Impact of sentencing on proven reoffending for young offenders

Insights from the MoJ – DfE data share

Summary

This paper analysed and compared one-year reoffending rates and frequency for children and young people aged 12 to 17 years old receiving different types of convictions at court or police cautions using reoffending data for the academic school years 2015/16 to 2017/18. Three cohorts were examined based on the school year of their index disposal: Year 8-9, Year 10-11, and After Year 11. Propensity score matching (PSM) analysis was conducted to estimate the impact of a particular sentence of interest, 'treatment sentence', instead of a 'comparison sentence' on reoffending outcomes whilst taking into account demographic characteristics, offence information, criminal history as well as the educational and social-care background. Importantly, results reveal what the impact was for someone receiving the 'treatment sentence'; for example, whether their reoffending rate would have been higher or lower than if they had instead received the comparison sentence. Thus, the below results do not allow to draw any conclusions about the expected impact for those receiving the 'comparison sentence'.

The results indicate that:

- Youth Rehabilitation Orders (of any type) were associated with both higher reoffending rates and frequency than matched referral orders across all three cohorts. For those in the Year 10-11 and After Year 11 offender cohorts, this was also the case when distinguishing between Youth Rehabilitation Orders with and without intensive supervision and surveillance.

- Discharges from court (regardless of type) were associated with lower reoffending rates than matched referral orders for those in the Year 10-11 and After Year 11 offender cohorts. For those in the Year 10-11 cohort, they were also associated with

lower reoffending frequency. No effects were found for those in Year 8-9 or when comparing discharges to cautions.

- Absolute discharges from court were associated with higher reoffending rates but not higher reoffending frequency compared to matched conditional discharges for those in the After Year 11 offender cohort.
- Referral orders were associated with similar reoffending outcomes compared to matched cautions across all three cohorts.

However, cautious interpretation should be applied as there may be important factors that could not be included within the analysis.

Introduction

Context

The number of children and young people under 18 years-old who have been cautioned or sentenced has steadily declined over the past decade. In the year ending March 2024, it was 67% lower than 10 years prior. However, at the same time the number of proven offences[1] committed by children saw a year-on-year increase for the second consecutive year, rising by 4% to around 35,600 (Youth Justice Board for England and Wales, 2025).

Similarly, results published in January 2025 for the most recent reoffending cohort[2] show that the proven reoffending rate for children has increased for the second consecutive year to 32.5%, a 0.3 percentage point increase on the previous year. Importantly for this cohort, both the number of children and the number of reoffenders increased for the first time in the last 10 years (Youth Justice Board for England and Wales, 2025).

Study aims and objectives

Reoffending rates are a key outcome when it comes to evaluating effectiveness of sentencing and as such provide an opportunity to analyse which sentences may result in a reduction of reoffending. Crucially, appropriate statistical analysis is required to address this question as the direct comparison of reoffending rates between different sentence groups would result in biased conclusions. This is because individuals do not only differ in their type of sentence but on a plethora of characteristics such as their demographics, offence and criminal history and wider life (family, education, socio-economic) circumstances.

In the youth justice context, a previous study (Ministry of Justice 2019) examined the impact of sentencing on reoffending in a cohort of 15-17-year-old juveniles who offended

[1] A proven offence is one which results in a caution or court sentence.

[2] Proven reoffending by children entering the cohort between April 2022 and March 2023. The published statistics presented annual data based on the aggregate of the four quarterly offender cohorts.

during 2012-2014. Its analysis focused on the comparison of various youth community and custodial sentences[3] controlling for offender demographics, as well as offence characteristics and criminal history. It found that custodial sentences were associated with higher reoffending rates than matched community orders. However, longer custodial sentences were associated with lower reoffending rates than shorter ones. This mirrors research on adult offenders (Eaton & Mews 2019, Mews, Hillier, McHugh, & Coxon, 2015). Finally, when comparing two types of community orders, youth rehabilitation orders were associated with higher reoffending rates than matched referral orders.

Although the study controlled for important characteristics that could explain differences between sentences, additional information on whether a defendant pled guilty and their wider socio-educational background were not available at the time. Furthermore, the focus on custodial sentences meant that a cohort of younger offenders was not analysed.

The current study aims to fill this gap by providing an updated analysis on the subject, however focussing on the comparison of non-custodial sentencing options. The objective is to provide policy makers with insights into the effectiveness of non-custodial sentencing on youth reoffending.

Methodology

Data sources and sample selection

Datasets for three year-group cohorts were constructed using data from the Ministry of Justice (MoJ) – Department for Education (DfE) data share. This share includes data from the Police National Computer (PNC) as well as the National Pupil Database.

