Review of compliance with the Code of Practice for Statistics

Justice Data Lab statistics

produced by the Ministry of Justice

December 2025

The Office for Statistics Regulation

Statistics should serve everyone, helping enhance knowledge about every section of society and the economy, and people's place within them.

The Office for Statistics Regulation (OSR) provides independent regulation of official statistics produced in the UK. Official statistics are statistics produced by Crown bodies and other organisations listed within an Official Statistics Order, on behalf of the UK Government or the devolved governments.

We aim to enhance public confidence in statistics produced by government by setting the standards that they must meet in the Code of Practice for Statistics. We ensure that producers uphold these standards by conducting reviews of statistics against the Code. We also report publicly on system-wide issues and on the way statistics are being used, celebrating when the standards are upheld and challenging publicly when they are not.

Compliance review

A compliance review is a review of any official statistics (or accredited official statistics) to establish the extent to which they meet the standards of Trustworthiness, Quality and Value in the Code of Practice for Statistics. We might undertake a review with a view to accrediting a set of statistics (this type of compliance review is called an assessment), focus on a specific issue or involve a high-level check of ongoing compliance.

A range of factors determine the scope of and approach to a compliance review, including the type of statistics, their accreditation status, the producer history, the profile of the statistics and the reasons for initiating the review.

Contents

Overview

At the time of this report, the Ministry of Justice (MoJ)'s Justice Data Lab (JDL) statistics are published as official statistics.

JDL is an analytical team within MoJ that provides free analysis to organisations offering interventions to reduce reoffending in England and Wales, to help them understand their impact. Each JDL report presents the analysis results for a specific intervention programme. JDL's work also contributes to MoJ's evidence base on reducing reoffending.

This review examined the extent to which the JDL statistics meet the standards of Trustworthiness, Quality and Value in the [Code of Practice for Statistics](#) (the Code), with a focus on statistical methods. It aims to make recommendations for improving the statistics.

Why we did this review

Between December 2024 and February 2025, a [concern was raised with us](#) about the statistical methods that MoJ uses to produce the JDL statistics, specifically its approach to statistical significance testing when comparing the difference in headline reoffending measures between the intervention group (the treatment group) and the comparison group.

We investigated some aspects of MoJ's statistical methods at the time. Although we were broadly satisfied with MoJ's approach to statistical significance testing, we committed to carrying out a review of the statistics' compliance with the standards of the Code. Accordingly, this review has considered, in more detail, MoJ's approach to statistical significance testing, as well as aspects of the data collection, the aims of the analyses and the presentation of the statistics.

Highlighted findings

- We are assured about MoJ's approach to statistical significance testing. The JDL team has sought expert advice on the appropriateness of applying multiple comparison correction (a technique used to mitigate the increased risk of false positive when carrying out many significance tests on the same dataset) and all measures tested for statistical significance are specified in advance of the tests being carried out.
- MoJ uses a consistent and rigorous methodology for JDL analyses, and peer review has helped MoJ refine JDL processes. The published information on JDL's methodology is comprehensive. To improve the accessibility of methods information, MoJ should consolidate existing information into a central document that covers all stages of the JDL process from start to finish.

- JDL reports contain clear and helpful advice for users on interpreting the statistics. Overall, uncertainty in the statistics and caveats and limitations are communicated well. To aid user interpretation, it would be helpful if more detail on key caveats and limitations was included in each report.
- All JDL reports contain some content written by the intervention programme provider, including a summary of the programme and their views on the analysis results. MoJ told us that including this material helps build trust with programme providers, and providers value its inclusion. Nevertheless, this content is not impartial or objective and should therefore be published separately from the main report.
- The summary spreadsheet of JDL publications helpfully brings together the findings of all JDL analyses to date. It also includes comparisons of findings across programmes, although one chart gives the impression that findings are directly comparable between programmes, when they are not.
- We welcome MoJ's renewed focus on user engagement. We encourage MoJ to monitor wider uses of the JDL statistics and consider how it can promote the statistics to a wider audience to maximise their value.

Our judgement

We have identified several recommendations that MoJ should address to enhance the Trustworthiness, Quality and Value of the statistics. Our recommendations ask MoJ to, among other things:

- publish content written by intervention programme providers separately from the main report while continuing to make the content easily accessible to users
- produce a central methodology document that outlines all stages of the JDL analysis process, including quality assurance arrangements
- strengthen its rationale for not applying multiple comparison correction if this is deemed appropriate and include and explanation in the report
- provide more detail on the most important general caveats and limitations in each report

Next steps

We expect MoJ to immediately implement the recommendation on the impartial and objective presentation of the statistics as this is critical to ensuring Code compliance. We will continue to engage with the JDL team as it implements the remaining recommendations.

