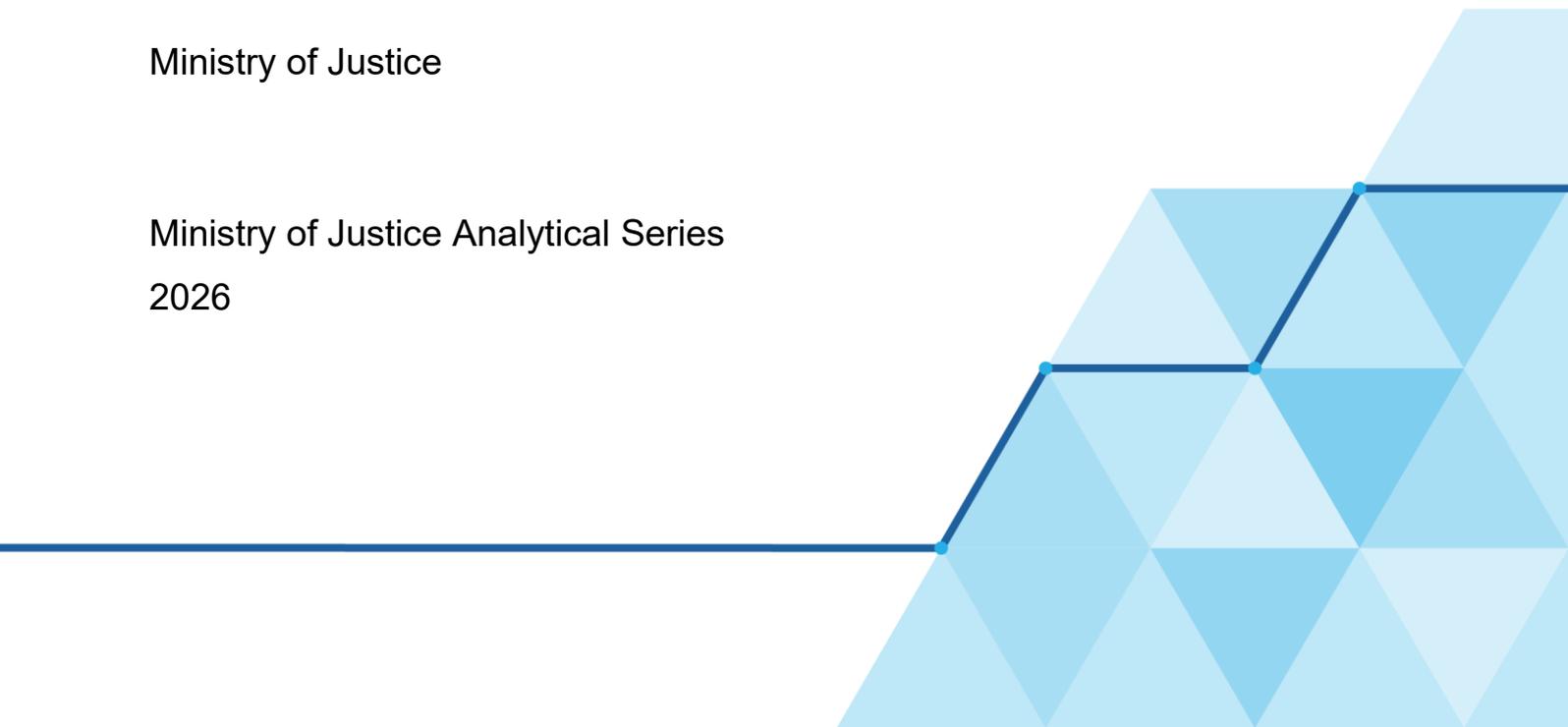# Integrated Offender Management

Impact Evaluation Technical Annex

Ministry of Justice

Ministry of Justice Analytical Series

2026

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.

Contents

List of tables

List of figures

Glossary

CAQDAS: Computer Assisted Qualitative Data Analysis Software

CMO(C): Context Mechanism Outcome Configuration – element of the realist evaluation approach. The CMOs are listed below:

Building Self-Belief: If IOM Practitioners give positive affirmation and support opportunities for people receiving IOM to build-up skills and pro-social activities or interests, **then** the people receiving IOM become more independent away from a criminal lifestyle, **because** they have internalised the positive affirmation and developed a greater level of self-belief and confidence in non-criminal identity

Control to Deter: If people receiving IOM want to avoid returning to prison, **then** they are likely to resist re-offending **because** they feel that they are likely to be caught if they offend

Flexibility: If staff have enough capacity and a holistic person-centred mindset, **then** they will provide responsive and practical support that can remove common obstacles to help avoid the people receiving IOM "being set up to fail" **because** they feel able to tailor their approach both as a team to the individual person receiving IOM and find creative solutions to deliver support.

Person Centred Approach: If people receiving IOM have something they want to change about their lives and staff have the capacity to deliver a person-centred approach to supervision, **then** people receiving IOM are more likely to engage, comply and make progress across desistance pathways, **because** they feel seen and supported, while experiencing IOM as a responsive source of stability that helps them to pursue meaningful change to their lifestyle and routine.

Swifter and Smoother Justice: If people receiving IOM re-offend or breach, and there is regular intel sharing between agencies, **then** there will be more responsive enforcement or control action to stop further offending **because** IOM Staff agree whether and how best to enforce using shared resources.

Trusting Relationships: If staff are honest and tailor supervision to the person receiving IOM's needs, **then** people receiving IOM are likely to buy in to IOM and be more open with their practitioner(s) **because** they have built trust in the practitioner(s) and know they are trying to help them.

Wrap Around Support: The overarching theory discussed in this section is that **If** staff have a holistic mindset, **then** there will be improved coordination of services which ultimately improve the offer available to people receiving IOM **because** collaboration is prioritised.

CSS: Crime Severity Score

Deselection: When an individual has been exited from IOM.

Fixed, flex and free: The model, made up of three distinct cohorts, laid out in the refreshed IOM strategy.

The fixed, flex and free model was designed to bring consistency to IOM selection, while also enabling schemes to retain their local flexibility and to work with other cohorts. For all cohorts, professional judgement is also factored in to determine whether police and probation believe there is 'value added' for someone by bringing them onto IOM.

- **Fixed**: Persistent offenders who have committed a neighbourhood crime and are considered as a high risk of reoffending (defined by their Offender Group Reconviction Scale (OGRS) and Crime Severity Scores (CSS)). Those in the fixed cohort are offered up to five appointments per week, with a target of three.

- **Flex**: Offenders who may not have committed a neighbourhood crime but who may have similar pathways to offending and would benefit from a neighbourhood crime IOM approach.

- **Free**: Freedom to use surplus resources to run IOM schemes for other cohorts, requiring a separate tailored approach with different pathways (e.g. serious violence, domestic abuse, and serious organised crime)

IOM: Integrated Offender Management. Those placed on IOM receive joint supervision from probation and police offender management teams. They also have access to various services funded by the Innovation Fund, which allows regions to address local needs and risks. IOM management emphasises enhanced information sharing, including co-located probation and police teams in some areas. Regular reviews are conducted through monthly Multi-Agency Case Conferences (MACC) and weekly Multi-Agency Practitioner (MAP) meetings.

IOM is delivered in the community but relates to offenders with index offences for both custody and community disposals.

MACC: Multi-Agency Case Conferences

MAP: Multi-Agency Practitioner Meetings

Neighbourhood crime: Crimes including domestic burglary, robbery, theft from the person, and vehicle and cycle crime. **OASys:** Offender Assessment System – used by the probation service to assess and manage offenders.

OCG: Organised Crime Group

OGRS: Offender Group Reconviction Score

PDU: Probation Delivery Unit

Person receiving IOM: An individual on probation who has been selected onto an IOM cohort, individuals are also referred to as IOM Nominals

PNC: Police National Computer

PPO: Prolific and Priority Offenders Programme – precursor to IOM

PSM: Propensity Score Matching

QED: Quasi-Experimental Design

Recall: When an individual is taken back to prison due to failing to comply with the conditions of their Licence.

SPO: Senior Probation Officer

TBE: Theory-Based Evaluation

ToC: Theory of Change, a model depicting how and why change is expected to happen from an intervention

1. Data Sources

The table below provides a broad overview of the various data sources included in the evaluation design. A more detailed summary of the quantitative data sources can be found in Sections 4 and 5.

Table 1.1: Data source breakdown by approach

Approach	Data Type	Data Source
CMO analysis	Primary Qualitative Data	Semi-Structured Interviews, Focus Groups, Non-Participant Observations
Changes in level of need	Secondary Quantitative Data	nDelius and OASys records
Effect on reoffending outcomes		Police National Computer (PNC)

2. Qualitative Methodology Technical Information

The following section provides an overview of the qualitative strand of the evaluation covering the sampling figures, topic guides[1], interview approach, and analytical process.

2.1 Qualitative sampling framework tables

Table 2.1:Number of interviews conducted by type and stage

Participant Type	Initial Stage: Site 1	Second Stage: Site 2	Third Stage: Site 1 & 2	Total
Probation Staff	6	5	8*	19
Police Staff	3	7	7*	17
Other Roles (including IOM funded roles and wider services)	3	2	4	9
People receiving IOM	1	2	9	12
Total Interviews	**13**	**16**	**28**	**57**

*All except one interview conducted with participants that were interviewed during one of the previous two stages.

Table 2.2: Number of observations conducted by type and stage

Observation Type	Initial Stage: Site 1	Second Stage: Site 2	Third Stage: Site 1&2	Total
People receiving IOM Supervision Observations	2	2	9	13
Multi-Agency Meeting Observations	2	3	0	5
Total Observations	**4**	**5**	**9**	**18**

---

[1] To enquire about the full detailed versions of the topic guides used in this research please contact IOMEvaluation@justice.gov.uk

Table 2.3: Breakdown of focus group representation

Stakeholder Group	Number of participants	Proportion of probation regions represented
Police Sergeants	10	83%
Senior Probation Officers	11	92%
Total Focus Group Participants	**21**	**100%**

2.2 Interview topic guides

Table 2.4: Topics covered during interviews broken down by participant groupings

Topic	People receiving IOM	IOM Practitioner (police and probation)	IOM Team Lead	Regional/Strategic Lead	Wider Agencies
Introduction and consent	Yes	Yes	Yes	Yes	Yes
Participant Background	Yes	Yes	Yes	Yes	Yes
General Reflections on IOM	Yes	Yes	Yes	Yes	Yes
Exploring Outcomes	Yes	Yes	Yes	Yes	Yes
Theory Testing:	Yes	Yes	Yes	Yes	Yes
Trusting Relationships	Yes	Yes	Yes		Yes
Building Self-Belief*	Yes	Yes	Yes		Yes
Control to Deter*	Yes	Yes	Yes		

Person-Centred Approach*	Yes	Yes	Yes		Yes
Flexibility		Yes	Yes		Yes
Wrap-Around Support*		Yes	Yes	Yes	Yes
Swifter and Smoother Justice*		Yes	Yes		
Exploring Context	Yes	Yes	Yes	Yes	Yes
Who IOM works best for	Yes	Yes	Yes	Yes	Yes
Unintended Outcomes		Yes	Yes	Yes	Yes
Areas for Improvement	Yes				
AOB and Wrap up	Yes	Yes	Yes	Yes	Yes

*Theory was discussed under different names and/or differing conceptualisations over the various phases of fieldwork. See Appendix B in the main report for details.

2.3 The Realist Interview

The design and delivery of the semi-structured interviews were informed by the principles of realist evaluation. Consistent with RAMESES II guidance, the interviews followed an iterative approach, with emerging findings used to adapt and refine CMO configurations and adjust the focus of subsequent interviews. This allowed for the continuous testing and development of programme theories throughout the data collection process.

The fieldwork was structured according to Manzano's (2016) three-phased realist interviewing framework:

- **Phase 1: Theory-gleaning interviews** – to surface initial theories and assumptions.

- **Phase 2: Theory-refinement interviews** – to interrogate, refine, and elaborate emerging CMO configurations.

- **Phase 3: Theory-consolidation interviews** – to test refined theories against participant experience and identify patterns of regularity or divergence.

At each stage, interviews were grounded in the teacher–learner model of theory building (Pawson and Tilley, 2004). In this dynamic process, the interviewer initially introduces programme theories (in the role of 'teacher'), which the interviewee ('learner') then engages with them by drawing on their experience. As the dialogue progresses, roles often reverse: the interviewee begins to 'teach' the interviewer by challenging, refining, or elaborating on the proposed theories. This creates a collaborative and reflexive space for theory development.

To support clarity and engagement, all interviews incorporated the use of "If…Then…Because" statements (Harris et al., 2024). These helped to clearly express the causal logic underpinning each theory in a concise and accessible format, allowing participants to confirm, refute, or build upon them based on their own experiences.

The same realist principles also shaped the design and facilitation of focus groups. Rather than aiming for consensus, these sessions were used to explore divergent experiences and understand how different contextual factors may be shaping variation in outcomes (Manzano, 2022). This approach enabled a richer understanding of the conditions under which programme mechanisms operate, and for whom.

2.4 Focus group topic guides

Table 2.5: Topics covered during focus group broken down by session

Topics	Focus Group 1	Focus Group 2	Focus Group 3	Focus Group 4
Introduction and consent	Yes	Yes	Yes	Yes
High level overview of findings from first two stages	Yes	Yes	Yes	Yes
Swifter and smoother Justice	Yes			
Person-Centred Approach		Yes		
Building Self-Belief			Yes	
Selection and De-selection processes				Yes
Shared Learning				Yes
Demand and Capacity				Yes
Trusting Relationships	Yes	Yes	Yes	Yes
Flexibility	Yes	Yes	Yes	Yes

2.5 Qualitative analysis

The qualitative analysis conducted in this evaluation followed a realist-informed approach: aiming to explore what works, for whom, in what contexts, and why. The analysis took point at various stages across the process of data collection, as is detailed in 'Qualitative Sampling Framework Tables'. Aligned with the key three stages of data collection, analysis was conducted iteratively and concurrently with data collection, allowing emerging insights to shape subsequent interview and focus group topic guides and the refinement of

developing programme theories. NVivo (version 14) was used as the primary tool for managing and analysing the qualitative data.

The process began with the development of initial CMO configurations, informed by Theory of Chage workshops, policy documents, operational guidance, and the researcher's early insights into the intervention. These initial CMOs were captured using NVivo's coding framework, with each code representing a hypothesised theory or pathway through which the intervention might lead to outcomes. Specific contexts, mechanism, and outcomes relating to each theory were recorded using child codes. A live model was created for each CMO using Miro[2]. These models were revisited and updated iteratively throughout the project to incorporate team reflections and shifts in theoretical assumptions.

As fieldwork progressed and recordings were transcribed, these transcripts were imported into NVivo and coded to relevant CMO codes and linked child codes. When data conflicted with current CMO models, it was coded as either falling under 'Refinement', requiring the team to consider alterations or additions to the CMO, or 'Refute', in instances where the data directly opposed current theoretical understandings. When new insights emerged, those not encompassed by existing CMO configurations, additional codes were created. All 'New', 'Refine', and 'Refute' codes were collaboratively reviewed and discussed at regular intervals along the analytical process and their subsequent decision making was logged and documented using Excel. This approach enhanced both the rigour and transparency of the qualitative analysis, aligning with best practice guidance for realist-informed evaluation.

To enable comparative analysis across data types, cases and classifications were used to distinguish between sources (for example, staff interviews, interviews with people receiving IOM, focus groups, and observations). This enabled flexible filtering and aggregation of coded content, supporting source-specific and cross-source analysis during theory refinement.

The research team met regularly to review coded data, identify areas where the evidence challenged or contradicted current theories, and collaboratively revise the CMO

---

[2] Miro is an online collaborative whiteboard tool used for visual planning, brainstorming, and team workshops.

configurations. Changes to the coding structure or theoretical propositions were documented through archived versions of the CMO models from which the codes were directly based upon.

This iterative approach facilitated the integration of multiple data sources and perspectives into the evolving explanatory framework. The use of Miro during the analysis process provided a collaborative analytical environment from which to work. This enabled all members of the research team, regardless of their direct involvement in data collection, to engage meaningfully in theory development.

3. Profile and Needs Technical Information

To better understand the characteristics and needs of the IOM cohort who receive custody and community disposals, and how these compare to offenders not on the IOM programme.

3.1 Background, Methodology and Coverage

Information is primarily derived from two data sources: the caseload management system (nDelius) and Offender Assessment System (OASys). nDelius is the National Probation Service's case management system. It is used to record information about offenders who are supervised by the Probation Service, including personal details, offending history, and risk assessments, throughout their sentence. OASys is an operational database used to assess and record the risks and needs of eligible offenders in prisons and the probation service across England and Wales.

For both the Profile and Needs sections, the remand population has not been included, and the data includes the following 13 regions: East Midlands Region, East of England, Greater Manchester, Kent Surrey Sussex, London, National Security Division, North East Region, North West Region, South Central, South West, Wales, West Midlands Region and Yorkshire and The Humber.

Profile

This section presents information on the criminogenic profile of the IOM and non-IOM cohort extracted from the caseload management system (nDelius) on a given date: 31 October 2024.

Data is broken down by the following three cohorts: IOM fixed, IOM flex and free (labelled IOM other) and non-IOM. Data includes both prison and probation populations across the three cohorts.