Specifically, analysis focused on individuals who convicted at court or cautioned by the police for an offence during the academic years 2015/16 to 2017/18[4]. Offenders were grouped into three cohorts based on the school year they were in at the time of receiving a court conviction or caution:

- Year 8 – 9
- Year 10 – 11
- After Year 11

[3] For an overview of youth community and custodial sentences please refer to https://www.sentencingcouncil.org.uk/sentencing-and-the-council/types-of-sentence/types-of-sentences-for-young-people/)

[4] In the UK, an academic year starts on 1 September and ends on 31 August. Thus, individuals were included if they offended between 1 September 2015 and 31 August 2018. This time window was chosen to avoid any impact of the Covid-19 pandemic on the results. The latest end date for the 1-year reoffending follow-up plus 6-months waiting period was 29 February 2020.

Aligning reoffending cohorts to academic years enabled consistent use of educational background information in the school year(s) prior to the conviction or caution that led to someone being included in a given sentencing comparison (also called index disposal[5]). This ensured that only information available prior to the sentencing decision was used.

To ensure the sentencing comparisons included a large enough number of cases, they were performed across the three-year period rather than separately (please note that MoJ reoffending statistics use 3-months cohorts). However, only the first sentencing occasion per disposal category of interest in a given school year was included in the analysis. Thus, when an offender was given the same sentence more than once in any given academic year, only the first sentence was considered in the analysis[6].

Analysis of proven reoffending followed the standard MoJ approach of a 1-year follow up plus 6-months wait period to allow for delays in sentencing (for details see Guide to proven reoffending statistics, section 2 Measuring reoffending, Ministry of Justice 2025)

Sentencing comparisons

This analysis focuses on comparisons involving cautions, discharges and youth community orders, specifically Referral Orders (ROs) and Youth Rehabilitation Orders (YROs). For YROs, we further distinguished between YROs with Intensive Supervision and Surveillance (with ISS) and without (YRO only). The MoJ-DfE data share did not contain sufficient information on the timing of release from custody for all types of youth custodial sentences (crucial information required for the reoffending follow-up). Therefore, no comparisons involving custodial sentences could be included in the analysis.

Listed below are all comparisons analysed in this study. Due to low counts of sentencing occasions, certain sentencing comparisons were not run for those whose index disposal occurred during Year 8-9 (marked with an asterisk, i.e. *, below).

- Discharge vs Cautions
- RO vs Cautions
- Absolute Discharge vs Conditional Discharge*
- Discharge vs RO
- Youth Rehabilitation Order vs RO
- YRO only vs RO*
- YRO w ISS vs RO*

[5] The index disposal of the offender is the type of sentence the offender received for their index offence. An index offence is the proven (through caution or court convictions) offence that leads to an offender being included in the cohort. An offence is only counted as an index offence if it is recordable, committed in England or Wales, prosecuted by the police, and not a breach offence.

[6] In a small number of cases where individuals received the same sentence on the same date, the data was simplified by counting only one of them. That is, if someone received the same sentence, for example two cautions, on the same date for the same offence, only one of the sentencing occasions was kept.

- YRO w ISS vs YRO only*

Propensity score matching

Propensity score matching (PSM) is an impact evaluation analysis technique for observational data where using a randomised control trial is not feasible because random assignment to treatment and control group cannot be justified in the criminal justice context. Comparing the impact of different sentences on reoffending is such a case and PSM has been used previously within MoJ to address this (see Eaton & Mews 2019, Mews, Hillier, McHugh, & Coxon, 2015).

The purpose of the PSM is to find offenders for the control group who are similar to those receiving the sentence of interest in terms of their offender and offence characteristics. The aim is to enable a fair comparison of outcomes between the two groups. In PSM, offenders in the treatment group are matched to non-treated offenders with similar propensity scores. In brief, propensity scores are derived for each offender using a logistic regression model, which is used to estimate the conditional probability (value between 0 and 1) of receiving the sentence of interest (treatment). Values closer to 1 represent an increased likelihood. The regression model should be constructed such that all factors associated with both the likelihood of the offender receiving a sentence and the likelihood of the outcome of interest (reoffending) are included. Sentencing occasions in the treatment group are then matched to sentencing occasions in the comparison group using the estimated propensity scores (for details, see the Appendix ‘Propensity score matching approach’ section). Provided the treatment and control groups are well-matched on the characteristics included in the logistic regression model (see matching quality results below), the mean difference in the outcome variables can be compared to infer the average treatment effect for those who received the treatment sentence (see impact effect estimators below).

The PSM approach assumes a level of variation in sentencing decisions. This assumption imposes its own limitations to how PSM should be used, since similar cases should be given different sentences only where sentencing decisions are marginal. Following cases being matched, the PSM approach assumes that the choice is, in effect, random – i.e. all non-random variation is controlled. However, as unmeasured factors may influence both the sentencing decision and reoffending outcomes, the conclusions of such analyses cannot be regarded as definitive.

Control variables

The variables used in the logistic regression estimating the propensity score fell into eight main categories and are listed below. For more details on the variable selection process please see the Appendix ‘Variable selection’ section, and for more details on data missingness and data imputation see the Appendix ‘Data imputation’ section.

