-violent property offences.

Barnes et al. (2017, *n* = 994)[34] randomly-allocated high-risk adult male probationers in the U.S. to either a cognitive-behavioural programme or standard intensive probation and examined reconvictions after 12 months. They reported positive statistically-significant differences between the groups, but only for a combined "any offences" outcome with no statistically-significant differences on violent or other specific offences.

[34] This paper contains data also cited in a doctoral thesis by Hyatt (2013).

Kubiak, Brenner, Bybee, Campbell, and Fedock (2016, *n* = 35**)** randomly-allocated women from a female-only U.S. prison to a short trauma-informed programme or treatment-as-usual (a longer violence-focused programme) and examined violent reconvictions after 12 months. There was no non-treatment control for comparison. They found that no women in either group were returned to prison for a serious offence or parole violation but did report a significant positive difference in "arrest/jail stays".

Watt, Shepherd, and Newcombe (2008, *n* = 269) randomly-allocated adult males with a prior conviction for an alcohol-related violent offence to a brief alcohol intervention or a non-treatment control group and examined differences on new criminal charges for violent or other offences 3 and 12 months after sentence. They reported no statistically-significant differences between the groups on either new violent or non-violent offences.

Bowes et al. (2014, *n* = 115) randomly-allocated adult male prisoners in the U.K convicted of alcohol-related violence to either HMPPS's COVAID programme plus treatment-as-usual or treatment-as-usual-only groups and examined reconviction outcomes after a fixed 6 months and an exploratory period of approximately 18-months. No statistically-significant group differences were found for violent reconvictions at either 6 or 18 months but a statistically-significant positive effect was found for general offending at 18 months (but not 6 months).

Davidson et al. (2009, *n* = 52**)** randomly-allocated U.K. community-based adult males with a diagnosis of antisocial personality disorders to either a cognitive-behavioural plus treatment-as-usual or treatment-as-usual-only groups and examined self-reported offending behaviour (not official reconvictions). They reported no statistically-significant group differences on 'incidents of any acts of verbal or physical aggression'.

Quasi-experimental studies

When a randomised controlled trial is not possible, quasi-experimental methods allow us to conduct comparison studies to estimate the impact of an intervention, albeit with limitations. Some quasi-experimental methods, like propensity score matching, use observed data and statistical theory to generate comparison groups that are more comparable than groups that are obtained through simple case-by-case variable matching. Although these techniques reduce bias on known variables, findings from quasi-experimental studies cannot account for unknown or confounding variables as randomisation can and as such, we cannot draw conclusions from the estimates they generate with the same confidence that we can from well-implemented RCTs.

Cortoni et al. (2006, *n* = 966) is a propensity score matching study with a large sample size taken from Correctional Services Canada. It controlled for a series of critical variables including attendance on other treatment programmes. They reported no statistically-significant main effects of treatment, on an ITT basis, on either general or violent reconvictions, but did report a statistically-significant effect of for those who had

completed the programme. The Motiuk et al. (1996, *n* = 120) study is an earlier evaluation of a programme also from Correctional Services Canada with a smaller sample. This used only case-matching of treatment participants, which controlled on release date, age, and sentence length, and were found to be balanced on risk. No statistically-significant differences between the two groups were found. The authors note, however, that the treated group had twice as many homicide offenders, and the control group included eleven individuals with historical sexual offences (compared with zero in the treatment group).

Since the publication of the 2019 Gannon et al. and Papalia et al. meta-analyses, this analysis has been replicated by **Higgs, Cortoni, and Nunes (2019, n = 683).** This study also utilises a propensity score matching design, with a longer-follow-up period, comparing 345 treated and 338 non-treated participants. An ITT design, they used a series of Cox regression analyses examining the association between treatment participation and violent reoffending. For violent reoffending, they found no statistically-significant difference between the treated and comparison groups but did find a significant difference for general reoffending. A series of sub-analyses found a significant effect for violent reoffending in participants from the minority indigenous Canadian community, but no similar effect for general reoffending, and a contrasting significant effect of general offending for non-indigenous-Canadian participants but no similar effect for violent reoffending. They also found a significant lower likelihood of violent reoffending for treatment completers relative to non-completers.

