Ministry of Justice

Analytical Summary 2018

Out of Court Disposals Pilot: Cautions Reoffending Analysis

Rachel Sturrock and Aidan Mews

The Ministry of Justice (MoJ) launched a12-month pilot in November 2014 to test the impact of a simplified out-of-court disposals (OOCD) framework in 3 police force areas (PFAs). The framework reduced the number of disposals available to the pilot police forces from 6[1] to only 2 OOCD types – community resolutions and Conditional Cautions. There was also a requirement that conditions were attached to both disposals.

Three counterfactual PFAs were identified using the Home Office 'Most Similar Groups' matching system.[2] They were required to operate 'business as usual' during the fieldwork pilot period. While Conditional and Simple Cautions could be used in the counterfactual areas, in practice, Simple Cautions were by far the most common type of caution given.[3]

This paper presents an analysis of the 12-month proven reoffending outcomes between the pilot and the counterfactual areas, focusing on those receiving cautions due to data availability.[4] A proven reoffence was defined as any offence committed in the 12 months after an initial caution, that led to a court conviction, caution, reprimand or warning within the one year follow-up or within a further 3 month waiting period (to allow the offence to be proven in court). [5] [6]

This study used a quasi-experimental methodology. Counterfactual areas were chosen to be a good comparison for the pilot areas and some differences between the areas were accounted for through modelling. There are, however, likely to be unobserved differences between the areas compared.

1 These were Simple Cautions, Conditional Cautions, Community Resolutions, Penalty Notices for Disorder (PNDs), Fixed Penalty Notices, Cannabis Warnings and Khat Warnings.
2 This aligned PFAs in terms of demographics (e.g. population density and deprivation) and throughput of disposals.
3 The PNC does not contain information on whether a caution is Conditional or Simple. The monitoring data indicates that all the initial cautions in the pilot areas were Conditional Cautions. In the counterfactual areas there were 32 individuals who received an initial Conditional Caution, otherwise they all received initial Simple Cautions.
4 As the data available on the Police National Computer does not contain full information on lower tier OOCDs (i.e. Community resolutions) this analysis focused on those receiving higher tier disposals – Simple and Conditional Cautions.
5 The standard methodology used by MoJ is to include a 6 month follow-up period, however, this was not possible due to time constraints. Logistic regression was used to generate an odds ratio summarising the odds of one reoffending outcome to another. In this case, all the odds ratios summarise the likelihood of reoffending.
6 Proven reoffending also excludes all OOCDs except cautions.

Key findings

There were no statistically significant[7] differences between pilot areas (i.e. Conditional Cautions) and counterfactual areas (i.e. predominately Simple Cautions) relating to:

- the likelihood of reoffending
- a survival analysis,[8] taking into account time to reoffend
- severity scores for the reoffence based on the Cambridge Crime Harm Index (the index has been developed by Cambridge University to give an indication of how much harm an offence has caused)[9]
- the likelihood of reoffending for those with domestic violence flagged initial offences[10]

Within the pilot areas, there were no statistically significant differences in the likelihood of reoffending associated with:

- different types of condition associated with the caution
- the number of conditions associated with the cautions

Within the pilot areas, there was some evidence that the reoffending impact of conditions may vary by gender, previous offences and age – in particular, rehabilitative conditions showed a statistically significant association with lower offending for men and higher offending for women.

The views expressed in this Analytical Summary are those of the author, not necessarily those of the Ministry of Justice (nor do they reflect government policy).

7 A result is considered to be statistically significant if (upon applying a statistical test) it is unlikely to have occurred simply by chance. Tests with associated levels of significance of $p<.05$ are taken to be statistically significant.

8 Survival analysis uses information on time taken to reoffend, and whether an individual reoffends, to calculate survival probabilities that show how likely it is that an offender will remain offence-free during the follow-up period.

9 The index has been developed by Cambridge University to add another dimension to assessment of reoffending, by giving an indication of how much harm an offence has caused. See for example: Sherman, Neyroud & Neyroud, The Cambridge Crime Harm Index: Measuring Total Harm from Crime Based on Sentencing Guidelines (2016), available at https://academic.oup.com/policing/article/10/3/171/1753592/The-Cambridge-Crime-Harm-Index-Measuring-Total

10 While domestic violence was flagged in the monitoring data collected by PFAs, reoffences were picked up in the PNC which does not have a domestic violence flag. Therefore, while it is possible to compare the overall reoffending of those who committed an initial domestic violence flagged offence, it is not possible to assess whether any reoffences were domestic violence related.