Related links

OSR to Sai: Methodology of Justice Data Lab Statistics

Introduction

Background to the statistics

1.1　The [Justice Data Lab (JDL)](#) is a small analytical team within the Ministry of Justice (MoJ) that supports organisations that provide offender services in England and Wales. It was set up in 2013 with the aim of giving organisations access to reoffending data to help them estimate the effectiveness of interventions to reduce reoffending.

1.2　Organisations supply data to JDL on the offenders who received the intervention. JDL uses these data to match offenders to records in the Police National Computer (PNC) database. It then uses a statistical technique called propensity score matching (PSM) to create a comparison group of offenders who did not receive the intervention. Offenders are matched on a range of variables, including demographic characteristics and offending history, so that the comparison group is as similar as possible to the intervention group (treatment group). Once matched, JDL conducts statistical significance testing to compare the reoffending outcomes of the treatment and comparison groups.

1.3　Analyses include headline results on reoffending outcomes (whether individuals have reoffended and the time and frequency of reoffending), typically over a one-year period after their release from prison. If the sample size is large enough and the matching quality is high enough, a series of sub-analyses are also conducted to investigate the impact of other factors, such as attendance levels, method of delivery of intervention programmes and participant demographics, on those reoffending outcomes.

1.4　JDL support is used by a wide range of organisations. These include both 'internal' programme providers, for example, His Majesty's Prison and Probation Service (HMPPS), and 'external' programme providers, encompassing mainly organisations in the voluntary and community sector, such as charities, educators and public and private bodies. As JDL is an analysis service, it is not involved in the design or delivery of programmes.

1.5　MoJ publishes a separate report for each programme it analyses. It has published over 250 JDL reports, all of which are published as official statistics. MoJ is one of the few UK Government departments that routinely publishes impact evaluation analysis as official statistics.

1.6　JDL's work contributes to MoJ's evidence base on reducing reoffending.

Why we did this review

1.7　Between December 2024 and February 2025, a [concern was raised with us](#) about a statistical method that MoJ uses to produce the JDL statistics, specifically its

approach to statistical significance testing. The concern focused on whether MoJ should be applying multiple comparison correction (MCC) methods when carrying out many significance tests on the same dataset to account for family-wise error – the probability of incurring at least one Type I error (a false positive) when performing multiple statistical hypothesis tests. MCC mitigates the increased risk of false positives (the likelihood of finding a statistically significant result by chance) and 'p-hacking' (persistent statistical testing until a statistically significance result is found). It does this by adjusting the statistical significance threshold, making it more difficult to achieve statistical significance.

1.8   We investigated some aspects of the statistical methods at the time the concern was raised. Although we were broadly satisfied with MoJ's approach to statistical significance testing, we committed to carrying out a review of the statistics' compliance with the standards of Trustworthiness, Quality and Value in the [Code of Practice for Statistics](#). Accordingly, this review has considered, in more detail, MoJ's approach to statistical significance testing, as well as aspects of the data collection, the aims of the analyses and the presentation of the statistics.

1.9   For this report, we initially reviewed a sample of JDL reports and all methodology reports and user guidance documents (see Annex for full list). The JDL reports we reviewed were published from 2013 to 2025 and cover both internal and external programmes. We examined the consistency of the presentation of the statistics, methods descriptions and terminology between reports.

Findings

Impartial and objective commentary

2.1   All JDL reports contain some content written by the programme provider. This typically includes a summary of the intervention programme and the provider's views on the analysis results.

2.2   MoJ told us that including this material helps build trust with programme providers and allows MoJ to verify that providers correctly understand the data analysis process and its findings.

2.3   During our compliance review, MoJ undertook a small review of user needs. It gathered views from several providers on the benefits of including their commentary in the report. MoJ found that providers value the opportunity to acknowledge the report's findings and to add context or clarification. Providers also consider that this demonstrates transparency in the process.

2.4   Statistics, data and explanatory material should be presented impartially and objectively, in line with the requirements of the Code of Practice for Statistics. In our view, the content in the reports written by the programme provider is neither impartial nor objective. For example, we found that some providers expressed disappointment if a statistically significant result had not been found, and some were openly critical of MoJ's methods or analysis approach. Therefore, this content should be published separately from the main report, with the nature of the commentary clearly indicated. To continue meeting the needs of providers, MoJ should ensure this content remains easily accessible.