Case-by-case matched studies

The collection of four **Polaschek** and **Berry** studies (Polaschek, Wilson, Townsend & Daly, 2005, *n* = 82; Polaschek, 2011, *n* = 224; Polaschek, Yesberg, Bell, Casey, & Dickson, 2016, *n* = 275; Berry, 1998/2003[35], *n* = 128) appear to be possibly overlapping iterations of samples sourced from one programme at the Rimutaka Prison in New Zealand and reporting mixed outcomes. The samples are of a medium-to-high risk population, where New Zealanders of Maori ethnicity are over-represented compared to their proportion in the New Zealand population. These studies all use case-to-case variable matching and basic post-matching assessment of balance between groups. The largest study in this collection (Polaschek et al., 2016) reports statistically-significant post-matching imbalance between the treatment and control groups. The imbalance was present on age and sentence length (the control group was younger and received shorter sentences), whilst 70% of the control group had also attended another form of programme. The study found statistically-significant reductions in both general and violent reoffending.

[35] These appear to be the same analyses, one reported as a government report the other in an academic journal.

The remaining studies, mainly cited in Papalia et al. but not in the other meta-analyses, are typically lower-quality studies. These have small samples and poorly-matched or unmatched comparison groups. These include:

Dowden, Blanchette and Serin (1999, *n* = 110): case-by-case matching of Canadian male federal offenders on demographic and offence-related variables to an anger-management programme or no-treatment comparison group. Participants in both groups were divided into high and low-risk sub-groups (of unreported sizes). They report statically-significant positive group differences in reconvictions after 3 years for the high-risk group on non-violent and violent offences, but no statistically-significant differences for the low-risk group.

Hatcher et al. (2008, *n* = 106**):** case-by-case matching of U.K. community-based males with violent convictions on risk assessment scores to an aggression replacement programme. They report a 13.3% lower reconviction rate in the experimental group after a 10-month follow-up period. No statistical-significance test was presented but the odds ratio for the outcomes in this study[36] appears to be 0.58 [95% CI: 0.27, 1.27], which does not indicate a statistically-significant group difference.

Henning and Frueh (1996, n = 124**)**: incidental matching of male prisoners on demographic and offence-related variables to a cognitive behavioural treatment or no-treatment comparison group. Significant baseline differences between the two groups are reported. They report statistically-significant positive group differences for any new criminal charges or violation of conditions after 2 years.

O'Brien and Daffern (2016, n = 114**)**: Unmatched comparison of Australian prisoners allocated to a violence-focused cognitive-behavioural treatment programme. Treatment completers and non-completers compared each-other and a non-treatment group. Significant baseline differences between the groups are reported. No statistically-significant group differences were reported for new violent reconvictions after approximately 3.5 years.

Serin, Gobeil and Preston (2009, n = 256**)**: Unmatched comparison of 256 Canadian males with convictions for violence to groups of treatment, alternative treatment, or a comparison "attrition group" (drop-outs and refusals). There was no no-treatment comparison group. Significant baseline differences between the groups are reported. No statistically-significant group differences were reported for returns to custody because of any reoffence after approximately 3.2 to 5 years (sample sizes were not sufficient for an analysis of violent reconvictions specifically).

Wong, Gordon, Gu, Lewis, and Olver (2012, n = 64**)**: case-by-case matching of 64 Canadian male high-risk psychopathic patients on demographic, offence-related, and clinical variables. Significant baseline differences between the groups are reported.

[36] The authors of this report, calculated an odds ratio using the numbers cited in the paper: 'A total of 27 (50.9%) of the comparison group were reconvicted compared to 20 (39.2%) of the experimental group' (Hatcher et al., p. 523).

No statistically-significant group differences were reported on all reconviction outcome variables, including incidence of violent and non-violent reconvictions.

General conclusions

Based on the individual studies reviewed, it could be suggested that there is limited evidence on which to reject the position that, at present, there is no effect of cognitive skills programmes on violent reconvictions. Six randomised trials and two well-matched propensity score matched studies demonstrate no effect of treatment on violent reconvictions, although they vary widely in how they define violent behaviour and test interventions targeted at very different criminal cohorts. Findings for the effect of violence-focused treatments on non-violent offending, however are more positive. Nevertheless, once results from the individual studies have been combined, three of the four meta-analyses demonstrate positive effects on both violent and non-violent reconvictions for violence-focused treatments.