Offender demographics

- Gender

- Ethnic group
- English as first language
- School term of birth (Summer = born between June - August, Spring = born between March - May, Autumn = born between September - November, Winter = born between December – February)

Index disposal (the sentence for the index offence)

- Age at conviction
- Academic year of conviction
- Plea (Pleading guilty, not guilty, or whether no plea was taken)
- Police force region where offence was processed
- Primary offence group

Offending history

- Age at first contact with the criminal justice system
- Number of previous custodial sentences
- Number of previous offences with breakdown by severity of offence (four variables for severity indicatable only, triable either way, summary offences, unknown severity; the sum of these four variables would be equivalent to the total number of previous offences)
- Whether the offender was a first time, second time or further time offender based on the number of previous convictions and cautions

Demographic characteristics, social care and educational experience in the year(s) prior to the index disposal

For those who received a conviction or caution in Year 8-9, this information was based on the school year prior to the index disposal. For those who received a conviction or caution in Year 10-11 or After Year 11, this information was based on the earliest information available going as far back as two school years prior to the index disposal. This approach was taken to address issues around missing education data for children in later school years, especially those that left school after Year 10 or Year 11.

- Absence from school in the year(s) prior to the index disposal, with focus on persistent absence defined as missing more than 10% of sessions per school year:
    - persistently absent for all reasons
    - persistently absent for unauthorised reasons (proxy for truancy)
- Having been permanently excluded from school (binary indicator)
- Having experienced suspension from school
- Being eligible for free school meals

- Index of Income Deprivation Affecting Children. Index decile rank based on the pupil's postcode (1 = living in most deprived 10% of neighbourhoods, 10 = living in least deprived 10% of neighbourhoods)
- Having special education needs (with a statement, without a statement) or not
- Being classed to social services as a child in need
- Being classed to social services as a child on a child protection plan
- Being classed to social services as a child who is being looked after

Outcome variables

The outcomes of interest in this study were as follows:

1) The 1-year proven reoffending rate, i.e. the percentage of individuals who reoffended
2) The average number of proven reoffences per offender (sentencing occasion)

Impact effect estimators

PSM effectively allows to estimate what the outcome in the treatment group would have been had they received the comparison sentence (average treatment effect on the treated, ATT, see the Appendix 'Impact effects hypothesis testing' for details). Crucially, estimated effects represent the average treatment effect of those sentencing occasions that could be matched in the treatment sentence group. Thus, the effects should not be generalised to the entire population of those receiving the treatment sentence. Furthermore, the ATT focuses exclusively on the treated population; it does not estimate what the impact of treatment would have been for those who were not treated; i.e, the comparison group.

In practice, impact of sentencing was quantified primarily using mean differences; i.e. the difference in mean reoffending rate and mean reoffending frequency between treatment and control group. Positive mean differences indicate that the treatment group reoffended more/ more frequently than those who received the comparison sentence. Negative mean differences indicate that the treatment group reoffended less/ less frequently than those who received the comparison sentence.

Furthermore, for reoffending rates as a binary outcome, the impact of sentencing was also estimated using odds ratios and risk ratios.

Odds ratios (OR) are a measure of association between an exposure (sentence type) and an outcome (reoffending), which help us to understand if a difference in exposure increases or decreases the likelihood of that event. ORs compare the odds of reoffending occurring in the treatment sentence group versus the comparison sentence group, and can be interpreted as follows:

- **OR = 1:** There is no association between sentence and reoffending rate.
- **OR > 1:** The 'treatment' sentence is associated with a higher likelihood of reoffending.

- **OR < 1:** The ‘treatment’ sentence is associated with a lower likelihood of reoffending.

Risk ratio (RR) is the ratio of the probability of an outcome in an exposed group to the probability of an outcome in an unexposed group. In the context of this analysis, it is the ratio of the mean reoffending rate of the treatment group relative to the mean reoffending rate of the comparison group. Therefore, it quantifies directly how much more likely reoffending is in the treatment group:

- **RR = 1:** The risk of reoffending in both sentencing groups is the same.
- **RR > 1:** The risk of reoffending in the treatment group is X times greater with X representing the value of the RR; e.g. a RR = 1.5 would indicate 1.5 times higher risk of reoffending in the treatment group.
- **RR < 1:** The risk of reoffending in the treatment group is X times lower with X representing the value of the RR; e.g. RR = 0.5 would indicate that the risk of reoffending in the treatment group is 0.5 times lower.

Results

Matching quality

Matching quality for each sentencing comparison was assessed using standardised mean differences (SMDs, see the Appendix ‘Propensity score matching approach’ section for details). This evaluated how closely comparable the matched groups were on the characteristics selected in the logistic regression model. Matching quality was evaluated as follows in line with established MoJ practice.

- Absolute SMDs of 5% or less indicate closely matched groups on a given characteristic.
- Absolute SMDs between 5-10% indicate a reasonably close match.
- Absolute SMDs above 10% are considered indicative of poor matching quality which could alter the interpretation of the results.