Needs

This section presents information on the needs of offenders on the probation caseload as at 31st October 2024, based on their latest OASys assessments. Information is derived from an analytical dataset that combines three Ministry of Justice data sources: the probation case management system (National Delius or nDelius), the prison case management system (P-NOMIS) and OASys. To create the underlying dataset, caseload listings from nDelius and P-NOMIS, as they appear on 31 October 2024, were matched with needs assessments sourced from OASys. Therefore, the dataset is a snapshot of

both the prison and probation population. There are two types of OASys assessments: a basic assessment (Layer 1), or a full assessment (Layer 3). A basic Layer 1 assessment is a reduced assessment. When matching to OASys assessments for individuals in the caseload, the most recent Layer 3 assessment is retained if it was completed during the current sentence and / or within the past two years. OASys coverage is defined as the proportion of all offenders with a complete Layer 3. [3]

This section presents information for those with a full OASys Layer 3. The data presented is therefore not representative of the full prison or probation caseload on 31 October 2024. Inferences should not be made about those without an assessment, and caution is needed when interpreting results where OASys assessment coverage (herein referred to as 'coverage') is lower or where numbers are small. The IOM cohort however has a relatively high coverage compared to the non-IOM cohort. 98% of the IOM cohort is covered in Layer 3 compared to 88% of the non-IOM cohort.

As the data is anonymised and does not contain the IOM category of nominal, the distinction between IOM fixed and the other categories of IOM could not be provided. Therefore, the needs section data is broken down into two groups: IOM and non-IOM.

3.2 Key findings

The profile of the IOM cohort and non-IOM cohort

This section shows the IOM cohort has more previous offences, a higher number of previous disposals, shorter custodial sentences, and a higher Risk of Serious Harm (RoSH) rating compared to the non-IOM cohort. The difference in demographics is much less pronounced, where white males under the age of fifty are slightly more prevalent in the IOM cohort than the non-IOM cohort.

Table 3.1 shows the proportion of the caseload, split by IOM and non-IOM that had at least one disposal recorded in nDelius prior to the latest recorded sentence. In counting disposals[4], only disposals resulting in a statutory probation and statutory custodial order are included, as nDelius does not consistently record fines or unsupervised convictions unless received during a period of supervision.

The IOM cohort is far more likely to have had a previous disposal compared to the non-IOM cohort. 98% of the IOM fixed cohort have previous sentences at disposal compared to 61% of the non-IOM cohort.

---

[3] While there may be some overlap, this process of sample selection differs to the approach implemented for the analysis of changes in OASys needs overtime due to their distinct analytical purposes.

[4] Data recorded prior to 2013 has have been migrated from a prior recording system and may be incomplete. These data are indicative only.

Table 3.1: Proportion of previous sentences at disposals by cohort

Cohort	More than one disposal	One disposal
IOM – Fixed	98%	2%
IOM Other Categories	92%	8%
Non-IOM	61%	39%

Source: nDelius case management system

Table 3.2 shows the average number of disposals and neighbourhood crimes disposals, split by IOM and non-IOM.

The IOM cohort has a higher number of disposals and neighbourhood crime disposals compared to the non-IOM cohort. The average number of previous disposals4 in the IOM fixed cohort is 15 of which 5 are neighbourhood crimes, compared to 5 of the non-IOM cohort of which 3 are neighbourhood crimes.

Table 3.2: Average number of disposals and neighbourhood crimes disposal by cohort

Cohort	Average number of disposals	Average number of neighbourhood crime disposals
IOM – Fixed	15	5
IOM Other Categories	11	4
Non-IOM	5	3

Source: nDelius case management system

Table 3.3 shows the proportion of those in the probation caseload with a custodial sentence[5], split by IOM and non-IOM.

The IOM fixed cohort has a much higher proportion with a custodial sentence[5] compared to the non-IOM cohort. 88% of the IOM fixed cohort has a custodial sentence compared to 56% of the non-IOM cohort.

Table 3.3: Proportion of those with custodial sentences by cohort

Cohort	Custodial Sentence	Non-Custodial Sentence
IOM – Fixed	88%	12%
IOM Other Categories	86%	14%
Non-IOM	56%	44%

Source: nDelius case management system

Table 3.4 shows the average custodial sentence length[6], split by IOM and non-IOM.

---

[5] Count of current custodial sentences include those on pre-release and post-release (on licence).

[6] Average length of custodial sentences is based on active statutory custodial sentences being served as at 31st October 2024 as reported in the monthly snapshot from nDelius MIS and will therefore differ from sentencing data, which is based on those sentenced over this period. Those with indeterminate sentences and life sentences are excluded.

Of those with a custodial sentence, the IOM fixed cohort has a relatively shorter average sentence length compared to the non-IOM cohort. The average sentence length for the IOM fixed cohort with a custodial sentence is 3 years compared to 5 years of the non-IOM cohort.

Table 3.4: Average sentence length in years by cohort

Cohort	Average Custodial Sentence in Years
IOM – Fixed	3
IOM Other Categories	4
Non-IOM	5

Source: nDelius case management system

Table 3.5 shows the top five most common offence types based on last active offence[7], split by IOM and non-IOM. Burglary (Domestic) was the most common offence type in the IOM fixed cohort, accounting for more than a quarter of offence types in IOM fixed. Violence was the most common offence type in both the IOM other categories and the non-IOM categories accounting for more than a quarter of offences both cohorts. In the non-IOM cohort, after violence, none of the remaining top five offence types individually accounted for more than 9% of the cohort, suggesting there is a wider spread of offence types in the non-IOM cohort compared to both the IOM fixed and IOM other categories cohort.

Table 3.5: Top five most common offence type based on last active offence.

Rank	IOM – Fixed		IOM Other Categories		Non-IOM	
1	Burglary (Domestic)	27%	Violence	27%	Violence	33%
2	Robbery	19%	Robbery	18%	Sexual (against child)	9%
3	Burglary (Other)	15%	Burglary (Domestic)	14%	Other offence	7%
4	Violence	14%	Theft (Non-motor)	8%	Drug import/export/production	7%
5	Theft (non-motor)	7%	Other offence	6%	Drug possession/supply	7%

Source: nDelius case management system

Table 3.6 shows the Risk of Serious Harm (RoSH) assessment rating distribution, split by IOM and non-IOM. Both the IOM fixed and IOM other categories are more likely to have a high or very high RoSH compared to the non-IOM cohort, with the IOM other categories having a slightly higher proportion of high or very high RoSH. 50% of the IOM fixed cohort and 58% of the IOM other categories is categorised as High or Very High for RoSH, compared to 35% in the non-IOM cohort.

---

[7] The last active offence is based on the main offence associated with a nominal's most recent active disposal as in the October 2024 snapshot. This excludes sentences not subject to supervision by Probation (e.g. sentences for summary offences are often not supervised).

Table 3.6: Risk Of Serious Harm (RoSH) rating by cohort

RoSH Description	IOM - Fixed	IOM Other Categories	Non-IOM
High & Very High RoSH - Total	50%	58%	35%
• High RoSH (RHRH)	48%	53%	33%
• Very High RoSH (RVHR)	2%	5%	2%
Low RoSH	1%	2%	12%
Medium RoSH	49%	41%	53%
Total	100%	100%	100%

Source: nDelius case management system

Figure 3.1 shows the age distribution across specified ranges, split by cohort. There are more people receiving IOM under the age of 50 in the IOM fixed and IOM other categories than in the non-IOM cohort. 91% of the IOM Fixed cohort is below the age of 50 compared to 81% of the non-IOM cohort.

Figure 3.1: Age group distribution by cohort
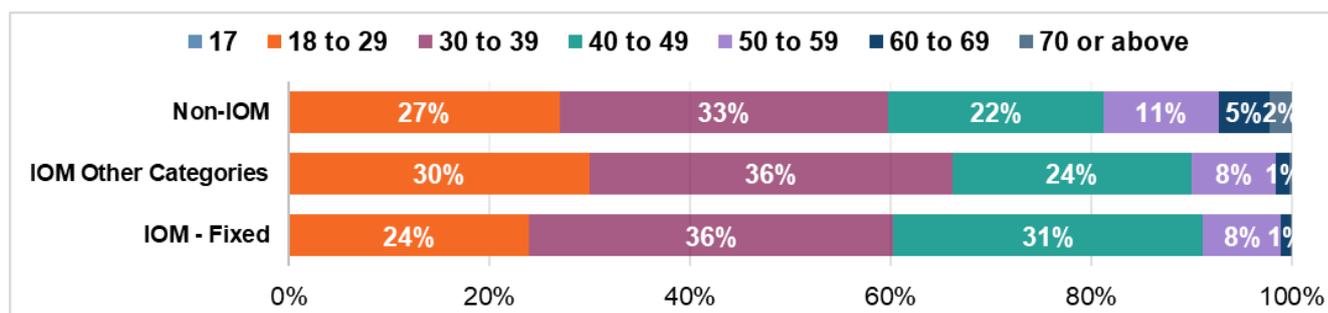
Source: nDelius case management system

Table 3.7 shows the gender distribution in each of the three cohorts. The IOM fixed cohort has a slightly higher proportion of male offenders compared to both the IOM other categories and non-IOM categories. 97% of the IOM fixed cohort is male compared to 91% of the non-IOM cohort.

Table 3.7: Gender Distribution by cohort

Cohort	Female	Male
IOM – Fixed	3%	97%
IOM Other Categories	6%	94%
Non-IOM	9%	91%

Source: nDelius case management system

Table 3.8 shows the ethnicity distribution in each of the three cohorts. The difference between the three cohorts is mainly in the proportion of white and Asian ethnicities. 82% of the IOM fixed cohort is White, compared to 72% of the non-IOM cohort. 3% of the IOM fixed cohort is Asian, compared to 12% of the non-IOM cohort. There is little difference in

the proportion of those with Black, mixed and other ethnic background between IOM fixed, IOM other categories and non-IOM cohort.

Table 3.8: Ethnicity Distribution by cohort

Ethnicities	IOM - Fixed	IOM Other Categories	Non-IOM
White	82%	79%	72%
Asian	3%	5%	12%
Black	8%	8%	9%
Mixed	6%	6%	5%
Other	1%	1%	2%
Not known	0%	1%	1%

Source: nDelius case management system

The needs of the IOM and non-IOM cohort

Categories of criminogenic needs in OASys assessment

Eight criminogenic needs are measured in OASys:

- accommodation,
- relationships,
- education, training and employment,
- lifestyle and associates,
- drug misuse,
- alcohol misuse,
- thinking and behaviour,
- pro-criminal attitudes

Table 3.9 and Table 3.10 present information in relation to the eight criminogenic needs. Mental health is also presented, giving nine needs in total.

All needs were derived from numbered questions in the OASys assessment. Further details on this can be found in Section 4[8].

This section shows there are a higher level of needs in the IOM cohort compared to the non-IOM cohort, for all nine need categories, with an overwhelming majority of the cohort having more than four or more concurrent needs. Table 3.9 shows the proportion of those with a need by each need category, split by IOM and non-IOM. The IOM cohort has a higher proportion of those with a need in all nine needs categories compared to the non-IOM cohort.

---

[8] This piece of analysis also includes mental health as a need. Details on how the mental health need is derived, can be found in section 4.1 of Identified offender needs, custody and community, 31 October 2024 - GOV.UK.

Lifestyles and Associates is the most prevalent need reported in OASys for both the IOM and non-IOM cohort, with 98% of the IOM cohort having the Lifestyles and Associates need. Pro-criminal attitudes was the second most prevalent need for the IOM cohort, with 95% of the IOM cohort having the Pro-criminal attitudes need. The biggest difference in needs between IOM and non-IOM is Drug Misuse, followed by Education, Training and Employment.

Table 3.9: Needs in the IOM and Non-IOM cohort

Cohort	IOM	Non-IOM
Accommodation	72%	46%
Relationships	85%	70%
Education, Training and Employment	84%	50%
Lifestyles & Associates	98%	76%
Drug Misuse	82%	44%
Alcohol Misuse	28%	23%
Thinking & Behaviour	88%	69%
Pro-criminal attitudes	95%	70%
Mental Health	75%	61%

Source: MoJ analytical dataset that combines nDelius, P-NOMIS and OASys

Table 3.10 shows the number of needs across specified ranges, split by cohort. The IOM cohort is likely to have more needs compared to the non-IOM cohort. 69% of the IOM cohort have 7 needs or above, out of which 14% have 9 needs, compared to 31% of the non-IOM cohort who have 7 needs or above, out of which 5% have 9 needs.

Table 3.10: Number of needs in the IOM and Non-IOM cohort.

Number of needs	IOM	Non-IOM
0 needs	0%	3%
1 to 3 needs	4%	25%
4 to 6 needs	27%	41%
7 to 8 needs	55%	26%
9 needs	14%	5%

Source: MoJ analytical dataset that combines nDelius, P-NOMIS and OASys

Other differences in needs by Cohort

This section presents information from the OASys assessment showing differences between the IOM and non-IOM cohort in areas other than criminogenic needs, to give more context of the IOM cohort. The selected questions have a high coverage within Layer 3, with at least 98% of questions completed across both cohorts.

Figure 3.2 shows the proportion with a disability[9], split by IOM and non-IOM. The IOM cohort has a higher proportion with a disability. 62% of the IOM cohort have a disability compared to 51% of the non-IOM cohort.

Figure 3.2: Proportion with a disability by Cohort
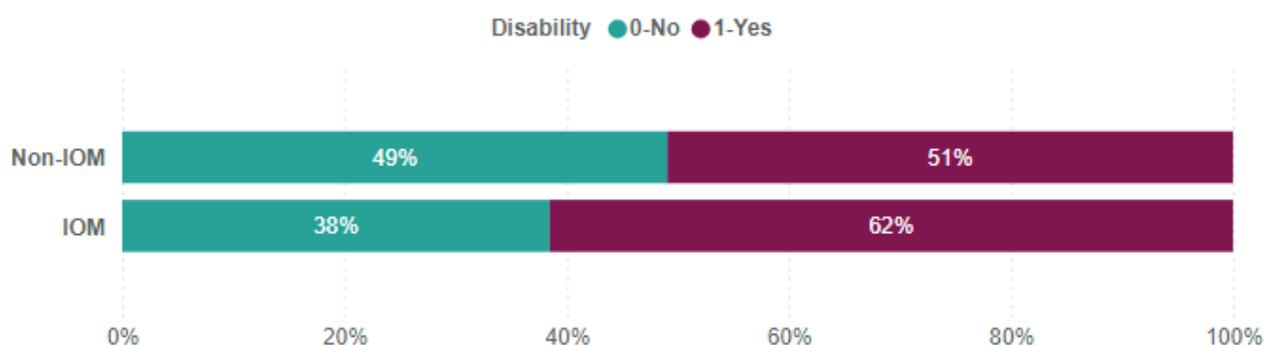
Source: MoJ analytical dataset that combines nDelius, P-NOMIS and OASys

Figure 3.3 shows the proportion with issues with self-harm, attempted suicide and suicidal thoughts or feelings, split by cohort. The IOM cohort has a higher proportion with such issues. 50% of the IOM cohort have issues with self-harm, attempted suicide, suicidal thoughts or feelings compared to 40% of the non-IOM cohort.

Figure 3.3: Proportion with issues with self-harm, attempted suicide, suicidal thoughts or feelings by Cohort
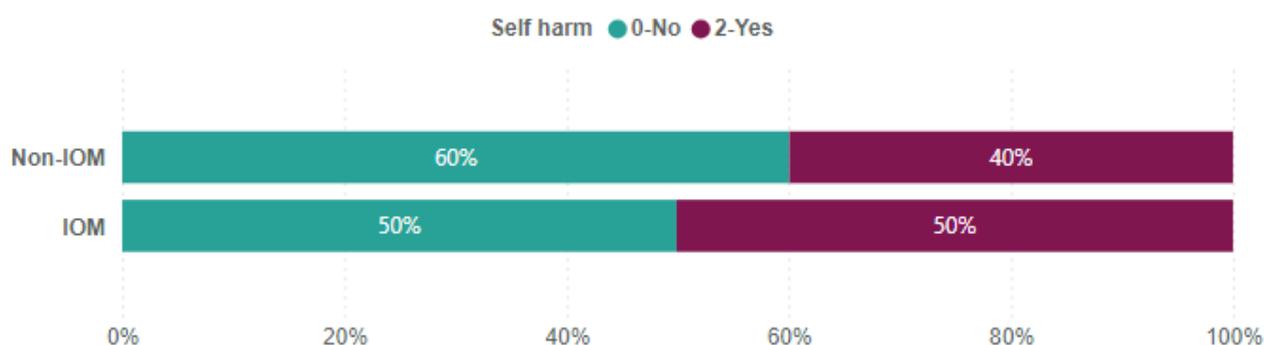
Source: MoJ analytical dataset that combines nDelius, P-NOMIS and OASys

Figure 3.4 shows the proportion with financial motivations involved in the offence, split by cohort. The IOM cohort has a significantly higher proportion with financial motivations

---

[9] Disability information does not come from the OASys assessment but is recorded on the nDelius case management system and relies on individuals on probation self-reporting a need. It can encompass a broad range of physical, cognitive, developmental and mental health disabilities.

compared to the non-IOM cohort. 76% of the IOM cohort have financial motivations involved in the offence compared to 34% of the non-IOM cohort.

Figure 3.4: Proportion with financial motivations by Cohort

Financial motivations ● 0-No ● 1-Yes

Non-IOM	66%	34%
IOM	24%	76%
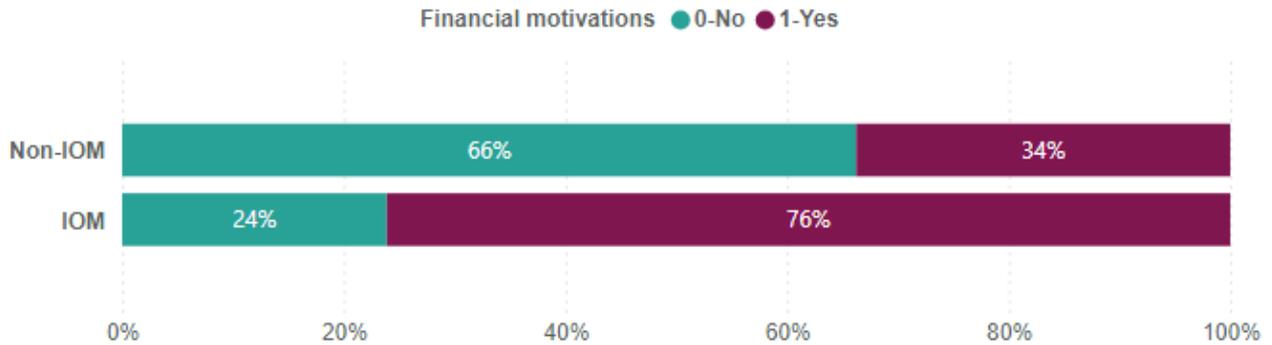
Source: MoJ analytical dataset that combines nDelius, P-NOMIS and OASys

Severity of needs by cohort

This section presents information in relation to specific questions in the OASys Assessment which show the difference in the severity of needs between the IOM and Non-IOM cohort. The questions selected highlight the intensity of the problems experienced by the IOM cohort. It shows that within the IOM cohort, among those with problems, those with "significant problems" outnumber those with "some problems" across various areas. The selected questions have a high coverage within Layer 3, with at least 98% of questions completed across both cohorts.

