



Ministry
of Justice

Data First: Criminal Courts Linked Data

**An exploratory analysis of returning
defendants and the potential of linked
criminal courts data from 2011 to 2019
in England and Wales**

**Tom Jackson, Caris Greyson, Dr Ian Rickard and
Professor Andromachi Tseloni**

Ministry of Justice

Ministry of Justice Analytical Series

2022



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.

Disclaimer

The views expressed are those of the authors and are not necessarily shared by the Ministry of Justice (nor do they represent Government policy).

First published 2022



© Crown copyright 2022

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit nationalarchives.gov.uk/doc/open-government-licence/version/3

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

Any enquiries regarding this publication should be sent to us at researchsupport@justice.gov.uk

This publication is available for download at <http://www.justice.gov.uk/publications/research-and-analysis/moj>

ISBN 978-1-84099-980-8

Acknowledgements

We would like to thank our colleagues Kylie Hill, Daisy Ward, Toby Ogun, Amy Summerfield, Georgina Eaton, Robin Linacre, David Wall and Robert Reeve within Data and Analysis at the Ministry of Justice for their support and contributions towards this report.

This report is part of the Data First programme of work, a Ministry of Justice led investment funded by Administrative Data Research UK (ADR UK).

The authors

Tom Jackson, Caris Greyson and Dr Ian Rickard are based within Data and Analysis at the Ministry of Justice. Andromachi Tseloni is Professor of Quantitative Criminology at Nottingham Trent University and the Academic Lead (ES/V002198/1) for the Ministry of Justice Data First programme.

Contents

List of tables

List of figures

1. Summary	1
2. The Magistrates' and Crown Court Data	5
2.1 Overview	5
3. Returning criminal court defendants	7
3.1 What is a returning defendant?	7
3.2 Effectiveness of identifying returning defendants	7
3.3 The extent of returning defendants to the magistrates' courts	9
3.4 What offence types lead to the highest return rates?	10
3.5 What offences do defendants return for?	11
4. The potential of linked criminal courts data	14
4.1 Linked data showing cases from the magistrates' courts to the Crown Court	14
4.2 Linking Crown Court data to the Index of Multiple Deprivation	16
5. Implications of the findings	21
References	24
Appendix A	27
Quality of deduplication and linking	27

List of tables

Table 1: Estimated number of distinct defendants	30
--	----

List of figures

Figure 1: Percentage of defendants proceeded against in the magistrates' courts in 2011 who later returned to the magistrates' courts by December 2019	9
Figure 2: Percentage of defendants who return to the magistrates' courts for any offence up to December 2019, by initial 2011 principal offence	10
Figure 3: Heatmap showing the proportion of defendants proceeded against by initial 2011 principal offence, to the subsequent return to the magistrates' courts	12
Figure 4: Percentage of defendants who return to the Crown Court following an initial 2013 case in the magistrates' courts	15
Figure 5: Percentage of defendants dealt with in the Crown Court from 2013 to 2019 by principal offence from the most deprived small areas in England (top 20% of the Index of Multiple Deprivation)	17
Figure 6: Percentage of defendants dealt with in the Crown Court from 2013 to 2019 by principal offence from the most deprived small areas in Wales (top 20% of the Index of Multiple Deprivation)	19
Figure 7: Receiver operating characteristics curve, Splink compared with clerical and rules-based matching	29

1. Summary

This report is the first analytical output of the Ministry of Justice's (MoJ) pioneering data-linking programme Data First, funded by [ADR UK](#) (Administrative Data Research UK) to harness the potential of the wealth of data generated when people interact with justice services. The programme links administrative datasets from across the justice system and with other government departments to provide powerful new insights on justice system users, their pathways, and outcomes across a range of public services. Data First enables accredited researchers to access this deidentified data via secure platforms under the Digital Economy Act 2017 (DEA), providing a rich new resource to develop the evidence base for government policy and practice.

The findings in this report are based on analysis of magistrates' and Crown Court defendants; the first two datasets shared under Data First.¹ These datasets have been cleaned, deidentified, deduplicated and linked using the [Fellegi-Sunter](#) model of data-linking to provide a joined-up picture of criminal court defendant and case journeys.² Derived from court case management information systems, the data includes information on defendant characteristics, offence categorisations, case outcomes, and sentencing information from 2011 to 2020 (the Crown Court data is provided from 2013). This report covers data reporting up to December 2019 to improve the likelihood of including cases completing in 2020.