Matching quality was high across all cohorts and comparisons with standardised mean differences being closely matched (smaller than 5%) for 95.29% or more of SMDs (see Table 1). There was a small number of SMDs that indicated reasonable matching (absolute SMD between 5-10%, see Table 2). For the Year 8-9 cohort, there were two comparisons with two and one comparison with three SMDs above 5% with the largest difference being no more than 8.55%. For the Year 10-11 cohort, there were six comparisons with one SMD above 5% respectively with the largest difference being 7.85%. Finally, for the after Year 11 cohort, there were two comparisons with one, one comparison with three, and one with four SMDs above 5% with the largest difference being 7.40%. Thus, overall, the chosen matching approach resulted in well-matched groups.

Impact estimates

Year 8-9 cohort

For the Year 8-9 cohort (see Table 3), only one sentencing comparison was associated with a difference in reoffending outcomes. Specifically, YROs were associated with both a higher rate and frequency in reoffending compared to matched ROs. The remaining comparisons involving referral orders compared to cautions and discharges compared to referral orders or cautions did not reveal any significant differences in outcomes.

Year 10-11 cohort

For the Year 10-11 cohort (see Table 4), YROs (regardless of type with ISS, without ISS, or combined) were associated with higher reoffending rates and frequency compared to matched referral orders. Furthermore, discharges from court were associated with lower reoffending rates and frequency compared to matched ROs. The remaining comparisons involving referral orders and discharges compared to cautions, as well as absolute vs conditional discharges did not reveal any significant differences in outcomes.

After Year 11 cohort

For the After Year 11 cohort (see Table 5), YROs (regardless of type with ISS, without ISS, or combined) were associated with higher reoffending rates and frequency compared to matched ROs. Furthermore, discharges from court were associated with lower reoffending rates albeit at the same frequency compared to matched ROs. Finally, absolute discharges from court were associated with higher reoffending rates but the same frequency compared to matched conditional discharges from court. The remaining comparisons involving referral orders and discharges compared to cautions did not reveal any significant differences in outcomes.

Table 1 Percentage of SMDs that are <= 5% and thus represent closely matched treatment and control groups by comparison and cohort

Cohort	YRO vs RO	YRO only vs RO	YRO w ISS vs RO	YRO w ISS vs YRO only	RO vs Cautions	Discharge vs RO	Absolute Discharge vs Conditional Discharge	Discharge vs Cautions
Year 8-9	97.62%	N/A	N/A	N/A	100%	97.62%	N/A	96.47%
Year 10-11	98.82%	98.82%	98.82%	98.82%	100.0%	98.82%	100.0%	98.82%
After Year 11	98.82%	96.47%	95.29%	100.0%	100.0%	98.82%	100.0%	100.0%

Table 2 Count of absolute SMD above 5% and largest SMD by comparison and cohort

Cohort	YRO vs RO	YRO only vs RO	YRO w ISS vs RO	YRO w ISS vs YRO only	RO vs Cautions	Discharge vs RO	Absolute Discharge vs Conditional Discharge	Discharge vs Cautions
Year 8-9	2 (7.79%)	N/A	N/A	N/A	N/A	2 (8.55%)	N/A	3 (7.88%)
Year 10-11	1 (5.40%)	1 (5.50%)	1 (6.05%)	1 (7.85%)	N/A	1 (5.14%)	N/A	1 (5.12%)
After Year 11	1 (7.40%)	3 (6.38%)	4 (7.38%)	N/A	N/A	1 (5.15%)	N/A	N/A

Table 3 Year 8-9 sentencing comparison results

Treatment group sentence	Control group sentence	Reoffending outcome	Treatment group mean (SE)	Control group mean (SE)	Impact estimate: Mean difference [CI LB-UB]	Impact estimate: Odds ratio [CI LB-UB]	Impact estimate: Risk ratio [CI LB-UB]
YRO	RO	Rate	73.94% (1.68%)	54.65% (1.09%)	19.29% [13.62%, 24.96%] ***	2.35 [1.83, 3.03] ***	1.35 [1.23, 1.49] ***
YRO	RO	Frequency	4.10 (0.19)	2.43 (0.08)	1.68† [1.17, 2.18] ***	N/A	N/A
RO	Cautions	Rate	40.18% (1.02%)	41.25% (0.65%)	-1.07% [-3.97%, 1.82%]	0.96 [0.85, 1.08]	0.97 [0.91, 1.05]
RO	Cautions	Frequency	1.50 (0.06)	1.59 (0.04)	-0.09 [-0.29, 0.11]	N/A	N/A
Discharge	RO	Rate	49.13% (2.5%)	50.52% (1.17%)	-1.40%† [-8.01%, 5.22%]	0.95 [0.73, 1.23]	0.97 [0.85, 1.11]
Discharge	RO	Frequency	2.29 (0.2)	2.35 (0.09)	-0.06 [-0.59, 0.48]	N/A	N/A
Discharge	Cautions	Rate	51.34% (2.59%)	50.95% (0.71%)	0.40%† [-6.95%, 7.75%]	1.02 [0.76, 1.36]	1.01 [0.87, 1.16]
Discharge	Cautions	Frequency	2.46 (0.22)	2.26 (0.05)	0.2 [-0.41, 0.81]	N/A	N/A