Recommendation 1: To ensure that the statistics in the main reports are presented impartially and objectively, MoJ should immediately publish all content written by programme providers separately from report.

Intervention programme data

2.5   To understand how an intervention works, MoJ works closely with programme providers and gathers a range of information from them. Providers must submit a data upload template with personal information on programme participants, including their name, date of birth and the date their sentence started or the date they were sentenced. These data enable MoJ to match individuals with their records in the Police National Computer (PNC) database. The template also asks questions about the type and nature of the intervention and how participants were selected. This information helps MoJ identify the factors that it needs to consider when creating a comparison group.

2.6 MoJ told us that it holds a validation meeting with the provider and its analysts at the start of the project to discuss how the programme was run and how data were collected.

2.7 It is important to note that MoJ does not collect or analyse any data on the performance or delivery of programmes. MoJ only uses the data on programme participants submitted by the provider, and JDL analyses only examine reoffending outcomes; no other behavioural outcomes are considered. Information on offending behaviour is drawn from MoJ's own administrative datasets.

2.8 Some individuals may be excluded from the analyses. There are a range of reasons why participants are excluded, for example if their details cannot be linked to the PNC; their details cannot be linked to JDL's reoffending data; or they cannot be matched to anyone in the comparison group. MoJ is transparent about the characteristics of participants: JDL reports include a profile of the treatment group that breaks down the participants included and not included in the analysis by demographic characteristics such as sex and ethnicity.

Methods

2.9 MoJ employs a consistent, well-structured and rigorous approach for JDL analyses. The JDL methodology begins with a systematic process that focuses primarily on three reoffending measures: whether an individual reoffends, the rate at which reoffending occurs and the frequency of reoffending. These headline metrics form the backbone of the analyses.

2.10 MoJ uses propensity score matching (PSM), a recognised statistical technique that creates a comparison group that is closely matched to the treatment group on key characteristics. PSM helps isolate the effect of the intervention and gives MoJ increased confidence in attributing differences in reoffending between the treatment and comparison groups to the provider's programme. The JDL team has developed expertise in PSM and provides protocols and training on the method to analysts within MoJ.

2.11 MoJ is transparent about how well the matching process worked. All reports contain a flow chart that clearly illustrates the number of programme participants that were submitted for analysis and explains how many individuals were excluded from the analysis and why. In addition, match scores are published alongside the report to provider further transparency.

2.12 Peer review is a core component of JDL's work. Between 2016 and 2017 the methodology was peer-reviewed. In addition, the methods for two of the largest intervention programmes (Resolve and the Thinking Skills Programme (TSP)) were peer-reviewed. The reviews invited experts, independent of the JDL team, to challenge the methods, methodology and consistency of analyses, and findings were transparently reported. While the reviews did not recommend any major methodological changes, they did provide valuable feedback that helped MoJ refine

JDL processes. We welcome the extent of peer review that has been carried out and are assured about the JDL team's independence from the peer review process.

2.13 The published information about JDL's methodology is comprehensive. Methods are explained across a range of different documents, including the [methodology paper](#), [pilot study work](#), research and peer reviews, individual reports and the summary spreadsheet of JDL publications. However, because information is published in different documents, it may be difficult for users to navigate and understand the full process for data collection, analysis and producing the statistics.

Recommendation 2: To improve the accessibility of methods information, MoJ should consolidate existing information into a central methodology document that covers all stages of the JDL analysis process from start to finish.

2.14 As we set out earlier, this review considered MoJ's methods for statistical significance testing. A specific concern had been raised with us between December 2024 and February 2025 regarding MoJ carrying out many statistical significance tests on the same dataset without applying multiple comparison correction (MCC). MCC addresses the increased risk of false positives (the likelihood of finding a statistically significant result by chance).

2.15 [Our response](#) at the time to this concern recognised that MCC is widely considered best practice when carrying out many comparisons on the same dataset. We emphasised that the responsibility for determining statistical methods lies with statistics producers, and that it is not our role as statistics regulator to require producers to apply a specific method. We can confirm that our position on this issue remains unchanged.

2.16 As part of this review, we have examined JDL's methods in detail and discussed them with MoJ. We heard that the JDL team carefully considered whether it would be appropriate to apply MCC to the analyses of the Resolve and TSP programmes, both of which involved many statistical significance tests. The team told us that it consulted MoJ methodologists and an internal expert panel on the use of MCC. The methodologists and panel advised against using MCC, as they thought that corrected results may not be easily understood by users, and it could potentially mask interesting results. The JDL team followed this advice and explained the rationale for not applying MCC in the reports for those programmes: "While multiple correction methods can be applied to reduce the risk of incorrectly finding a positive treatment effect, they can also increase the likelihood that real differences will not be detected. The results presented in this report have therefore not undergone multiple correction methods."