These datasets have enabled, for the first time, the extent and nature of repeat users to be explored at scale, including the type of offences repeat users are most likely to enter the criminal courts for. It also includes the proportion of defendants who reside in the most deprived areas of England and Wales using the 2019 Index of Multiple Deprivation.

¹ The MoJ Data First defendant level magistrates' courts (lower criminal courts in England and Wales) and the MoJ Data First defendant level Crown Court (higher criminal courts hearing more serious cases) datasets.

² Defendant journeys reflect all contacts with the justice system whether they refer to the same or different cases. Case journey refers to all contacts with the justice system for the same case against a defendant. Each defendant may have multiple case journeys.

As a newly available resource for accredited researchers, the report seeks to provide a foundational understanding of the scope and potential of this data. Academic researchers are encouraged to explore this dataset's potential further by making applications to securely access and use the data available for research.³

The key findings from this report are:

- Upon a defendant's first known case in 2011, over half (56%) of defendants had returned to the magistrates' courts by December 2019. 17% returned only once within this period and 18% of defendants returned six or more times.
- The return rate to the magistrates' courts varied by the defendant's first offence. The highest rate of return was for defendants initially proceeded against for a theft offence in 2011 (82% returned to the court by December 2019). The lowest rate of return across all offence groups was for sexual offences (43%). This could be due to longer custodial sentences imposed for sexual offences, reducing the opportunity to reappear again in front of the courts during the timeframe of this analysis.
- Regardless of a defendant's first offence group, many defendants later returned to the magistrates' courts for some form of summary offence.
- Defendants were most likely to return for the same offence group as their initial 2011 case, suggesting that the defendant could be specialising in their offence group.
- The percentage of defendants who were sent to the Crown Court increased with the number of times they returned to the magistrates' courts following their initial case. This might indicate an escalation in returning defendants' offending behaviour and subsequent sentences.
- In 2019, 43% of defendants dealt with in the Crown Court resided in the 20% most deprived neighbourhoods in England. This is an over-representation of the

³ The application form to access Data First datasets can be found here: <https://www.gov.uk/government/publications/moj-data-first-application-form-for-secure-access-to-data>

general population, where according to 2011 census population data, 20% of England's population lived in the top quintile of deprived areas (Ministry of Housing, Communities and Local Government, 2020).

- A similar finding was found for Wales. In 2019, 41% of defendants resided in the 20% most deprived Welsh neighbourhoods.
- Across all offence groups, defendants dealt with for a sexual offence in the Crown Court in 2019 were least likely to reside in the most deprived small areas in England (36%). Robbery (50%) and theft offences (50%) had the highest percentage of defendants who resided in the most deprived areas. Similar patterns were found in Wales.

The findings present important implications for policy and practice. They identify that overall, more than half of defendants returned to the courts within the data period, but this was highest for specific offence groups, including theft, robbery and drug offences. Defendants returned disproportionately to the courts for the same offence, and those who returned to the court multiple times were more likely to be directed to the higher Crown Court. These findings suggest both a specialisation and escalation in offending behaviour and point towards the importance of targeting interventions to prevent criminal careers.

Locality-based analysis on Crown Court defendants provide important insights on the backgrounds of justice system users, showing an over-representation of defendants residing in the most deprived areas in England and Wales compared to the general population. Further analysis show links between neighbourhood deprivation and crime type-specific offending and offender characteristics. The over-representation of defendants dealt with for acquisitive offences such as theft and robbery residing in the most deprived areas of England and Wales suggests that these offences are likely driven by financial need. Policy and practice services and interventions for this type of offender rehabilitation are therefore likely to be most effective when targeted by area. They could include initiatives to address and improve residents' health, education and employment needs and outcomes. In contrast, the relative under-representation of defendants residing in the most deprived areas for sexual and fraud offences suggest that these may be more likely to be individually motivated, though this warrants further investigation. Further implications of the

findings are drawn out throughout the report. They provide only an initial indication of policy insights that the Data First datasets can offer to inform national crime prevention initiatives, including those to tackle fraud, burglary, and vehicle crime.