Background

Between November 2014 and October 2015, 3 police forces piloted a simplified OOCD framework, A final evaluation report for the pilot was produced by Sheffield Hallam University in conjunction with Ipsos MORI and published in 2018.[11]

The original evaluation included an analysis of 3-month proven reoffending in the pilot compared to the counterfactual areas for offenders given an OOCD during the first 3 months of the pilot. The impact evaluation did not find a statistically significant difference between reoffending in the counterfactual and pilot areas.

This summary provides a follow-up analysis of the proven 12-month reoffending outcomes between the pilot and counterfactual areas for those receiving Simple or Conditional Cautions.[12]

Research objectives

There were 4 key objectives for the reoffending analysis:

1. Did increased use of Conditional Cautions in the pilot PFAs result in a difference in proven reoffending following cautions? This was in comparison to counterfactual areas who were heavily reliant on the use of Simple Cautions.
    - Was there a difference in the likelihood of reoffending following a caution in the pilot compared to the counterfactual areas?
    - Was there a difference in the frequency of reoffending, the severity of reoffending, or the time taken to reoffend following a caution in the pilot compared to the counterfactual areas?
2. Did pilot areas improve their practice over time and did the likelihood of reoffending following a caution change during the course of the pilot?
3 Were Conditional Cautions associated with differences in the likelihood of reoffending for those with domestic violence flagged offences?
4. How effective were the conditions attached to Conditional Cautions, and who were they are most effective for?

Approach

In the main evaluation, the police in the 3 pilot and 3 counterfactual PFAs collected monitoring data on OOCD cases between 3 November 2014 and 31 October 2015. The monitoring data was then matched to the Police National Computer (PNC) to provide proven reoffending and criminal history for each individual.

The PNC does not contain information on all lower tier OOCDs (i.e. Community Resolutions, PNDs etc), this analysis therefore focused on higher tier OOCDs (Simple and Conditional Cautions).

Overall, 25,628 individuals were in the monitoring data (including lower tier OOCD). 17,245 individuals were dropped after matching to PNC, cleaning and filtering the data, leaving 8,383 individuals.

Table 1: Number and proportion of individuals by pilot and counterfactual areas following matching

	Individuals	%
Pilot	4670	56%
Counterfactual	3713	44%
Total	8383	100%

As Table 1 indicates, after matching, 44% of individuals in the sample were from the counterfactual areas and 56% from the pilot areas.[13]

Analysis was undertaken on the matched data to investigate differences between pilot and counterfactual areas. Standardised (mean) differences between the groups in terms of number of previous offences, and the ethnicity categories of unknown, Asian and white North European were above 10%, differences in the age at which individuals received their first caution during the pilot period were above 5%.[14] These differences were accounted for using multivariate regression (see below).

Reoffending analysis

The reoffending analysis presented in this report focused on 12-month proven reoffending. This was defined as any offence committed which led to a court conviction, caution, reprimand or warning in the one-year follow-up, or within a further 3-month waiting period to allow the offence to be proven in court.[15]

[11] See 'Adult Out of Court Disposal Pilot Evaluation – Final Report' published by the MoJ in 2018.

[12] The reoffending analysis in this evaluation is based on a 12-month reoffending follow-up period after the first OOCD was given, plus a 3-month lag for any offending to be proven and recorded.

[13] The slightly smaller proportion in the counterfactual areas is mainly due to the fact that one counterfactual area did not provide a full date of birth for each offender and therefore many of these individuals could not be matched.

[14] Standardised differences can be interpreted as follows: <5% = groups are closely matched on that particular offender or offence characteristic, 5–10% = a reasonable match quality, >10% = a poor quality of matching which could alter the interpretation of the final result.

[15] Proven reoffending excludes all OOCDs except cautions.