Although those meta-analyses typically include varying combinations of high- and low-quality programmes, their aggregated positive findings are compelling. Few of the RCT studies were included in any of the meta-analyses other than Papalia et al. This is likely to be due to more conservative inclusion criteria in the Jolliffe and Farrington, Henwood et al., and Gannon et al. meta-analyses that would have excluded those high-quality RCTs on the basis that they target specific forms of violence (e.g., alcohol-related), types of participants (e.g., forensic psychiatric patients, homicide cohorts, female cohorts), non-proven violent outcomes (e.g., arrests, charges), or the lack of a non-treatment control group (e.g., comparisons with another treatment).

It is also fair to conclude that there are currently very few experimental and quasi-experimental studies on which to base conclusions. Each of the meta-analyses contain studies that were methodologically-compromised and in which substantial bias is likely to exist. We did not, however, examine or determine here the direction in which those biases might act – depending on the influence of bias, the likelihood of a treatment effect can either increase or decrease. Furthermore, because of the lack of high-quality studies, there is also insufficient evidence on what programme factors characterise effective programmes for this population. Although it appears from the existing literature that, at a minimum, programme completion, staff supervision, and service quality are potential factors that should be explored.

Annex 4: Description of CSAAP

The Correctional Services Accreditation and Advice Panel (CSAAP) comprises independent international academics and expert practitioners who advise HMPPS on accrediting programmes and provide independent, evidence-based advice on programme development and practice. The Ministry of Justice uses accreditation to provide confidence that its offending behaviour programmes are designed based on the best available evidence, will be delivered as intended, and will be evaluated to show the outcomes that are being met.

Once an accredited programme has been running for a sufficient amount of time, CSAAP considers the impact of the programme when deciding whether to recommend that the programme maintains accreditation. If CSAAP do not recommend that the programme maintains accreditation, HMPPS may withdraw the programme.

Programmes are assessed using the evidence-based principles for effective interventions. The Accreditation Criteria are laid out below.

The requirements for accreditation state that programmes and services must demonstrate that they:

1. Are evidence-based and/or have a credible rationale
2. Address factors relevant to reoffending and desistance
3. Are targeted at appropriate participants
4. Develop new skills (as opposed to only raising awareness)
5. Motivate, engage and retain participants
6. Are delivered as intended by staff with appropriate skills and quality assured, via:
    a. a quality assurance plan, and
    b. by providing quality assurance findings
7. Are evaluated, via:
    a. an evaluation plan, and
    b. by providing results of evaluation every 5 years

Annex 5: Further explanation of HMPPS programme integrity data collection

Quality Assurance Approach Summary

Quality assurance for accredited programmes is completed through the Interventions Integrity Framework (IIF). This explores evidence of practice in relation to whether the programme delivered met the guidelines set out in the programme and management manual. Evidence is collected from a variety of sources, including self-assessment and questionnaires; centrally held data such as starts and completions and training records; video recordings of sessions and clinical evidence such as supervision notes and post programme reports. Two iterations of the IIF have been used since it was first introduced in 2014. Whilst similar evidence for RESOLVE is used in both iterations, the headings under which the evidence is reported and scored are different. Due to these changes and a change in how scores are applied, the cycles are not comparable.

2014 – 2016 IIF

Part 1: Quality Completions (QC)

A specified dataset outlined the information required from providers to allow HMPPS Intervention Services (IS) to provide feedback on each offender attending the intervention. For example, whether the offender received the intervention by trained facilitators; at the appropriate dosage and frequency; was in a group of an appropriate size and met the selection criteria for the programme. Sites provided pre/post programme completion data (collated in a completions dataset) as well as session-level data on delivery and programme attendance (analysed by IS).

Part 2: Quality of Delivery

This component was assessed by specialist clinicians within IS for each programme being delivered, and reviewed quality of delivery in the group room as well as quality of treatment management. A sample of products was requested based on the site's volume of delivery. The products selected depended upon the requirements of the programme but included a sample of session recordings, reports and other material to provide a holistic assessment of performance once every two years. The quality of delivery section was also underpinned by an element of self-assessment focused on the key processes which support programme delivery, for example, the composition of groups and the frequency of supervision.