Figure 3.5 shows the severity of the problem with accommodation location, split by cohort. The IOM cohort has a slightly higher proportion of those with some problems compared to the non-IOM cohort. However, the IOM cohort has a much higher proportion with "significant problems". 46% of the IOM cohort have "significant problems" for accommodation location compared to 26% of the non-IOM cohort.

Figure 3.5: Problem with accommodation location by cohort
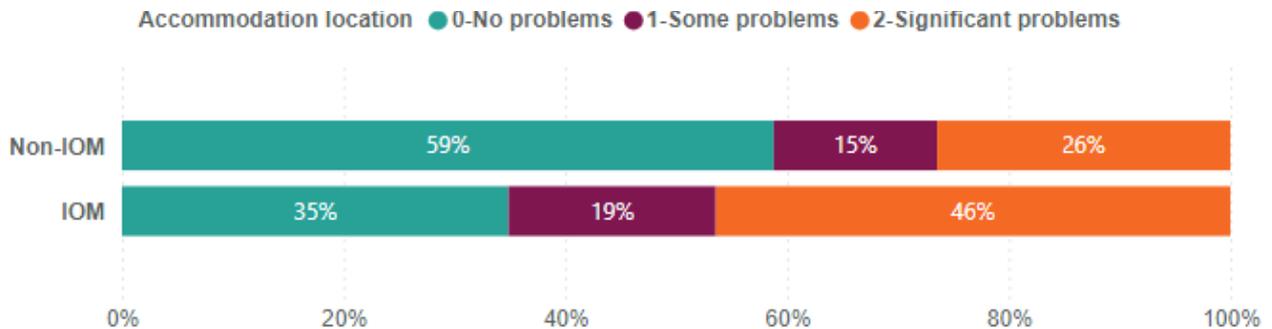
Source: MoJ analytical dataset that combines nDelius, P-NOMIS and OASys

Figure 3.6 shows the severity of the problem with accommodation permanence, split by cohort. The IOM cohort has a much higher proportion with "significant problems". 50% of the IOM cohort have "significant problems" for accommodation permanence compared to the 27% of the non-IOM cohort.

Figure 3.6: Problem with accommodation permanence by cohort
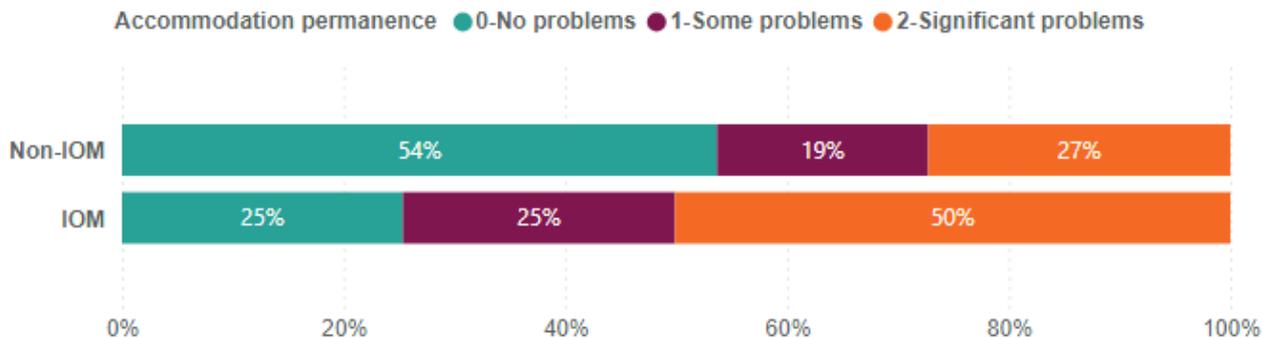
Source: MoJ analytical dataset that combines nDelius, P-NOMIS and OASys

Figure 3.7 shows the severity of the problem with reckless behaviour, split by cohort. The IOM cohort has a much higher proportion with "significant problems". 65% of the IOM cohort have "significant problems" for reckless behaviour compared to the 39% of the non-IOM cohort.

Figure 3.7: Problem with reckless behaviour by cohort
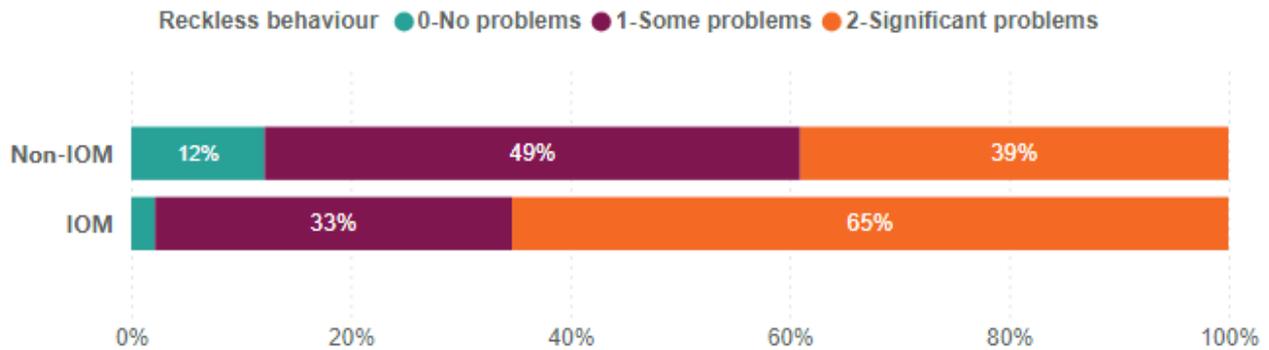
Source: MoJ analytical dataset that combines nDelius, P-NOMIS and OASys

Figure 3.8 shows the severity of the problem with impulsivity, split by cohort. The proportion in the IOM cohort with "significant problems" with impulsivity is twice as high as in the non-IOM cohort. 56% of the IOM cohort have "significant problems" for impulsivity compared to the 28% of the non-IOM cohort.

Figure 3.8: Problem with impulsivity by cohort
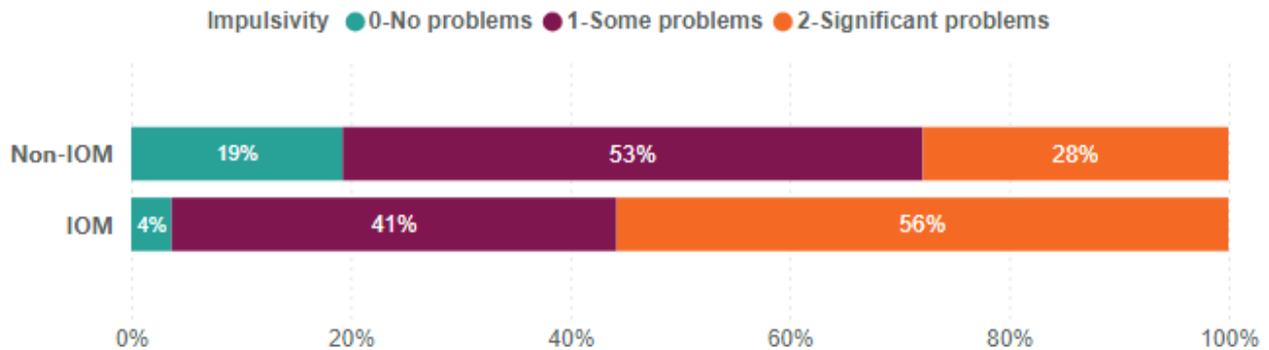
Source: MoJ analytical dataset that combines nDelius, P-NOMIS and OASys

Figure 3.9 shows the severity of the problem with financial management, split by cohort. The proportion in the IOM cohort with "significant problems" with financial management is more than twice as high as compared to the non-IOM cohort. 43% of the IOM cohort have "significant problems" with financial management compared to 16% of the non-IOM cohort.

Figure 3.9: Problem with financial management by cohort
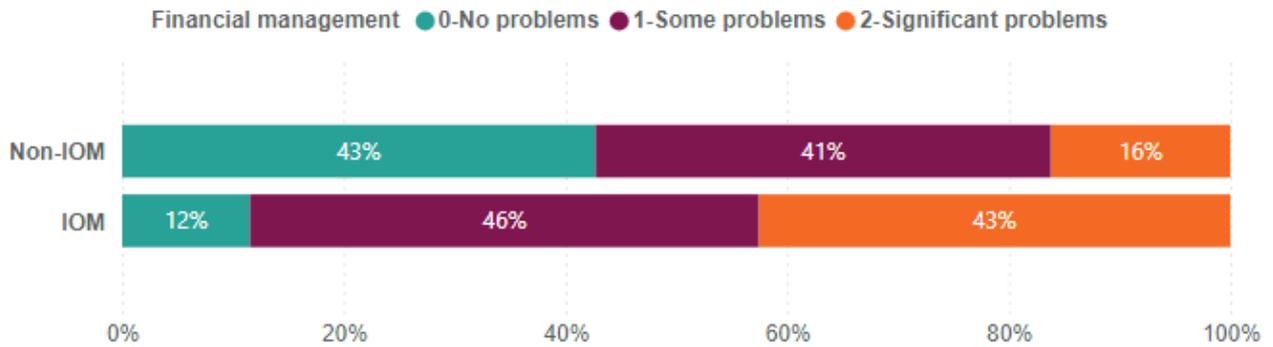
Source: MoJ analytical dataset that combines nDelius, P-NOMIS and OASys

4. OASys Criminogenic Needs Analysis Technical Information

4.1 Overview

This appendix outlines the technical specification and methodology for the quantitative analysis conducted using OASys and nDelius administrative data. The analysis focused on addressing Research Question 4: *'How and to what extent does the IOM programme influence individuals' level of support needs over 12 and 24 months?'.*

The inclusion of this quantitative analysis also enhanced the validity, depth, and reliability of the report's findings in several key areas as it provided a means of data triangulation. Specifically, it enabled further examination of the outcomes identified in the qualitative data collection for *'Person Centred Approach'* and *'Building Self-Belief'*, through applying a complementary quantitative lens. Furthermore, this analysis extends the evidence base provided by the PSM analysis conducted by the JDL by examining all IOM cohorts (Fixed, Flexed, and Free), including both male and females receiving IOM, and investigating a different, and longer, period of time (PSM analysis looked at IOM fixed cohort nominals registered between April 2021 – June 2022, whereas, this full cohort analysis looked at cases on IOM at any point between April 2022-May 2024). Together, these contributions broaden the scope of the evaluation and improve the generalisability of the findings.

4.2 Methodology

Data sources:

This analysis was conducted using data sourced from the probation case management system nDelius and the Offender Assessment System (OASys).

Dataset 1 – National Delius (nDelius)

National Delius (nDelius) is the established case management system used by the probation service. Data used for this analysis includes registration date onto IOM, IOM category, and deregistration date from IOM.

Dataset 2 – Offender Assessment System (OASys) dataset - England & Wales

OASys is a risk assessment and management system developed and used by the prison and probation services of England and Wales. It includes analysis of static (criminal history and demographic) and dynamic (social and personal) risk factors, risk of serious harm, sentence planning, a self-assessment (i.e. offender-completed) questionnaire and a summary sheet. This analysis was based on changes in the eight criminogenic needs that measure factors linked to offending behaviour, assessed using OASys: Accommodation, Employability, Relationships, Lifestyle, Drug misuse, Alcohol misuse, Thinking & Behaviour, and Attitudes.

Sample selection and data cleaning:

The sample consists of individuals on IOM between April 2022 and May 2024[10], in any of the three IOM-refresh cohort categories: Fixed, Flex or Free. People receiving IOM may have been selected and de-selected from IOM multiple times during that period, or moved between cohort types. Where people receiving IOM have multiple periods on IOM, their longest continuous period spent on IOM was selected as their 'main' IOM experience, and its start date used for the registration and subsequent follow-up periods. Similarly, where individuals moved between cohort categories within their 'main' IOM experience, they were classed as a 'mixed' cohort. As such, where findings are broken down by cohort type, there are four groups: 'Fixed', 'Flex', 'Free', and 'Mixed'.

Only individuals on IOM who had a sufficiently recent OASys assessment at registration and 12 months post-registration were included within the sample. Individuals were excluded if they had no OASys assessment completed within the six-month window[11] of their registration, and one year later, resulting in a reduction in sample to 14,752. [12]The 2-year OASys analysis also excluded those without an OASys assessment within the 6-month window of the 24-month post-registration date, as such the sample for that analysis is substantially reduced to 9,270. Further, to ensure the scores were meaningful, checks

---

[10] This time period was selected to reflect the start date of the IOM refresh funding up to the point of Probation Reset.

[11] As an OASys assessment is not likely to be completed more regularly than this, it was considered to provide a suitable window without reducing the sample excessively.

[12] It should be noted that not all sections of the OASys assessments were necessarily completed.

were put in place to ensure the OASys assessments were unique to each time period (so it would not be possible to use the same assessment for registration, and the one year follow-up).

The sample does not control for whether individuals may have:
- been de-selected from IOM within the treatment period[13] - time spent on IOM during the period varied (see table 4.5)
- had multiple periods on IOM[14] - although separate periods on IOM where there was less than 28 days gap between de-registration and re-registration were considered to be part of the same IOM period and merged.
- moved between the cohort groups during the treatment period – these cases are categorised as a 'mixed' cohort
- been on IOM before the treatment period – they may have received earlier IOM treatment prior to, or in the early stages of the IOM refresh

Outcome measures:

The primary outcomes of interest included in this analysis were the eight criminogenic needs recorded in OASys at various points at and following IOM registration. These measures provided both categorical indicators of need types and ordinal-level scores reflecting the severity or extent of each need.

Overall needs are calculated from the scores of specific questions from across assessments. Each of these questions is scored on a scale from 0 to 2 (some being scored 0 or 2, and others 0, 1 or 2) with 0 denoting 'no need', 1 'some need' and 2 'significant need'. Each set of questions has a maximum score, and a need is therefore identified when the score equals or exceeds a cut-off value (see Table 4.1).

---

[13] This could be for either positive or negative reasons.
[14] Start date refers to the start of the longest period the individual was on IOM during the treatment period.

Table 4.1: OASys criminogenic needs and relevant scored questions

Criminogenic needs	Scored questions	Scale range	Cut-off
Accommodation	3.3, 3.4, 3.5, 3.6	0-8	2+
Employment	4.2, 4.3, 4.4, 4.5	0-8	3+
Relationships	6.1, 6.3, 6.6	0-6	2+
Lifestyle and Associates	7.2, 7.3, 7.5	0-6	2+
Drug Misuse	8.4, 8.5, 8.6, 8.8, 8.9	0-10	2+
Alcohol Misuse	9.1, 9.2, 9.3, 9.5	0-8	4+
Thinking & Behaviour	11.5, 11.6, 11.7, 11.9	0-8	4+
Attitudes	12.1, 12.4, 12.5, 12.8	0-8	2+

To evaluate change in criminogenic needs over time, OASYS data relating to three time points relative to individuals' selection onto IOM were explored:

- **Baseline**: Defined as the first OASys assessment recorded within 6 months before or after registration onto IOM.
- **12-Month Follow-Up**: Defined as the first OASys assessment recorded within 6 months of the 12-month post-registration date.
- **24-Month Follow-Up**: Defined as the first OASys assessment recorded within 6 months of the 24-month post-registration date.

Not all IOM cases within our sample have a two-year OASYS assessment; this could be due to a number of reasons including completion of sentence/order, return to custody, the case being too recent to have a two year follow up date, or simply no OASYS assessment being completed. There are therefore two sets of analysis reported, one set of IOM cases with OASys scores at registration and one year follow up (n=14752), and a subset of those cases who have OASys scores at a two year follow up point (n=9270). Where relevant, findings are presented separately for both sets of cases.

Analytical approach:

Using each of the above data points analysis was conducted to assess the change in the criminogenic needs scores. It is important to note that as people receiving IOM may not have all parts of an OASYS assessment completed, changes in scores could not be calculated for people receiving IOM without OASys scores for the relevant sections at both registration and the subsequent sample points. As a result, there are varied base sample sizes depending on criminogenic need.

Statistical significance testing was performed on the observed changes in criminogenic need to assess whether these differences were statistically reliable, rather than attributable to random variation, thereby validating the evidence of the IOM's impact. It is important to note that there is no counterfactual analysis for the full IOM sample statistical analysis, and further exploratory statistical analysis was conducted to partly remedy this, detailed in Sub-analysis using the PSM counterfactual dataset:.

The need scores explored in this testing were derived through combining scores across multiple questions, and as such the scores are ordinal, not continuous. This, and that initial analysis indicated that the data was not normally distributed, therefore made the commonly used paired t-test unsuitable. To account for this, the Wilcoxon Signed-Rank tests, a non-parametric test alternative, was carried out.

The results of the statistical significance testing can be found in the *'Result Summary'*.

In addition to these measures of change over time, descriptive measures were calculated for the constructed sample as a whole, as well as separately for each of the respective cohort types.