Multivariate analysis was used to investigate reoffending for those in the pilot or counterfactual areas (see Appendix A for more detail). To control differences between groups which may account for some of the variation in reoffending, the following factors were included in each model:

- whether the individual was in the pilot or counterfactual area
- ethnicity,[16] gender and age at first caution during the pilot period
- number of previous proven offences
- offence type of the first caution received during the pilot period
- whether the first caution had a domestic violence flag[17]

Likelihood of reoffending

Likelihood of reoffending following a caution was the key outcome measure. The odds of reoffending during the one-year follow-up period in the pilot and counterfactual areas were compared. This was assessed using logistic regression which allows an odds ratio to be generated, summarising the odds of one reoffending outcome to another. In this case, all the odds ratios summarise the likelihood of reoffending.

Frequency of reoffending

The number of proven reoffences following a caution committed during the one-year follow-up period was assessed using negative binomial regression. Negative binomial models are used where a dependent variable is a count variable.

Severity of reoffending

The Cambridge Crime Harm Index was used to gain an understanding of the severity of reoffences committed by those in the pilot and counterfactual areas following a caution.[18] This index scores offences on the basis of the number of days needed to complete the minimum sentence recommended by the Sentencing Council of England and Wales in Sentencing Guidelines.[19] The starting point is the sentence the crime itself deserves, without any consideration of the prior offending history (or lack of it) on the part of the offender, or the aggravating or mitigating circumstances of the offence.[20]

Propensity score matching (PSM) was found to be the most robust and appropriate methodology for assessing differences in the severity score. PSM is a way of constructing matched comparison groups, allowing the outcomes of the offenders in the pilot areas who received the 'treatment' to be compared with similar offenders in the counterfactual areas.

PSM was conducted on those who reoffended following a caution in the pilot and counterfactual areas (see Appendix B for more detail).

The time taken to reoffend

This was analysed using survival analysis. This analysis uses information on time taken to reoffend following a caution, and whether an individual reoffends, to calculate survival probabilities that show how likely it is that an offender will remain offence-free during the follow-up period. In order to conduct this analysis, matched groups were created using PSM. Groups were matched in terms of ethnicity, gender, age, number of previous offences, type of offence that led to the initial caution and whether the offence was domestic violence flagged. The closeness of the matched groups on characteristics selected for the model was tested using standardised (mean) differences. The groups were well matched.

Key caveats and limitations

The analysis assessed the combined impact of all the interventions used in the Conditional Cautions. It did not assess the effectiveness of individual conditions and interventions. It is possible that the effectiveness of individual interventions are offset by the ineffectiveness of other interventions when considered at a global level.

While a range of variables were controlled for in the modelling, there may be important variables that are unobserved, creating bias in the result.

One of the counterfactual areas did not provide a full date of birth for each individual in the monitoring data. This affected the matching process and reduced the size of the counterfactual group. This was managed through controlling for any differences in the pilot and

[16] Ethnicity was taken from the PNC, therefore it is officer-identified.
[17] Although data were collected on the number of hate crime flagged offences, there was a very small sample. This was not a statistically significant factor in the model and worsened the fit of the model, therefore it was not included in the final model.
[18] See footnote 9.
[19] For example, a prison sentence of a year would lead to a score of 365. If the sentence is a fine or a community order, the score is calculated as the number of hours worked to complete the community sentence or a pay a fine, and then this is converted into days.
[20] The complete list of the starting point sentences by offence types is posted on the University of Cambridge website here: https://docs.google.com/spreadsheets/d/12uwGjz2HhoqKXsHB-F-J2vvWgHvLJ7ukrDIVP8YjlpQ/edit#gid=0.

counterfactual areas such as gender, age, ethnicity and number of previous offences

The monitoring data from PFAs were manually populated, drawing on records from different police systems including free text case management records. This was time consuming and the data was dependent on how comprehensively and accurately it had been recorded by officers and support staff. These therefore required extensive cleaning and any data gaps were resolved where possible.

The MoJ has a standard 12-month reoffending period, with a 6-month recording lag to ensure that records have been processed and recorded on the PNC. Due to time constraints, the lag period was reduced to 3 months which may lead to under-counting of reoffences, particularly those that are more severe

The severity analysis involved PSM. PSM provides an unbiased estimate where all the relevant variables which can account for the probability of being assigned to the treatment group, and those which are associated with the outcome of interest are included. In this case, our information is limited to the monitoring data and information from the PNC. Therefore, as with the other analyses, it is not possible to control for all the differences that might account for variations in outcomes between the pilot and counterfactual groups.

