An impact evaluation of the Drug Recovery Prison at HMP Holme House

Barnaby Elwes and Giles Stephenson

Ministry of Justice

Ministry of Justice Analytical Series

2025

Data and Analysis 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.

Acknowledgements

The authors

Barnaby Elwes

Giles Stephenson

Contents

List of tables

1. Executive Summary

1.1 Background to the Drug Recovery Prison

The Drug Recovery Prison (DRP) was a joint Ministry of Justice and National Health Service intervention piloted at HMP Holme House between 2017 and 2020 to test a 'whole prison approach' to tackling substance misuse. The pilot was launched as part of the Reform Prisons Programme (Ministry of Justice, 2016) and involved an additional spend of up to three million pounds per year. The DRP aimed to test new ways to restrict supply and reduce demand for drugs in prisons, and to support recovery and improve health of prisoners. Prior to the implementation of the DRP, Holme House had a substantial problem with drugs, particularly new psychoactive substances.

1.2 Methodology

A process evaluation of the DRP has already been conducted by Ayres *et al*. (2023) which explored the implementation of the DRP. This found that the DRP was successfully implemented, albeit with some delays, and that generally prisoners and staff had positive views of the model. This new research builds upon the existing process evaluation by investigating the reoffending outcomes of prisoners who resided at the DRP at any point over the pilot period. It aimed to detect whether the DRP reduced reoffending and by what magnitude, and whether this varied by drug need[1]. Reoffending outcomes were:

- the rate of proven reoffending within one year of release from prison;
- the number of proven reoffences within one year;
- the number of days to first proven reoffence; and
- the frequency of proven reoffences resulting in a custodial sentence.

The analysis used Propensity Score Matching (PSM) to compare the reoffending outcomes of DRP prisoners against similar prisoners who resided in other Category C

---

[1] 'Drug need' in this study is defined using criminogenic needs data from OASys assessments (Appendix C). For brevity this group is called the 'drug cohort'. This data is not a formal assessment on whether an individual requires structured drug treatment. Prisoners in both Holme House and the comparator prisons will have received medical treatment for substance misuse. This study does not measure the impact of medical treatment.

prisons. PSM accounts for differences between the groups by using many individual-level variables and characteristics to match similar prisoners together. Over 100 variables were used to match prisoners, reducing the risk of bias and confounding factors, and allowing for a 'like-for-like' comparison. After matching, the two groups were similar across the variables used to match them, meaning robust comparison of their outcomes could be conducted. The sample used in the analysis had 2,451 prisoners in the treatment group and 36,082 prisoners in the comparison group.

The analysis used linked data from three sources: Prison National Offender Management Information System (NOMIS), Police National Computer (PNC) and Offender Assessment System (OASys). The methodology used was based on the Ministry of Justice's existing approach for propensity score matching (Justice Data Lab, 2013), which has been used in prior analyses (Brinn *et al*., 2023; Eaton & Mews, 2019). PSM is considered a rigorous impact evaluation technique (Farrington *et al*., 2002).

1.3 Main Findings

Headline Reoffending Outcomes

The analysis found that prisoners who were housed in Holme House prison were more likely to have a proven reoffence following release than prisoners housed in alternative Category C prisons. All results below were statistically significant.

- For DRP prisoners, 47.8% went on to commit a proven reoffence within one year, compared to 44.6% of comparison prisoners, a difference of 3.2 percentage points.

- DRP prisoners committed 2.4 proven reoffences on average, compared to 2.1 proven reoffences on average for comparison prisoners, a difference of 0.3 proven reoffences within one year following release.

- DRP prisoners took 111 days to commit their first proven reoffence on average, compared to 118 days for comparison prisoners, a difference of seven days.

Drug Cohort Reoffending Outcomes

The analysis found that prisoners in the DRP with a drug need were more likely to have a proven reoffence following release than prisoners in alternative Category C prisons with a drug need. All results below are statistically significant.

- For the drug cohort in the DRP, 57.4% went on to commit a proven reoffence within one year, compared to 52.7% of the drug cohort in the comparison group, a difference of 4.7 percentage points.

- The drug cohort in the DRP committed 3.1 proven reoffences on average, compared to 2.7 proven reoffences on average for the drug cohort in the comparison group, a difference of 0.4 proven reoffences within one year following release.

Non-Drug Cohort Reoffending Outcomes

The analysis found that there was no statistically significant difference for reoffending rate and days to first proven reoffence for prisoners in the DRP without a drug need and prisoners in alternative Category C prisons without a drug need. It found that the non-drug cohort in the DRP committed more proven reoffences following release than the non-drug cohort in alternative C prisons.

1.4   Conclusion

The findings of this impact evaluation indicate that DRP was ineffective at reducing reoffending, compared to other similar prisons. The results show statistically significant higher reoffending outcomes for prisoners in Holme House compared to comparison prisoners. This result is the same for the drug cohort. The analysis shows no statistically significant difference in the reoffending rate for the non-drug cohort. All standardised effect sizes were small, implying that the DRP's impact on the reoffending outcomes for prisoners was limited.

There are a range of methodological and intervention-specific considerations which may partially explain and contextualise the results found:

- **Unobserved variable bias –** whilst over 100 variables were used to match prisoners together, PSM cannot account for unmeasured or hidden variables. As a result, there may be factors related to reoffending which this evaluation does not account for, such as recovery from substance misuse. Furthermore, the evaluation was unable to control for the effects of engagement in substance misuse treatment as this data is not held by the department.

- **Prison-level factors** – since this evaluation compares one specific prison to essentially a 'national average' of all other prisons of the same category, it may miss intrinsic differences between Holme House and all other Category C prisons. For instance, if Holme House has a different regime, different post-release support services or different probation provision to the rest of the country, this may influence the results. Given the different results for the drug cohort and the non-drug cohort, these differences (if they existed) would have to only effect the drug cohort.

- **Regional inequality** – deprivation is a key determinant of reoffending and other factors associated with reoffending, such as homelessness, mental health difficulties and childhood adverse experiences. Whilst the PSM model includes variables on individual measures of deprivation (such as employment and financial status) it may be that regional differences in deprivation drive reoffending at Holme House more than at other Category C prisons. The evaluation attempted to account for regional inequality by comparing the DRP to a subset of prisons in other high deprivation areas. Prisoners in the DRP still reoffended more than prisoners in these comparison prisoners.

- **Implementation** – if the DRP model was not fully implemented or not implemented as intended, this may influence the results found. The process evaluation of the DRP found that there were some difficulties with implementation, such as delays and inefficiency of some equipment, but broadly it found that the model was implemented successfully by the end of the pilot. Prisoners in at Holme House during the earlier stages may not have fully benefited from the DRP.

- **Post-release drivers of reoffending** - there are many factors which increase the risk of reoffending beyond those within the control of the DRP, such as adverse childhood experiences, personal relationships, unemployment, financial problems, homelessness and mental health problems. Whilst the analysis accounts for some of these variables by using OASys assessments completed when prisoners enter prison, it may be that these factors are driving reoffending upon release. The process evaluation of the DRP highlighted a 'cliff edge' of post-release support which was felt to risk prisoner's continued recovery If there is particular lack of post-release support around Holme House to help prisoners with these criminogenic

risks, or greater exposure to drugs in the community, it may drive reoffending more at Holme House than other Category C prisons.

- **Dynamics of Cultural Change** – the DRP sought to overhaul the culture of Holme House. It is likely that the complete reform and cultural change needed to enable a whole prison approach takes many more years than the initial pilot ran for. It is possible that different results may be found if the evaluation is repeated once sufficient time for wholescale cultural change had passed (e.g. 5-10 years).

2. Introduction

2.1 Context

Drug misuse is a substantial problem for the criminal justice system. Half of prisoners have an identified substance misuse need (MoJ, 2025), and an estimated 44% of acquisitive crime is committed by those regularly using heroin and/or crack cocaine (Home Office, 2013). Within prisons, drug misuse contributes to violence, crime and vulnerability, which threatens the safety of prison staff and their ability to deliver effective regimes (HMPPS, 2019). Supporting prisoners with recovery is essential for preventing reoffending and reducing harm.

2.2 The Drug Recovery Prison

In 2016, HMP Holme House[2] was earmarked as the site of a combined Ministry of Justice (MoJ) and Department of Health (DH) pilot to test a 'whole prison approach' to substance misuse. Its aim was to explore new ways of tackling substance misuse in prisons, to support the rehabilitation of prisoners and deter reoffending. In April 2017, a three-year pilot of the 'Drug Recovery Prison' (DRP) officially began. The pilot was funded by an additional spend of up to three million pounds per year, divided between Her Majesty's Prison and Probation Service (HMPPS) and NHS England (NHSE). During the pilot period, Holme House was also changed from a Category B (Remand, training, and resettlement) to a Category C (training and resettlement) prison, but this was separate from the DRP programme. The three-year pilot was divided into year-long implementation phases: developing and initiating; implementing and progressing; consolidating and maintaining practices following the culmination of the pilot (Wheatley, 2019).