The Sample

Table 4.2: Sample breakdown by gender

Gender	Year One Sample	Year Two Sample
Male	13942	8821
Female	810	449
Total	14752	9270

Table 4.3: Sample breakdown by ethnicity

Ethnicity	Year One Sample	Year Two Sample
Arab	37	20
Asian or Asian British: Bangladeshi	95	56
Asian or Asian British: Chinese	3	2
Asian or Asian British: Indian	71	31
Asian or Asian British: Other	69	44
Asian or Asian British: Pakistani	210	124
Black or Black British: African	270	164
Black or Black British: Caribbean	568	368
Black or Black British: Other	177	113
Mixed: Other	99	62
Mixed: White and Asian	80	53
Mixed: White and Black African	82	46
Mixed: White and Black Caribbean	580	407
Other Ethnic Group	70	44
White: Irish	144	97
White: Other	205	91
White: British/English/Welsh/Scottish/Northern Irish	11722	7389
White: Gypsy or Irish Traveller	19	15
White: Gypsy, Irish Traveller, Romany	176	109
White: Roma	4	2
Refusal	59	27

Table 4.4: Age of sample

	Age[15] of one-year sample	Age of two-year sample
Mean (Standard Deviation)	36.8 (9.59)	36.93 (9.4)
Median (Min - Max)	36 (18 - 88)	37 (19-82)

Table 4.5: Length of time on IOM[16]

	One-year sample (n=14752)	Two-year sample (N=9270)
Up to 3 months	7%	3%
Over 3-6 months	9%	5%
Over 6 to 9 months	12%	8%
Over 9 to 12 months	14%	10%
Over a year	58%	74%

---

[15] Age is recorded at the point the data is updated, as such it refers to the age of the individual on the snapshot date of 01 April 2025.

[16] Where individuals were on IOM for multiple separate periods, this represents the longest period they spent on IOM during April 2022 - May 2024.

Table 4.6: Distance in time between IOM registration and OASys completion date for one-year sample

One year Sample (n=14,752)	1 month	2 months	3 months	4 months	5 months	6 months
Registration	46%	22%	14%	9%	5%	4%
One-year follow up	33%	23%	16%	12%	9%	7%

Table 4.7: Distance in time between IOM registration and OASys completion date for two-year sample

Two year sample (n=9,270)	1 month	2 months	3 months	4 months	5 months	6 months
Registration	43%	22%	14%	10%	6%	4%
Two-year follow up	30%	23%	15%	12%	11%	9%

4.3 Results

Descriptive data on changes to criminogenic needs over time

Figure 4.1: Percentage of sample with criminogenic needs by year

Proportion with Identified Needs by Year

Need	Registration	One year	Two year
Attitudes	91%	94%	95%
Thinking	85%	88%	89%
Alcohol	29%	31%	32%
Drugs	74%	74%	74%
Lifestyle	94%	96%	96%
Relationships	82%	86%	87%
Employment	67%	67%	68%
Accommodation	69%	71%	73%

■ Need at two year outcome (n = 9270)
■ Need at one year outcome (n = 14752)
■ Need at registration (n = 14752)
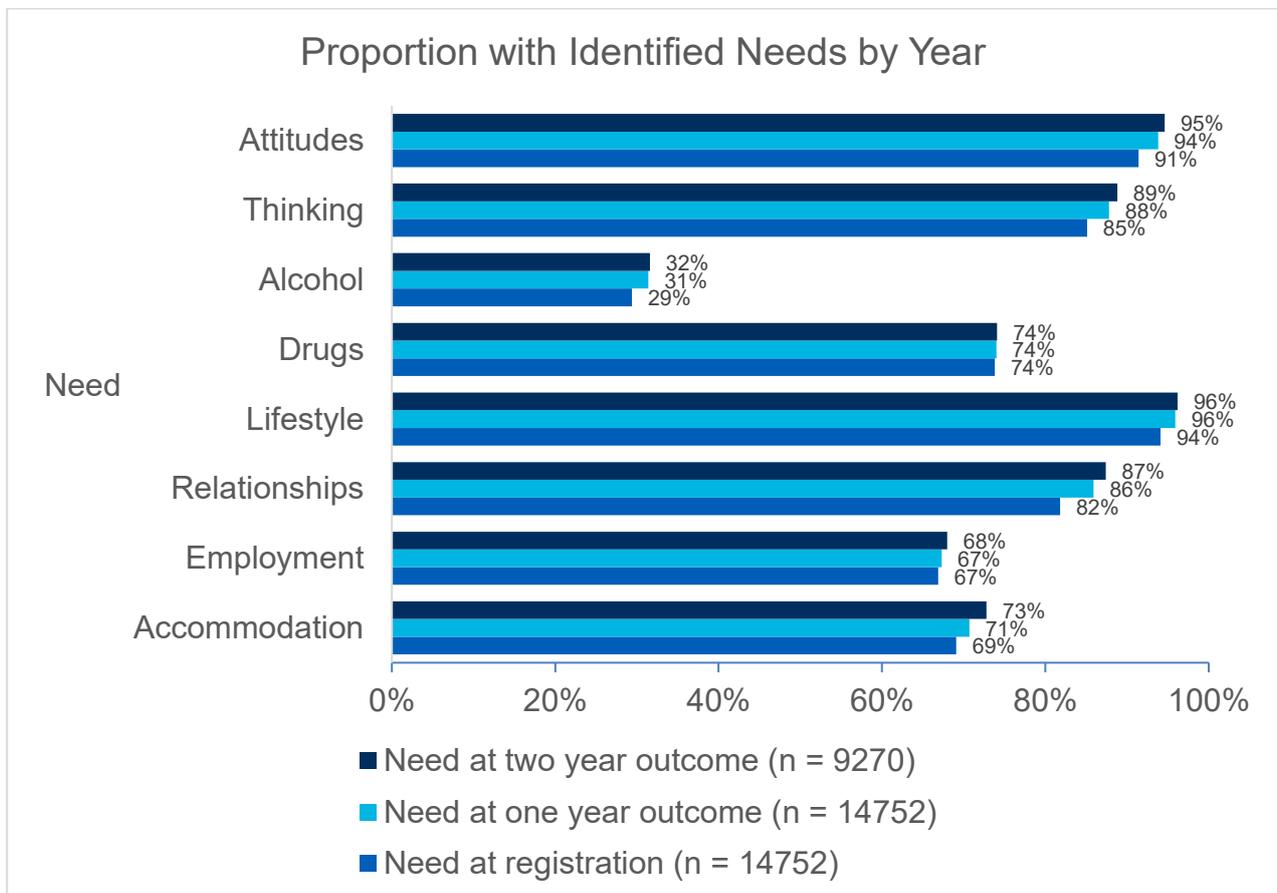
Figure 4.2: Change in Severity of Needs (measured by their OASYs need score) at one-year post-registration

Change in Severity of Needs One-Year Post-Registration

Need	No need at either time point	Increased	No change	Decreased
Attitudes (n=14260)	6%	20%	59%	15%
Thinking (n=14264)	13%	12%	67%	9%
Alcohol (n=9273)	54%	6%	30%	11%
Drugs (n=12795)	19%	17%	49%	14%
Lifestyle (n=14231)		12%	77%	9%
Relationships (n=14138)	15%	10%	70%	5%
Employment (n=10957)	14%	7%	69%	10%
Accommodation (n=14262)	29%	10%	41%	20%
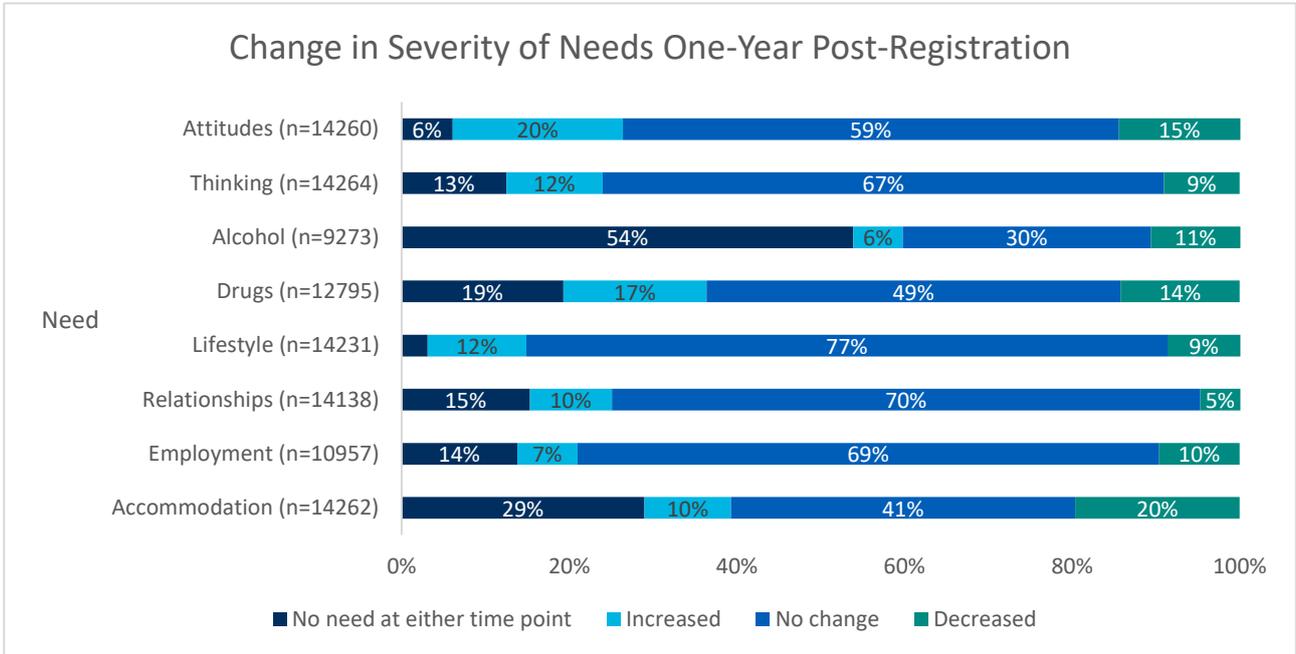
Figure 4.3: Change in Severity of Needs (measured by their OASYs need score) at two-year post-registration

Change in Severity of Needs Two-Year Post-Registration

Need	No need at either time point	Increased	No change	Decreased
Attitudes (n=8933)		33%	44%	21%
Thinking (n=8935)	7%	23%	55%	14%
Alcohol (n=5760)	45%	18%	22%	15%
Drugs (n=7817)	11%	31%	37%	21%
Lifestyle (n=8915)		20%	65%	14%
Relationships (n=8862)	10%	22%	60%	8%
Employment (n=6631)	8%	17%	61%	14%
Accommodation (n=8938)	16%	28%	33%	24%
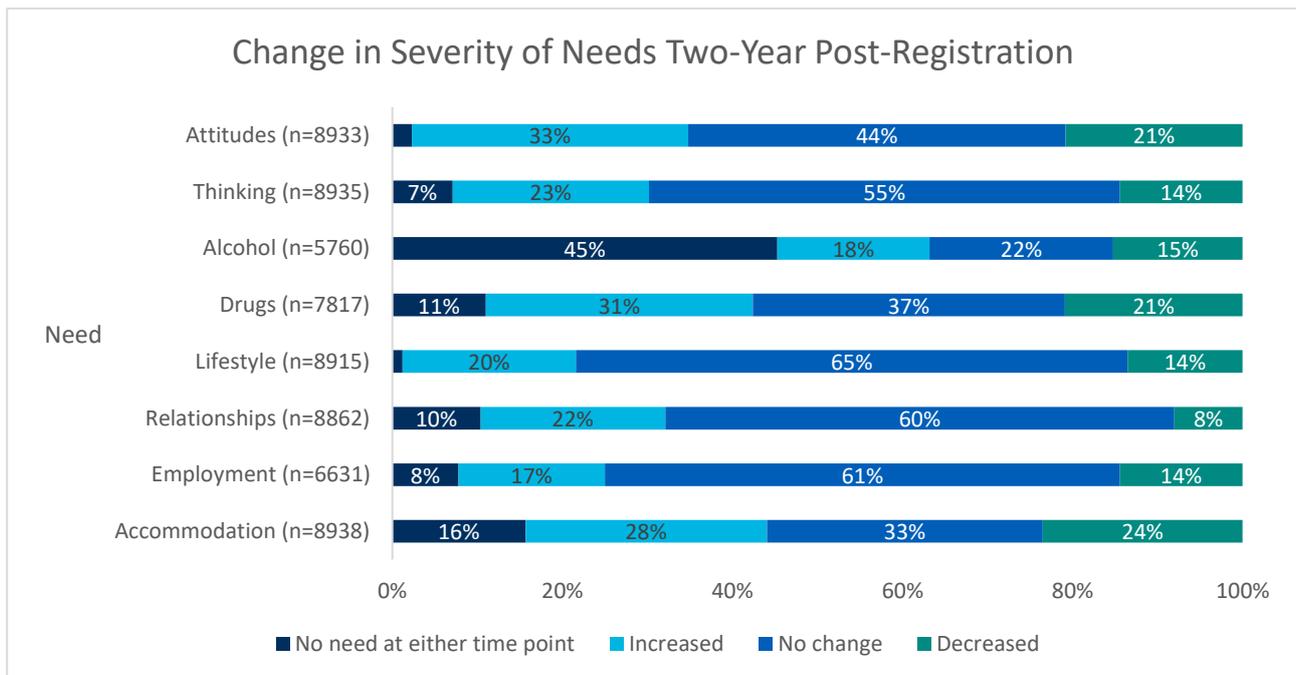
Significant Testing Results

Table 4.8: Paired Wilcoxon Signed Rank Test: Change in Severity of Needs between IOM registration and one year post-registration

Need	Sample size	Start Median	End Median	Test Statistic	Significance (p-value)	Effect Size
Accommodation	14262	4	4	9068218	0.996	0.001
Employment	10957	5	5	1413983	0.348	0.007
Relationships	14138	3	4	833233	< 0.001	0.197
Lifestyle and Associates	14231	4	4	2068969	< 0.001	0.086
Drug Misuse	12795	4	4	4336564	< 0.001	0.125
Alcohol Misuse	9273	3	3	1664963	0.001	0.028
Thinking and Behaviour Skills	14264	6	6	2564671	< 0.001	0.117
Pro-Criminal Attitudes	14260	5	5	5479564	< 0.001	0.135

Table 4.9: Wilcoxon Signed Rank Test: Change in Severity of Needs between IOM registration and two years post-registration

Need	Sample size	Start Median	End Median	Test Statistic	Significance (p-value)	Effect Size
Accommodation	8938	3	5	5514209	< 0.001	0.071
Employment	6631	5	5	1186081	0.054	0.041
Relationships	8862	3	4	934634	< 0.001	0.260
Lifestyle and Associates	8915	4	5	1952445	< 0.001	0.113
Drug Misuse	7817	4	5	3364259	< 0.001	0.148
Alcohol Misuse	5760	3	3	1397512	< 0.001	0.041
Thinking and Behaviour Skills	8935	6	6	2402255	< 0.001	0.139
Pro-Criminal Attitudes	8933	5	5	4411048	< 0.001	0.162

Table 4.10: Wilcoxon Paired Signed Rank Test: Change in severity of needs over one year follow up period, broken down by cohort type

Cohort Type	Need	Sample size	Start Median	End Median	Test Statistic	Significance (p-value)	Effect Size
Fixed	Accommodation	6087	4	4	1641349	0.430	0.016
	Employment	4700	5	5	219272	0.521	0.359
	Relationships	6041	3	3	165554	<0.001	0.187
	Lifestyle and Associates	6072	5	5	335276	<0.001	0.089
	Drug Misuse	5598	4	5	815692	<0.001	0.149
	Alcohol Misuse	3995	3	3	246614	<0.001	0.064
	Thinking and Behaviour Skills	6092	6	6	416648	<0.001	0.141
	Pro-Criminal Attitudes	6091	5	5	945675	<0.001	0.137
Flex	Accommodation	4370	3	3	835528	0.177	0.024
	Employment	3372	5	5	143108	0.293	0.004
	Relationships	4327	3	3	69525	<0.001	0.207
	Lifestyle and Associates	4366	4	4	197197	<0.001	0.091
	Drug Misuse	3899	4	4	392451	<0.001	0.110
	Alcohol Misuse	2737	3	3	149308	0.162	0.033
	Thinking and Behaviour Skills	4375	5	6	234167	<0.001	0.091
	Pro-Criminal Attitudes	4372	5	5	513022	<0.001	0.121
Free	Accommodation	2148	4	3	213835	0.550	0.017
	Employment	1615	4	4	35122	0.977	0.024
	Relationships	2130	4	4	15521	<0.001	0.204
	Lifestyle and Associates	2146	4	4	52511	0.027	0.070
	Drug Misuse	1770	4	4	84630	0.009	0.073
	Alcohol Misuse	1504	4	4	68961	0.001	0.087
	Thinking and Behaviour Skills	2147	6	6	72246	0.096	0.088
	Pro-Criminal Attitudes	2148	5	5	135323	<0.001	0.117
Mixed	Accommodation	1657	3	3	125389	0.198	0.028
	Employment	1270	5	5	24322	0.875	0.016
	Relationships	1640	3	4	13898	<0.001	0.198
	Lifestyle and Associates	1647	4	4	34896	0.425	0.079
	Drug Misuse	1528	4	5	68542	<0.001	0.135
	Alcohol Misuse	1037	3	3	20603	0.019	0.059
	Thinking and Behaviour Skills	1650	5	6	40704	<0.001	0.137
	Pro-Criminal Attitudes	1649	5	5	81023	<0.001	0.179

Table 4.11: Wilcoxon Paired Signed Rank Test: Change in severity of needs over a two year follow up period, broken down by cohort type

Cohort Type	Need	Sample size	Start Median	End Median	Test Statistic	Significance (p-value)	Effect Size
Fixed	Accommodation	3922	4	5	1010272	<0.001	0.098
	Employment	2932	5	5	203262	0.306	0.046
	Relationships	3893	3	4	198199	<0.001	0.257
	Lifestyle and Associates	3910	5	5	346007	<0.001	0.121
	Drug Misuse	3497	4	5	665981	<0.001	0.174
	Alcohol Misuse	2527	3	3	235786	<0.001	0.065
	Thinking and Behaviour Skills	3922	6	6	407189	<0.001	0.166
	Pro-Criminal Attitudes	3923	5	5	781615	<0.001	0.171
Flex	Accommodation	2663	3	4	497387	0.013	0.055
	Employment	1967	5	5	111417	0.051	0.049
	Relationships	2638	3	4	77550	<0.001	0.277
	Lifestyle and Associates	2660	4	5	184047	<0.001	0.113
	Drug Misuse	2317	4	5	309377	<0.001	0.110
	Alcohol Misuse	1660	3	3	127244	0.158	0.019
	Thinking and Behaviour Skills	2665	6	6	227643	<0.001	0.106
	Pro-Criminal Attitudes	2664	5	5	428773	<0.001	0.150
Free	Accommodation	1158	4	5	95545	0.268	0.029
	Employment	829	5	5	21789	0.356	0.004
	Relationships	1148	4	4	13148	<0.001	0.228
	Lifestyle and Associates	1155	4	4	40829	0.328	0.057
	Drug Misuse	957	4	4	51095	0.002	0.107
	Alcohol Misuse	825	4	4	41607	0.075	0.060
	Thinking and Behaviour Skills	1158	6	6	48823	0.052	0.105
	Pro-Criminal Attitudes	1158	5	5	78292	<0.001	0.127
Mixed	Accommodation	1195	3	4	108058	0.021	0.062
	Employment	903	5	5	24820	0.202	0.050
	Relationships	1183	3	4	16611	<0.001	0.262
	Lifestyle and Associates	1190	4	5	31892	0.001	0.140
	Drug Misuse	1046	4	5	55761	<0.001	0.179
	Alcohol Misuse	748	3	3	18339	<0.001	0.127
	Thinking and Behaviour Skills	1190	5	6	45623	<0.001	0.157
	Pro-Criminal Attitudes	1188	5	5	79655	<0.001	0.187

4.4 Limitations

Analysis based on monitoring information data must be interpreted with the caveat that it has been extracted from systems designed to administer or monitor an operational service rather than for research purposes. Data is therefore subject to clerical and input errors, which has implications on the quality of the data, linking and deduplication of records.