*Notes: *** = significant at 0.001 level, ** = significant at 0.01 level, * = significant at 0.05 level, † =Due to rounding effects, the impact estimate mean difference may deviate from the presented difference between treatment and control group shown in the table by 0.01*

Table 4 Year 10-11 sentencing comparison results

Treatment group sentence	Control group sentence	Reoffending outcome	Treatment group mean (SE)	Control group mean (SE)	Impact estimate: Mean difference [CI LB-UB]	Impact estimate: Odds ratio [CI LB-UB]	Impact estimate: Risk ratio [CI LB-UB]
YRO	RO	Rate	61.63% (0.85%)	52.53% (0.58%)	9.10% [5.70%, 12.51%] ***	1.45 [1.26, 1.67] ***	1.17 [1.10, 1.25] ***
YRO	RO	Frequency	2.77 (0.07)	2.30 (0.05)	0.47 [0.15, 0.79] **	N/A	N/A
YRO only	RO	Rate	60.85% (0.88%)	52.13% (0.58%)	8.72% [5.16%, 12.29%] ***	1.43 [1.24, 1.65] ***	1.17 [1.09, 1.25] ***
YRO only	RO	Frequency	2.70 (0.07)	2.18 (0.05)	0.52 [0.22, 0.81] ***	N/A	N/A
YRO w ISS	RO	Rate	67.27% (2.81%)	52.05% (0.74%)	15.21%† [8.05%, 22.38%] ***	1.89 [1.39, 2.58] ***	1.29 [1.15, 1.46] ***
YRO w ISS	RO	Frequency	3.34 (0.26)	2.25 (0.06)	1.09 [0.40, 1.78] **	N/A	N/A
YRO w ISS	YRO only	Rate	67.96% (2.55%)	64.62% (0.93%)	3.34% [-2.39%, 9.07%]	1.16 [0.90, 1.51]	1.05 [0.97, 1.15]
YRO w ISS	YRO only	Frequency	3.34 (0.23)	2.88 (0.08)	0.46 [-0.05, 0.97]	N/A	N/A
RO	Cautions	Rate	36.19% (0.56%)	35.62% (0.44%)	0.57% [-1.13%, 2.26%]	1.03 [0.95, 1.10]	1.02 [0.97, 1.07]
RO	Cautions	Frequency	1.20 (0.03)	1.17 (0.03)	0.03 [-0.07, 0.13]	N/A	N/A

Treatment group sentence	Control group sentence	Reoffending outcome	Treatment group mean (SE)	Control group mean (SE)	Impact estimate: Mean difference [CI LB-UB]	Impact estimate: Odds ratio [CI LB-UB]	Impact estimate: Risk ratio [CI LB-UB]
Discharge	RO	Rate	45.69% (1.16%)	50.92% (0.59%)	-5.22%† [-8.62%, -1.82%] **	0.81 [0.71, 0.93] **	0.90 [0.84, 0.96] **
Discharge	RO	Frequency	1.75 (0.08)	2.14 (0.05)	-0.39 [-0.68, -0.10] **	N/A	N/A
Absolute Discharge	Conditional Discharge	Rate	49.47% (2.96%)	47.59% (1.29%)	1.88% [-4.96%, 8.71%]	1.08 [0.82, 1.42]	1.04 [0.90, 1.20]
Absolute Discharge	Conditional Discharge	Frequency	2.69 (0.27)	2.09 (0.10)	0.60 [-0.01, 1.21]	N/A	N/A
Discharge	Cautions	Rate	46.93% (1.20%)	50.30% (0.47%)	-3.37% [-7.52%, 0.77%]	0.87 [0.74, 1.03]	0.93 [0.86, 1.02]
Discharge	Cautions	Frequency	1.86 (0.09)	1.95 (0.03)	-0.09 [-0.38, 0.21]	N/A	N/A

*Notes: *** = significant at 0.001 level, ** = significant at 0.01 level, * = significant at 0.05 level, † =Due to rounding effects, the impact estimate mean difference may deviate from the presented difference between treatment and control group shown in the table by 0.1*