2. The Magistrates' and Crown Court Data

2.1 Overview

The magistrates' courts and Crown Court defendant level datasets used as the basis for analysis in this report are two unique datasets created through Data First. Magistrates' courts data is sourced from extracts of Libra, and Crown Court data from extracts of XHIBIT, the administrative databases used by the magistrates' courts and Crown Court to manage cases across England and Wales. Both datasets have been made securely available to accredited researchers under the Digital Economy Act (2017) via the Office for National Statistics (ONS) Secure Research Service (SRS).

The magistrates' courts dataset includes information from January 2011 to December 2020 and the Crown Court dataset covers the period from January 2013 to December 2020. There were 13,357,982 records in the magistrates' courts dataset and 1,000,827 in the Crown Court dataset. They include information on the dates of hearings, type of offences, arrests and the initiation date of proceedings. This data denotes how proceedings were dealt with and provides information on the final plea of the defendant. It also includes details on bail, remand, and whether the case (and the defendant) was sent to the Crown Court. The data is structured with one record per defendant per case.⁴ Full details of the data included are contained within the published [data catalogues](#).

Both datasets have been deduplicated using the open-source statistical [Splink software](#). The deduplication process allows data users to identify which defendants have re-entered the criminal courts more than once by establishing defendant records that are believed to belong to the same person, using a probabilistic linkage method. This is the basis of the analysis conducted in this report. Further detail of the data-linking methodology is outlined in section 3.2.

⁴ The unit of analysis in each dataset is defendants dealt with in either the magistrates' courts or Crown Court. For example, defendants who have been dealt with twice appear in the data twice and so on and so forth.

The two datasets have been linked to identify records between the two courts that belong to the same defendant. This data linkage adds value to the underlying administrative data sources by enabling questions to be explored on defendants that are frequently the subjects of the criminal courts. These datasets enable research exploring returning-defendant characteristics, outcomes, and journeys, to be conducted at this scale for the first time.

Analysis based on this 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. The findings in this report are not comparable to other published statistics or research, including the quarterly MoJ Criminal Justice System Statistics and Criminal Court Statistics, due to different units of data, processing and analysis.⁵

⁵ The latest Criminal Court Statistics at the time of writing can be found here: <https://www.gov.uk/government/collections/criminal-court-statistics> and the latest Criminal Justice System Statistics can be found here: <https://www.gov.uk/government/collections/criminal-justice-statistics>

3. Returning criminal court defendants

3.1 What is a returning defendant?

A defendant who has entered the criminal courts on more than one occasion may be defined as a returning defendant. Any return to the court could be for a different offence type; for example, a defendant appearing in the court for a theft-related offence, could later return for a drug-related offence. The ability to identify the various interactions an individual has with the criminal courts is valuable information which can improve our understanding of journeys within the criminal justice system. Understanding the frequency with which individuals return to the courts, and the reasons why, can inform targeted interventions to reduce offending and return rates. Furthermore, the ability to look beyond the frequency of returning defendants to analyse, for example, changes in offending patterns, diversity or specialisation of offending, can help inform the effectiveness of different sentencing options. Socio-demographic information such as age-cohort, sex of the defendant, and ethnicity (although data is limited in some areas) can also tell us more about the characteristics of the profile of returning defendants.

3.2 Effectiveness of identifying returning defendants

The original magistrates' courts and Crown Court administrative data (extracts of Libra and XHIBIT respectively) do not include a unique defendant identifier which indicates court records that belong to the same person. Instead Data First data scientists created Splink – a package that implements the Fellegi-Sunter model (Fellegi and Sunter, 1969) of record linking, to deduplicate each dataset and produce an estimated defendant identifier (ID) to identify which defendants may (with a high degree of confidence) be the same person.

This ID is used as a unique person identifier for the basis of analysis in this report. It is recommended that all analysis based on these datasets treats records sharing an ID as belonging to the same person, however, like any data linkage technique it is expected that a small amount of under and over-matching will have occurred. For example, some

records may have been grouped together which really belong to two or more people – increasing the number of repeat appearances and reducing the number of unique defendants counted; while other records belonging to the same person may not have been connected – decreasing the number of repeat appearances and falsely increasing the number of unique defendants counted (see Appendix A for more information on the data-linking methodology).