Prior to the pilot, Holme House had a substantial problem with drugs. An inspection by Her Majesty's Inspectorate of Prisons (HMIP, 2017) found that 36% of mandatory drug tests results were positive, 26% of prisoners had developed a drug problem in Holme House,

---

[2] Henceforth referred to as Holme House.

and 60% of prisoners thought that it was easy to get drugs into the prison[3]. The high prevalence of spice (a colloquial term for synthetic cannabinoids) was leading to frequent medical emergencies and many prisoners reported developing a substance misuse problem whilst in Holme House (HMIP, 2017). There were, however, areas where Holme House was excelling prior to the pilot – the therapeutic community (TC), a special wing which provides intense support for prisoners with a substance misuse problem, was regarded very highly and thought to be a model for how to encourage community-mindedness and personal responsibility for those with an addiction (HMIP, 2017).

The DRP aimed to move away from punitive approaches to substance misuse within prison, and to acknowledge the trauma, and psychological and emotional distress that prisoners who misuse substances often suffer from (Wheatley, 2019). In line with the HM Government Drug Strategy (Home Office, 2017) and Prisons Drug Strategy (HMPPS, 2019), the DRP had four strategic aims:

- Promote safety and security (restrict supply of illicit items)
- Enhance care and wellbeing (reducing demand for substances)
- Develop the prison environment, making it a more positive place (building recovery)
- Strengthen continuity of care post release by creating vibrant and sustainable links to communities (building recovery) (Wheatley, 2019).

Promote Safety and Security
The DRP aimed to reduce the supply of drugs and illicit items into the prison through the different supply routes: 'over the wall' drops, visitors, corruption, and prisoners themselves. It enhanced its ability to deter, detect and disrupt illicit items through investment in physical, procedural and interpersonal security (Wheatley, 2019). The DRP invested in physical measures through a new staff searching facility, an incoming goods search area with cargo and baggage x-ray machines, a full body x-ray facility in reception, millimetre-wave security scanners in the visitor entry area and visits hall, a magnetic resonance spectrometer, a mobile phone interrogation device, new mobile phone detectors, an

---

[3] 26% of prisoners at Holme House reported developing a drug problem, compared to a national average of 13% (Question 13.4) (HMIP 2017, HMIP 2018).

Automatic Number Plate Recognition camera, metal window opening restriction apparatus, toilet waste examination equipment and a new security building. In addition, a dedicated Drugs Crime Reduction Unit (DCRU) was created to improve procedural security within the prison, encompassing checks and intelligence gathering whilst using principles of procedural justice (Wheatley, 2019).

Enhance Care and Wellbeing

The DRP aimed to expand its treatment offer for prisoners with substance misuse problems alongside medication assisted recovery such as opiate substitution therapy (OST). The DRP introduced six other health improvements to support medication assisted recovery and enhance care and wellbeing:

- Introduce Rec-Cap assessment, a recovery planning and evaluation tool, to improve assessment and development of personal care plans.
- Appoint a nurse consultant to strengthen clinical leadership and strengthen a community of practice to improve service delivery.
- Appoint an applied psychology team to develop a new range of evidence-based interventions that promote health and understanding.
- Appoint a speech and language therapist to improve communication and understanding with prisoners.
- Appoint healthcare support workers and pharmacy technicians to enable wing-based care.
- Appoint a paramedic to improve response to urgent care and accidents and emergencies (Wheatley, 2019).

Develop the Prison Environment

The DRP aimed to improve the physical environment and culture within the prison, acknowledging that the prison infrastructure and architecture, alongside its regime and interpersonal relationships between staff and prisoners are all crucial for improving the rehabilitative capacity of the prison. To do this, the DRP invested in refurbishing its house blocks, treatment rooms, grounds and gardens, and walkways between house blocks. The refurbishments improved the conditions and decency of the prison, as well as aiming to bring in more nature and a sense of the outside world, based on evidence that this can

promote recovery (Moran, 2019). The DRP also introduced a broader range of purposeful activities, such as music, arts, social activities, woodwork and gardening, to utilise the skills of the prisoners in developing the prison environment, to counteract boredom, disorder and to focus prisoner's time on a worthwhile activity (Wheatley, 2019). The DRP expanded the use of peer mentors ('Recovery Navigators') and prisoner Democratic Councils to expand prisoner's sense of input into prison reform, collaboration with prison officers, and to support recovery of others by utilising prisoner's lived experiences (Wheatley, 2019).

Strengthen Continuity of Care

Ensuring that prisoners continue to engage in treatment within the community following release is vital for interventions to be successful in supporting prisoners in their recovery and to reduce reoffending (de Andrade *et al*., 2018; Butzin, Martin and Inciardi, 2005). To strengthen this, the DRP invested in a new team, the Connecting Communities Team, to better help prisoners' transition from custody to the community. This team worked with probation, to help link up prisoners in treatment with support services in the community.

2.3   Existing evidence on the Drug Recovery Prison

The DRP attempted to implement a whole prison approach to substance misuse. However, prior to implementation of the DRP there very little evidence on whether these approaches work to reduce substance misuse, prisoner health and reoffending. A process evaluation of the DRP was conducted by Ayres *et al*. (2023) to assess its implementation and early indication of outcomes. The evaluation found that the model was successfully implemented, albeit with some delays, and that generally prisoners and staff had positive views of the model. Staff and prisoners felt that the additional security elements had reduced drug usage and had created a calmer and safer environment. However, there was an indication that greater security had led to displacement of substances, with more alcohol being brewed as a result. There was some evidence that prisoners had seen a slight improvement in physical and mental health. Peer mentors and the continuity of care team were seen as particular successes, although a 'cliff edge' of post-release support was still a concern which was felt to risk prisoner's continued recovery. Generally, both staff and prisoners thought the DRP had a positive effect on prisoners, staff, the culture,

feelings of safety, relationships, healthcare provision, recovery, rehabilitation, violence, and the supply and use of substances (Ayres *et al*., 2023).

2.4   Research Aims

Building upon the process evaluation of the DRP (Ayres *et al*., 2023), this evaluation aims to assess whether the DRP had a successful effect upon the reoffending outcomes of prisoners who resided there over the pilot period. Ayres *et al*. (2023) identified two long-term outcomes that the DRP might theoretically produce: enhanced health and wellbeing and reduced reoffending. This evaluation measures the impact of the DRP on the second of these outcomes. Data on health and wellbeing outcomes is not held by the Ministry of Justice and as such is out of scope for this evaluation.

The evaluation aims to answer two research questions:

1) Has the DRP had a positive or negative impact on the proven reoffending rate of prisoners who experienced the DRP pilot period?

2) What is the scale of this impact?

3. Methodology

This evaluation aims to assess the impact of the DRP on proven reoffending outcomes. This requires comparing reoffending outcomes of prisoners who experienced the DRP (the treatment group) against prisoners who did not (the comparison group). However, there may be differences in the characteristics of prisoners who were in Holme House compared to those in other prisons. If these groups were compared directly, it would not be clear whether any differences in proven reoffending were due the DRP, or due to other characteristics. To account for these differences, propensity score matching was used.

This methodology is based on an MoJ standard methodology for propensity score matching (Justice Data Lab, 2013), used in previous MoJ analyses (Brinn *et al*., 2023; Eaton & Mews, 2019). Methodological development, analysis and reporting were undertaken by MoJ analysts, in line with GSR standards.

3.1 Propensity Score Matching

Propensity score matching (PSM) is a statistical technique which uses factors associated with both receiving the intervention and the outcome variable (reoffending) to calculate a 'propensity score'[4]. Such characteristics include gender, ethnicity, age at offence and criminal history. A full list of variables including in PSM can be found in Appendix A.

The propensity score reflects the likelihood that a prisoner received the intervention, given their recorded characteristics. Prisoners in the treatment group are then matched to prisoners with similar propensity scores in the comparison group. Compared with other approaches to matching, PSM allows for the inclusion of more variables with the analysis and affords a tighter comparison of 'like-for-like' prisoners. Prisoners in the treatment group are matched to comparison prisoners whose scores are in a pre-defined close proximity[5].

---

[4] A propensity score is a value between 0 and 1 which represents the likelihood of receiving treatment. A value closer to 1 reflects increased likelihood. It is calculated using a logistic regression model with stepwise elimination.

[5] This approach is known as Radius Matching (with replacement) (Dehejia & Wahba, 2002).

3.2 Datasets and Data Linking

Prisoners were sampled from the National Offender Management Information System (NOMIS) based on the dates within which they were recorded as residing in Holme House or an alternative Category C prison. Details on prisoners' demographics, sentences and movements were extracted. The NOMIS data were then linked to Police National Computer (PNC) data to match prisoners to the correct sentence from which to measure reoffending and to extract their reoffending histories and reoffending outcomes. Prisoners were linked based on multiple unique identifiers and names.[6] The linking process reduces the sample size where the corresponding sentences of prisoners extracted from NOMIS cannot be identified on the PNC (due to aliases, alternative spellings or differing records), or where there are anomalies in the original sample (e.g. female or juvenile offenders). Duplicate offenders are removed to isolate the correct sentence to measure their reoffending. More information on the criteria with which records are matched can be found in Appendix B. Where the matching process creates 'suspect matches' (records which are likely to relate to the same sentence) these are manually screened for accuracy and accepted or rejected. Once the NOMIS data were linked to PNC data, information from the Offender Assessment System (OASys) was added which includes data relating to criminogenic needs, such as drug use, accommodation status, health status, employment history, neurodiversity and behavioural factors.[7] A full list of variables can be found in Appendix A.