In the case of the nDelius dataset, there are known issues with the administrative reporting of registration dates. Specifically, when individuals move between registration categories, the new category is sometimes incorrectly backdated to the start date of the previous one. This can result in inaccurate timelines that can distort analyses of programme engagement and progression. Additionally, the dataset lacks detailed information on the reasons for de-selection from IOM, for example whether this is due to more negative reasons such as returning to custody or non-engagement, or more positive ones such as progression across desistance pathways or reduction in perceived level of risk.

The use of OASys data also incurs several key limitations which could bias the representativeness of the sample.

- There is variation in the timing between assessment the follow-up periods used for outcome measurement (e.g. 12-month and 24-month post-selection) (see tables 4.7 and 4.8). The majority of baseline OASYS assessments were conducted within 1-2 months of the people receiving IOM's registration onto the programme for both sets of the sample (68% for one year sample, and 66% for two year sample), this is slightly more varied for the 12- (56%) and 24- (52%) month follow up dates. This variability presents the risk of introducing inconsistencies across the sample.

- The exclusion of individuals who do not have an assessment completed within the set parameters may bias the sample also. OASys assessment may be completed at intervals, but can also be triggered by an event such as recall, release, change in needs, that may mean those individuals with more regular OASys assessment are different from those with less regular assessments.

- The analysis also does not control for the duration of time individuals spend on IOM, which may be a significant factor influencing observed outcomes.

- Consistent with past publications this piece of analysis only focuses on the eight listed criminogenic needs of OASys and does not include Mental Health as a need.[17]

There is a plan for practitioners to transition away from using OASys, and as such analysis may not be replicable using future samples due to its planned replacement with the Assess Risks, Needs and Strengths (ARNS) Programme.

Finally, the method used for testing statistical significance of changes in criminogenic needs over time carry its own limitations. The Wilcoxon Signed-Rank test, while appropriate in such contexts where data cannot be assumed to be normally distributed, is generally less powerful than parametric alternatives such as a paired t-test. This means it has a reduced ability to detect statistically significant differences when they exist, particularly in smaller samples or when effect sizes are modest.

Sub-analysis using the PSM counterfactual dataset:

To address the limitations of OASys outcomes analysis comparing initial and follow up scores without a non-IOM counterfactual, a supplementary sub-analysis was conducted using the dataset derived from the PSM analysis, as detailed in Section 5. This analysis aimed to examine the differences in changes to OASys scores between those on IOM Fixed cohort and a matched counterfactual group. While this approach allows for a more robust examination of statistical significance, the analysis remains subject to the same caveats regarding its scope, as the PSM-derived treatment groups consist solely of males within the fixed cohort, and within the very early stages of the IOM refresh (the PSM cases were registered on IOM between April 2021 and June 2022, while the other sample could have been registered as on IOM anytime between April 2022 and May 2024, the extent to which the two sets of samples overlap is not explored).

The weighted treatment and control cohorts from the one-year proven reoffending analysis were further employed to test the hypothesis that there might be a significant difference in the change in aggregated OASys scores between people receiving IOM and those who

---

[17] However, Mental Health is reported on in Appendix E

didn't receive IOM following a one or two-year period since registration for particular areas of criminogenic need.

Treatment and matched counterfactuals' intervention/pseudo intervention registration dates, and associated weights, was linked with OASys responses. These were restricted to the same six-month window, meaning only cases with an assessment within a six-month window of their start date, and the same window of their one- and two-year anniversaries of registration (or pseudo-registration date), were retained.

The initial dataset consisted of 3,836 registrations relating to 3,835 people receiving IOM in the treatment group and 2,294 people on probation in the control group, a total of 6,130 rows. Following linking with OASys, assessments were found for 4,855 nominals within the specified window around registration/pseudo-registration. Of these, 3,652 had assessments within the one-year window, and 2,514 within the two-year window

Changes in scores could not be calculated for people receiving IOM without OASys scores for the relevant sections at both registration and the subsequent sample points. This resulted in moderate data loss, which varied across OASys sections and aggregated criminogenic need scores. Consequently, the sum of the weights applied to the control group ceased to match that of the treatment group. Additionally, information regarding which members of the treatment and control group were matched could not be provided. As a result, when considering weighted variation in scores, it could only be assumed that those lost from the treatment and control groups were similar. The remaining weights were proportionately adjusted to re-equate the weights between the treatment and control groups. This is a significant, non-robust assumption.

The difference in scores was calculated, and both the unweighted and weighted contrast between groups were evaluated via a weighted/unweighted Wilcoxon rank sum test for unpaired, non-normally distributed, continuous data.

Key findings:

One and Two Year Paired-samples outcomes:

While there are general similarities between the smaller PSM fixed cohort compared to the larger IOM fixed sample within the wider analysis, differences remain when comparing the relative change within each cohort over the one- and two-year outcomes.

Comparing Table 4.10 with Table 4.12 both demonstrated very small but statistically significant decreases in criminogenic scores across the same needs, except for the lifestyle and associates need for which only the larger cohort showed a small significant decline over the year.

There are slightly more differences between the two group's two-year outcomes. While the PSM cohort demonstrated very small statistically significant decreases in scores in all needs except for lifestyle and associated need; the larger fixed cohort showed decreases for all but employment.

However, as all of these significant effect sizes are small or very small, we continue to treat these two groups of fixed cohorts as comparable for the purposes of incorporating a counterfactual-based analysis on OASYS outcomes. The key finding to note is that across one and two years, the majority did change their score.

Table 4.12: Fixed Cohort cases included in the PSM analysis: Wilcoxon Signed Rank Test exploring change in eight criminogenic needs in the one year following registration on to IOM

Need	Sample size	Start Median	End Median	Test Statistic	Significance (p-value)	Effect Size
Accommodation	1471	3	3	94557	0.113	0.041
Employment	1148	5	5	13800	0.535	0.023
Relationships	1459	3	3	12053	<0.001	0.202
Lifestyle and Associates	1465	5	5	23523	0.522	0.032
Drug Misuse	1428	4	4	59024	<0.05	0.082
Alcohol Misuse	928	3	3	14209	<0.05	0.075
Thinking and Behaviour Skills	1470	6	6	26409	0.050	0.083

Pro-Criminal Attitudes	1469	5	5	56765	<0.001	0.130

Table 4.13: Fixed Cohort cases included in the PSM analysis: Wilcoxon Signed Rank Test exploring change in eight criminogenic needs in the two year following registration on to IOM

Need	Sample size	Start Median	End Median	Test Statistic	Significance (p-value)	Effect Size
Accommodation	1069	3	4	76708	<0.001	0.126
Employment	799	5	5	14752	<0.05	0.109
Relationships	1062	3	3	53726	<0.001	0.291
Lifestyle and Associates	1065	5	5	31266	0.180	0.067
Drug Misuse	1031	4	4	72410	<0.01	0.096
Alcohol Misuse	676	3	3	19051	<0.05	0.101
Thinking and Behaviour Skills	1066	6	6	37862	<0.01	0.132
Pro-Criminal Attitudes	1066	5	5	61991	<0.001	0.194

Change in Severity of Needs: A Comparison between the PSM fixed and matched counterfactual samples

The analysis showed no evidence of differences between the Fixed Cohort and the matched and weighted counterfactual in regards to change in the severity of criminogenic needs over one year or two years.

Table 4.14: Weighted Wilcoxon Signed Rank Test: Change in Severity of Needs between PSM IOM Fixed Cohort and matched counterfactual, over one year post-IOM registration

Need	IOM: Start Median*	IOM: End Median*	Comparator: Start Median*	Comparator: End Median*	Test Statistic	Significance (p-value)	Effect Size
Accommodation	3	3	3	4	1475698	0.746	0.005
Employment	5	5	5	5	895074	0.384	0.014
Relationships	3	3	3	3	1457101	0.731	0.006
Lifestyle and Associates	5	5	4	5	1436080	0.080	0.029
Drug Misuse	4	4	4	4	1364613	0.665	0.007
Alcohol Misuse	3	3	3	3	609389	0.552	0.010
Thinking and Behaviour Skills	6	6	5	5	1475517	0.751	0.005
Pro-Criminal Attitudes	5	5	4	5	1464355	0.461	0.012

* Weighted Medians

Table 4.15: Weighted Wilcoxon Signed Rank Test: Change in Severity of Needs between PSM IOM Fixed Cohort and matched counterfactual, over two year post-IOM registration

Need	IOM: Start Median*	IOM: End Median*	Comparator: Start Median*	Comparator: End Median*	Test Statistic	Significance (p-value)	Effect Size
Accommodation	3	4	3	5	720066	0.979	0.001
Employment	5	5	5	5	393043	0.444	0.015
Relationships	3	3	3	3	693614	0.583	0.011
Lifestyle and Associates	5	5	4	5	671693	p<0.01	0.057
Drug Misuse	4	4	4	4	637620	0.306	0.020
Alcohol Misuse	3	3	3	3	280145	0.384	0.017
Thinking and Behaviour Skills	6	6	5	5	693098	0.095	0.033
Pro-Criminal Attitudes	5	5	4	5	691263	0.101	0.033

*Weighted Medians

This suggests that reduction in severity of need demonstrated above, may happen regardless of IOM support. This is only indicative, and further analysis would be required,

as described above this sub-analysis has considerable limitations, in addition for its focus on the fixed cohort only, and the different samples and time periods – the extent of the overlap between the samples has not been assessed. There are also indications that the sub-analysis of the PSM treatment and control group, with its additional filters added to remove missing OASYS scores, may have reduced the quality of the comparability between the treatment and comparator group. Further, many of the samples were reporting extreme need scores, which limits the utility of comparing medians.

5. PSM Technical Information

Reoffending impact evaluation of the Integrated Offender Management programme (fixed cohort)

Analytical Report

Justice Data Lab

May 2025

5.1 JDL evaluation

Following a feasibility study, which determined that Propensity Score Matching (PSM) would be a suitable methodological approach for assessing the programme's impact, The Justice Data Lab were commissioned by the Probation Research team to conduct an impact evaluation of IOM.

Due to the flexibility of the flex and free cohorts, the commission was limited to an impact evaluation of the fixed cohort only. The Probation Analysis Team provided JDL with data for both the treatment and possible candidates for a comparison group.

Consultation with key stakeholders

This analysis included extensive consultation between JDL and Probation Analysis Team. This included input from and discussions with probation staff involved in the operational delivery of IOM. JDL was also provided with a range of documentation and reports relevant to the programme (including a national operational guidance document and a process evaluation report), to inform the analytical approach.

5.2 Programme eligibility criteria for the fixed cohort

The IOM Refresh introduced greater consistency in selecting individuals for the IOM fixed cohort. According to national guidance, the fixed cohort should consist of individuals who have committed neighbourhood crime offences and are assessed as having a high, very high or prolific risk of reoffending. This assessment is based on the individual's OGRS (Offender Group Reconviction Scale) score and CSS (Crime Severity Score).

The national guidance does not set a specific threshold on these scores, instead encouraging local regions to set their own. However, it offers an example threshold of an OGRS score of 75 and a CSS of 1500. Staff interviewed during the recent process evaluation of IOM frequently referred to this national threshold.

5.3 Key results

Summary

This analysis measures proven reoffences in a one-year period for a treatment group of men with IOM fixed registration dates falling between 1 April 2021 and 30 June 2022, and a comparison group of similar offenders who did not receive IOM. It covers offenders with both custody and community disposals.

The overall analysis does not provide clear evidence on whether support from IOM (for fixed cohort, males only) either increases or decreases:

a) ***the proportion of those who commit a proven reoffence in a one-year period; or***
b) ***the number of proven reoffences in a one-year period***

There may be a number of reasons for this, and it is possible that an analysis of more participants would provide such evidence.

However, ***for those men with a proven reoffence in a one-year period, support from IOM (for fixed cohort) shortens the average time to first proven reoffence for its participants. This is a statistically significant result.***

Sample sizes were not sufficient and matching quality was too poor to include an analogous analysis for females.

Detailed results

Table 5.1 presents the sample sizes for both the treatment group and the comparison group for the male cohort. It includes the unweighted and weighted number of reoffenders in the comparison group, with the weighted numbers being used to calculate the reoffending rate in Table 5.2.

Tables 5.2-4 show the one-year measures for proven reoffending for both the treatment and the comparison group. Rates are expressed as percentages and frequencies expressed per person. The estimated differences shown are the 95% confidence intervals for the differences between the relevant treatment and comparison group measures. Standardised effect sizes are also included, suggesting that any treatment effects observed (should they exist) are generally very small. More detail is provided in the Effect sizes section below.

The one-year reoffending period was counted from the index date: This is defined as the IOM registration date for non-custodial sentences, and the later of IOM registration date and release date for custodial sentences. For the comparison group, a pseudo intervention start date has been created to estimate when an individual would likely have begun treatment.

Table 5.1: Sample sizes for male cohort after matching for one-year reoffending analysis.

Treatment group size	Comparison group size	Reoffenders in treatment group	Reoffenders in comparison group (weighted number[18])
2297	3839	894	1335 (1479)

Table 5.2: Proportion of males who committed a proven general reoffence in a one-year period (reoffending rate) after support from IOM (fixed cohort), compared with a matched comparison group.

Number in treatment group	Number in comparison group	Treatment group rate (%)	Comparison group rate (%)	Estimated difference (% points)	Standardised effect size (Cohen's d)	Statistically significant difference?	p-value
2297	3839	38.9 (36.9 to 40.9)	38.5 (37.0 to 40.1)	0.4 (-2.1 to 2.9)	0.01	No	0.75

---

[18] Each comparison group member is given a weight as part of the PSM process so that the treatment and comparison groups can be readily compared. More detail is included in the Methodology section of this document

Table 5.3: Number of proven general reoffences committed in a one-year period (reoffending frequency) by men who received support from IOM (fixed cohort), compared with a matched comparison group.

Number in treatment group	Number in comparison group	Treatment group frequency	Comparison group frequency	Estimated difference	Standardised effect size (Cohen's d)	Statistically significant difference?	p-value
2297	3839	1.5 (1.4 to 1.7)	1.5 (1.4 to 1.5)	0.08 (-0.08 to 0.24)	0.03	No	0.32

Table 5.4: Average time (days) to first proven general reoffence in a one-year period for men who received support from IOM (fixed cohort), compared with a matched comparison group.

Number in treatment group	Number in comparison group	Treatment group time (days)	Comparison group time (days)	Estimated difference (days)	Standardised effect size (Cohen's d)	Statistically significant difference?	p-value
894	1335	133 (126 to 140)	142 (137 to 148)	-10 (-18 to -1)	-0.09	Yes	0.03

When interpreting the results set out above, please refer to the methodology section of this document for details (including key limitations) of how this evaluation was conducted, and the glossary of terms in the technical appendices

5.4 Methodology

Treatment Window and Period of Interest

Treatment Window: IOM fixed registration dates must fall between 1 April 2021 and 30 June 2022.

Period of Interest: Neighbourhood crime standard index dates must fall between 1 April 2019 and 30 June 2022.

The IOM refresh was officially rolled out to all regions on 1 October 2021. By this date, all regions were required to align their existing IOM cohorts with the new approach, and official performance measures began. However, the data shows that registrations to the fixed cohort started increasing from April 2021 and the operational team confirmed that some regions may have implemented the new approach before the official rollout in October. As a result, it was agreed that the treatment window would span from 1 April 2021 to 30 June 2022. The end of June 2022 was chosen as the cutoff date given that this was the latest point for which proven reoffending data (including index offence and criminal history details) were available at the time the data were being sourced and linked.