Any exploratory analysis on the extent of returning defendants should be interpreted with some caveats. Firstly, the time-bound coverage of the data means the true extent of returning defendants cannot be known (a defendant may have (re-)entered the court prior to 2011 or since 2019). All analysis in this report will examine the extent of returns over the nine-year period which is useful in providing a snapshot of returns during this time. When referring to a defendant's first time in the criminal courts, this is true to be the first time from 2011, but not necessarily in their life.

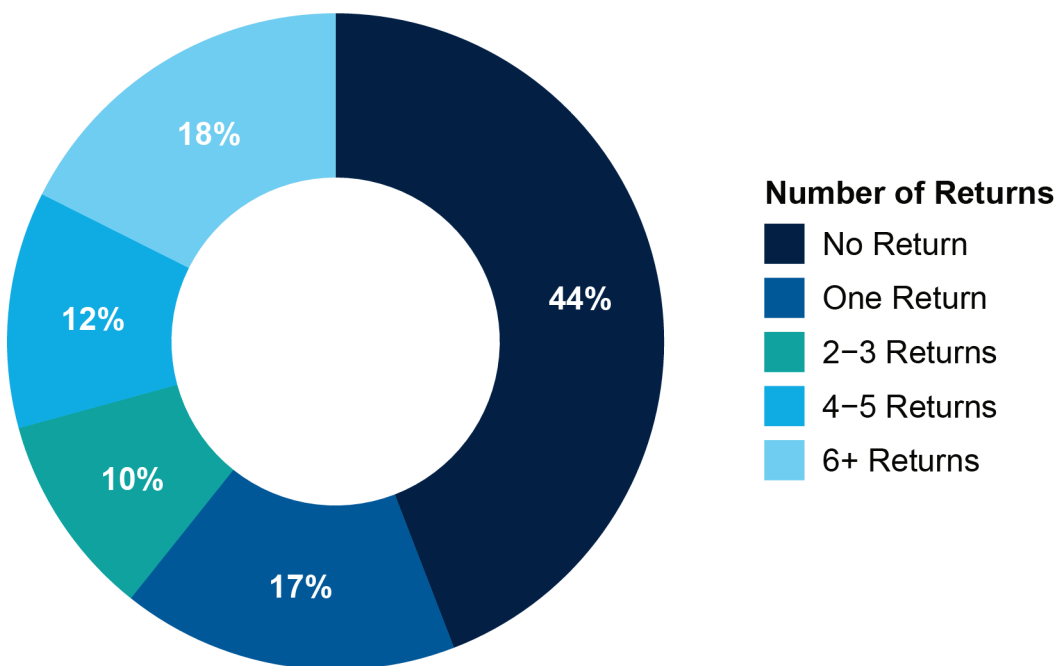
Secondly, this analysis does not account for custodial periods which, given reduced opportunities to offend, can significantly distort the picture of defendants' return rate to the Crown Court. This is likely to have fewer implications for defendants' return rates to the magistrates' courts due to the short length of prison sentences magistrates can impose (Francis, Soothill and Fliegelstone, 2004).⁶ Relatedly, the datasets of this report do not include actual but only imposed (by the courts) incarceration periods. Data First has released [prisoner data](#) on served custodial periods which can be used to build a more comprehensive picture of pathways through the criminal courts and prison.

⁶ Up to 12 months, and in practice less than that as according to the Offender Rehabilitation Act 2014 these offenders are released on licence after serving half of their sentence.

3.3 The extent of returning defendants to the magistrates' courts

In 2011, 1.1 million defendants appeared before the magistrates' courts in England and Wales. Upon a defendant's first known case in 2011, 44% of defendants proceeded against did not return to the magistrates' courts as a defendant before the end of 2019, whilst 56% of defendants returned at least once (as shown in Figure 1). Seventeen percent of defendants returned to the magistrates' courts within this period once, 10% of defendants returned two or three times, 12% of defendants returned four to five times, and 18% of defendants returned six or more times.