Table 5 After Year 11 sentencing comparison results

Treatment group sentence	Control group sentence	Reoffending outcome	Treatment group mean	Control group mean	Impact estimate: Mean difference [CI LB-UB]	Impact estimate: Odds ratio [CI LB-UB]	Impact estimate: Risk ratio [CI LB-UB]
YRO	RO	Rate	54.93% (0.77%)	47.00% (0.57%)	7.93% [4.32%, 11.54%] ***	1.37 [1.19, 1.59] ***	1.17 [1.08, 1.26] ***
YRO	RO	Frequency	2.15 (0.05)	1.69 (0.03)	0.46 [0.23, 0.69] ***	N/A	N/A
YRO only	RO	Rate	55.35% (0.81%)	45.99% (0.57%)	9.36% [5.80%, 12.92%] ***	1.46 [1.26, 1.68] ***	1.20 [1.12, 1.30] ***
YRO only	RO	Frequency	2.15 (0.06)	1.62 (0.03)	0.53 [0.32, 0.73] ***	N/A	N/A
YRO w ISS	RO	Rate	51.69% (2.55%)	43.77% (0.70%)	7.92% [1.12%, 14.72%] *	1.37 [1.04, 1.81] *	1.18 [1.02, 1.36] *
YRO w ISS	RO	Frequency	2.17 (0.19)	1.42 (0.04)	0.75 [0.32, 1.17] ***	N/A	N/A
YRO w ISS	YRO only	Rate	54.25% (2.33%)	56.28% (0.82%)	-2.04%† [-7.22%, 3.15%]	0.92 [0.75, 1.14]	0.96 [0.88, 1.06]
YRO w ISS	YRO only	Frequency	2.41 (0.18)	2.21 (0.06)	0.20 [-0.19, 0.59]	N/A	N/A
RO	Cautions	Rate	31.24% (0.53%)	29.94% (0.47%)	1.30% [-0.45%, 3.05%]	1.06 [0.98, 1.15]	1.04 [0.99, 1.11]
RO	Cautions	Frequency	0.93 (0.03)	0.90 (0.02)	0.02† [-0.06, 0.10]	N/A	N/A

Treatment group sentence	Control group sentence	Reoffending outcome	Treatment group mean	Control group mean	Impact estimate: Mean difference [CI LB-UB]	Impact estimate: Odds ratio [CI LB-UB]	Impact estimate: Risk ratio [CI LB-UB]
Discharge	RO	Rate	40.02% (0.91%)	44.05% (0.56%)	-4.03% [-7.24%, -0.82%] *	0.85 [0.74, 0.97] *	0.91 [0.84, 0.98] *
Discharge	RO	Frequency	1.40 (0.05)	1.58 (0.04)	-0.19† [-0.39, 0.02]	N/A	N/A
Absolute Discharge	Conditional Discharge	Rate	50.70% (2.96%)	40.76% (0.97%)	9.94% [3.4%, 16.47%] **	1.49 [1.15, 1.94] **	1.24 [1.09, 1.43] **
Absolute Discharge	Conditional Discharge	Frequency	1.86 (0.18)	1.54 (0.06)	0.32 [-0.08, 0.71]	N/A	N/A
Discharge	Cautions	Rate	42.18% (0.92%)	44.38% (0.52%)	-2.20% [-5.96%, 1.55%]	0.91 [0.78, 1.06]	0.95 [0.87, 1.04]
Discharge	Cautions	Frequency	1.53 (0.06)	1.89 (0.04)	-0.36 [-0.73, 0.01]	N/A	N/A

*Notes: *** = significant at 0.001 level, ** = significant at 0.01 level, * = significant at 0.05 level, † =Due to rounding effects, the impact estimate mean difference may deviate from the presented difference between treatment and control group shown in the table by 0.1*

References

Austin P. C. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. Pharmaceutical statistics, 10(2), 150–161. https://doi.org/10.1002/pst.433

Eaton, G., & Mews, A. (2019). *The impact of short custodial sentences, community orders and suspended sentence orders on reoffending.* (Ministry of Justice Analytical Series, unnumbered). London: Ministry of Justice.

Mews, A., Hillier, J., McHugh, M. & Coxon, C. (2015). *The impact of short custodial sentences, community orders and suspended sentence orders on re-offending.* (Ministry of Justice Analytical Series, unnumbered). London: Ministry of Justice.

Ministry of Justice. (2019). *Annex A: Impact of sentencing on proven reoffending for young offenders in England and Wales, 2012 to 2014.* London: Ministry of Justice. https://assets.publishing.service.gov.uk/media/5cc1893a40f0b6403c3c014d/PRSQ_Annex_A_April_2019.pdf.

Ministry of Justice. (2025). *Guide to proven reoffending statistics*. London: Ministry of Justice. https://assets.publishing.service.gov.uk/media/6807bdea3bdfd1243078e6ee/Technical_guide_to_proven_reoffending.pdf.

Ridgeway, Greg. 2006. “Assessing the Effect of Race Bias in Post-Traffic Stop Outcomes Using Propensity Scores.” Journal of Quantitative Criminology 22 (1): 1–29. https://doi.org/10.1007/s10940-005-9000-9.

Stuart, E. A. (2008). *Developing Practical Recommendations for the Use of Propensity Scores: Discussion of ‘A Critical Appraisal of Propensity Score Matching in the Medical Literature Between 1996 and 2003’ by Peter Austin, Statistics in Medicine*. Statistics in Medicine 27 (12): 2062–65. https://doi.org/10.1002/sim.3207.

Stuart, E. A. (2010). *Matching Methods for Causal Inference: A Review and a Look Forward*. Statistical Science 25 (1): 1–21. https://doi.org/10.1214/09-STS313.

Youth Justice Board for England and Wales. (2025). *Youth Justice Statistics 2023 to 2024 England and Wales. (Statistics bulletin).* https://www.gov.uk/government/statistics/youth-justice-statistics-2023-to-2024/youth-justice-statistics-2023-to-2024.