An exploration of the treatment group revealed that there could be a substantial 'waiting time' between an individual's standard index date (defined as the disposal date for those serving a community sentence and the release date for those serving a custodial sentence) and their IOM start date. Several factors could contribute to this delay, including

administrative data errors related to the registration date, individuals being moved between different IOM categories (which can result in a new registration date), or reoffending/recalls occurring between their first release date and registration date. Since 96% of index dates were after 1 April 2019, it was agreed that the period of interest would span from April 2019 to June 2022. The June 2022 cutoff was again selected given the availability of reoffending data[19]. The follow-up period for the reoffending analysis has also been tailored to allow for these waiting times (see the section on the starting point for measuring reoffending in this document for more information).

Eligibility criteria applied for analysis

Following in-depth consultation with analytical teams and IOM operational staff, as well as exploration of the treatment group data, the following eligibility criteria were agreed for the treatment group:

1. Must have an IOM fixed registration start date between April 2021 and June 2022.
2. Must have committed an eligible neighbourhood crime index offence.
3. Standard index date for neighbourhood crime offence must be between April 2019 and June 2022.
4. Order must be active at the time of IOM registration date.

Because there is regional flexibility on threshold scores for OGRS and CSS, no initial filters were applied in relation to OGRS or CSS. However, sensitivity analyses were run based on specific threshold scores to identify whether applying such filters would lead to different results, with more details on these sensitivity analyses provided in Appendix 8. The lack of CSS data is a limitation of this study, although efforts have been made to proxy for this[20].

An exploration of the data revealed the offence codes in Table 5.5 were present in the treatment group data. After review, the operational team agreed to use these 18 offence codes as the eligible neighbourhood crime offences.

---

[19] Treatment group records with IOM registration dates before July 2022 and index dates in Q3 or Q4 2022 (where registration was prior to release for custodial sentences) were also added, given that proven reoffending data for this period became available during JDL's analysis. These accounted for 7.5% of the treatment group for the headline analysis. The comparison group did not include records with index dates after Q2 2022 by design, so there is a slight mismatch between the two groups. A sensitivity analysis was run excluding these additional treatment group records and the results were very similar to the headline analysis, so it was decided to retain these records in the results to boost the sample size.

[20] As previously noted, CSS figures cannot be accessed by MoJ analysts. The methodology used to calculate CSS is known, though, and the standard index offence and criminal history variables used by JDL (including previous offence counts in relation to a wide range of offence categories, and the Copas rate, which measures how an individual has built up convictions over time) should provide for a good proxy to CSS, significantly mitigating the risks of not having CSS data. The criminal history variables will also ensure that prolific/persistent offenders are well-matched between the treatment and comparison groups.

Table 5.5: Eligible neighbourhood crime offences

	HO offence code	Description
1	02800	Burglary in a dwelling
2	02801	Burglary (dwelling) with intent to commit, or the commission of an offence triable only on indictment
3	02802	Burglary with violence or the threat of violence
4	02803	Other burglary in a dwelling
5	02900	Aggravated burglary in a dwelling (including attempts)
6	03000	Burglary, other than a dwelling
7	03001	Burglary (non-dwelling) with intent to commit, or the commission of an offence triable only on indictment
8	03002	Other burglary other than a dwelling
9	03100	Aggravated burglary in a building other than a dwelling
10	03400	Robbery and assaults with intent to rob
11	03401	Robbery
12	03402	Assault with intent to rob
13	03900	Stealing from the person of another
14	04500	Stealing from vehicles
15	04510	Stealing from motor vehicles
16	04511	Stealing from other vehicles
17	04800	Theft of motor vehicle
18	04801	Stealing motor vehicle

Analysis split by gender

Due to known differences in their reoffending behaviours[21], the standard JDL approach to reoffending impact evaluations is to separate males and females into two distinct cohorts and run separate PSM analyses for them. Sample sizes for the pre-matched treatment and comparison groups for the female cohort were low, and the PSM models run for females resulted in poor matching quality. Hence, the results presented in this report only relate to males.

Construction of treatment and comparison groups

Summary

The treatment group comprised male individuals who met the eligibility criteria as described above. The comparison group comprised male individuals who met the same eligibility criteria as the treatment group, with the only difference being that they had not been registered to any IOM cohort between 1 April 2021 and 30 June 2022. It was theorised that these individuals would have potentially been eligible for IOM based on their OGRS score and index offence, making them suitable comparison group candidates.

Context and challenges

In theory, everyone who is suitable/eligible for IOM fixed should be assigned to treatment, given that this is a comprehensive programme rolled out across all probation

---

[21] https://publications.parliament.uk/pa/cm5803/cmselect/cmjust/265/report.html

regions/areas. This poses a methodological challenge when constructing the comparison group, given that if the programme has been implemented in line with its design, there should be no-one who is suitable/eligible for treatment who has not been treated. In practice, though, there will be operational reasons why some of those who are suitable/eligible for IOM fixed will not have been registered, and therefore this analysis assumes as a premise that a large enough cohort of such individuals exists as candidates for the comparison group.

The JDL held in-depth conversations with the operational team to determine why these individuals may not have received IOM despite meeting some of the eligibility criteria. The primary reason proposed was the role of professional judgement in the selection process. As outlined in the operational guidance:

> "Once the calculation (OGRS/CSS) has been made, the final assessment should be specific to the effectiveness of the arrangement and therefore the question asked should be 'What can this IOM arrangement add in terms of value to the reducing reoffending process, which is not already statutory via licence conditions/statutory Community Order/Suspended Sentence Order?'"

As part of this consideration, the following factors might also play a role:

- Engagement history – has the individual received IOM in the past and failed to engage?
- Pathway and needs – does the individual have risks/behaviours associated with neighbourhood crime offending or can their offending be attributed to other factors?
- Compliance – is the individuals compliant with orders and is it likely that they'll comply with the increased supervision of IOM?
- Capacity – is there sufficient staff to support them?

Another possible reason for these individuals not being placed on IOM is the role of CSS. Because of the lack of CSS data, it is possible that their scores were not deemed high enough to place them in the fixed cohort.

Therefore, it is possible that the reason the comparison group was not selected for IOM could be due to factors not captured in the eligibility criteria, such as professional judgement or CSS scores. This introduces the risk that the comparison group may differ from the treatment group in ways that affect the outcomes being measured, which could potentially lead to biased estimates of the treatment effect. It is important to be aware of the limitation associated with this feature of how the comparison group was constructed. Steps taken to mitigate any biases that may exist from not including unobserved characteristics are set out in the section below.

Approach

To mitigate the above risks, a wide range of variables have been included in the regression models to help reduce bias from these unobserved factors and to ensure that the characteristics of both the matched treatment and comparison groups are as similar as possible.

As previously noted, variables related to offending history, and the severity of previous offences can serve as proxies for CSS. Additionally, several variables from OASys are included to control for differences in risks and needs, including those related to motivation and engagement with probation/police staff, all of which should be factors considered when applying professional judgment in the selection process.

The treatment group also includes IOM fixed registrations in the period prior to official rollout in October 2021, so the comparison group will include some records from regions where the programme was not in force at their expected treatment start date (see section on pseudo-start dates later in this document for more detail). For such records, the risks of there being a limited comparison group pool available of those suitable/eligible for treatment will be much lower, given the programme had not been fully implemented at this stage.

Starting point for measuring reoffending

The approach used in this analysis measures reoffending from the intervention start date (which is the IOM registration date for this evaluation). This avoids measuring any non-intervention effects during the follow-up period and means that no participants need to be excluded from the treatment group due to delays between release or conviction date and selection on to IOM.

Since individuals in the comparison group do not participate in IOM, they will not have an intervention start date. To address this, a pseudo intervention start date has been calculated, which estimates when an individual would likely have been registered for treatment based on observed pre-treatment characteristics in the data. The pseudo intervention start date can then be used to derive a new index date (the starting point for measuring reoffending from) for the comparison group, which is the later of the pseudo intervention start date and release date (for custody disposals only). This date has also been used to select appropriate OASys records for analysis. Detail on the derivation of pseudo intervention start dates is included in appendix 6.

Controlling for regional variation (spatial confounding)

To account for any regional variation in reoffending behaviour, the Police Force Area (PFA)[22] in which individuals live has been identified for records in the treatment and comparison groups, and the overall reoffending rates and frequencies observed at a PFA level have been included as variables in the regression models used in the PSM analysis[23].

---

[22] There are 43 Police Force Areas (PFA) in England and Wales. Records have been mapped to a PFA based on information taken from the following variables in the Delius database: Probation Delivery Unit, probation region,

[23] The PFA-based reoffending rates and frequencies are included in the geographic data tool published as part of quarterly proven reoffending statistics (https://www.gov.uk/government/collections/proven-reoffending-statistics)

The aim is to ensure that any differences observed between the post-matched groups can be attributed to treatment effects alone, as opposed to differences in reoffending outcomes that may be explained by regional factors.

Propensity Score Matching (PSM)

Offenders in the treatment group were matched to untreated offenders using propensity score matching (PSM). This allows offenders in the treatment group to be matched to non-treated offenders with similar propensity scores.

The propensity score reflects the likelihood that an offender received the intervention, given the recorded characteristics. It is a value between 0 and 1 and is calculated using a logistic regression model used to predict likelihood of treatment. Treatment group members were matched to similar untreated offenders, where their propensity scores were within a certain tolerance level (calliper).

The PSM process involves a combination of matching with replacement and one-to-many. Where multiple comparison group members had propensity scores within the required calliper for a given treatment group member, the comparison group records all received the same weighting factor. For example, if 10 comparison records were matched to a single treatment group record, each comparison group record would have a weight of 1/10 applied, with the treatment group record having a weight of 1. Where treatment group records had no corresponding comparison group record within the tolerance level, they were excluded from the analysis (their weight was set to 0).

In summary, all non-treated offenders whose scores are in a pre-defined proximity to scores of treated offenders will be matched to the treated offenders, even if they have previously been matched to other treated offenders, and these non-treated offenders will form the matched control group. This approach is known as radius matching (with replacement).

In addition, because IOM is delivered to both those with a custody and community disposal, and given the differences between these sentence types, comparison group records with a custody disposal could only be matched to treatment group records with a custody disposal. Similarly, comparison group records with a community disposal could only be matched to treatment group records with a community disposal.

Using the post-matched groups, the weighted reoffending rates for the treatment and comparison groups were compared. PSM can provide a robust quasi-experimental approach, although offenders can only be matched on observable variables. While extensive efforts were undertaken in identifying relevant factors, it is possible that unobserved factors could have affected the results of this research.

Time to first reoffence during the follow-up period

The male reoffending analysis produced a statistically significant difference in the average time to the first reoffence between the two groups, with those reoffenders in the IOM fixed group reoffending on average 10 days earlier than the comparison group. The chart below provides some insight into the timing of first reoffences during the follow-up period,

showing the proportion of reoffenders within each group that have yet to reoffend at any point over the 12-month period.

Figure 5.1: Weighted proportion of post-matched reoffenders who haven't reoffended, by days to first reoffence: male reoffending analysis
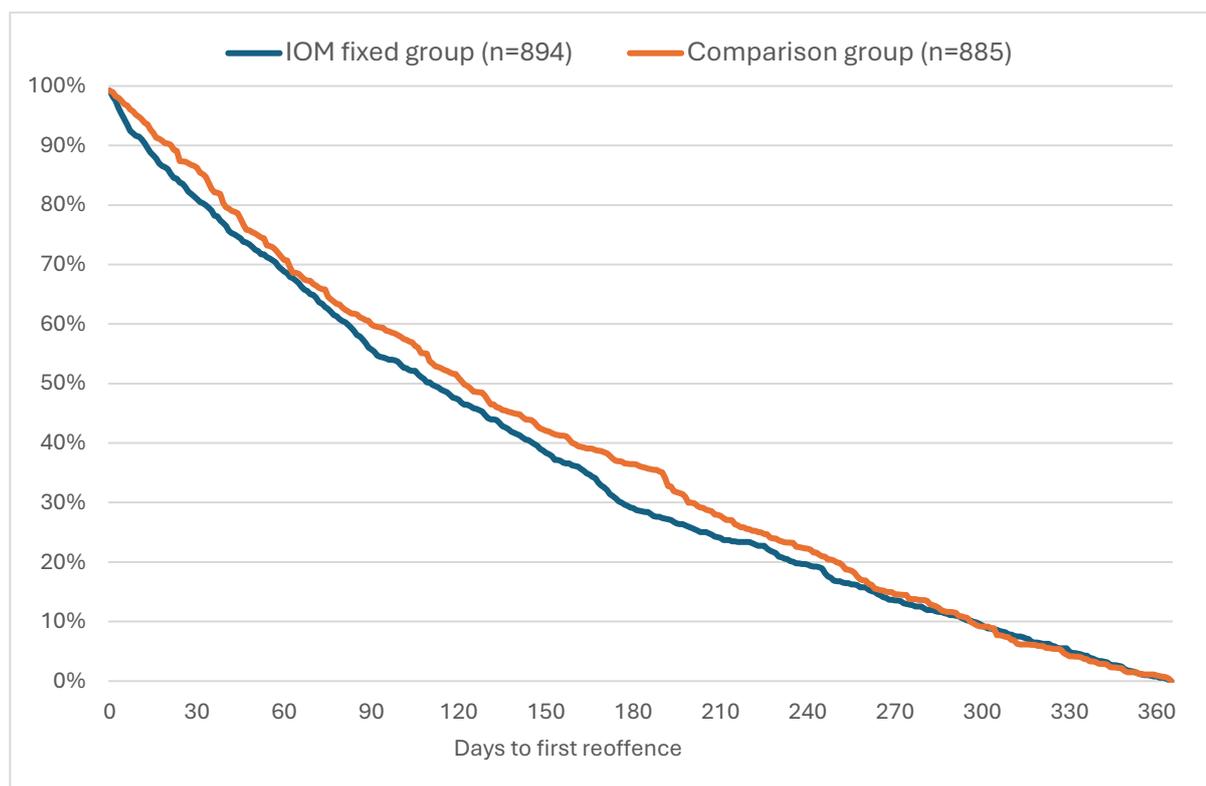
The median time to the first reoffence is 111 days for the treatment group and 122 days for the comparison group. This compares with mean periods to the first reoffence of 133 days and 142 days, respectively (as detailed in Table 5.4 of the Detailed Results).

Recalls and the interaction of recalls and reoffending

The analysis includes both custody and community disposals, with 91% of the treatment group for the male headline analysis having a custody disposal. The custody disposals are also split between those registered onto IOM in relation to their initial release from prison, and those with registrations following a recall (or after multiple recalls) to prison for breach of their licence conditions. Of the custody disposals in the treatment group, 27% have had at least one recall prior to IOM registration date. The split of disposals and mix of recalls prior to IOM registration (or pseudo-registration) is well balanced in both the matched treatment and comparison groups in this analysis.

In relation to the reoffending outcomes measured, the results of this analysis do not explicitly factor in recalls during the follow-up period. This is because proven reoffending, by definition, does not account for recalls. However, to an extent, recalls and reoffending are competing risks, in the sense that a recall event may significantly affect the likelihood

or frequency of a reoffending event, and vice versa. Moreover, the order of these events will affect their relative frequencies.

For example, if someone is recalled to prison during the reoffending follow-up period having breached a licence condition but without having reoffended up to that point, the likelihood of a reoffence during the rest of the follow-up period will change. Any misconduct in prison is dealt with by an adjudications process[24] and, while it is possible for offences occurring in custody to be convicted, the factors that determine whether a proven reoffence takes place in prison are clearly different to those in the community.

While the reoffending rates for the treatment and comparison groups are very similar in this analysis, it is possible that these groups exhibit different recall rates. As a crude assessment of this, the incidence of recalls within the post-matched groups from the male reoffending analysis has been compared, with a higher one-year recall rate observed in the treatment group (treatment rate = 43%, comparison rate = 35%).

Effect sizes

Effect sizes can be considered useful tools when evaluating the strength of a statistical relationship between two groups, or any differences between them. A standardised effect size calibrates the difference between outcomes of the treatment and comparison groups, in terms of the standard deviation. By standardising the measure, the units of measurement are removed, with the aim of making them easier to evaluate and compare (including comparisons with other similar interventions). Cohen's d has been selected as the measure of standardised effect size.

Whether an effect size is categorised as big or small will depend on various factors, including having regard to the type of intervention being evaluated. The effect sizes of all three outcome metrics included in this analysis (as detailed in tables 4.2-4 in the detailed results section of this document) are less than 0.1, indicating that any treatment effects observed (should they exist) are likely to be very small for this programme.

Key limitations

As with all analysis and evaluations, it is important to identify any limitations or caveats that should be attached to the results. The main factors are summarised below.

Unobserved variables

While propensity score matching is a recognised evaluation methodology and can provide a robust quasi-experimental approach, it can only match, and therefore reduce bias, on observed factors (information that is recorded). Despite efforts to include all observed factors known to be predictive of selection onto the IOM fixed cohort programme and of reoffending risk, the importance of information that is not recorded cannot be known and it is possible that unobserved factors could influence these results.

---

[24] https://assets.publishing.service.gov.uk/media/664f4730ae748c43d3794155/adjudications-pf.pdf

Two specific examples in this evaluation relate to the lack of CSS data available, and the use of professional judgment in the selection process. However, as previously described, proxy variables have been included in this analysis to mitigate this risk.

Constructing the comparison group

Given the comprehensive roll-out of the intervention, there were challenges in constructing a suitable comparison group pool of records for this analysis. These are discussed, together with the approach used to mitigate these risks, in the Methodology section of this document.

Relative treatment effects

Any treatment effects observed in this analysis are likely to reflect differences relative to other interventions, services or monitoring in place, as opposed to the absence of any treatment at all. While some variables relating to other 'treatments' have been included and are balanced in the post-matched groups, it has not been possible to collect full details on other programmes or interventions that individuals in either the treatment or comparison group have participated in. This is another example of an unobserved variable risk as described above.