3.3 Sampling

Period of Study

The study looks at the impact of the DRP on prisoners who were housed in Holme House from 2018 to 2020. The DRP funding became available in 2017, and many of the changes and investments of the DRP took around a year to implement. As a result, the first year of the pilot is excluded from the analysis[8], and prisoners were sampled from April 2018 to ensure the DRP had started to embed. The end period of the study is April 2020, as this

---

[6] The variables used for matching prisoners to a PNC record were: NOMIS number, PNC number, first name, surname and date of birth.

[7] The most complete OASys record which was closest to the sentence record was selected.

[8] This also helps exclude any residual effect from the re-roll of Holme House, changing from a Category B to Category C prison.

was the official end of the pilot period and allows us to assess the impact of the DRP prior to Covid-19 restrictions. The DRP model has been embedded into Holme House's business-as-usual regime and continues to this day.

Sample Groups

The study used two groups to measure impact:

1) **The treatment group[9]** – all prisoners who were sentenced to a custodial sentence and were housed at Holme House at any point over the DRP pilot period once the DRP had become fully embedded. This period covers April 2018 – April 2020. There was not a minimum period within which prisoners had to be at Holme House, although they could not have been released from a different prison more than six months after leaving Holme House.[10] Duplicate prisoners were permitted where they had independent custodial sentences in Holme House over the time period[11]. Sex offenders were excluded.[12]

2) **The comparison group** – all prisoners who were sentenced to a custodial sentence and were housed in alternative Category C male prisons over the period April 2018 – April 2020. Comparison prisoners were not allowed to be housed at Holme House at any point. There was not a minimum period within which prisoners had to be at a Category C prison, although they could not be released from a different prison more than six months after leaving their Category C prison.[13] Duplicate prisoners were permitted where they had independent custodial

---

[9] Treatment in this instance is used within the meaning of technical impact evaluation, referring to the group that receives the DRP intervention (i.e. treatment and control group). It should not be understood as receipt of any medical treatment for substance misuse, and this study does not measure the impact of medical treatment. Prisoners in both Holme House and the comparator prisons will have received medical treatment for substance misuse.

[10] This ensures that we are measuring the impact of the prison prisoners are in for most of their sentence and reduces contamination from other prisons they may be transferred to. No minimum period was set on the assumption that the whole prison approach would start to benefit prisoners from the initial date of exposure. It is plausible that longer exposure to the intervention may result in a greater effect, although dose-response is not within the scope of this analysis.

[11] Each 'dosage' of the treatment is therefore captured, to reduce confounding of prisoners who were in Holme House multiple times.

[12] Sexual offenders were excluded because of evidence that they have different reoffending behaviour, including notably lower reoffending rates (Falshaw *et al.*, 2003, Hanson, 2018).

[13] This matches the requirement for the treatment group.

sentences over the time period.[14] Sex offenders were excluded. A full list of Category C prisons can be found in Appendix D.

Since prisoners could appear in the datasets more than once where they had distinct custodial sentences within the selected prisons, the base unit from which reoffending is calculated is sentences rather than individuals.

Sample Sizes

Records were drawn from NOMIS before being matched to records on the PNC. Table 1 shows the sample sizes and the rate at which records were matched on the PNC. The final samples used in the analysis are highly powered given their size. Power calculations show that the samples have a power of 1.0 to detect an effect size of 0.8 (Cohen's d) with a confidence level of 99%.

Table 1: Number of distinct sentences in treatment and comparison groups

	Treatment Group (Holme House)	Comparison Group (Alternative Cat C prisons)
Number of records sampled from NOMIS	4,683	94,385
Number of records taken off the PNC[15]	2,459	36,207
Number of records included in analysis[16]	2,451	36,082
Percentage of original sample included in analysis	52%	38%

---

[14] This matches the requirement of the treatment group and is in line with prior MoJ analysis.

[15] Records were excluded because they could not be identified on the PNC, they did not have the relevant adjudication result, they did not have a record on the PNC that corresponded with their period of custody in Holme House/alternative Category C prisons, or they did not meet the inclusion criteria for analysis (e.g. no previous sexual offences, a custody sentence, aged over 18, and male).

[16] Records were excluded where errors were found in the data cleaning process e.g. an individual's first contact with the criminal justice system was below age 10 (the age of criminal responsibility) or records were for female offenders.

Sample Characteristics

Demographic and sentence characteristics of the pre-matched sample groups were analysed and are presented in Table 2 and Table 3. Broadly, the pre-matched sample groups were very similar, although the treatment group had a higher proportion of white prisoners and a higher proportion of prisoners with a drug or alcohol need. Additionally, the treatment group had a higher population of prisoners with short-term custodial sentences (0-12 months) and a higher population of prisoners in custody for theft offences. It is important to note that these differences are prior to the calculation of, and matching on, propensity scores. The matching of individual units within the sample, rather than the original sample, is intended to create the like-for-like comparison of outcomes.

Table 2: Demographic characteristics of treatment and comparison samples prior to matching

Group	Characteristic	Treatment Group (%)	Comparison Group (%)
Age	Young Adult (18 – 20)	2.6	2.1
	Adult (21 – 23)	7.1	9.3
	Adult (24+)	90.3	88.6
Ethnicity	White	96.7	75.9
	Asian	1.6	7.3
	Black	0.7	10.9
	Mixed	0.8	4.5
	Other	0.2	1.0
Drug Need[17]	Yes	57.3	45.9
	No	42.7	54.1
Alcohol Need	Yes	28.5	20.0
	No	71.5	80.0

---

[17] Appendix C contains definitions and further information on how Drug and Alcohol needs are calculated.

Table 3: Sentence characteristics for treatment and comparison samples prior to matching

Group	Characteristic	Treatment Group (%)	Comparison Group (%)
Drug offences as part of sentence	Yes	63.4	67.8
	No	36.6	32.2
Main offence group[18]	Theft offences	31.5	23.3
	Violence against the person	18.7	21.4
	Summary non-motoring	8.8	16.0
	Possession of weapons	8.5	8.4
	Miscellaneous crimes against society	7.0	8.2
Prison sentence length band	0-6 months	29.8	10.6
	7-12 months	12.9	7.3
	1-4 years	43.9	55.2
	4+ years	12.9	26.1
	Indeterminate	0.5	0.7

3.4   Outcomes

Reoffending refers to proven reoffending – offences that are recordable, committed in England or Wales, prosecuted by the police, proven through caution or court conviction

---

[18] Offence groups reflect the classification used by the Home Office crime statistics (MoJ, 2024).

and are not breach offences (MoJ, 2023). Proven reoffending is the standard measure of reoffending for the Ministry of Justice.

After data linking and cleaning, treated prisoners were matched to comparison prisoners to allow comparison of their reoffending outcomes. Data were cleaned to improve its use in the propensity score models. This included reducing categories of ethnicity, nationality and offence type to higher-level categories. For offence type, criminal justice offence codes were recategorised using Home Office offence groups.[19]

Logistic regression models were developed that predicted the likelihood of receiving a custodial sentence in Holme House. To be included in the model, variables needed to be related to the likelihood of receiving a custodial sentence in Holme House. Over 100 variables were included in the initial models, which were then simplified using Backwards Stepwise Elimination – where the least significant variable was dropped and the model run again, then this process repeated until the model stopped improving. Propensity scores were then calculated.

Each prisoner in the treatment group was matched with all prisoners in the comparison group who had a propensity score within a certain range of the treated individual's propensity score, using a one-to-many radius matching approach with replacement.[20] This meant treated prisoners were matched to similar comparison prisoners.

The models and matching were checked for quality. This included reviewing the variables included in the models and their p-values and estimated coefficients, visually checking histograms of logit propensity scores, and calculating the receiver operating characteristic statistic to check it was above 0.7. To check the quality of the control group matching, mean standard differences and matching rates were calculated. Each variable was checked to ensure the treatment and matched control group were closely matched.[21]

---

[19] Home Office offence groups reflect the classification used in crime statistics (MoJ, 2024).

[20] A one-to-many radius matching approach with replacement means that prisoners in the treatment group can be matched with multiple prisoners in the comparison group based on proximity of propensity scores and a prisoner from the comparison group can be matched to multiple prisoners from the treatment group.

[21] Standardised differences of 5% (0.05) or less indicate that the groups were closely matched for that variable. Standardised differences of between 6% and 10% (0.06-0.1) suggest the groups were reasonably matched for that variable.

Table 4 and Table 5 show demographic and sentence characteristics of the treatment and comparison groups after the matching has taken place. Mean standard differences for included variables were low and matching rates were high. This indicated well balanced, representative groups which are suitable for comparison. Appendix E shows the summary of the propensity score models and their quality, and Appendix F outlines the quality checks performed on models.