Part 3: Rehabilitative Culture

This component of the IIF was aimed at measuring the rehabilitative environment which should have supported the aims and values of the intervention so that staff and offenders felt fully supported to address their offending and reach their potential. This supports one of HMPPS' key overarching commissioning intentions which is to 'Enhance public protection and ensure a safe, decent environment and rehabilitative culture'. This intention focuses on the active development of an environment which is safe, secure and decent, and one which assists offenders towards rehabilitation. This requires a culture where authority is exercised confidently, consistently and fairly in order to build trust and improve safety. Providing a safe and decent delivery environment is fundamental to achieving outcomes and is an essential foundation for building a supportive and rehabilitative culture that motivates and enables offenders to make positive changes in their lives. A number of key processes were reviewed which if followed, would help to support a rehabilitative culture, for example, having the appropriate management structure. Staff and offender questionnaires further supported this.

2016 – 2019 IIF

Key Line of Enquiry (KLOE) 1: Is the intervention being delivered as designed?

This reviewed selection, attrition, and rate and dosage of delivery from central and local data sources. Research shows that the effectiveness of interventions is related to careful matching of the intervention to the assessed risks of reoffending, criminogenic needs and learning styles of those who participate.[37]To maintain momentum in learning and ensure motivation, scheduling and attendance must be at the appropriate dosage and rate.

KLOE 2: Is the learning environment safe, constructive and effective?

In order for learning to be effective the delivery style should be engaging, motivational and supportive, and in line with the core competency framework. Materials including session recordings, supervision notes, and treatment planning information were reviewed to ensure the programme was delivered with integrity, and responsively in a way that all individuals could understand the key learning points and practise new skills as appropriate. Group dynamics and boundaries were also reviewed to support an effective learning environment.

[37] See, for example, Andrews and Bonta (2010).

KLOE 3: Is the team enabled to effectively deliver the programme?

Facilitation of effective rehabilitative activities require well-trained and appropriately supported staff. Delivery staff should be supervised and encouraged to maintain and continually develop their skills. This KLOE reviewed evidence including self-assessment, session monitoring reports, supervision notes, and post-programme reports to assess the quality of treatment management.

KLOE 4: Does the culture/environment support and enable change?

Providing a safe and decent delivery environment is fundamental to achieving outcomes and is an essential foundation for building a supportive and rehabilitative culture that motivates and enables individuals to make positive changes in their lives. The rehabilitative environment should authenticate the aims and values of the intervention so that participants feel fully supported to address their offending and reach their potential. This reinforces one of HMPPS' key overarching commissioning intentions which is to 'Enhance public protection and ensure a safe, decent environment and rehabilitative culture'. To review this, self-assessment and staff and participant questionnaires were used.

RAG Rating (Red – Amber – Green)

Across both cycles, following a review of the evidence, each section was awarded a RAG rating indicative of the integrity of the programme in each of the key areas. All scores except Rehabilitative Culture are allocated by programme, as well as an overall score for the site. Rehabilitative Culture is scored only for the site as a whole:

Green	Programme integrity is maintained effectively. There are no risks to programme integrity – minor development areas may be noted.
Amber/Green	Improvements are required in order to meet the required standard, although programme integrity is not threatened to a significant degree.
Amber/Red	Programme integrity is under significant threat. Significant improvements need to be made in key areas in order to meet the required standard.
Red	Programme integrity is compromised. Critical areas of improvement have been identified. Immediate remedial action is required to minimise the threat and restore programme integrity

Annex 6: Details of matching criteria

Below is a table of variables that were used for PSM. The name of each variable, its type and categories are shown.

Variable	Type	Categories
Demographics		
Sex	Categorical	Male
Ethnicity	Categorical	White; Black; Asian; Other; Unknown
Nationality	Categorical	UK; Non-UK; Unknown
Criminal history		
Age at index (release) date	Continuous (integer)	-
Age at first contact with criminal justice system	Continuous (integer)	-
Primary index offence group	Categorical	Violence; Robbery; Public disorder; Possession of weapons; Criminal damage; Miscellaneous crimes against society; Burglary and Theft; Fraud; Drugs offences; Summary non-motoring offences; Summary motoring offences
Primary index offence severity	Categorical	Indictable only; Triable either way; Summary only
Index sentence type	Categorical	Custody
Index custodial sentence type (sentence lengths include licence period after release)	Categorical	Unknown; Less than or equal to 6 months; More than 6 months to less than 12 months; 12 months to less than 4 years; 4 years to 10 years; More than 10 years; Mandatory Life sentence; Other Life sentence; Imprisonment for Public Protection
Year of release from prison from index offence	Categorical	2011; 2012; 2013; 2014; 2015; 2016; 2017; 2018
Number of previous conviction events	Continuous (integer)	-
Number of previous court order events	Continuous (integer)	-
Number of previous custodial sentences	Continuous (integer)	-
Number of previous offences	Continuous (integer)	-
Number of previous indictable only offences	Continuous (integer)	-
Number of previous triable either way offences	Continuous (integer)	-
Number of previous summary only offences	Continuous (integer)	-
Number of previous offences of unknown severity	Continuous (integer)	-