Outcome definition

This evaluation measures a treatment effect using proven reoffending outcomes. As such, the study does not measure treatment effects on crimes that are committed but not recorded by the police or do not lead to a caution or conviction.

Interaction of reoffending and recalls

The results of this analysis do not explicitly factor in recalls to custody during the follow-up period. This is because proven reoffending, by definition, does not account for recalls. This is considered in more detail in the Methodology section above.

Technical Appendices

Appendix 1: Data used – sources, variables, linking and quality

Data sources

The data for this impact evaluation was selected from a large range of criminal justice-related data sources, summarised in the table below:

Data source	Description/Usage	Key variables
Delius (or nDelius)	This is the national probation case management system. The key evaluation data was sourced from the curated Delius databases on the MoJ's analytical platform and provided to JDL. This identified the initial pre-matched cohorts of potential records for both the treatment and comparison groups. In addition, JDL has sourced some extra variables from Delius.	Personal information and IDs used for linking Index offence disposal details Demographic information Probation and licence periods IOM registration details Regional/spatial variables Details on other programmes/services/risk measures/previous IOM registrations OGRS score
Proven Reoffending database	This database is used to produce the quarterly proven reoffending statistics publications. JDL uses information from this database to derive index offence and criminal histories for all records. This database also includes proven reoffending data, which was used to validate the bespoke reoffending measure used for this evaluation	Index offence information Criminal History Demographic information Reoffending data
Police National Computer (PNC)	The MoJ is provided with regular extracts from the Home Office's PNC database, which includes detailed information on offenders' criminal histories.	Personal information and IDs Data used in the calculation of the bespoke proven reoffending metrics [PNC data also includes the source variables used to derive many of the proven reoffending database variables detailed above]
Offender Assessment System (OASys)	OASys assessments include a wide range of information on the criminogenic risks and needs of offenders, together with predictions of reoffending risks	Offender risks and needs in key categories OGRS score

Data Linking and Manipulation

As detailed above, a range of data sources were used to pull in all the required variables for this evaluation, which included linking the data sets together. The key stages were:

1) The key Delius evaluation data (provided separately for the treatment and comparison groups) was merged and some additional Delius variables were added
2) The data was then matched to identify the individuals on the PNC
3) Additional variables were pulled in from the proven reoffending data
4) A pseudo-start date was derived for comparison group records
5) OASys assessment data was linked in
6) The bespoke reoffending outcome metrics were produced
7) Extra variables were derived to control for regional/geo-spatial variation in reoffending
8) The merged data set was then filtered to exclude certain records, and a few additional variables were derived or adjusted so that the final data is ready and formatted correctly for the PSM stage

A detailed list of all variables used in the final evaluation data is provided in appendix 4. The methodology used to select the OASys and OGRS data is described in appendix 5.

Data quality, completeness and sense checks

Offender identifier variables common to multiple data sets were used to link records from the data sources listed above. Consistency and sense checks were carried out on the data, as part of the data linking process. Additionally, records were excluded where no match was found to the data on the PNC. As with all large administrative data sets, some degree of inaccuracy can be expected, but the risk of errors is mitigated as part of the linking process and the checks as described above. More detail on records removed is included in appendix 2 below, showing how the treatment and comparison group were constructed. Analytical decisions in relation to the data are also set out in the Methodology section of this document.

Appendix 2a: Treatment group flow diagram (chevron)

4416 records submitted (4250 male, 166 female) with index dates from April 2019 to December 2022

4416

27 records (1%) excluded as they could not be identified on the Police National Computer (PNC), or were aged under 18 at the index date

4389

447 records (10%) excluded records excluded because the index offence could not be identified in the proven reoffending data

3942

176 records (4%) excluded because the index offence in the proven reoffending data did not correspond with a neighbourhood crime

3766

150 records (3%) excluded because the probation period ends outside the expected timeframes

3616

For those with multiple records for the same disposal, keep the first IOM registration and exclude any subsequent participations (255 records (6%) excluded)

3361

85 records (2%) excluded with previous sexual offences[25]

3276

125 records (3%) excluded for females

3151

11 records (0%) excluded with missing OGRS data

3140

---

[25] Individuals with previous sexual offences are excluded from JDL analyses as standard practice, on the basis that these offenders have different criminogenic risks and needs, and reoffending patterns.

843 records (19%) excluded who were not matched at the PSM stage

2297

2297 records matched at PSM stage (54% of
the original male records submitted)

Appendix 2b: Comparison group flow diagram (chevron)

21213 records submitted (19134 male, 2079 female) with index dates from April 2019 to June 2022

21213

387 records (2%) excluded as they could not be identified on the Police National Computer (PNC), or were aged under 18 at the index date

20826

2895 records (14%) excluded records excluded because the index offence could not be identified in the proven reoffending data

17931

1622 records (8%) excluded because the index offence in the proven reoffending data did not correspond with a neighbourhood crime

16309

4907 records (23%) excluded because the probation period ends outside the expected timeframes

11402

6 records (0%) excluded because the disposal type isn't common to any of the treatment group records

1139

285 records (1%) excluded with previous sexual offences

1111

1090 records (5%) excluded for females

1002

507 records (2%) excluded with missing OGRS data

9514

5675 records (27%) excluded who were not matched at the PSM stage

3839

3839 records matched at PSM stage

Appendix 3: Profile of the treatment group

Of the 4,416 treatment group records submitted for analysis, 4,411 (99.9%) were identified on the PNC. Of these, 2,297 were included in the final post-matched group for the male reoffending analysis (see appendix 2a for details of the various stages where records were excluded). To assess whether these 2,297 records are representative of the original cohort submitted for analysis, the table below includes basic information for the 4,411 records identified, split into three different groups: males in the post-matched group; males not included in the post-matched group; and females.

	Male participants included in analysis (2,297 records)	Male participants not included in analysis (1,948 records)	Female participants (166 records)
Age			
Average age at standard index date	34.9	34.9	35.5
Average age at first contact with the criminal justice system	14.1	13.9	16.0
Ethnicity (as recorded by police officer)			
White	85%	86%	88%
Black	10%	10%	9%
Asian	3%	2%	2%
Other	0%	0%	0%
Unknown ethnicity	2%	2%	1%
Nationality			
UK nationality	98%	99%	98%
Non-UK nationality	2%	1%	2%
Unknown nationality	0%	0%	1%
Disposal type (sentence)			
Custody	91%	89%	89%
Community (Community order)	4%	5%	6%
Community (Suspended Sentence Order)	5%	5%	5%
Community (Youth Rehabilitation Order)	0%	1%	1%
Recall incidence (custody index disposals only)			
Proportion with a recall between standard index date and IOM registration date	30%	28%	31%
Proportion with no recall between standard index date and IOM registration date	70%	72%	69%

While there is a small amount of variation, the profile of these variables is generally similar across all three groups.

Appendix 4: Variables for propensity score models

Variable	Type	Categories/Notes
Demographics		
Ethnicity (as recorded by police officer)	Categorical	White; Black; Asian; Other; Unknown
Nationality	Categorical	UK; Non-UK; Unknown
Age at standard index date	Continuous (integer)	-
Index offence information		
Index Disposal/Sentence type	Categorical	Custody; Community Order; Suspended Sentence Order
Index offence category	Categorical	Robbery; Aggravated Burglary; Burglary+; Burglary (dwelling); Burglary (non-dwelling); Stealing
Index offence severity	Categorical	Indictable only; Triable either way; Summary only
Index offence custodial sentence length (only applies to custody disposals)	Categorical	Less than or equal to 6 months; More than 6 months to less than 12 months; 12 months to less than 4 years; 4 years to 10 years; More than 10 years; Mandatory Life Sentence; Other Life Sentence; Imprisonment for Public Protection
Reoffending cohort year (year of standard index date)	Categorical	-
Criminal history		
Age at first contact with the Criminal Justice System	Continuous (integer)	-
Number of previous prison events	Continuous (integer)	-
Number of previous convictions	Continuous (integer)	-
Number of previous court orders	Continuous (integer)	-
Number of previous offences	Continuous (integer)	-
Number of previous indictable only offences	Continuous (integer)	-
Number of previous triable either way offences	Continuous (integer)	-
Number of previous summary offences	Continuous (integer)	-
Number of previous violent offences	Continuous (integer)	-
Number of previous robbery offences	Continuous (integer)	-
Number of previous public order offences	Continuous (integer)	-
Number of previous domestic burglary offences	Continuous (integer)	-
Number of previous other burglary offences	Continuous (integer)	-
Number of previous theft offences	Continuous (integer)	-
Number of previous handling offences	Continuous (integer)	-

Number of previous fraud or forgery offences	Continuous (integer)	-
Number of previous theft of vehicles offences	Continuous (integer)	-
Number of previous drink driving offences	Continuous (integer)	-
Number of previous criminal damage offences	Continuous (integer)	-
Number of previous drug import/export/production/supply offences	Continuous (integer)	-
Number of previous drug possession or supply offences	Continuous (integer)	-
Number of previous sexual offences	Continuous (integer)	-
Number of previous breach offences	Continuous (integer)	-
Copas rate (logarithmic rate of convictions and cautions over time)	Continuous (number)	-
Delius data (various information)		
Community sentence length (community disposals only)	Categorical	Less than or equal to 12 months; More than 12 months to less than or equal to 18 months; 19 months to less than or equal to 2 years; More than 2 years
Any type of electronic monitoring conditions in force at the standard index date	Categorical (binary)	Yes; No
Any multi-agency public protection arrangements (MAPPA) at the standard index date	Categorical (binary)	Yes; No
Latest OGRS3 score (2yr) prior to intervention start date (source: Delius)	Categorical	1-24; 25-49; 50-64; 65-74; 75-84; 85+; Unknown
Most recent Risk of Serious Harm categorisation at standard index date	Categorical	Unknown; Low; Medium; High; Very High
The number of previous IOM registrations prior to standard index date (excluding the current registration for the treatment group)	Continuous (number)	-
The number of recalls to custody prior to the intervention start date (or pseudo- start date)	Continuous (number)	-
The waiting period from the standard index date to the intervention start date (or pseudo- start date)	Continuous (number, in days)	
Remaining probation period at the intervention start date (or pseudo- start date)	Continuous (number, in days)	
Variables used to control for regional variation (at a Police Force Area level)		
Police Force Area 5-yr average reoffending rate (banded)	Categorical	Unknown; Low; Low-Mid, Mid-High; High

Police Force Area 5-yr average reoffending frequency per offender (banded)	Categorical	Unknown; Low; Low-Mid; Mid-High; High
OASys		
OASys assessment captured in data	Categorical (binary)	Yes; No
Latest OGRS3 score (2yr) prior to intervention start date (source: OASys)	Categorical	1-24; 25-49; 50-64; 65-74; 75-84; 85+; Unknown
Does the offender have problems with being in no fixed abode or in transient accommodation?	Categorical	Unknown; No; Some; Significant
Does the offender currently have problems with the suitability of accommodation?	Categorical	Unknown; No; Some; Significant
Does the offender currently have problems with a permanent place of accommodation?	Categorical	Unknown; No; Some; Significant
Does the offender currently have problems with the suitability of the location of accommodation?	Categorical	Unknown; No; Some; Significant
Does the offender have accommodation issues that are linked to their risk of serious harm, risks to the individual, and other risks?	Categorical	Unknown; Yes; No
Does the offender have accommodation issues that are linked to their offending behaviour?	Categorical	Unknown; Yes; No
Does the offender have problems with being unemployed or being unemployed upon release?	Categorical	Unknown; No; Some; Significant
Does the offender have problems with either reading, writing, or numeracy?	Categorical	Unknown; No; Some; Significant
Does the offender have problems with work-related skills?	Categorical	Unknown; No; Some; Significant
Does the offender have issues with attitudes to employment?	Categorical	Unknown; No; Some; Significant
Does the offender have employment and/or education issues that are linked to their risks of serious harm, risks to the individual, and other risks?	Categorical	Unknown; Yes; No
Does the offender have employment and/or education issues that are linked to their offending behaviour?	Categorical	Unknown; Yes; No
Does the offender currently have any problems with their financial situation?	Categorical	Unknown; No; Some; Significant
Does the offender have issues with illegal earnings as a source of income?	Categorical	Unknown; No; Some; Significant

Does the offender have financial management issues that are linked to their risks of serious harm, risks to the individual, and other risks?	Categorical	Unknown; Yes; No
Does the offender have financial management issues that are linked to their offending behaviour?	Categorical	Unknown; Yes; No
Does the offender currently have problems having a relationship with their close family members?	Categorical	Unknown; No; Some; Significant
Does the offender have issues with childhood experiences?	Categorical	Unknown; No; Some; Significant
Does the offender have relationship issues that are linked to their risks of serious harm, risks to the individual, and other risks?	Categorical	Unknown; Yes; No
Does the offender have relationship issues that are linked to their offending behaviour?	Categorical	Unknown; Yes; No
Does the offender have lifestyle and associates issues that are linked to their offending behaviour?	Categorical	Unknown; Yes; No
Does the offender have lifestyle and associates issues that are linked to their risks of serious harm, risks to the individual, and other risks?	Categorical	Unknown; Yes; No
Has the offender ever misused drugs?	Categorical	Unknown; Yes; No
Does the offender have issues with the level of drug use?	Categorical	Unknown; Yes; No
Does the offender have issues with motivation to tackle drug misuse?	Categorical	Unknown; No; Some; Significant
Does the offender have issues with drug use and obtaining drugs being a major activity?	Categorical	Unknown; No; Some; Significant
Does the offender have issues with drug misuse that are linked to their risk of serious harm, risks to the individual, and other risks?	Categorical	Unknown; Yes; No
Does the offender have issues with drug misuse that are linked to their offending behaviour	Categorical	Unknown; Yes; No
Does the offender currently have problems with alcohol misuse?	Categorical	Unknown; No; Some; Significant
Does the offender have issues with motivation to tackle alcohol misuse?	Categorical	Unknown; No; Some; Significant
Does the offender have issues with alcohol misuse that are linked to their risk of serious harm, risks to the individual, and other risks?	Categorical	Unknown; Yes; No

Does the offender have issues with alcohol misuse that are linked to their offending behaviour	Categorical	Unknown; Yes; No
Does the offender have psychological problems, including depression?	Categorical	Unknown; No; Some; Significant
Does the offender have issues with wellbeing and mental health that are linked to their risk of serious harm, risks to the individual, and other risks?	Categorical	Unknown; Yes; No
Does the offender have issues with wellbeing and mental health that are linked to their offending behaviour?	Categorical	Unknown; Yes; No
What level of interpersonal skills problems does the offender possess?	Categorical	Unknown; None; Some; Significant
Does the offender have issues with their thinking and behaviour that are linked to their risk of serious harm, risks to the individual, and other risks?	Categorical	Unknown; Yes; No
Does the offender have issues with their thinking and behaviour that are linked to their offending behaviour?	Categorical	Unknown; Yes; No
Does the offender have problems with pro-criminal attitudes?	Categorical	Unknown; No; Some; Significant
Does the offender have problems with attitudes toward staff?	Categorical	Unknown; No; Some; Significant
Does the offender have problems with attitudes towards supervision/licence?	Categorical	Unknown; No; Some; Significant
Does the offender have problems with attitudes towards community/society?	Categorical	Unknown; No; Some; Significant
Does the offender have problems with their motivation to address their offending?	Categorical	Unknown; No; Some; Significant
Does the offender have attitudes that are linked to their risks of serious harm, risks to the individual, and other risks?	Categorical	Unknown; Yes; No
Does the offender have attitude issues that are linked to their offending behaviour?	Categorical	Unknown; Yes; No

Appendix 5: Selecting OASys assessments, variables and OGRS scores

The Offender Assessment System (OASys) is a key data source, which provides standardised assessments of offenders' risks and needs, together with actuarial methods of predicting reoffending risk. OASys assessments for offenders can happen at various points during their interaction with the criminal justice system and, for the purposes of this PSM analysis, the aim is to identify the most recent and complete assessment prior to treatment (so that it best reflects offenders' risks and needs in both the treatment and comparison groups at this point).

The Offender Group Reconviction Scale (OGRS) predicts the likelihood of committing any offence leading to reconviction (proven reoffending) over either one or two years. This is based on static factors such as age, gender and criminal history. The key metric used by probation practitioners is the two-year OGRS3 prediction, measuring the probability of a proven reoffence within two-years of conviction for a community sentence, or at discharge from custody for custodial sentences. Any mention of OGRS or OGRS3 scores in this report and the accompanying tables relate to this two-year probability. OGRS scores are produced automatically as part of an OASys assessment.

The OASys assessment chosen for this evaluation was the most complete assessment in the period from six months prior to conviction date up to one month after IOM registration date (or the pseudo-start date for the comparison group). The most complete assessment was also the most recent assessment in 92% of records where an assessment was found.

For the post-matched groups in the male headline analysis, 93% have an OASys assessment included and, of these, 87% had an assessment date no more than 12 months prior to the IOM registration date/pseudo-start date, and 96% were no more than two years prior to the start date. Those 7% of records without an OASys assessment were retained in the headline male analysis, but a sensitivity analysis was also run with these records excluded. The results of this sensitivity were very similar to the headline analysis (see appendix 8 for more detail).

Given the importance of OGRS scores in the context of selection for the IOM fixed programme, the following approach was used to identify OGRS data:

Two separate OGRS scores were used in the regression models, one derived directly from OASys assessment data, and one taken from Delius. These scores were obtained as follows:

- OGRS3 score sourced from OASys data: The most recent non-blank OGRS3 score in the period up to one month after the IOM start date. For the post-matched groups in the male headline analysis, 97% had a non-blank score for this variable. Of these, 76% have scores captured within 12 months prior to IOM start date, and 87% were no more than two years prior to IOM start date
- OGRS3 score sourced from Delius data: The most recent non-blank OGRS3 score in the period up to the later of IOM start date and standard index date: For the post-matched groups in the male headline analysis, 96% had a non-blank score for this

variable. Of these, 41% have scores captured within 12 months prior to IOM start date, and 64% were no more than two years prior to IOM start date.