Figure 1: Percentage of defendants proceeded against in the magistrates' courts in 2011 who later returned to the magistrates' courts by December 2019⁷



These findings present interesting questions for further analysis, such as: what is the profile of defendants that return (frequently) to court? What offences feature the highest proportion of returning users? How, if at all, does the offence a defendant appears in court for change on each return to the court? How does their sentence change when they return

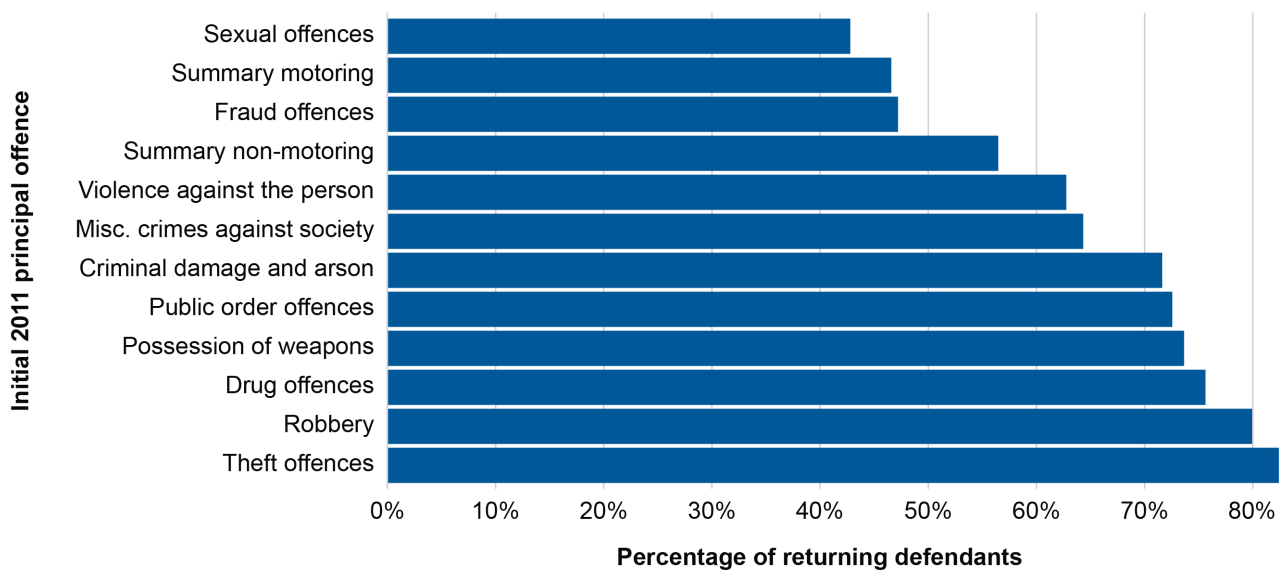
⁷ Due to rounding, the total percentage within the chart exceeds 100%.

to the court for similar (or more serious) offences? The following sections draw insight on these questions.

3.4 What offence types lead to the highest return rates?

The return rate of unique defendants varied by the offence for which a defendant first appeared. Figure 2 shows the percentage of defendants who returned for any offence to the magistrates’ courts by their initial 2011 principal offence. The lowest return rate was for sexual offences, where 43% of defendants proceeded against in the magistrates’ courts for an initial 2011 principal sexual offence, returned to the court (for any offence) by December 2019.⁸

Figure 2: Percentage of defendants who return to the magistrates’ courts for any offence up to December 2019, by initial 2011 principal offence



Similarly, low rates were observed for summary motoring offences (47%) and fraud offences (47%). The return rate for defendants charged with more serious (indictable) offences may not be accurately represented due to likely longer custodial sentences, meaning they could not have returned to the magistrates’ courts within the data period.

⁸ The percentage for sexual offences was calculated by dividing the total number of defendants who returned to the courts for any offence by the end of 2019 following an initial 2011 sexual offence, by the total number of defendants in the magistrates’ courts for a sexual offence in 2011.