Table 4: Demographic characteristics of treatment and comparison samples after matching

Group	Characteristic	Treatment Group (%)	Comparison Group (%)
Age	Age at first contact with CJS (years)	15.2	15.2
	Age at index offence (years)	33.5	33.7
Ethnicity	White	96.6	95.8
	Asian	1.6	1.7
	Black	0.7	0.9
	Mixed	0.8	0.9
	Other	0.2	0.2
Drug Need	Yes	56.9	56.5
	No	43.1	43.5
Alcohol Need	Yes	28.6	28.9
	No	71.4	71.1

Table 5: Sentence characteristics for treatment and comparison samples after matching

Group	Characteristic	Treatment Group (%)	Comparison Group (%)
Drug offences as part of sentence	Yes	62.8	61.9
	No	37.2	38.1
Main offence group[22]	Theft offences	30.5	30.1
	Violence against the person	19.2	19.2
	Summary non-motoring	9.1	8.8
	Possession of weapons	8.1	8.5
	Miscellaneous crimes against society	7.1	7.1
Prison sentence length band	0-6 months	28.3	28.7
	7-12 months	13.0	13.2
	1-4 years	44.9	44.4
	4+ years	13.3	13.2
	Indeterminate	0.5	0.5

A statistical test (Welch's t-test)[23] was then used to identify statistically significant differences in outcomes between treatment and matched control groups. Cohen's d was also calculated to help interpret the effect size. Typically, a Cohen's d statistic of 0.2 is

---

[22] Offence groups reflect the classification used by the Home Office crime statistics (MoJ, 2024).

[23] Welch's t-test was used to test for significance at the 0.05 level.

considered a small effect size, 0.5 is considered a medium effect size and 0.8 is considered a large effect size (Cohen, 1988).

The outcomes of interest were:

1) **Reoffending** – whether prisoners were proven to have reoffended within one year of release date from prison.

2) **Frequency of reoffending** – the number of incidences of proven reoffending in one year following release from prison.

3) **Days to first reoffence** - the number of days between release from custody and the first proven reoffence, for those who had reoffended within one year of release.

4) **Frequency of reoffending resulting in custody** – the number of incidences of proven reoffending in one year following release from prison that result in a custody sentence.

The methodological approach, data linking, analysis, and results were quality assured by other MoJ analysts, who had not worked on the project previously.

3.5   Subgroup Analysis

Drug Cohort

Additional analysis was conducted to investigate whether the impact of the DRP differed for different groups of prisoners, such as those with a drug need. Questions from OASys assessments were used to categorise prisoners into those with an identified drug need and those without (see Appendix C for further details on this process). For brevity these cohorts are henceforth referred to as the 'drug cohort' and the 'non-drug cohort'. It is important to note that this data only shows a criminogenic need relating to drugs and is not a formal assessment on whether an individual requires structured drug treatment. The reoffending outcomes of each of these groups were then analysed. This means that only the drug cohort in the treatment and comparison groups are compared against each other, and then similarly for those in the non-drug cohort. Since these groups had their propensity scores and matching conducted independently of each other, we cannot directly compare their reoffending outcomes. We can, however, comment on whether the DRP had a positive or negative impact on either of these groups.

Prisons in Areas of High Deprivation

Additional analysis was undertaken to attempt to account for regional differences in deprivation between the area surrounding Holme House and the areas in which comparison prisons are in. Indices of multiple deprivation were linked to each record based on the prison that prisoner was in. Whilst this does not account for individual-level inequality, it was used to verify whether headline results were affected by regional inequality. Reoffending outcomes were then calculated to compare Holme House prisoners against those in a prison with a similar level of regional deprivation to Holme House (see Appendix G).

3.6   Limitations

This study adopts a high-quality methodological approach to impact analysis that has been successfully implemented here and in many previous MoJ evaluations. It uses a large, well-matched sample and it is likely to produce a reliable and valid measure of the impact of the intervention. However, as with all research, there are important limitations that should be considered when interpreting the findings.

- Over a hundred variables were used in this analysis (see Appendix A). Whilst this considerable number allows us to robustly match treated and untreated prisoners together to allow tight comparison, there may be many variables which significantly impact reoffending outcomes which are not present in the data or cannot be measured (e.g. long-term recovery). This is a consideration for any PSM analysis and can only be overcome through random allocation (i.e. a randomised control trial).

- The research team intended to include a variable on specific drug needs (e.g. opiates, alcohol or other drugs) and investigate a subgroup analysis of how the DRP may differently impact the reoffending behaviour of each of these groups. However, the recording rate for this question in OASys assessments was too low to include in analysis, which precludes investigating whether certain cohorts of prisoners with a drug need are differently impacted by the DRP.

- The evaluation was unable to control for the effects of engagement in substance misuse treatment on reoffending behaviours as this data is not held by the

department. As a result, it may be the case that the DRP is more effective for those that engage in substance misuse treatment but we are unable to comment on this.

- There is a risk of contamination within this evaluation where the DRP shared good practice or ideas with other prisons used within the samples. This may impact results in an uncertain direction and the impact of contamination cannot be assessed.

- During 2017, Holme House also went through a re-roll and changed from a Category B to a Category C. Whilst aspects of this re-roll would happen immediately, there may be some aspects that take longer to change. If aspects of the Category B prison lingered despite the re-roll, this could potentially bias the result.

- Variables that relate to the chance of a drug need were crucial for the analysis so only those with OASys records where the relevant questions were completed were included. Those who are higher risk or have offended more frequently may be more likely to receive a more complete L3 OASys assessment meaning that there is a risk of bias to the sample.

- The evaluation assessed any prisoner who resided in the DRP over the pilot period against similar prisoners who did not. In statistical terms, the analysis is based on the Intention-to-treat (ITT) rather than per-protocol (PP). It does not take into account differing levels of engagement with the DRP ethos and community, engagement with purposeful activities or engagement with treatment that likely to impact individual reoffending.

- The environment, condition and regime of prisons within the comparison group is likely to vary significantly, particularly due to geography, prison population and condition of the prison infrastructure. These factors will likely impact reoffending outcomes. The pooling of these prisons within the comparison group is likely to reduce the risk that this biases results.

Ultimately, this evaluation is a focused assessment of the impact of the DRP on reoffending outcomes. There are other indicators of the DRP's success which are not captured here due to the exclusive focus on reoffending outcomes.

4. Findings

4.1 Headline Reoffending Outcomes

This section covers the reoffending outcomes of the entire treatment and comparison groups. Mean standard differences for these groups were low and the matching rate was high, which indicated well balanced, representative groups. This implies that the prisoners being compared were broadly similar on a range of characteristics.

After conducting propensity score matching, the analysis found that prisoners who were housed in Holme House prison were more likely to have a proven reoffence following release than prisoners housed in alternative Category C prisons.

Three main indicators of reoffending outcomes were used: one year proven reoffending rate, the number of proven reoffences within one year following release, and the number of days from release to first proven reoffence for those that did reoffend within one year.

- For treated prisoners, 47.8% went on to commit a proven reoffence within one year, compared to 44.6% of comparison prisoners, a difference of 3.2 percentage points.

- Treated prisoners committed 2.4 proven reoffences on average, compared to 2.1 proven reoffences on average for comparison prisoners, a difference of 0.3 proven reoffences within one year following release.

- Treated prisoners took 111 days to commit their first proven reoffence on average, compared to 118 days for comparison prisoners, a difference of seven days.

- All results were statistically significant at the 95% confidence level. Indicators one and two were statistically significant at the 99% confidence level. There was no statistically significant difference in the number of proven reoffences that resulted in a custody sentence between treatment and comparison prisoners.

- Cohen's d statistics were calculated for all reoffending outcomes to provide an understanding of standardised effect size. All effect sizes for reoffending outcomes that were statistically significant were small.

Table 6: Headline reoffending outcomes for treatment and comparison groups

Reoffending indicator	Treatment mean	Comparison mean	Difference in means	p-value	Standardised effect size (Cohen's d)
1. One year proven reoffending rate (%)	47.76	44.60	3.16	0.003***	0.255
2. Number of proven reoffences	2.40	2.08	0.32	< 0.001***	0.298
3. Days to first proven reoffence	111.27	118.26	-6.99	0.025**	-0.287
4. Number of proven reoffences resulting in custody sentence	2.73	2.73	0.0	0.999	0.166

** Indicates statistical significance at the 95% confidence level.
*** Indicates statistical significance at the 99% confidence level.
Treatment group n = 2,364; comparison group n = 28,019 for indicators 1 & 2. Treatment group n = 1,129; comparison group n = 9,955 for indicators 3 & 4.

4.2   Drug Cohort Reoffending Outcomes

Reoffending outcomes for the drug cohort (see Appendix C) in the treatment group were compared to the drug cohort in the comparison group. Mean standard differences for these groups were low and the matching rate was high, which indicated well balanced, representative groups. This implies that the prisoners being compared were broadly similar on a range of characteristics.

The analysis found that the drug cohort in Holme House were more likely to have a proven reoffence following release than the drug cohort in alternative Category C prisons.

- For the drug cohort in the treatment group, 57.4% went on to commit a proven reoffence within one year, compared to 52.7% of the drug cohort in the comparison group, a difference of 4.7 percentage points.

- Treated drug cohort prisoners committed 3.1 proven reoffence on average, compared to 2.7 proven reoffences on average for comparison drug cohort prisoners, a difference of 0.4 proven reoffences within one year following release.

- These results were statistically significant at the 99% confidence level. All other results were not statistically significant. These results were considered to be a small effect size.