Number of previous offences of violence	Continuous (integer)	-
Number of previous offences of robbery	Continuous (integer)	-
Number of previous offences of public order	Continuous (integer)	-
Number of previous offences of criminal damage	Continuous (integer)	-
Number of previous offences of domestic burglary	Continuous (integer)	-
Number of previous offences of other burglary	Continuous (integer)	-
Number of previous offences of theft not from vehicles	Continuous (integer)	-
Number of previous offences of theft from vehicles	Continuous (integer)	-
Number of previous offences of handling stolen goods	Continuous (integer)	-
Number of previous offences of fraud and forgery	Continuous (integer)	-
Number of previous offences of drug import, export, production and supply	Continuous (integer)	-
Number of previous offences of drug possession	Continuous (integer)	-
Number of previous offences of drink-driving	Continuous (integer)	-
Number of previous offences of breach of sentence conditions	Continuous (integer)	-
Number of previous offences of breach of a restraining order	Continuous (integer)	-
Any restraining order imposed within three years before conviction	Categorical (binary)	No; Yes
Any restraining order imposed within five years before conviction	Categorical (binary)	No; Yes
Any restraining order imposed within ten years before conviction	Categorical (binary)	No; Yes
Copas rate (logarithmic rate of convictions and cautions over time)	Continuous	-
Employment and benefits		
Any Pay As You Earn (PAYE) employment within one month before conviction	Categorical	Unknown; No; Yes
Any PAYE employment within one year before conviction	Categorical	Unknown; No; Yes
Any out-of-work benefits received within one year before conviction	Categorical	Unknown; No; Yes
Any Job Seeker's Allowance received within one year before conviction	Categorical	Unknown; No; Yes

Any Incapacity Benefit or Income Support received within one year before conviction	Categorical	Unknown; No; Yes
Accredited Programmes other than Resolve		
Any other Accredited Programme taken during sentence	Categorical (binary)	No; Yes
Thinking Skills Accredited Programme or Enhanced Thinking Skills taken during sentence	Categorical (binary)	No; Yes
Other Accredited Programme to address violence taken during sentence	Categorical (binary)	No; Yes
OASys assessment (usually taken within a year before treatment start date)		
OASys assessment used	Categorical (binary)	No; Yes
OVP Score	Continuous (integer)	For the purposes of matching, these have been banded (see standardised differences document for details)
OGRS3 Score	Continuous (integer)	For the purposes of matching, these have been banded (see standardised differences document for details)
Problems with reading, writing and/or numeracy	Categorical	Unknown; None; Some; Severe
Problems with reading	Categorical	Unknown; No; Yes
Problems with writing	Categorical	Unknown; No; Yes
Problems with numeracy	Categorical	Unknown; No; Yes
Learning difficulties	Categorical	Unknown; None; Some; Severe
Perpetrator of domestic violence or abuse towards a partner at any time	Categorical	Unknown; No; Yes
Problems with regular activities that encourage offending	Categorical	Unknown; None; Some; Severe
Problems with being easily influenced to offend	Categorical	Unknown; None; Some; Severe
Misused drugs (except alcohol) at any time	Categorical	Unknown; No; Yes
Problems with motivation to tackle drug misuse	Categorical	Unknown; None; Some; Severe
Drug misuse linked to risks to themselves or others	Categorical	Unknown; No; Yes
Drug misuse linked to offending behaviour	Categorical	Unknown; No; Yes
Problems with alcohol misuse within one year before assessment	Categorical	Unknown; None; Some; Severe
Problems with binge drinking within six months before assessment	Categorical	Unknown; None; Some; Severe
Violent behaviour related to alcohol misuse at any time	Categorical	Unknown; No; Yes