Nonetheless, the return rate is accurate for summary offences and is still a useful indicator for non-summary offences with less severe custodial sentences.⁹

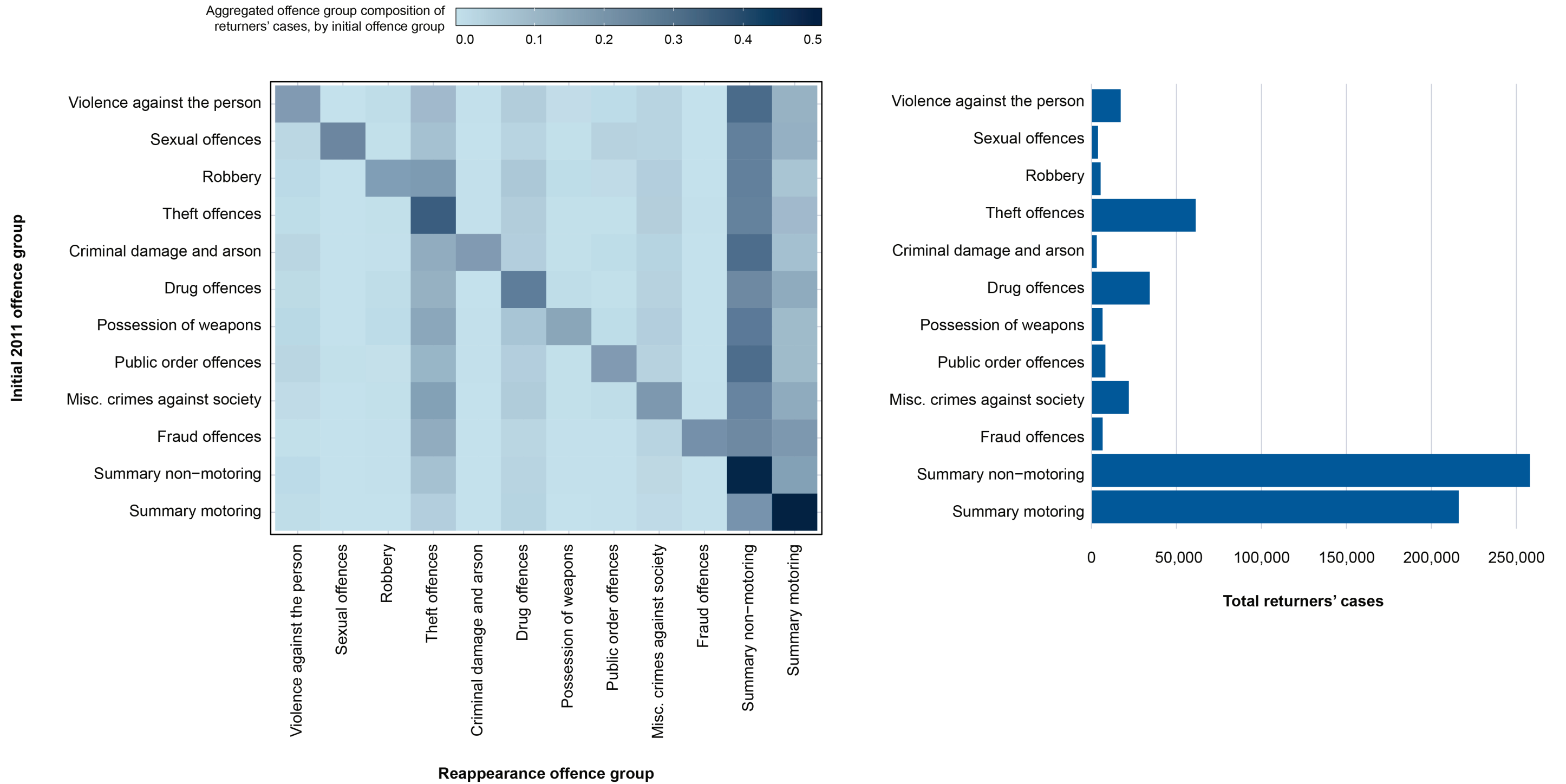
The highest return rate was for theft offences (82%), with a similarly high rate for robbery offences (80%) followed by drug offences (76%), possession of weapons (74%) and public order offences (73%). These findings are in line with wider research on drivers of such offences. The causes of drug and public order offences – the latter mostly relating to alcohol abuse – remain unless treated with long-term mental health support and post-treatment care (Brunton-Smith and Hopkins, 2013). Further, theft offenders may re-offend to support drug abuse (Hunter, Garius, Hamilton and Wahidin, 2018).