Appendix

This appendix contains additional technical details and results regarding the analysis presented in this report.

Propensity score matching approach

To match, treatment and comparison sentencing occasions on propensity scores, Epanechnikov Kernel matching was used on the logit of propensity score with a bandwidth of 0.2 of the pooled standard deviation of the logit of the propensity score (Austin, 2011). This algorithm involves treatment observations being matched to as many comparison observations as possible with the latter being weighted according to the proximity of their propensity scores to those of the treatment observations; the closer the propensity scores the higher the weighting. Other methods considered were radius matching which is similar to Kernel matching but without weighting as well as 3:1 nearest-neighbour matching with replacement and a caliper.

During matching, common support restrictions were implemented. This meant 'treatment' sentencing occasions which did not have common support in the control group, that is which didn't have a similar propensity score to that of any 'comparison' sentencing occasion, were excluded. Thus, the results of this study can only be considered representative for the matched treatment group of sentencing occasions.

Matching quality for each sentencing comparison was assessed using SMDs. This evaluated how closely comparable the matched groups were on the characteristics selected in the logistic regression model to estimate the propensity scores. The SMD is calculated by standardising the absolute weighted mean difference between treatment and comparison group for each characteristic using a standardisation factor; i.e. dividing the mean difference by a standard deviation. In this analysis, the standard deviation of a given variable in the treated group was used as implemented in the cobalt R package, which implements recommendations by Stuart (2008, 2010) for this calculation. A key detail is that the standard deviation, no matter how it is computed, is always computed using the unadjusted sample; in this case to estimate the average treatment effect on the treated, the standard deviation of the treated group was used.

Variable selection

The control variables listed in the main body of this report do not represent the entire set of variables that were originally considered for inclusion in the PSM analysis. For instance, for all DfE indicator variables listed in the section *Demographic characteristics, social care and educational experience in the year(s) prior to the index disposal* above, similar versions were tested that aggregated the same kind of information across school years 1-6; for example, if someone was ever eligible for free school meals in school years 1-6.

Control variables were selected using a data-driven approach which focussed on reducing multi-collinearity between control variables. Specifically multi-collinearity was assessed using pair-wise correlation analysis across all available control variables as well as analysis of the variance inflation factor (VIF). In simple terms, the VIF gives a numerical value that indicates how much of the variance of a regression coefficient is inflated due to multicollinearity. The adjusted generalized standard error inflation factor (aGSIF) was used instead of the generalised variance inflation factor. The main benefit thereof is that it adjusts for the number of levels for non-binary categorical variables allowing comparability with other variables. aGSIF was assessed in an iterative process removing variables where the aGSIF was greater than the square root of 5 (~2.24) which indicated the presence of high multi-collinearity until all aGSIFs were below this cut-off.

Variables were also removed if they had perfect or near perfect collinearity (r > 0.99) based on the correlation analysis. This resulted in the removal of the squared versions of 'age at conviction' and 'age at first contact with the justice system' variables.

The 'total number of previous offences' variable was not included as it was the same as the sum of four variables coding total number of previous offences by offence severity level, resulting in perfect multi-collinearity.

Analysis of the variance inflation factor led to the exclusion of:

- Squared versions of the four 'number of previous offences by offence severity' variables
- Copas rate[7] and Copas rate squared
- Number of previous prison events squared
- Number of previous conviction events and its squared version
- Number of disabilities recorded
- Number of disabilities recorded squared

Finally, educational achievement measures could not be included as control variables. This was due to policy changes impacting attainment measures so they were not comparable in our cohorts.

Data imputation

Analysis checked missingness across included control variables. Missingness for individual variables was less than 4.5% for the Year 8-9 cohort, less than 5.5% for the Year 10-11, and less than 11.5% for the After Year 11 cohort.

[7] The Copas rate controls for the rate at which an offender has built up convictions. The higher it is, the more likely the offender is to reoffend. The formula is as follows:

$$copas\ rate = \log_e\left(\frac{Number\ of\ court\ appearances\ or\ cautions + 1}{Length\ of\ criminal\ career\ in\ years + 10}\right)$$

Across the three year-group cohorts, complete case analysis revealed that across the variable selection, case completeness was as shown in Table 6.

Table 6. Complete case counts and proportions for each cohort after the final variable selection

Cohort	N complete	N incomplete	% complete	% incomplete
Year 8-9	34,291	1,928	94.68	5.32
Year 10-11	92,335	5,312	94.56	5.44
After Year 11	89,258	13,086	87.21	12.79

Instead of removing control variables from the analysis to achieve around 95% or more of complete cases for each year-group cohort (thus dropping 5% of sentencing occasions), missing data was imputed using the K-nearest-neighbours method using the R package VIM (function kNN). Imputation was applied to each year group cohort separately but not separately by treatment status for the various sentencing comparisons. This ensured that imputation preserved covariate comparability and sample size prior to propensity score estimation, avoiding potential bias from within-group imputations. A potential limitation of this approach is that it may attenuate genuine differences between treatment and control sentence groups if missingness patterns differed by treatment status. To impute missing values for a given sentencing occasion, the algorithm finds the top five similar neighbours (sentencing occasions) and then uses the control variable values of the neighbours to impute the missing ones. The values of the nearest neighbours are aggregated using their weighted average, meaning the nearer (more similar) neighbour will be given a larger weight in the calculation.