2.17 One aspect of the concern raised with us relates to how MoJ structures its analyses; the measures to be evaluated must be specified in advance of the statistical tests being carried out to mitigate the risk of 'p-hacking'. This is a legitimate concern where exploratory research is conducted with no clear hypotheses or analysis plan. We found that JDL analyses are clearly and consistently structured, and all measures tested for statistical significance are set out ahead of time. The three headline

reoffending measures (reoffending occurrence, rate and frequency) are included in every report, and the four additional sub-analyses, which are included if the sample size is large enough, are also pre-defined, as outlined in the general annex (PDF). Analysis of the Resolve and TSP programmes involved further sub-analyses, but again, each analysis was specified in advance and explained in the reports.

2.18 Having reviewed the statistical methods in greater depth, we are assured about MoJ's approach to statistical significance testing. While it is good that MoJ has been transparent about its choice not to apply MCC in some reports, we consider that MoJ's rationale for not applying MCC should be strengthened. Providing such an explanation in all reports that involve many statistical significance tests will enhance transparency about the statistical methods and assure users about MoJ's approach to statistical significance testing.

Recommendation 3: To assure users about its approach to multiple comparisons in statistical significance testing, MoJ should strengthen its rationale for not applying multiple comparison correction, if this is deemed appropriate for the analyses. If this is the case, an explanation should be included in the report.

Development of new methods

2.19 MoJ is currently developing new methods for analysing the outcomes of longer-term and more-complex programmes. It is hoped that these methods will enable MoJ to evaluate other outcomes of programmes, for example, those related to accommodation and employment, as well as reoffending.

2.20 MoJ is also exploring methods for conducting between-programme comparisons to further assess JDL's value for money. MoJ told us that it is aware of the challenges of doing between-programme comparisons and that it is seeking expert methodological advice on the best approach. We welcome MoJ's caution, given the range of limitations, assumptions and constraints associated with such analyses. Differences in the intervention design, timeframe, setting and other factors can make it difficult to draw conclusions across interventions.

2.21 To ensure that the methods are robust and are clearly communicated to users, we encourage MoJ to have both methods peer-reviewed and to publish its methods and analysis plans.

Quality assurance

2.22 We found that the level of quality assurance applied to analyses of internal and external programmes differs. For external programmes, MoJ told us that quality assurance is limited to sense checks; the data-matching process is important for identifying low-quality or improbable-looking data. The quality assurance of data from internal programmes is more extensive because MoJ can validate the data against internal databases rather than the PNC. As a result, match rates for internal

programmes are substantially higher than those for external programmes. A higher match rate increases the sample size of the treatment group and allows MoJ to carry out a wider range of analyses.

2.23 MoJ publishes no information about how JDL analyses are quality-assured. MoJ should explain how it checks and validates results.

Recommendation 4: To demonstrate transparency about all aspects of quality, MoJ should publish a summary of the quality assurance arrangements for JDL analyses.

Quality information

2.24 Overall, uncertainty in the statistics is communicated well. Headline findings are accompanied by charts that illustrate the difference between the treatment and comparison groups and the 95% confidence intervals (the range of values that contains the true value for the population 95% of the time). Both the methodology paper and the general annex provide advice for users on interpreting confidence intervals and statistically significant results.

2.25 The methodology paper, general annex and summary spreadsheet of JDL publications highlight caveats and limitations that should be considered when reading a JDL report. These cover features of the treatment group (such as sample size and potential bias in the selection of participants), methods (such as the match quality) and results (such as comparability). While these general caveats are clear and aid user interpretation of the statistics, each document contains different information. As a result, some users may miss some key caveats and limitations.

Recommendation 5: To aid user interpretation of the statistics, MoJ should provide more detail on the most important general caveats and limitations in each report.

2.26 Additional caveats and limitations specific to an analysis are presented in each report. These cover different aspects of the programme and data, such as the COVID-19 pandemic leading to delays in reoffence convictions.

Clarity and insight

2.27 JDL reports follow a standard format. Reports always present the statistical significance of differences between the treatment group and comparison group and the estimated scale of the difference for all headline reoffending measures. For the largest interventions, including TSP and Resolve, effect sizes are also reported and explained. The effect size indicates the strength of the impact of a programme on an offender's behaviour and provides helpful context when a statistically significant result has been found.