Table 7: Drug cohort reoffending outcomes

Reoffending indicator	Treatment mean	Comparison mean	Difference in means	p-value	Standardised effect size (Cohen's d)
1. One year proven reoffending rate (%)	57.37	52.65	4.72	< 0.001***	0.262
2. Number of proven reoffences	3.06	2.70	0.35	0.007***	0.278
3. Days to first proven reoffence	101.80	108.20	-6.41	0.082	-0.304
4. Number of proven reoffences resulting in custody sentence	2.87	3.06	-0.18	0.211	0.126

** Indicates statistical significance at the 95% confidence level.

*** Indicates statistical significance at the 99% confidence level

Treatment group n = 1,363; comparison group n = 14,044 for indicators 1 & 2. Treatment group n = 782; comparison group n = 6,227 for indicators 3 & 4.

4.3 Non-Drug Cohort Reoffending Outcomes

Reoffending outcomes for the non-drug cohort (see Appendix C) in the treatment group were compared to the non-drug cohort in the comparison group. Mean standard differences for these groups were low and the matching rate was high, which indicated well balanced, representative groups. This implies that the prisoners being compared were broadly similar on a range of characteristics.

The analysis found that the there was no statistically significant difference across three indicators between the non-drug cohort in Holme House and the non-drug cohort in

alternative Category C prisons. It found that the non-drug cohort in Holme House committed more proven reoffences following release than the non-drug cohort in alternative C prisons, but there was no statistically significant difference in proven reoffending rate.

- Treated non-drug cohort prisoners committed 1.6 proven reoffences on average, compared to 1.3 proven reoffences on average for comparison non-drug cohort prisoners, a difference of 0.4 proven reoffences within one year following release.

- This result was the only result statistically significant result. It was significant at the 99% confidence level. The effect size is considered small.

Table 8: Non-drug cohort reoffending outcomes

Reoffending indicator	Treatment mean	Comparison mean	Difference in means	p-value	Standardised effect size (Cohen's d)
1. One year proven reoffending rate (%)	35.80	33.23	2.58	0.092	0.239
2. Number of proven reoffences	1.62	1.25	0.37	< 0.001***	0.375
3. Days to first proven reoffence	127.82	135.30	-7.48	0.168	-0.268
4. Number of proven reoffences resulting in custody sentence	2.41	2.08	0.33	0.060	0.255

** Indicates statistical significance at the 95% confidence level.
*** Indicates statistical significance at the 99% confidence level.
Treatment group n = 1,039; comparison group n = 18,451 for indicators 1 & 2. Treatment group n = 372; comparison group n = 4,679 for indicators 3 & 4.

5. Discussion

The findings of this impact evaluation indicate that DRP was ineffective at reducing reoffending, compared to other similar prisons. The results show statistically significant higher reoffending outcomes for prisoners in Holme House compared to comparison prisoners. This result is the same for the drug cohort. The analysis shows no statistically significant difference in the reoffending rate for the non-drug cohort. All standardised effect sizes were small, implying that the DRP's impact on the reoffending outcomes for prisoners was limited.

Whilst this impact evaluation has used a rigorous study design, with an established PSM approach that has been used repeatedly by the MoJ and is considered a level 4 on the Maryland Scientific Methods Scale (Farrington *et al*., 2002)[24], there are factors relating to the methodology and the intervention which are worth considering and may partially explain the results.

5.1 Methodological considerations

Unobserved Variables

Propensity score matching works by reducing confounders between a treated group and a comparison group to allow as close as possible 'like-for-like' comparison between those groups. The more individual-level factors that are included, the more robust it is. This evaluation used over 100 variables in the propensity score models which accounts for a large number of individual-level factors that may influence reoffending. However, it cannot account for unobserved variables which may influence reoffending, such as recovery from substance misuse.

Prison-level Factors

Since this evaluation compares one specific prison to essentially a 'national average' of all other prisons of the same category, it cannot adequately account for prison-level factors. If Holme House had unique factors (separate from the DRP) that were not present in any

---

[24] This is a five-point scale of methodological rigour, ranging from 1 for correlation analysis, to 5 for randomised controlled trials.

other Category C prisons, then the evaluation would be unable to separate the impact of these factors.

There is limited evidence to support this theory. The inspection prior to the introduction of the DRP (HMIP, 2017) found that levels of violence and use of force were comparable to other prisons. However, there was a high prevalence of new psychoactive substances in Holme House, which was leading to frequent medical emergencies and lockdowns. Performance issues were not uncommon for comparator prisons, with 38% of Category C prisons being scored 2 (out of 4[25]) for overall prison performance in 2017-18 (including Holme House) (MoJ, 2018).

Similar themes were also identified during contemporaneous research at Holme House in 2018 (MoJ, 2020b). This highlighted the interaction between high-levels of 'spice' use alongside issues such as staff-shortages and lack of funding, which destabilised the prison and created issues running the daily prison regime.

The inspection at the end of the pilot period (HMIP, 2020) found that conditions in the prison had not substantially improved. In 2020, the use of force was deemed to be at times disproportionate, and self-harm had increased significantly. The rate of positive drug tests had fallen substantially over this period, likely reflecting the introduction of the DRP.

In 2019-2020, Holme House was rated 3 for overall prison performance, along with 61% of Category C prisons (MoJ, 2020a). Across both periods, the Measuring Quality of Prison Life (MQPL) survey found an overall rating of 3 (out of 4) for safety and decency, which largely reflects similar prisons (MoJ, 2018; MoJ, 2020a).

Whilst Holme House had some issues with drugs, purposeful activity and release planning over the study period, it does not appear that these difficulties were unique to the prison. Additionally, as this evaluation measures the whole prison as a single intervention (reflecting the whole prison approach), any structural differences between Holme House

---

[25] Overall performance for each prison is graded into one of four bands. These bands are: 4: Exceptional performance; 3: Meeting majority of targets; 2: Overall performance is of concern; and 1: Overall performance is of serious concern. Seventeen performance measures are used to create the combined score, including data on mandatory drug tests, assaults, self-harm.

and other Category C prisons as a result of the DRP implementation will be baked into the intervention and will therefore be detected in the outcomes.

Regional Inequality

There are many factors which may increase the risk of reoffending beyond those within the control of the DRP. Such factors related to reoffending include pro-criminal attitudes, adverse childhood experiences, personal and family relationships, unemployment, financial problems, homelessness and mental health problems (MoJ, 2013). There is also a strong link between deprivation and poor health and/or substance misuse, which are in turn associated with higher reoffending (Public Health England, 2017; MoJ and PHE, 2017). Whilst the inclusion of OASys variables accounts for many of these risk factors when prisoners enter prison, the evaluation is blind to prisoners' experiences post-release. In particular, the area in which prisoners are released may heavily influence their risk of reoffending, based on the life opportunities that the area affords them. Where prisoners are released into more deprived areas, where job opportunities are lower and risk of homelessness higher, they may be more likely to reoffend. Linking the indices of multiple deprivation (Ministry of Housing, Community, and Local Government, 2019; StatsWales, 2019) onto the area in which Holme House is in showed that it is in the top decile for multiple deprivation. Furthermore, Middlesborough has one of the highest recorded crime rates in England and Wales (ONS, 2024). As a result, it may be that prisoners' experiences upon release from Holme House contribute to their risk of reoffending.

We attempted to account for regional inequality by linking indices of multiple deprivation to all prisons in the sample based on their area and filtering the comparison group to only those prisons in areas with high deciles of deprivation. By comparing treatment prisoners with comparison prisoners in prisons in areas of high deprivation, this acted as a check of whether regional inequality was driving the results. Appendix G shows that reoffending outcomes were statistically significantly higher for treatment prisoners compared to comparison prisoners in prisons within areas of high deprivation. This suggests that the regional deprivation of Holme House is not driving reoffending, although this does not account for individual-level deprivation and post-release experiences which may still drive reoffending.

5.2   Intervention considerations

Implementation

Implementation fidelity is crucial to the effectiveness of a policy. Any barriers to implementation may have hindered the ability for the DRP to positively impact prisoners. The higher level of reoffending among the drug cohort (with no effect on the non-drug cohort) suggest that drug-specific implementation issues may be the driving the observed results.

Findings from the process evaluation show that, broadly, the security element of the DRP was successfully implemented, although, over time, some prisoners began to question the effectiveness of some equipment to detect drugs and new supply routes may have opened (Ayres *et al*., 2023). Implementation of cultural change was more complex and difficult, with a lack of awareness of the goal of the DRP amongst prisoners and staff, less support for the demand reduction element of the project, and some stigma towards prisoners with substance misuse problems persisted. The care and wellbeing goal of the DRP had similar mixed levels of implementation: despite the move towards a house block-based healthcare model, few prisoners noticed an improvement in accessing healthcare and waiting lists were still long, whilst the access to and visibility of the Drugs and Alcohol Recovery Team had drastically improved. The Connecting Communities team was seen as a large improvement in continuity of care for prisoners, although concerns persisted about the level of post-release support available to prisoners given the variability of provision by CRCs. Generally, the DRP model was successfully implemented as intended, although there were differences in implementation fidelity across the DRP's aims.