There was one exception to this; data was not imputed for the ethnic group variable which was missing for less than 0.1% for each year group, and incomplete cases were removed as a result.

Impact effects hypothesis testing’

Following the PSM procedure, treatment effects were estimated using weighted outcome models applied to the matched sample. For continuous outcomes (reoffending frequency), this model was estimated using weighted least squares linear regression, and for binary outcomes (reoffending rate) using weighted logistic regression. To estimate, the ATT, the matching weights derived from the PSM procedure were applied during the modelling to ensure that the analysis reflected the weighted matched sample. For both outcome measures, the regression specification included only the treatment indicator (sentence of interest vs comparison sentence) as an independent variable. This approach was chosen after checks confirmed a good quality of matching on the propensity score.

Using the computed models described above, contrasts of predicted outcomes between the treatment sentence and comparison sentence were calculated. Two-sided Wald tests

(z-tests) were used to assess statistical significance, and 95% confidence intervals were computed. This was implemented using the marginaleffects package in R.

For the reoffending frequency outcome, the average marginal mean difference in predicted outcomes between treated and matched comparison units was calculated.

For binary reoffending rate outcomes, the analysis computed average marginal effects (mean difference, odds ratio, and risk ratio) by comparing predicted probabilities for the treatment sentence and comparison sentence for each observation.

PSM Results: Matched sample sizes

Table 7 gives an overview of the number of sentencing occasions in both treatment and control group by match group. For the treatment group, we report the number of sentencing occasions that could be matched on their propensity score as well as the number of sentencing occasions that could not be matched with a control group case during kernel matching. For the control group, we report the total (unweighted) number of sentencing occasions used during matching. Please note that the weighted sample size of the control group after kernel matching is the equivalent to the size of the matched treatment group. In addition, for the matched control group, we report the effective sample size (ESS), which is a measure of quantifying the effect of weighting the sample during matching. It represents the size of an unweighted sample that would provide the same level of precision as the weighted sample (Ridgeway, 2006). We also report the number of sentencing occasions not used during matching.

Table 7 Sample sizes (N) by match group for each cohort and sentence comparison respectively

Cohort	Treatment group sentence	Control group sentence	N Matched treatment group	N Unmatched treatment group	N Matched control group (N ESS)	N Unmatched control group
Year 8-9	YRO	RO	683	117	2,101(453)	304
Year 8-9	RO	Cautions	2,332	73	5,823 (2,251)	79
Year 8-9	Discharge	RO	401	50	1,818 (496)	587
Year 8-9	Discharge	Cautions	372	79	4,946 (347)	956
Year 10-11	YRO	RO	3,310	332	7,422 (1,060)	199
Year 10-11	YRO only	RO	3,050	255	7,393 (981)	228
Year 10-11	YRO w ISS	RO	278	58	4,585 (460)	3,036
Year 10-11	YRO w ISS	YRO only	334	2	2,635 (1,004)	670
Year 10-11	RO	Cautions	7,409	212	11,617 (5,454)	56
Year 10-11	Discharge	RO	1,834	192	7,145 (1,499)	476
Year 10-11	Absolute Discharge	Conditional Discharge	285	6	1,502 (765)	233
Year 10-11	Discharge	Cautions	1,741	285	11,416 (825)	257
After Year 11	YRO	RO	4,156	490	7,679 (920)	304

Cohort	Treatment group sentence	Control group sentence	N Matched treatment group	N Unmatched treatment group	N Matched control group (N ESS)	N Unmatched control group
After Year 11	YRO only	RO	3,754	428	7,753 (960)	230
After Year 11	YRO w ISS	RO	385	77	5,085 (466)	2,898
After Year 11	YRO w ISS	YRO only	459	3	3,684 (1,581)	498
After Year 11	RO	Cautions	7,608	375	9,544 (4,350)	35
After Year 11	Discharge	RO	2,896	360	7,724 (1,412)	259
After Year 11	Absolute Discharge	Conditional Discharge	286	4	2,590 (1,017)	376
After Year 11	Discharge	Cautions	2,895	361	9,198 (937)	381

Further information

Contact

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

Tel: 020 3334 3536

Email: newsdesk@justice.gov.uk

Other enquires about these statistics should be directed to the Data Linking Hub of the Ministry of Justice:

Data Linking Team: PNC and External Shares

Ministry of Justice, 10th Floor, 102 Petty France, London, SW1H 9AJ

Email: datalinkingteam@justice.gov.uk

General enquiries about the statistical work of the Ministry of Justice can be e-mailed to:

statistics.enquiries@justice.gsi.gov.uk

General information about the official statistics system of the UK is available from:

http://statisticsauthority.gov.uk/about-the-authority/uk-statistical-system

Produced by the Ministry of Justice