Problems with motivation to tackle alcohol misuse	Categorical	Unknown; None; Some; Severe
Alcohol misuse linked to risks to themselves or others	Categorical	Unknown; No; Yes
Alcohol misuse linked to offending behaviour	Categorical	Unknown; No; Yes
Difficulties in coping with life	Categorical	Unknown; None; Some; Severe
Psychological problems around time of assessment	Categorical	Unknown; None; Some; Severe
Actual self-harm or suicidal thoughts at any time	Categorical	Unknown; No; Yes
Psychiatric problems around time of assessment	Categorical	Unknown; None; Some; Severe
Problems with interpersonal skills	Categorical	Unknown; None; Some; Severe
Problems with impulsive behaviour	Categorical	Unknown; None; Some; Severe
Problems with aggressive behaviour	Categorical	Unknown; None; Some; Severe
Problems with temper control	Categorical	Unknown; None; Some; Severe
Difficulties in recognising own problems	Categorical	Unknown; None; Some; Severe
Difficulties in solving own problems	Categorical	Unknown; None; Some; Severe
Difficulties in considering consequences of own actions	Categorical	Unknown; None; Some; Severe
Difficulties in understanding other people's feelings	Categorical	Unknown; None; Some; Severe
Difficulties in adapting own views	Categorical	Unknown; None; Some; Severe
Attitudes in favour of criminal behaviour	Categorical	Unknown; None; Some; Severe
Problems with behaviour towards staff in criminal justice system around time of assessment	Categorical	Unknown; None; Some; Severe
Actual or predicted lack of co-operation with supervision in the community around time of assessment	Categorical	Unknown; None; Some; Severe
Attitudes in favour of antisocial behaviour	Categorical	Unknown; None; Some; Severe
Difficulties in recognising reasons for own offending	Categorical	Unknown; None; Some; Severe
Problems with motivation to change offending behaviour around time of assessment	Categorical	Unknown; None; Some; Severe
Any long-term health condition, disability or pregnancy around time of assessment	Categorical	Unknown; No; Yes
Unwillingness to complete programmes to change offending behaviour around time of assessment	Categorical	Unknown; No; Yes

Glossary of Terms

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.

Common support: This occurs when there is sufficient overlap in the range of propensity scores across treatment and comparison groups.

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.

Effect size: A value measuring the strength of the relationship between two variables in a statistical population.

Imputation: The process of replacing missing data with substituted values.

Index date: The prison release date and the date from which the follow up period for measuring reoffending begins.

Index offence: The primary offence for which the offender was convicted and received a custodial sentence (specifically, the index sentence).

Index sentence: For the treatment group, the period of custody in which the offender participated in RESOLVE. For the comparison group, the period of custody in which the offender was matched to a treatment group record.

Interquartile range (IQR): A measure of variability that divides the dataset into quartiles. It is defined as the range of values between the first and third quartile. It is often used to show a more representative spread of values around a given variable as the IQR is resistant to outliers that may skew the mean of the treatment group.

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) will be a key output for Justice Data Lab analyses as the reoffending rates for the treatment and control 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.

OASys Violence Predictor (OVP): Percentage likelihood of committing any violent proven reoffence within 2 years. This is based on static and dynamic factors including age, gender and criminal history. This includes minor violent offences like common assault, harassment and criminal damage and more serious violent offences. An OVP

score of 30%+ is the criterion for accredited programmes that address violent offending behaviour. The more intensive programmes specify an OVP score of 60% or above.

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.

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. OGRS scores can be used to target those resources designed to reduce reoffending. Accredited offending behaviour programmes often require particular OGRS scores as part of their eligibility criteria.

One-year proven reoffending rate: The proportion of offenders in a cohort who committed an offence during a 12-month period starting on the index date 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.

Police National Computer (PNC): An administrative data system used by all police forces in England and Wales, managed by the Home Office. The PNC records offender, crime and disposal details.

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

P-value: The p-value is the probability of obtaining results at least as extreme as the observed results of a statistical hypothesis test, assuming that the null hypothesis is correct.

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

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

Treatment group: The group of offenders that the provider delivered their intervention to. In other words, the offenders who received 'the treatment'.

Two-year proven reoffending rate: The proportion of offenders in a cohort who committed an offence during a 24-month period starting on the index date 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.

References

Anderson, C. A., & Bushman, B. J. (2002). Human aggression. Annual review of psychology, 53(1), 27-51.

Andrews, D. A. & Bonta, J. (2010). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy, and Law*, 16, 39-55.

Barnes, J. C., El Sayed, S. A., TenEyck, M., Nedelec, J. L., Connolly, E. J., Schwartz, J. A., et al. (2017). Estimating relative stability in developmental research: A critique of modern approaches and a novel method. *Journal of quantitative criminology*, *33*(2), 319-346.