There were 6 individuals in the pilot areas who committed serious crimes which attracted very high outlier scores in the severity analysis. As these individuals were not the intended target of this intervention, and these scores distorted the distribution considerably, these were removed from the severity analysis.

In total, 50 individuals were removed from the severity analysis as none of their recorded offences were possible to score. Sixteen of these individuals were in the counterfactual and 34 were in the pilot areas. Just under 60% of the sample of those who reoffended were in the pilot area, so the distribution of missing values is reasonably similar to the overall distribution of the sample.

While domestic violence was flagged in the monitoring data collected by PFAs, reoffences were picked up in the PNC which does not have a domestic violence flag. Therefore, while it is possible to compare the overall reoffending of those who committed an initial domestic violence flagged offence, it is not possible to assess whether any reoffences were domestic violence related.

Only a small number of offences (89) linked to hate crime were recorded in the monitoring data. This limited any analysis in this area.

Results of the offending analysis

Table 2 presents the 12-month proven reoffending outcomes of those who received a Simple or Conditional caution during the pilot period. This shows that a slightly higher proportion of those in the pilot reoffended, and committed more reoffences on average than those in the counterfactual areas. However, as elaborated below, these differences were not statistically significant.

Table 2: Number and proportion of offenders reoffending within 12 months of the receiving an OOCD in the pilot, average number of reoffences and average time to first reoffence

	Counterfactual (n=3713)	Pilot (n=4670)
Number reoffending	602	806
Proportion reoffending	16.2%	17.3%
Average number of reoffences	0.26	0.29
Average time to first offence	124 days	131 days

Note: Differences were not statistically significant.

Likelihood of reoffending

There was no statistically significant difference in the odds of reoffending following a caution in the pilot and counterfactual areas after multivariate analysis. This indicates there was no association between the increased use of Conditional Cautions in the pilot areas and the likelihood of reoffending following a caution (see Appendix A for more detail).

Frequency of reoffending

Analysis of the frequency of reoffending investigated whether the number of reoffences committed by an individual following a caution was significantly different in the pilot and counterfactual areas. There was no statistically significant difference in the frequency of reoffending between the pilot and counterfactual areas (see Appendix A for more detail).

Severity of reoffending

Total Cambridge Crime Harm Index scores were calculated for those who reoffended following a caution during the pilot period. After creating matched groups, scores for the pilot and counterfactual areas were compared. This study found there was no statistically significant difference between severity scores for the pilot and counterfactual areas.

Time to reoffence

A survival analysis was conducted to investigate any differences between the pilot and counterfactual areas in the probability of not reoffending over time (see methodology for details).[21] There was a very similar pattern of reoffending following a caution in both the pilot and counterfactual areas, with no statistically significant difference. Over time, the probability of lasting without offending reduced by a very similar amount, and at a very similar rate in the pilot and counterfactual areas.

Reoffending by quarter

In order to understand whether there was learning over time during the course of the pilot, it was necessary to explore whether reoffending following a caution varied according to when the caution was received.

There were no statistically significant differences between the likelihood of reoffending for those receiving a caution in quarter 1, 2, 3 or 4.

There were also no statistically significant differences in the proportion of those reoffending following a caution in the pilot compared to the counterfactual areas in quarter 1, 2, 3 or 4.

Domestic violence flagged offences

As noted in the key limitations, while it was possible to compare the reoffending of those who committed an initial domestic violence flagged offence, it was not possible to assess whether any reoffences were domestic violence related. This section therefore focuses on overall reoffences (shown in Table 3).

Table 3: Number and proportion of individuals whose initial OOCD during the pilot period was a domestic violence flagged offence

	Pilot	Counterfactual
Domestic violence flagged	985 (21%)	785 (21%)
Not domestic violence flagged	3,685 (79%)	2,928 (79%)
Total	4,670	3,713

The general reoffending rate for domestic violence flagged offences in the pilot areas was 15.3% and 15.6% in the counterfactual areas.