Post-release Drivers of Reoffending

The section before pointed to multiple risk factors associated with increased reoffending, such as accommodation, employment and personal relationships. The DRP could only influence offenders' reoffending behaviour up until the point of release, after which these factors may be more critical to their risk of reoffending than any positive work done whilst within the prison. The process evaluation of the DRP highlighted a 'cliff edge' of post-release support which was felt to risk prisoner's continued recovery. Whilst staff within the prison can work on post-release planning with offenders and continuity of care for those with a substance misuse problems, ultimately the experiences of offenders upon release

are beyond control of the intervention being measured and yet are likely to be significant on their reoffending behaviour. As a result, both the intervention and the evaluation cannot account for post-release drivers of reoffending.

DRP Culture Change

A key consideration is how long large-scale reforms to a prison take to embed within the prison culture. The process evaluation highlighted that the first year of the pilot was at times turbulent due to the prison's re-roll and delays to funding and equipment (Ayres *et al*., 2023). Moreover, the DRP encompassed not just reforms to the infrastructure and regime of the prison, but also its culture too. Prison staff acknowledged that significant cultural change could take many years to develop and relies upon the nurturing of positive relationships between staff and prisoners, the development of social capital across the prison, and enthusiasm for the model across staff of all grades (Ayres *et al*., 2023). It is likely, therefore, that by the end of the pilot, the culture of the DRP was still developing and had not fully embedded itself and self-perpetuate. This is essentially Ayres *et al*.'s (2023) conclusion.

It is supported by the HMIP (2017, 2020) reports which highlight the difficulties the prison still faced at the end of the pilot. In the latest HMIP (2023) inspection, the prison had shown multiple notable improvements, although purposeful activity remained at "not sufficiently good". Notable improvements included stronger partnership working with the local council, healthcare services and local organisations, and an improved prison community, encompassed by park runs, excellent family days and a new democratic Prison Council. Many innovative advancements to the prison had been implemented, such as brightly coloured private rooms for family video calls, improvements to the infrastructure to bring more nature into the prison and a menagerie of animals for prisoners to look after.

The time taken to implement the DRP and embed cultural change means that prisoners in at Holme House during the earlier stages of the pilot may not have fully benefited from the DRP improvements.

Repeating this analysis in the future, once eighteen months (the necessary time-lag to detect reoffending) has elapsed, would help to provide further evidence on whether reoffending outcomes have changed following further implementation of the DRP concept.

6. Conclusion

By comparing the reoffending outcomes of prisoners who resided in the DRP over the latter two years of its pilot against prisoners in alternative Category C prisons, this impact evaluation has found that the DRP was ineffective at reducing reoffending. The results show that prisoners in Holme House were more likely to reoffend than prisoners in other prisons. Prisoners in Holme House committed more proven reoffences than comparison prisoners and committed their first proven reoffence in fewer days. Similar results were found in the sub-group of prisoners with a drug need, although there was no evidence of a DRP effect on the non-drug cohort. The standardised effect sizes show that these impacts were small.

This impact evaluation has used a robust propensity score matching methodology, with well-matched treatment and comparison groups, which afforded a strong "like-for-like" comparison. However, there are multiple considerations which should be acknowledged when interpreting these results. Firstly, the evaluation lacked data on specific structural differences that Holme House may possess compared to all other Category C prisons, although it is unclear whether these exist. Secondly, there are many factors which drive reoffending such as employment and accommodation which are beyond the control of the DRP, and which cannot be accounted for in the evaluation. Finally, the analysis was conducted on the latter two years of the pilot period, before the DRP became embedded in the prisons' business-as-usual regime. It is likely that significant reform and cultural change to a prison takes more than a couple of years to embed and perpetuate.

These considerations prevent a definitive conclusion on the effectiveness of the DRP model. Further research is required, and it would be beneficial to repeat this analysis once the DRP has been given more time to become embedded, and with the inclusion of wider data sources covering structured drug treatment.

References

Baltieri, D. A. (2014). Order of onset of drug use and criminal activities in a sample of drug-abusing women convicted of violent crimes. *Drug and Alcohol Review, 33*, 202-210. DOI: 10.1111/dar.12107

Black, C. (2020). *Review of drugs: evidence pack.* Home Office. [Review of drugs: phase one report - GOV.UK (www.gov.uk)](www.gov.uk)

Brinn, A., Preston, J., Costello, R., Opoku, T., Sampson, E., Elliott, I. & Sorbie, A. (2023). *An impact evaluation of the prison-based Thinking Skills Programme (TSP) on reoffending*. Ministry of Justice. [TSP_Reoffending_Report.pdf (publishing.service.gov.uk)](publishing.service.gov.uk)

Butzin, C. A., Martin, S. S. & Inciardi, J. A. (2005). Treatment during transition from prison to community and subsequent illicit drug use. *Journal of Substance Abuse Treatment, 28,* 351-358. DOI: 10.1016/j.jsat.2005.02.009

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

De Andrade, D. (2018). The "Drugs-Crime Nexus". In Kinner, S. A. & Rich, J. D. (Eds.), *Drug Use in Prisoners: Epidemiology, Implications, and Policy Responses* (pp. 1-16). Oxford University Press.

De Andrade, D., Ritchie, J., Rowlands, M., Mann, E. & Hides, L. (2018). Substance Use and Recidivism Outcomes for Prison-Based Drug and Alcohol Interventions. *Epidemiological Reviews*, *40,* 121-133. DOI: 10.1093/epirev/mxy004

Eaton, G. & Mews, A. (2019). *The impact of short custodial sentences, community orders and suspended sentence orders on reoffending.* Ministry of Justice. [The impact of short custodial sentences, community orders and suspended sentence orders on reoffending (publishing.service.gov.uk)](publishing.service.gov.uk)

European Monitoring Centre for Drugs and Drug Addiction [EMCDDA]. (2007). *Drugs and crime — a complex relationship.* European Monitoring Centre for Drugs and Drug Addiction. https://www.euda.europa.eu/html.cfm/index36331EN.html_en

Falshaw, L., Bates, A., Patel, V., Corbett, C., & Friendship, C. (2003). Assessing reconviction, reoffending and recidivism in a sample of UK sexual offenders. *Legal and Criminological Psychology, 8*(2), 207-215. https://doi.org/10.1348/135532503322362979

Farrington, D. P., Gottfredson, D. C., Sherman, L. W. & Welsh, B. C. (2002). The Maryland Scientific Methods Scale. In Sherman, L. W., Farrington, D. P., Welsh, B. C. & MacKenzie, D. L. (Eds.), *Evidence-Based Crime Prevention* (pp. 13-21). Routledge.

Hanson, R. K. (2018). Long-term recidivism studies show that desistance is the norm. *Criminal Justice and Behavior, 45*(9), 1340-1346. https://doi.org/10.1177/0093854818793382

Hedrich D, Alves P, Farrel M, Stöver H, Møller L, Mayet S. (2012). The effectiveness of opioid maintenance treatment in prison settings: a systematic review. *Addiction, 107*(3), 501-517. DOI: https://doi.org/10.1111/j.1360-0443.2011.03676.x

Her Majesty's Inspectorate of Prisons [HMIP]. (2017). *Report on an unannounced inspection of HMP Holme House.* Her Majesty's Inspectorate of Prisons. https://www.justiceinspectorates.gov.uk/hmiprisons/wp-content/uploads/sites/4/2017/12/Holme-House-Web-2017.pdf

Her Majesty's Inspectorate of Prisons [HMIP]. (2018). HM Chief Inspector of Prisons annual report: 2017 to 2018. Her Majesty's Inspectorate of Prisons. HM Chief Inspector of Prisons annual report: 2017 to 2018 - GOV.UK

Her Majesty's Inspectorate of Prisons [HMIP]. (2020). *Report on an unannounced inspection of HMP Holme House.* Her Majesty's Inspectorate of Prisons. HMP Holme House – HM Inspectorate of Prisons

Her Majesty's Prison and Probation Service [HMPPS]. (2019). *Prisons Drug Strategy.* Her Majesty's Prison and Probation Service. Prison Drugs Strategy (publishing.service.gov.uk)

Home Office (2014). Understanding organised crime: estimating the scale and the social and economic costs Research Report 73. Home Office. Understanding organised crime: estimating the scale and the social and economic costs

Home Office. (2017). *2017 Drug Strategy.* Home Office. Drug strategy 2017 - GOV.UK (www.gov.uk)

Justice Data Lab [JDL]. (2013). *Justice Data Lab: Methodology Paper.* Ministry of Justice. Accessing the Justice Data Lab service - GOV.UK (www.gov.uk)

Malta, M., Varatharajan, T., Russell, C., Pang, M., Bonato, S., & Fischer, B. (2019). Opioid-related treatment, interventions, and outcomes among incarcerated persons: a systematic review. *PLoS medicine*, *16*(12), e1003002. https://doi.org/10.1371/journal.pmed.1003002

Ministry of Housing, Communities & Local Government [MHCLG]. (2019). *English indices of deprivation 2019*. English indices of deprivation 2019 - GOV.UK (www.gov.uk)

Ministry of Justice (2020b) *Exploring substance use in prisons: a case study approach in 5 closed male English prisons*. Exploring substance use in prisons: a case study approach in 5 closed male English prisons - GOV.UK

Ministry of Justice [MoJ] & Public Health England [PHE]. (2017). *The impact of community-based drug and alcohol treatment on re-offending*. Ministry of Justice & Public Health England. The effect of drug and alcohol treatment on re-offending - GOV.UK (www.gov.uk)