For probation practitioners involved in the selection process onto IOM fixed, it is not clear what OGRS scores will have been available to them when making decisions. The following tables summarise how the OGRS scores taken from different sources compare for the post-matched groups in the male headline analysis (where both data sources have non-blank scores):

Post-matched:	Treatment group		Comparison group	
Distribution of OGRS score	Delius-sourced	OASys-sourced	Delius-sourced	OASys-sourced
mean	75.5	71.8	76.0	72.5
1st quartile	69	64	69	64
median	78	75	79	76
3rd quartile	85	82	85	83

% of OGRS scores taken from both sources where:	Treatment group	Comparison group
the scores match	35.7%	36.3%
the dates match	24.4%	24.6%
both the scores and dates match	24.2%	24.5%
The Delius date is earlier than the OASys date	64.4%	63.7%
The OASys date is earlier than the Delius date	12.2%	11.6%

In summary:

1) The Delius-sourced data is on average at an earlier date than the OASys-sourced data (the median difference is approximately 7 months, the mean difference is approximately 13 months)
2) The Delius-sourced data has on average a higher score (approximately 3-4 points higher) than the OASys-sourced data
3) Scores from both sources are identical in 36% of records
4) Where both sources capture OGRS scores at the same date (in about one-quarter of records), the scores match in almost all cases (so they are consistent, in this sense)
5) Based on spot checks undertaken for records where the dates did not match from the two sources, the date and score taken from the Delius data was also found in the wider OASys assessment data (it's just that a different assessment was chosen for the OASys-derived data used). As such, no inconsistencies were identified.

Given the differences identified and the uncertainty in relation to the timing of the OGRS score used to assess eligibility, it was decided to add both the OASys-derived and Delius-derived OGRS scores into the regression modelling. Sensitivity analyses were also run to

test the impact of which data source is used for OGRS scores, with results obtained for all combinations being similar to the headline model (see appendix 8 for details).

The list of variables selected from the OASys assessment is included in appendix 4.

Appendix 6: Derivation of pseudo intervention start dates

Since individuals in the comparison group do not participate in IOM, they will not have an intervention start date. To address this, a pseudo intervention start date has been calculated, which estimates when an individual would likely have been registered for treatment based on observed pre-treatment characteristics in the data. This has been done using imputation, a statistical technique where individual sentencing and demographic information is analysed to estimate this pseudo start date for comparison group records, using the treatment group (where the intervention start date is known) as training data.

The key variables in the imputation process are set out below:

Variable	Missing variable in comparison group?
Gender	
Disposal type (custody or community)	
Period from conviction date to index date	
Period from conviction date to end of probation period	
Year of conviction date	
Year of index date	
Age at index date	
Period from index date to end of probation period	
Period from conviction date to start of period of interest (01/04/2021)	
Period from index date to start of period of interest (01/04/2021)	
Period from index date to end of period of interest (30/06/2022)	
Period from earliest possible start date to latest possible start date (given the period of interest, the conviction date, and the end of the probation period)	
Period from conviction date to IOM registration date	YES
Waiting proportion (the period from the earliest possible start date to the IOM registration date, divided by the period from earliest possible start date to latest possible start date)	YES

The imputed period from conviction date to IOM registration date and waiting proportion were calculated first and then used to calculate the pseudo intervention start date.

An alternative model was tested which, in addition to those variables set out above, included a range of criminal history and index offence data. However, the extra variables in this model were adding *noise*[26] to the estimation process and hence these were dropped in favour of the imputation approach using fewer variables.

Appendix 7: Common support

As part of the validation of the analyses, checks were performed to ensure there was a sufficient level of common support between the post-matched groups. The propensity scores represent the probability of being assigned to treatment given an offender's

---

[26] The approach was validated by a range of sense checks, including estimating the intervention start dates for the treatment group and comparing them with the actual intervention start dates. To do this, a test data set was used, including two copies of the treatment data (one with the intervention start dates included and one without) and the imputation process estimated the dates for those records with missing values. The estimates were found to be more accurate for this test data set when using the model with fewer variables.

recorded characteristics. There was a large region of common support (where the propensity scores for the treatment and comparison groups overlap), which implies they can be matched. After matching, the distributions of propensity scores in the two groups were closely aligned, showing that the likelihood of receiving treatment as predicted by the PSM model is well matched across the groups.

Appendix 8: Sensitivity analyses

The principal aim of the sensitivity analyses performed in this evaluation is to test whether changing some of the assumptions or methods used will have a material impact on the key (headline) results. When assessing the results of these sensitivity analyses, more weight should be attached to those with a better matching quality (namely those with lower mean absolute standardised differences (MASD)[27]). In summary, the results of all sensitivity analyses were similar to those of the headline analysis. The key findings are detailed below.

For the proportion of proven reoffences (reoffending rate) during the 12-month follow-up period:

- None of the headline or sensitivity analyses resulted in a statistically significant difference between the groups, with high p-values observed throughout
- The central estimate of the difference between the groups was generally very small, being less than 0.5 percentage points in all 7 analyses with a MASD of less than 2%. In contrast with the headline analysis, the estimate of the reoffending rate was slightly higher in the comparison group than the treatment group in 4 of these 7 analyses, suggesting that if any treatment effect does exist then it is likely to be very small and could go in either direction.

For the number of proven reoffences (reoffending frequency) during the follow-up period:

- None of the analyses resulted in a statistically significant difference between the groups, with high p-values observed throughout
- The difference between the groups was generally very small, being less than 0.1 reoffences per offender in all 7 of the analyses with a MASD of less than 2% (all of which resulted in a slightly higher frequency in the treatment group)

For the average days to first proven reoffence among reoffenders during the follow-up period:

- 10 of the 13 analyses resulted in a statistically significant difference between the two groups (the highest p-value among the 3 non-significant results was 0.13). The baseline result was statistically significant for this outcome with a p-value of 0.03[28].

---

[27] The MASD is the average of the absolute values of the standardised difference in means between the treatment and comparison groups for all variables in the final logistic regression model used in the PSM process. The lower the MASD, the better the matching quality of the model.

[28] The standard JDL approach uses a significance level of 5%, with any p-values below 0.05 interpreted as being statistically significant. A p-value of 0.03 means that there is a 3% risk of the observed difference being calculated by chance where there is actually no underlying difference.

- The average time to the first reoffence was shorter in the treatment group in all 13 analyses (including the headline analysis, and also for those with non-significant results).
- While most of the sensitivities produced statistically significant differences in line with the headline analysis, these analyses provide some context around the uncertainty associated with the findings for this outcome, while noting that the headline approach selected included the most comprehensive set of variables and resulted in the best matching across all the sensitivities.

The table below provides more detail on the sensitivity analyses performed.

| Sensitivity	Rationale for sensitivity	Findings
1: Headline analysis (as per the key results presented in this report) = baseline	The chosen (baseline) approach for this analysis is displayed here for comparison against other sensitivities. Radius matching (with replacement) was applied.	119 variables were included in the final model with a mean absolute standardised difference (MASD) of 1.75%.
2: No matching conditions (other than having sufficiently close propensity scores)	The baseline analysis only allowed treatment group units to be matched with comparison group units if their sentence type (custody or community) was the same. This condition was dropped to see if the matching quality improved.	Matching quality (MASD = 1.77%) and results were similar to baseline, although there was no statistically significant difference in the average days to first reoffence outcome (p=11%)
3: Additional matching condition: presence of OASys assessment	Only allow treatment group units to be matched with comparison group units if both their sentence type and OASys inclusion indicator (i.e., whether OASys assessment data is included) are the same.	Matching quality was worse MASD (1.94%) but the results were similar to the baseline.
4: Exclude records with any missing OGRS data	The baseline analysis only excluded records if no OGRS data was present in either data source (OASys or Delius). This sensitivity required both sources to have non-blank OGRS data.	Very similar results to the baseline (MASD = 1.74%)
5: Exclude records with any missing OGRS data and use continuous variable	As per sensitivity 4 but also include OGRS scores as continuous (numeric) variables in the regression model. This contrasts with the baseline approach of using banded categorical variables for OGRS scores.	Matching quality worsened (MASD = 2.43%), including the OGRS scores not being well matched, so this approach was discounted (the results were broadly similar to the headline)
6: Use only the OASys-sourced OGRS score (as a continuous variable)	The Delius-sourced OGRS score was ignored and only those with a non-blank OASys-sourced OGRS score were retained (with the OGRS score included as a continuous variable in the regression model).	Very similar results to baseline, but worse matching quality (MASD = 2.31%).
7: Use only the Delius-sourced OGRS score (as a	The OASys-sourced OGRS score was ignored and only those with a non-blank Delius-sourced OGRS score were	Similar results to baseline, although matching quality is worse (MASD = 2.02%) and

continuous variable)	retained (with the OGRS score included as a continuous variable in the regression model).	there was no statistically significant difference in the average days to first reoffence outcome (p=13%)
8: At least one OGRS score ≥75 and previous offences ≥ 16	The OGRS score and number of previous offences are important factors when considering eligibility/suitability for IOM fixed. This sensitivity checks to see how including a filter on these metrics impacts the overall results. In this case, at least one of the OGRS scores (either sourced from OASys or Delius) must be ≥75, and the offender must have at least 16 previous offences (as a proxy for prolific offending).	As expected (given OGRS and previous offence thresholds), the reoffending rates in both groups are higher than the baseline, but the differences between the two groups are very similar. The matching quality is worse, though (MASD = 2.38%).
9: Both OGRS scores ≥75 and previous offences ≥ 16	As per sensitivity 8, but in this case both the OGRS scores (sourced from OASys and Delius) must be ≥75.	As expected (given OGRS and previous offence thresholds), reoffending rates in both groups are higher than baseline, but differences between two groups are very similar. The matching quality is worse, though (MASD = 3.29%).
10: Exclude treatment group records with index dates between 01/07/22 and 31/12/2022	During the course of the analysis, more reoffending data became available for Q3 and Q4 2022 index dates, which was used for the treatment group in the headline analysis. However, owing to time constraints, the comparison group data was not adjusted to span these index dates. This sensitivity is included to test that the addition of these records does not materially impact/bias the results	Very similar results to baseline, although matching quality is slightly worse (MASD = 1.83%).
11: Addition of squared terms in regression models	It is possible that the variables in the regression model affect the likelihood of treatment in a non-linear way. For all continuous (numeric) variables in the regression model, a squared term was included in this sensitivity (in addition to the linear term) to test if a quadratic relationship exists.	Very similar results to baseline, although matching quality is slightly worse (MASD = 1.78%).
12: Exclude all records without OASys assessment data	Approximately 7% of records in the baseline analysis did not have OASys assessment data included. These records were removed from this sensitivity.	Very similar results to baseline, although matching quality is worse (MASD = 2.02%).
13: Exclusion of OASys variables in the model	To test whether a simplified regression model (with fewer terms) without OASys variables included affects the evaluation results. Note that OGRS scores were retained in this sensitivity.	Similar results to baseline, although matching quality is worse (MASD = 1.87%) and there was no statistically significant difference in the average days to first reoffence outcome (p=12%)

Appendix 9: Breakdown of reoffences in the post-matched groups

One of the additional outputs produced by JDL is a breakdown of the offence types for those with proven reoffences in the post-matched groups. This provides the following information:

- the input data used for an evaluation of the economic impact of the programme (out of the scope of JDL's analysis)
- insight into any impacts of the programme on the types of proven reoffences
- insight into the types of reoffences that are common to the male IOM fixed cohort
- a sense check that there is no evidence of sizeable differences in the overall distribution of reoffences between the post-matched groups, which could suggest that the PSM process may not have adequately balanced the post-matched groups

Figure 5.2: Distribution of the 10 most common reoffences in the post-matched groups (male reoffending analysis)
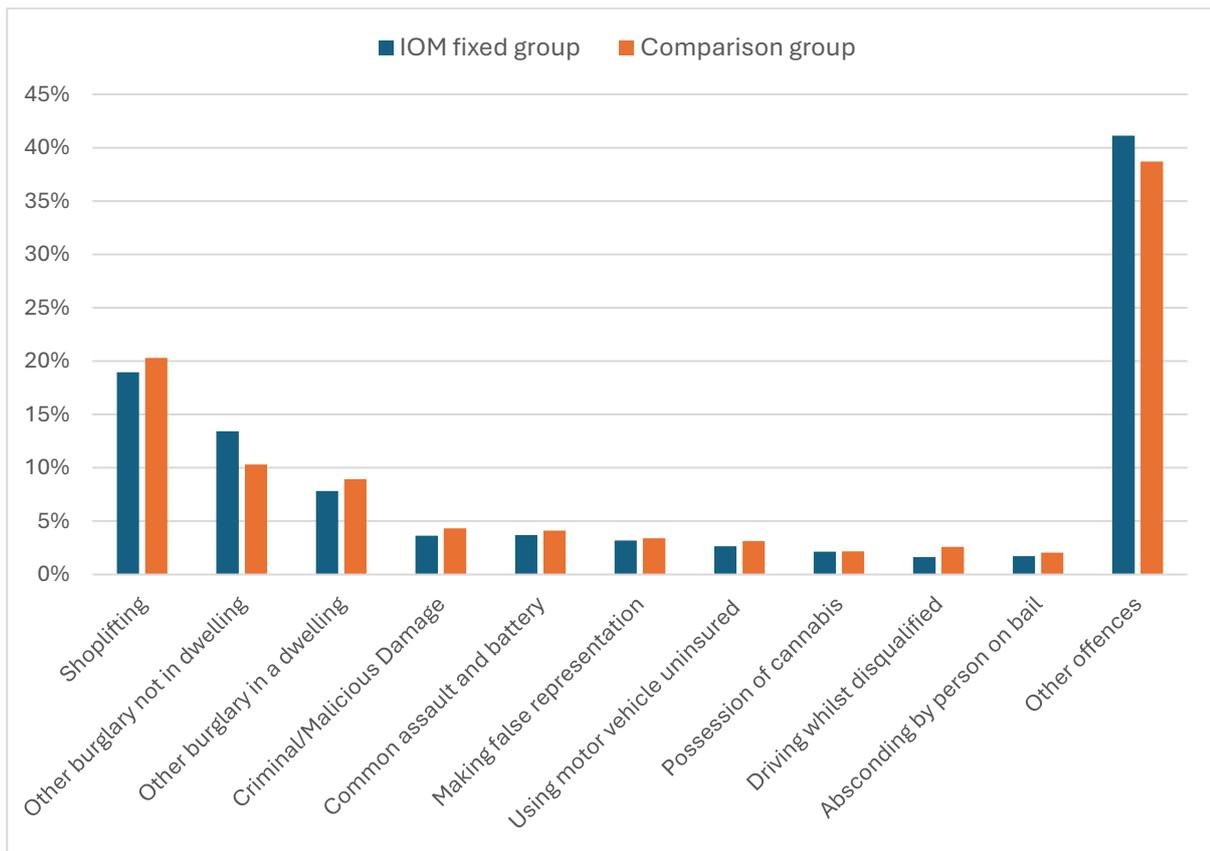
The chart shows that there is a similar distribution of reoffences in the post-matched groups, with 'shoplifting' being the most common reoffence among both groups.

Glossary of terms

Competing risks: Multiple potential outcomes that may occur simultaneously, where the occurrence of one outcome may preclude the occurrence of others.

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

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

Follow-up period: The duration of time over which individuals are observed or monitored for outcomes after receiving an intervention.

Frequency of reoffences: The number of proven reoffences committed per person within the group.

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

Index date: The date at which the follow-up period for measuring reoffending begins. For this evaluation, the index date is the intervention start date. See also *standard index date.*

Index offence: The primary offence for which the offender was convicted, specifically the index sentence (which is then linked to the intervention being evaluated for these offenders)

Intervention start date: The date on which an individual starts receiving support from IOM

Licence: Permission granted by authorities for an offender to be released from custody under specific conditions.

Logistic regression: A technique used to predict a binary, categorical outcome; for the Justice Data Lab this will mainly be used to ascertain the likelihood of an offender receiving treatment or not. Predictions are based on the variables used in the regression.

nDelius: The probation case management system used for managing offenders and their risk assessments.

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

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

One year proven reoffending rate: The MoJ defines this as the proportion of offenders in a cohort who commit an offence in a one-year follow-up period and were disposed of within 18 months from the start of this period, having received a court conviction, caution, reprimand or warning.

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

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

Propensity Score Matching (PSM): The methodology used for constructing a matched control group in Justice Data Lab analyses. Uses logistic regression to predict the likelihood of each offender receiving treatment; these predicted probabilities called

propensity scores. Treated and non-treated offenders are matched based on the closeness of their propensity scores.

Propensity scores: Scores representing the likelihood of individuals receiving a treatment or intervention in observational studies.

Proven reoffending: Occurs when an individual commits an offence that leads to a court conviction, caution, reprimand, or warning within a specified follow-up period after being released from custody, receiving a non-custodial conviction, or starting a court order.

Pseudo intervention start date: A hypothetical or simulated start date used for analytical purposes, often in studies where actual start dates are unavailable or inconsistent.

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

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

Standard index date: The standard starting point for measuring reoffending in JDL evaluation. This is the earliest point an individual is at risk of reoffending: either the release date for those serving custodial sentences or the conviction date for those serving non-custodial sentences. Note that for this evaluation a different index date has been used (see *index date* above)

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

Statistically significant difference: A difference between groups or conditions that is unlikely to have occurred by chance alone, based on statistical analysis and typically indicated by a p-value below a predetermined threshold.

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

Time to first reoffence: The average number of days from the start of the follow-up period (index date) to the date of the first proven reoffence, calculated only for those who reoffend.

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

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

Unobserved variable: A variable that influences the outcome of interest but is not directly measured or accounted for in the analysis.

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