For those who committed a domestic violence flagged offence, likelihood of reoffending was modelled against whether the individual was in the pilot or the counterfactual area, controlling for age, gender, ethnicity, offence type, and number of previous offences. There was no statistically significant difference in the likelihood of reoffending following a caution between the pilot and counterfactual areas.

The effectiveness of conditions

There were 4 main categories of conditions: reparative, rehabilitative, punitive and restrictive. The most common condition type was rehabilitative, followed by reparative, restrictive and then punitive.

Multivariate analysis was used to investigate differences in reoffending following a caution between those with different types of condition within the pilot areas.

One individual could have multiple types of condition; therefore, it was difficult to isolate the effect of any specific type of condition. This analysis therefore focused on the variation in reoffending for those with at least one type of a condition in comparison to not having that condition.

There were no statistically significant differences in the likelihood of reoffending following a caution for those with different types of condition within the pilot areas. Similarly, there were no significant differences in likelihood of reoffending for those who had one, 2 or 3 conditions of one specific condition type in comparison to not having any conditions of that type.

In order to explore who Conditional Cautions are most effective for, analysis was conducted for each condition on whether reoffending varied by gender, ethnicity, age and number of previous offences.

There was a statistically significant interaction between gender and receiving a rehabilitative condition. Rehabilitative conditions seemed to be associated with lower offending for men and with higher offending for women. For example, with no rehabilitative condition, men had 49% higher odds of reoffending than women, however, with a rehabilitative condition, their odds of reoffending were 9% lower than those of a woman.

There was also a statistically significant interaction between having a restrictive condition and 2 previous offences. Restrictive conditions were associated with lower reoffending for those with 2 previous offences and with higher reoffending for those with no previous offences. For example, with no restrictive condition, the odds of reoffending were 132% higher for those with 2 previous offences than for those with no previous offences. With a restrictive condition, those with 2 previous offences were 2% less likely to offend than those with no previous offences.

[21] PSM was conducted to create matched groups used in this survival analysis (see page 3 for more details).

There was also a statistically significant interaction between being in the 40–59 age group and receiving a punitive condition. With no punitive condition, those who were 40–59 had 72% lower odds of reoffending in comparison to those who were 18–20. With a punitive condition, those aged 40–59 were 94% less likely to offend than those aged 18–20.

Conclusion

This study assessed whether the increased use of Conditional Cautions in 3 OOCD pilot PFAs was associated with changes in 12-month proven reoffending. This was in comparison to the more frequent use of Simple Cautions in the counterfactual areas.

The results do not indicate any statistically significant differences between the pilot and counterfactual areas in terms of the proportion who reoffended following caution, the frequency of reoffending, the time taken to reoffend or the severity of reoffending. Therefore, it is not possible to conclude that there is an association between the increased use of Conditional Cautions in the pilot areas compared to the counterfactual area and differences in reoffending behaviour.

There was also no evidence that being in the pilot areas was associated with differences in likelihood of reoffending following a caution for those with domestic violence flagged offences.

Exploring the impact of conditions was challenging given that individuals frequently had multiple types of condition – it was therefore not straightforward to isolate the effect of a condition. This analysis indicated that none of the conditions in isolation were associated with statistically significant reductions in reoffending following cautions. However, it did suggest that the benefits of the conditions in reducing reoffending may vary by gender, previous offences and age.

Appendix A

Results of the reoffending prevalence analysis

Table A shows the results of the multivariate analysis of the likelihood of reoffending following caution. It presents the odds of reoffending associated with each of the characteristics below, as compared to the chosen reference category. For example, the odds of reoffending for those who were 21–24 were 0.67, therefore this age group is 33% less likely to reoffend than the 18–20 age group.