Ministry of Justice [MoJ]. (2013). *Transforming Rehabilitation: a summary of evidence on reducing reoffending*. Ministry of Justice. Transforming Rehabilitation: a summary of evidence on reducing reoffending (publishing.service.gov.uk)

Ministry of Justice [MoJ]. (2016). *Biggest shake-up of prison system announced as part of Queen's Speech*. Biggest shake-up of prison system announced as part of Queen's Speech - GOV.UK (www.gov.uk)

Ministry of Justice [MoJ]. (2018). *Prison performance ratings 2017 to 2018*. Prison performance ratings 2017 to 2018 - GOV.UK (www.gov.uk)

Ministry of Justice [MoJ]. (2020a). *Prison performance ratings 2019 to 2020.* Prison Performance Ratings: 2019 to 2020 - GOV.UK (www.gov.uk)

Ministry of Justice [MoJ]. (2022). *Identified needs of offenders in custody and the community from the Offender Assessment System, 30 June 2021.* Identified needs of offenders in custody and the community from the Offender Assessment System, 30 June 2021 - GOV.UK (www.gov.uk)

Ministry of Justice [MoJ]. (2023) Proven reoffending statistics: July to September 2021. https://www.gov.uk/government/statistics/proven-reoffending-statistics-july-to-september-2021

Ministry of Justice [MoJ]. (2024). *A Technical Guide to Criminal Justice Statistics Quarterly (CJSQ).* Criminal Justice System statistics quarterly: December 2023 - GOV.UK (www.gov.uk)

Ministry of Justice [MoJ]. (2025) *Identified offender needs, custody and community, 31 October 2024.* Identified offender needs, custody and community, 31 October 2024 - GOV.UK

Mitchell, O., Wilson, D. B. & MacKenzie, D. L. (2012). The Effectiveness of Incarceration-Based Drug Treatment on Criminal Behavior: A Systematic Review. *Campbell Systematic Reviews*, *18,* 3-75. DOI: 10.4073/csr.2012.18

Moore, K. E., Hacker, R. L., Oberleitner, L., & McKee, S. A. (2020). Re-entry interventions that address substance use: A systematic review. *Psychological Services*, *17*(1), 93-101. https://doi.org/10.1037/ser0000293

Moran, D. (2019). How the prison environment can support recovery. *Prison Service Journal, 242,* 44-49.

National Institute for Health and Care Excellence [NICE]. (2007). *Methadone and buprenorphine for the management of opioid dependence: Technology appraisal guidance.* National Institute for Health and Care Excellence. Overview | Methadone and buprenorphine for the management of opioid dependence | Guidance | NICE

Office for Health Improvement & Disparities [OHID]. (2023). *Substance misuse treatment for adults: statistics 2022 to 2023.* [Substance misuse treatment for adults: statistics 2022 to 2023 - GOV.UK (www.gov.uk)](#)

Office for Health Improvement & Disparities [OHID]. (2024). *Substance misuse treatment in secure settings: 2022 to 2023.* [Substance misuse treatment in secure settings: 2022 to 2023 - GOV.UK (www.gov.uk)](#)

Office for National Statistics [ONS]. (2012). *2011 Census geographies*. [2011 Census geographies - Office for National Statistics (ons.gov.uk)](#)

Office for National Statistics [ONS]. (2024). *Crime in England and Wales: year ending March 2024*. [Crime in England and Wales - Office for National Statistics](#)

Pierce, M., Hayhurst, K., Bird, S. M., Hickman, M., Seddon, T., Dunn, G. & Millar, T. (2017). Insights into the link between drug use and criminality: Lifetime offending of criminally-active opiate users. *Drug Alcohol Dependence, 179,* 309-316. https://doi.org/10.1016%2Fj.drugalcdep.2017.07.024

Public Health England [PHE]. (2017). *An evidence review of the outcomes that can be expected of drug misuse treatment in England*. Public Health England. [Drug misuse treatment in England: evidence review of outcomes - GOV.UK (www.gov.uk)](#)

StatsWales (2019). *WIMD 2019*. [WIMD 2019 (gov.wales)](#)

Wheatley, M. (2019). Promoting Recovery in Prison: The Holme House Approach. *Prison Service Journal, 242,* 11-17.

Woodall, J. & Freeman, C. (2020). Promoting health and well-being in prisons: an analysis of one year's prison inspection reports. *Critical Public Health, 30,* 555-556. https://doi.org/10.1080/09581596.2019.1612516

World Health Organisation [WHO]. (2014). *Prisons and Health.* World Health Organisation. https://www.who.int/europe/publications/i/item/9789289050593

Glossary

Comparison group/prisoners – this refers to all prisoners who resided in a Category C prison, excluding Holme House, over the study period (April 2018 – April 2020).

DRP – The Drug Recovery Prison, piloted at Holme House between April 2017 – April 2020.

No significant difference – This means that, based on this analysis, it is not possible to say for sure whether the DRP had any effect (either positive or negative) on the outcome. There is a greater than 5% possibility that any differences between the groups were due to chance. A non-significant result might happen because there are not many people in the treatment group, the impact on the outcome is small and difficult to detect, or there is no impact on the outcome.

NPS – New Psychoactive Substances, newly available drugs that are designed to replicate the effects of established illicit drugs while evading legal restrictions. These substances are now controlled by the Psychoactive Substance Act 2016.

OST – Opiate Substitution Therapy, prescribed substitutes (such as methadone and buprenorphine) to reduce illicit opiate use amongst opiate users.

PSM – Propensity Score Matching, a statistical technique to match similar prisoners together based on a range of individual-level characteristics to allow comparison of outcomes.

Reoffending – proven reoffending one year after release from prison. A proven reoffence is any further offence committed within one year of release, that is detected and convicted.

Statistically Significant difference – This means that there is a less than 5% possibility that the impact of the DRP on the outcome is due to chance, and we can therefore be confident that the DRP had an impact (positive or negative) on that outcome.

Standardised effect size – This measures the strength of the relationship between two variables. This study calculated effect sizes using Cohen's d. Typically, a Cohen's d

statistic of 0.2 is considered a small effect size, 0.5 is considered a medium effect size and 0.8 is considered a large effect size (Cohen, 1988).

Treatment group/prisoners – this refers to all prisoners who resided in Holme House over the study period (April 2018 – April 2020). Treatment in this instance is used within the meaning of technical impact evaluation (i.e. treatment and control group). It should not be understood as receipt of any medical treatment for substance misuse, and this study does not measure the impact of medical treatment. Prisoners in both Holme House and the comparator prisons will have received medical treatment for substance misuse.

Appendix A
Variables used in propensity score matching

Over 100 variables were used in the propensity score models. For ease, the below list reflects groups of variables, rather than individual variables.

Demographics

- Gender
- Nationality
- Ethnicity
- Age at first contact with criminal justice system
- Age group during custodial sentence

Sentencing and criminal history

- Offence type based on Home Office offence groups, including whether offences were drug offences or sexual offences.
- Sentence type
- Sentence length
- Number of sentences
- Prison location
- Number of custodial sentences in that prison
- Severity of index offence (ranked 1 to 3 with 3 being the most severe)
- Offending history
- Number of previous offences, in total and broken down by severity and offence group, including sentence breaches, and their squared terms.
- Copas rate – the rate at which an offender has built up convictions.
- Number of previous custodial sentences, court orders, court convictions and cautions, and their squared terms.

OASys assessment

- Accommodation status and suitability of accommodation
- Unemployment status and employment history
- School attendance

- Literacy and numeracy difficulties

- Financial status, including incomes from illegal sources.

- Current relationship with family members

- Current relationship with partner

- Experience of childhood

- Regular activities which encourage offending

- Easily influenced by criminal associates

- Health status

- Drug use history, including level of drug use and if drugs have ever been injected[26]

- Motivation to tackle drug misuse

- Drug use as a major activity/occupation

- Alcohol use history, including regular binge drinking and past misuse

- Violent behaviour related to alcohol use

- Motivation to tackle alcohol misuse.

- Difficulties coping

- Psychological problems and/or depression

- Suicidal thoughts or self-harm

- Psychiatric problems

- Behavioural characteristics, including impulsivity, aggression and temper control

- Awareness of the consequences of one's actions

- Problem solving skills

- Ability to understand other people's views

- Concrete thinking – focus on immediate and tangible details, rather than abstract thinking

- Awareness of their own reoffending behaviour

---

[26] Excludes information on specific drug use due to poor completion rate.

Appendix B
PNC Linking Criteria

The below table shows the criteria which records from NOMIS were matched to the PNC and the inclusion and exclusion criteria for analysis.

Table 9: Criteria for linking records on the PNC.