Barnes, J. C., TenEyck, M. F., Pratt, T. C., & Cullen, F. T. (2020). How *power*ful is the evidence in criminology? On whether we should fear a coming crisis of confidence. *Justice Quarterly*, *37*(3), 383-409.

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R. et al. (2018). Redefine statistical significance. *Nature Human Behaviour*, 2(1), 6-10.

Berry, S. (1998). *The Montgomery House Violence Prevention Programme: An Evaluation*. Wellington, NZ: Department of Corrections Psychological Services.

Berry, S. (2003). Stopping violent offending in New Zealand: Is treatment an option? *New Zealand Journal of Psychology*, 32(2), 92.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. New York: Wiley.

Bowes, N., McMurran, M., Evans, C., Oatley, G., Williams, B., & David, S. (2014). Treating alcohol-related violence: a feasibility study of a randomized controlled trial in prisons. *The Journal of Forensic Psychiatry & Psychology*, *25*(2), 152-163.

Bushman, B. J., & Anderson, C. A. (2001). Is it time to pull the plug on the hostile versus instrumental aggression dichotomy? Psychological Review, 108, 273−279.

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

Cortoni, F. A., Nunes, K. L., & Latendresse, M. (2006). *An examination of the effectiveness of the Violence Prevention Program*. Ottawa, ON: Correctional Service Canada.

Davidson, K. M., Tyrer, P., Tata, P., Cooke, D., Gumley, A., Ford, I., et al. (2009). Cognitive behaviour therapy for violent men with antisocial personality disorder in the community: an exploratory randomized controlled trial. *Psychological Medicine*, 39(4), 569.

Dodge, K. A., Coie, J. D., & Lynam, D. (2006). Aggression and antisocial behaviour in youth. In W. Damon, R. M. Lerner, and N. Eisenberg (Eds.), *Handbook of child psychology, vol. 3.* (pp. 719−776) New York: John Wiley & Sons.

Dowden, C., Blanchette, K., & Serin, R. (1999). *Anger management programming for federal male inmates: An effective intervention* (Research Report R-82). Ottawa, Ontario, Canada: Correctional Service of Canada.

Finn, M. A., & Muirhead-Steves, S. (2002). The effectiveness of electronic monitoring with violent male parolees. *Justice Quarterly*, 19(2), 293-312.

Gannon, T. A., Olver, M. E., Mallion, J. S., & James, M. (2019). Does specialized psychological treatment for offending reduce recidivism? A meta-analysis examining staff and program variables as predictors of treatment effectiveness. *Clinical psychology review*, 73, 101752.

Hatcher, R. M., Palmer, E. J., McGuire, J., Hounsome, J. C., Bilby, C. A., & Hollin, C. R. (2008). Aggression replacement training with adult male offenders within community settings: a reconviction analysis. *The Journal of Forensic Psychiatry & Psychology*, *19*(4), 517-532.

Henning, K. R., & Frueh, B. C. (1996). Cognitive-behavioral treatment of incarcerated offenders: An evaluation of the Vermont Department of Corrections' Cognitive Self-Change Program. *Criminal Justice and Behavior*, 23(4), 523-541.

Henwood, K. S., Chou, S., & Browne, K. D. (2015). A systematic review and meta-analysis on the effectiveness of CBT informed anger management. *Aggression and violent behaviour*, 25, 280-292.

Higgs, T., Cortoni, F., & Nunes, K. (2019). Reducing violent risk: Some positive recidivism outcomes from Canadian treated high-risk offenders. *Criminal Justice and Behavior*, 46(3), 359-373.

Huesmann, L. R. (1998). The role of social information processing and cognitive schema in the acquisition and maintenance of habitual aggressive behaviour. In R. Geen and E. Donnerstein (eds.), *Human Aggression: Theories, Research, and Implications for Social Policy* (pp. 73-109). New York: Academic Press.

Huesmann, L. R., & Guerra, N. G. (1997). Children's normative beliefs about aggression and aggressive behaviour. *Journal of Personality and Social Psychology*, 72, 408−419.

Hyatt, J. M. (2013). *The impact of cognitive behavioral therapy on the recidivism of high-risk offenders: Results from a randomized trial*. Philadelphia: University of Pennsylvania.

Jolliffe, D. & Farrington, D. P. (2007). *A rapid evidence assessment of the impact of mentoring on reoffending: A summary*. London: Home Office.