Table A: Results of the logistic regression, predicting the likelihood of reoffending[22]

	Odds Ratio	95% Confidence Interval	
Counterfactual (reference)			
Pilot	0.98	0.87	1.11
No previous offences (reference)			
1 previous offence	1.51*	1.25	1.81
2 previous offences	1.61*	1.29	2.01
3 previous offences	2.23*	1.73	2.86
4 previous offences	2.50*	1.88	3.30
5 previous offences	4.54*	3.27	6.24
6 previous offences	4.00*	2.87	5.51
7 previous offences	4.04*	2.63	6.08
8 previous offences	3.80*	2.42	5.82
9 previous offences	3.99*	2.46	6.31
10 previous offences	8.09*	5.15	12.62
11–15 previous offences	5.29*	3.89	7.16
16–20 previous offences	7.70*	5.22	11.29
21 and above previous offences	10.88*	7.72	15.35
White – North European (reference)			
Asian	1.06	0.82	1.35
Black	1.42*	1.09	1.82
White – South European	1.36	0.90	2.02
Other ethnic and unknown	0.37*	0.21	0.61
Female (reference)			
Male	1.07	0.92	1.25
Not domestic violence flagged			
Domestic violence flag	1.17	0.99	1.37
Aged 18–20 (reference)			
Aged 21–24	0.67*	0.55	0.81
Aged 25–29	0.47*	0.39	0.58
Aged 30–39	0.48*	0.40	0.58
Aged 40–49	0.33*	0.26	0.42
Aged 50–59	0.20*	0.15	0.28
Aged 60+	0.16*	0.09	0.27
Summary offences excluding motoring (reference)			
Other offences	0.71	0.48	1.02
Violence against the person	1.00	0.66	1.46
Theft offences	2.40*	2.03	2.83
Drug offences	1.11	0.93	1.31
Possession of weapons	1.56	0.94	2.48
Public order offences	0.86	0.54	1.32
Sample size	8,330[23]		

[22] Statistically significant odds ratios are marked with an asterisk.
[23] 53 records were removed from the regression analysis due to gender not being identified

Results of the reoffending frequency analysis

Table B shows the results of the multivariate analysis of frequency of reoffending following cautions in the pilot and counterfactual areas. It presents the frequency of reoffending associated with each of the characteristics below, as compared to the chosen reference category. For example, the expected frequency of reoffending for those whose initial offence was theft is 2.53 times that for those whose initial offence was a summary offence excluding motoring.

Table B: Results of the negative binomial regression, predicting the reoffending incident rate[24]

	Reoffending incident rate	95% confidence interval	
Counterfactual (reference)			
Pilot	0.98	0.88	1.11
Summary offences excluding motoring (reference)			
Other offences	0.67*	0.47	0.94
Violence against the person	0.83	0.56	1.23
Theft offences	2.53*	2.16	2.96
Drug offences	1.12	0.95	1.32
Possession of weapons	1.25	0.78	2.02
Public order offences	0.88	0.58	1.33
Aged 18–20 (reference)			
Aged 21–24	0.70*	0.58	0.83
Aged 25–29	0.47*	0.39	0.57
Aged 30–39	0.51*	0.43	0.61
Aged 40–49	0.31*	0.25	0.39
Aged 50–59	0.22*	0.16	0.30
Aged 60+	0.22*	0.14	0.35
No previous offences (reference)			
1 previous offence	1.48*	1.25	1.75
2 previous offences	1.47*	1.18	1.82
3 previous offences	2.34*	1.85	2.96
4 previous offences	2.53*	1.94	3.30
5 previous offences	3.94*	2.88	5.38
6 previous offences	3.32*	2.41	4.57
7 previous offences	4.88*	3.31	7.20
8 previous offences	4.57*	3.03	6.88
9 previous offences	4.24*	2.73	6.59
10 previous offences	5.75*	3.75	8.82
11–15 previous offences	5.01*	3.74	6.69
16–20 previous offences	6.68*	4.64	9.61
21 and above previous offences	10.02*	7.33	13.71
White – North European (reference)			
Other Ethnic and unknown	0.47*	0.30	0.74
Asian	1.06	0.83	1.34
Black	1.31*	1.02	1.67
White – South European	1.29	0.88	1.89
Female (reference)			
Male	1.08	0.94	1.25
Not domestic violence flagged (reference)			
Domestic violence flag	1.12	0.96	1.31

[24] Statistically significant estimates are marked with an asterisk.

Appendix B

Severity analysis methodology

The Cambridge Crime Harm Index was used score reoffences according to their severity. The index has been developed by Cambridge University to add another dimension to assessment of reoffending, by giving an indication of how much harm an offence has caused.