2.28 JDL reports contain clear advice for users on interpreting the statistics, setting out what you can and cannot say about the results for each headline measure. When a

result is not statistically significant, the report explains that "there may be a number of reasons for this and it is possible that an analysis of more participants would provide such evidence". Such statements help ensure that readers understand the results.

2.29 Some visualisations do not follow best practice. For example, the charts for 'one-year proven reoffending frequency' and the 'average time to first proven reoffence' use icons which make it difficult to compare the figures for the treatment and comparison groups and to interpret the confidence intervals. MoJ should simplify the charts to improve their clarity and usability. We recommend consulting the Analysis Function's guidance on data visualisations.

2.30 As part of every publication round, MoJ produces a summary spreadsheet of JDL publications that helpfully brings together the findings of all JDL analyses to date. MoJ is developing an interactive tool to replace the spreadsheet, which will enable users to better explore JDL findings.

2.31 The summary spreadsheet includes comparisons of findings across programmes. For instance, it presents the number of programmes, grouped by type, which found a statistically significant decrease, inconclusive result or statistically significant increase in reoffending outcomes. It also includes a 'forest plot' to illustrate the relationship between the size of the matched treatment group and the effect on the one-year proven reoffending rate across all programmes. While the plot is well-explained, we consider that is not an appropriate tool for comparison, as it gives the impression that findings are directly comparable between programmes, when they are not. In addition, the user need for the plot is unclear.

Recommendation 6: To minimise the risk of misinterpretation of data comparison between intervention programmes, MoJ should remove the forest plot and any related material from the summary spreadsheet of JDL publications.

2.32 Many JDL reports have been added to the UK Government's Evaluation Registry, a repository for all planned, live and completed government evaluations. MoJ is in the process of adding the remaining JDL reports to the registry. Making the reports available via the registry enhances the potential insight and value of the statistics.

User engagement

2.33 In April 2015, MoJ ran a user feedback survey with customers who had used JDL during the pilot phase. MoJ published a summary of user feedback and a pilot summary report, which outlined its key user engagement activities with external organisations. MoJ is currently running a similar user engagement exercise with a sample of programme providers. We welcome this renewed focus on user engagement, given the time elapsed since the previous exercise. MoJ intends to publish the findings of this work and set out how it will continue to improve the JDL statistics.

2.34 For both external and internal programmes, the programme provider is the key user of the statistics; reports are tailored to each programme. MoJ told us that it works closely with the provider to discuss all aspects of the work, analysis and reports.

2.35 JDL statistics may have users other than the programme providers and MoJ and HMPPS. We encourage MoJ to monitor wider uses of the JDL statistics and consider how it can promote the statistics to a wider audience to maximise their value.

List of recommendations

Recommendation 1: To ensure that the statistics in the main reports are presented impartially and objectively, MoJ should immediately publish all content written by the programme providers separately from the report.

Recommendation 2: To improve the accessibility of methods information, MoJ should consolidate existing information into a central methodology document that covers all stages of the JDL analysis process from start to finish.

Recommendation 3: To assure users about its approach to multiple comparisons in statistical significance testing, MoJ should strengthen its rationale for not applying multiple comparison correction, if this is deemed appropriate for the analyses. If this is the case, an explanation should be included in the report.

Recommendation 4: To demonstrate transparency about all aspects of quality, MoJ should publish a summary of the quality assurance arrangements for JDL analyses.

Recommendation 5: To aid user interpretation of the statistics, MoJ should provide more detail on the most important general caveats and limitations in each report.

Recommendation 6: To minimise the risk of misinterpretation of data comparison between intervention programmes, MoJ should remove the forest plot and any related material from the summary spreadsheet of JDL publications.

Annex: List of documents reviewed

Justice Data Lab reports

- [Willowdene Residential](#) (2025)

- [Brighton Women's Centre](#) (2024)

- [HMPPS Creating Future Opportunities (CFO)](#) (2024)

- [Stepping Stones (Forward Trust)](#) (2024)

- [The Thinking Skills Programme (TSP) reoffending report](#) (2023)

- [RESOLVE](#) (2021)

Methodology and pilot reports

- [A peer review of existing methodology – Response](#) (2016)

- [Methodology paper](#) (2013)

- [Justice Data Lab: Pilot summary](#) (2015)

- [Justice Data Lab: User feedback summary](#) (2015)

- [Justice Data Lab: The pilot year](#) (2014)

User guidance and other supporting information

- [Justice Data Lab: User Guidance](#)

- [Data upload template](#)

- [General annex](#)

- [Summary spreadsheet of JDL publications](#)