Jolliffe, D. & Farrington, D. P. (2009). A systematic review of the relationship between childhood impulsiveness and later violence. *Personality, personality disorder, and violence*, 41-61.

Kubiak, S. P., Brenner, H., Bybee, D., Campbell, R., & Fedock, G. (2018). Reporting Sexual Victimization During Incarceration: Using Ecological Theory as a Framework to Inform and Guide Future Research. *Trauma, Violence, & Abuse*. 19(1), 94-106.

Lipsey, M. W., Landenberger, N. A., & Wilson, S. J. (2007). Effects of cognitive-behavioural programs for criminal offenders. *Campbell systematic reviews*, 3(1), 1-27.

Mews, A., Di Bella, L., & Purver, M. (2017). *Impact evaluation of the prison-based core sex offender treatment programme*. London: Ministry of Justice.

Mews, A., Hillier, J., McHugh, M. & Coxon, C. (2013). *The impact of short custodial sentences, community orders and suspended sentence orders on reoffending*. London: Ministry of Justice.

Ministry of Justice (2015). *The impact of short custodial sentences, community orders and suspended sentence orders on reoffending*. London: Ministry of Justice.

Motiuk, L., Smiley, C., & Blanchette, K. (1996). Intensive programming for violent offenders: A comparative investigation. *Forum on Corrections Research*, 8, 10-12.

O'Brien, K., & Daffern, M. (2016). The impact of pre-treatment responsivity and treatment participation on violent recidivism in a violent offender sample. *Psychology, Crime & Law*, *22*(8), 777-797.

Papalia, N., Spivak, B., Daffern, M., & Ogloff, J. R. P. (2019). A meta-analytic review of the efficacy of psychological treatments for violent offenders in correctional and forensic mental health settings. *Clinical Psychology and Practice*, 26(2), e12282.

Polaschek, D. L. (2011). High-intensity rehabilitation for violent offenders in New Zealand: Reconviction outcomes for high-and medium-risk prisoners. *Journal of interpersonal violence*, 26(4), 664-682.

Polaschek, D. L. (2017). *Prevention of Recidivism and Violent and Aggressive Offenders*. The Wiley Handbook of Violence and Aggression, 1-12.

Polaschek, D. L., Wilson, N. J., Townsend, M., & Daly, L. (2005). Cognitive-behavioral rehabilitation for high-risk violent offenders: An outcome evaluation of the Violence Prevention Unit. *Journal of Interpersonal Violence*, 20, 1611-1627.

Polaschek, D. L., Yesberg, J. A., Bell, R. K., Casey, A. R., & Dickson, S. R. (2016). Intensive psychological treatment of high-risk violent offenders: Outcomes and pre-release mechanisms. *Psychology, Crime and Law*, 22(4), 344-365.

Riley, R. D., Higgins, J. P., & Deeks, J. J. (2011). Interpretation of random effects meta-analyses. *British Medical Journal*, 342.

Robinson, D. (1995). *The impact of cognitive skills training on post-release recidivism among Canadian federal offenders.* Ottawa: Correctional Service of Canada (Research Report R-41).

Serin, R. C., Gobeil, R., & Preston, D. L. (2009). Evaluation of the persistently violent offender treatment program. *International Journal of Offender Therapy and Comparative Criminology*, 53(1), 57-73.

Sherman, L.W., Gottfredson, D.C., Mackenzie, D.L., Eck, J., Reuter, P., & Bushway, S.D. (1998). *Preventing Crime: What works, what doesn't, what's promising*. Washington, DC: National Institute of Justice

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3), 1-67.

Watt, K., Shepherd, J., & Newcombe, R. (2008). Drunk and dangerous: a randomised controlled trial of alcohol brief intervention for violent offenders. *Journal of Experimental Criminology*, *4*(1), 1-19.

Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). Assessing risk of violence to others. *Impulsivity: Theory, assessment, and treatment*, 251-277.

Wong, S. C., & Gordon, A. (2000). *Violence Risk Scale (Version 2)*: Regional Psychiatric Centre (Prairies) and University of Saskatchewan.

Wong, S. C., Gordon, A., Gu, D., Lewis, K., & Olver, M. E. (2012). The effectiveness of violence reduction treatment for psychopathic offenders: Empirical evidence and a treatment model. *International Journal of Forensic Mental Health*, 11(4), 336-349.