Criteria	Logic
Index disposal code must be prison.	Ensures that sentences included in analysis were custodial.
Age must be more than or equal to 18.	Juvenile offenders were excluded due to differing reoffending behaviour.
No previous sexual offences and the main offence must not be a sexual offence.	Sexual offenders were excluded due to differing reoffending behaviour.
Conviction date must be before or equal to index date.	A sense check to ensure that release from prison is after date of conviction for that crime.
Person IDs must match.	To ensure the correct records were being matched.
Entry to prison date must be after or equal to conviction date or no more than 30 days before and six months later.	A sense check to ensure that entry to prison is after date of conviction for that crime, whilst accounting for differences in recording time.
Entry to prison date must be before release date.	A sense check to ensure that records show a prisoner is in prison before release.
Difference between exit date from prison they serve their sentence in and release date must be less than or equal to six months	To ensure that we are measuring the impact of the prison prisoners are in for most of their sentence, and reduces contamination from other prisons they may reside in.
Duplicate offenders could not have the same person ID and release date.	To ensure duplicate offenders reflected independent custodial sentences.
Offenders in the comparison group could never have resided in Holme House	To reduce risk of contamination of the two groups.

Appendix C
Drug need score calculations

OASys assessments are conducted on individuals under HMPPS supervision to better understand the needs of and risks posed by them. OASys assessments include eight criminogenic needs measuring factors linked to offending behaviour: accommodation, employability, relationships, lifestyle and associates, drug misuse, alcohol misuse, thinking & behaviour, and attitudes.

Drugs needs were calculated using questions from OASys. Each question is scored on a scale from 0 to 2 (some being score 0 or 2, and others 0, 1 or 2) with 0 denoting 'no need', 1 'some need' and 2 'severe need'. Scores from each question are summed to give a total score for that section. Each set of needs questions has a maximum score, and a need is identified when the score equals or exceeds a cut-off. This study used 2+ as the score boundary to classify a prisoner as having a drug need. This is the boundary used in published statistics (MoJ, 2025).

Table 10: OASys variables used to calculate drug and alcohol needs.

	Questions	Score Boundary
Drug Need	Frequency of use Ever injected drugs Motivation to tackle drug misuse Main activities involve drugs	2+ = need 8+ = severe need

Appendix D
List of Category C prisons

Prisoners were eligible to be drawn into the comparison sample when located within other Category C prisons.

Table 11: List of Category C prisons in England and Wales

Cat C List	
Ashfield	Northumberland
Berwyn	Oakwood
Brixton	Onley
Buckley Hall	Parc
Bure	Portland
Channings Wood	Ranby
Coldingley	Risley
Dartmoor	Rochester
Erlestoke	Stafford
Featherstone	Stocken
Guys Marsh	Stoke Heath
Highpoint	Swinfen Hall
Hindley	Usk
Holme House	Verne
Humber	Warren Hill
Huntercombe	Wayland
Isis	Wealstun
Lancaster Farms	Whatton
Lindholme	Wymott
Littlehey	
Maidstone	
Moorland	
Mount	

Appendix E
Model statistics summary

Table 12 provides an overview of the model quality of the propensity score models used for each subgroup.

Table 12: Overview of the propensity score models for each subgroup.

Subset	No. of variable levels	Match rate	Mean standardised differences			
			Min.	Max.	No. of variables at -5% to -10%	No. of variables at 5% to 10%
Headline reoffending	244	78.84%[27]	-0.098	0.042	1	0
Drug cohort	241	85.69%	-0.081	0.060	2	2
Non-drug cohort	235	94.82%	-0.121	0.050	1	0
High IMD prisons	243	93.81%	-0.068	0.041	1	0

---

[27] The overall matching rate includes both treatment and comparison groups. The matching ratio for the treatment group was 96.5%; the matching ratio for the comparison group was 77.7%.

Appendix F

Assessing the quality of propensity score models and matching

The propensity score models and matching process underwent several checks to validate their quality, to ensure that the treatment and comparison groups were as similar as possible to allow tight comparison between similar prisoners. The following checks were performed:

Common support

The distribution of weighted propensity scores and weighted logit propensity scores were plotted in histograms and density plots to check for common support. Common support is the level of overlap in the distribution of propensity scores between the treatment and comparison groups. Where there is a lot of overlap, this shows that there are sufficient comparable units in each group to be matched. In addition, summary statistics of the weighted propensity scores were calculated to ensure sufficient similarity between the two groups. These statistics for the headline reoffending groups are shown in Table 13. They show a high level of similarity in the weighted propensity scores between the treatment and comparison group, implying a strong level of comparability between the two groups. Subgroups had similar levels of strong common support.

Table 13: Headline weighted propensity score summary

Group	Minimum	Quartile 1	Median	Mean	Quartile 3	Maximum
Treatment	0.001	0.056	0.107	0.154	0.205	0.856
Comparison	0.001	0.056	0.107	0.154	0.206	0.857

Receiver Operating Statistic and Area Under Curve

Another assessment of model quality is the receiver operating statistic (ROC) and associated area under curve (AUC). The ROC curve assesses how well the propensity score model distinguishes between the treatment and control groups. It represents the proportion of 'correct pairs' by showing the proportion of times where a unit in the treatment group correctly has a higher propensity score than a unit in the comparison

group. Where the ROC is between 0.7 – 0.9, this implies a high-quality model. For the headline propensity score model, the ROC was 0.792, which indicates good model quality.

Standard differences

Standard differences for all variables included in propensity score models were checked. Standard differences indicate the degree of similarity on a variable between the treatment and comparison groups. High quality models will maximise the number of variables with standard differences below 5% (0.05). Standard differences are presented in Appendix E.

Assessing p-values in the propensity score model

All variables were checked for the associated p-value in the propensity score models. High p-values indicates that the variable in question not significantly contribute to the model i.e. that the variable is not useful in predicting the probability of treatment. Variables with high p-values were checked to ensure they were protected. Protected variables are those that are deemed to be vital to the model, even if p-values are high. This included demographic characteristics (ethnicity, nationality), drug needs and key sentence characteristics (offence group, sentence length, offence severity and release year).

The Mean Average Standardised Difference

The mean average standardised difference (MASD) provides a single number which summarises the standard differences across all variables included in the propensity score model. The lower the MASD, the higher the degree of similarity between the treatment and comparison groups across all variables included in the propensity score model. During the matching process, the MASD was minimised so that treatment and comparison groups were kept as similar as possible in the aggregate. The MASD needed to be at least below 0.01 to imply similarity between the groups.

Matching ratios

The number of unmatched treatment units were kept as low as possible during the matching process to ensure they were included in the final reoffending outcomes. The number of unmatched treatment units needed to be below 5% of the total treatment group. Similarly, the ratio of matched treatment units to matched comparison units was kept as high as possible so that each treatment unit could be compared to several comparison

units. The ratio was kept at least 10:1. The matching rate (the proportion of matched units across both treatment and comparison groups) for each subset is shown in Appendix E.

Balancing MASD and Matching Ratios

A low MASD was balanced with a high matching ratio to ensure that groups were as similar as possible, whilst also including as many units as possible in the overall analysis. A range of matching weights were used to assess the trade-off between low MASD and the matching ratio and find a weight that best fit both criteria. The final MASD for the headline matching process was 0.0089, with a matching ratio between comparison and treatment group of 12:1, and 3.5% of the treatment group as unmatched treatment units. Subgroups had similarly low MASDs and matching ratios.

Appendix G
Accounting for Regional Inequality

To account for regional inequality between the areas in which prisons are located, indices of multiple deprivation were linked to the datasets. This limits the total sample so that it only includes prisoners in prisons which have similar levels of geographical deprivation. It does not account for individual-level deprivation and is not included in the propensity score models.

Prison postcodes were used to match a prisons' area to their ONS lower layer super output area (ONS, 2012). The Indices of Multiple Deprivation for England and Wales were then linked to the data based on prison's lower layer super output area (MHCLG, 2019; StatsWales, 2019). The treatment and comparison groups were then filtered to only those in a similar deprivation decile to Holme House. The lower layer super output area in which Holme House is in has a deprivation level in the top decile. However, reducing the comparison sample to only the top decile would have reduced the statistical power of the sample to too little to conduct meaningful analysis. To preserve statistical power, the comparison sample was filtered to prisons in the top three deciles of deprivation level based on their lower layer super output area. Propensity score models were then constructed for this smaller sample and matching conducted as per the main analysis. This approach was only used for the additional analysis on regional inequality, and the main analysis did not filter by deprivation.

The reoffending outcomes for prisoners in comparison prisons within the top three deciles of deprivation and for prisoners in the treatment prison are shown in Table 14. The results show statistically significant higher reoffending outcomes for the treatment group, once regional inequality was accounted for. Whilst this approach can only partially control for regional inequality, it does show that the statistically significant higher reoffending outcomes persist despite the additional controls for inequality. This suggest that other factors are driving the observed outcome rather inequality.

Table 14: High Indices of Multiple Deprivation prison areas reoffending outcomes

Reoffending indicator	Treatment mean	Comparison mean	Difference in means	p-value	Standardised effect size (Cohen's d)
1. One year proven reoffending rate (%)	48.25	44.77	3.48	0.002***	0.290
2. Number of proven reoffences	2.45	2.04	0.42	< 0.001***	0.354
3. Days to first proven reoffence	109.75	119.96	-10.21	0.003***	- 0.336
4. Number of proven reoffences resulting in custody sentence	2.73	2.45	0.28	0.026**	0.234

** Indicates statistical significance at the 95% confidence level.

*** Indicates statistical significance at the 99% confidence level.

Treatment group n = 2,404; comparison group n = 9,126 for indicators 1 & 2. Treatment group n = 1,160; comparison group n = 3,131 for indicators 3 & 4.