This index scores offences on the basis of the number of days needed to complete the sentence recommended by the Sentencing Council of England and Wales in Sentencing Guidelines as the 'starting point' for sentencing. For example, a prison sentence of a year would lead to a score of 365. If the sentence is a fine or a community order, the score is calculated as the number of hours worked to complete the community sentence or a pay a fine, and then this is converted into days. The starting point is the sentence the crime itself deserves, without any consideration of the prior offending history (or lack of it) on the part of the offender, or the aggravating or mitigating circumstances of the offence.

Each offender was assigned a score for every offence they committed in the 12 month follow up period after their initial caution. These scores were summed and each offender given a total score for the year. The scores of those who reoffended in the pilot and counterfactual areas were then compared.

A number of different methodologies for comparing the pilot and counterfactual areas were explored. Initially negative binomial regression was used, however, this model was found to be very sensitive to model specification (the way the variables were entered into the model). Propensity score matching was then used as this methodology is more robust to this issue.

Propensity score matching (PSM)

Propensity score matching is a methodology that allows a matched counterfactual group to be constructed. This matching allows the outcomes of the offenders in the pilot areas who received the 'treatment' to be compared with similar offenders in the counterfactual areas.

In order to achieve this, the probability of receiving the 'treatment' needs to be calculated (a propensity score between 0 and 1) using all the important factors which are associated with receiving the treatment and with the outcome of interest. Offenders receiving the treatment are then matched to offenders not receiving the treatment on the basis of their propensity scores. The two groups can then be compared to estimate the effect of the treatment. This approach cannot definitively control for all potential differences between the matched groups as there may be important factors that have not been included in the estimation of the propensity scores. This is, however, a robust way of reducing differences between groups.

- PSM was conducted on those who reoffended in the pilot and counterfactual areas, using the following variables:
- Whether the individual was in the pilot or counterfactual area.
- The severity score associated with their first caution received during the pilot period.[25]
- Ethnicity[26] gender and age at first caution during the pilot period.
- Number of previous proven offences.
- Offence type of the first caution received during the pilot period.
- Whether their first caution had a domestic violence flag.

Numerous PSM algorithms for generating a comparison group are available. This study found Kernal matching using a normal distribution (with a bandwidth of 0.03) provided the best balance between an excellent quality of matching and minimised the number of treatment individuals who couldn't be matched.

[25] This was included as continuous variable in the final analysis, when included as a categorical variable the results were similar.
[26] Ethnicity was taken from the PNC, therefore it is officer- identified.

The impact estimates exclude treatment cases where there is 'no common support' – that is, those for which there were no cases with a similar propensity score in the counterfactual areas. After matching, less than 1% of the treatment group were lost. Table C1 shows the number of individuals in the sample after matching. Treatment cases are excluded where there is 'no common support' – that is, those for which there were no cases with a similar propensity score in the counterfactual areas. After matching 8 individuals in the treatment group were lost.

Table C1: Number of individuals in the pilot and counterfactual groups after propensity score matching.

	Off support	On Support	Total
Counterfactual	0	586	586
Pilot	8	753	761
Total	8	1339	1347

Following matching, the closeness of the matched groups on characteristics selected for the model was tested using standardised (mean) differences (see p. 7 for detail on interpreting standardised mean differences). The matching quality was very high, with the average of the standardised mean differences being 1.5% and the highest being 3.4%.

Results of the severity analysis

Total Cambridge Crime Harm Index scores were calculated for those who reoffended following cautions during the pilot period. After creating matched groups, scores for the pilot and counterfactual areas were compared. The study found there was no statistically significant difference between severity scores for the pilot and counterfactual areas.

Table A shows the scores after matching, these were compared using a T-test and there was no statistically significant difference found.

Table C2: Mean and Cambridge Harm Index scores for those who reoffended in the pilot and counterfactual areas after matching.

	Number of individuals	Mean score	Difference	Standard error	T-statistic	p-value
Counterfactual	586	21.07	3.45	4.64	0.74	0.23
Pilot	753	24.52				

Analytical Services exists to improve policy making, decision taking and practice by the Ministry of Justice. It does this by providing robust, timely and relevant data and advice drawn from research and analysis undertaken by the department's analysts and by the wider research community.

First published 2018

ISBN 978-1-84099-817-7

Contact info: mojanalyticalservices@justice.gsi.gov.uk