3.5 What offences do defendants return for?

The nature of defendants who return to the magistrates' courts can be explored further by investigating which offence groups defendants return for. Figure 3 illustrates this in two parts. The first part (left hand-side) is a heatmap showing proportions of defendants who returned by each offence group. The heatmap shows only the first return from an initial 2011 offence, meaning any further reappearances following the first return are excluded (avoiding counting defendants multiple times). The second (right hand-side) part shows the number of cases of defendants who returned to the magistrates' courts across offence groups against their initial 2011 court appearance. For example, 17,121 defendants proceeded against for an initial violence against the person case returned to the magistrates' courts by the end of 2019 (right hand-side top bar). Out of these, 21% returned for another violence against the person offence, and 36% for a summary non-motoring offence (top horizontal heat map bar).

⁹ A summary offence is a criminal offence which is normally tried in magistrates' courts and which is generally considered to be less serious than other types of offences.

Figure 3: Heatmap showing the proportion of defendants proceeded against by initial 2011 principal offence, to the subsequent return to the magistrates' courts



A clear pattern is the dominance of the summary (less serious) offences. Regardless of the first offence group, many defendants later returned to the magistrates' courts for some form of summary offence (as shown by the darker shading in the respective penultimate and last vertical bars of Figure 3). These patterns may reflect the high frequency of returning defendants for these offence types (right hand side bar chart).

A second pattern is that returns from the initial 2011 offence disproportionately matched the same offence group, suggesting that the defendant was specialising in their offence category. Notably, there was a high proportion (40%, as can be seen in Table 3) of defendants who were proceeded against with an initial 2011 theft offence who later returned for a theft offence again.

Excluding initial 2011 summary offences, theft offences are the most common offence group that defendants returned for (see fourth vertical bar in Figure 3). For example, 21% of defendants initially proceeded against for a robbery offence later returned for a theft offence, with similarly high rates for initial miscellaneous crimes against society (19%) and possession of weapon offences (17%). While previous research has shown those serving a sentence for an acquisitive offence, such as robbery or theft, are likely to re-enter the justice system (Brunton-Smith and Hopkins, 2013), further exploration of defendant journeys from other offence groups to acquisitive theft or robbery offences would be an avenue for future research.

4. The potential of linked criminal courts data

Linked criminal courts data provides the opportunity to follow a case or defendant journey through the magistrates' courts and Crown Court. This section demonstrates some of the potential of linking these datasets for two important areas of investigation. First, the nature of cases which have progressed to the Crown Court from defendants who have been dealt with by the magistrates' courts multiple times. Second, linking geographical court data to the Index of Multiple Deprivation (IMD¹⁰) to provide contextual information on the socio-economic backgrounds of defendants.

4.1 Linked data showing cases from the magistrates' courts to the Crown Court

For each Crown Court record, there will be a corresponding magistrates' courts record as all cases start in the magistrates' courts regardless of seriousness. By identifying and linking those corresponding records, it is possible to use the data to see how cases flow through the court system. This builds on the analysis of returning defendants in section two by exploring whether defendants who returned following their initial 2013¹¹ case in the magistrates' courts later returned to the Crown Court. The ability to link data in this way allows questions to be asked such as, what percentage of defendants who were proceeded against in the magistrates' courts for an initial offence in 2013, returned and were dealt with in the Crown Court?

¹⁰ The 2019 Index of Multiple Deprivation (IMD) has been used in this time series analysis as a proxy for neighbourhood deprivation. As only the 2019 IMD measure has been used, the analysis does not account for changes in relative deprivation across iterations such as the 2010 and 2015 IMD for England. However, these were not extensive, as 88% and 83% of English neighbourhoods in the most deprived quintile according to the respective 2019 and 2015 IMD were so in the previous (2015 and 2010, respectively) iteration (MHCLG, 2019; DCLG, 2015).

¹¹ Crown Court data begins from 2013, therefore when linking between the two datasets, magistrates' courts data starts from 2013 rather than 2011 as used in section two of the report.

Figure 4: Percentage of defendants who return to the Crown Court following an initial 2013 case in the magistrates' courts

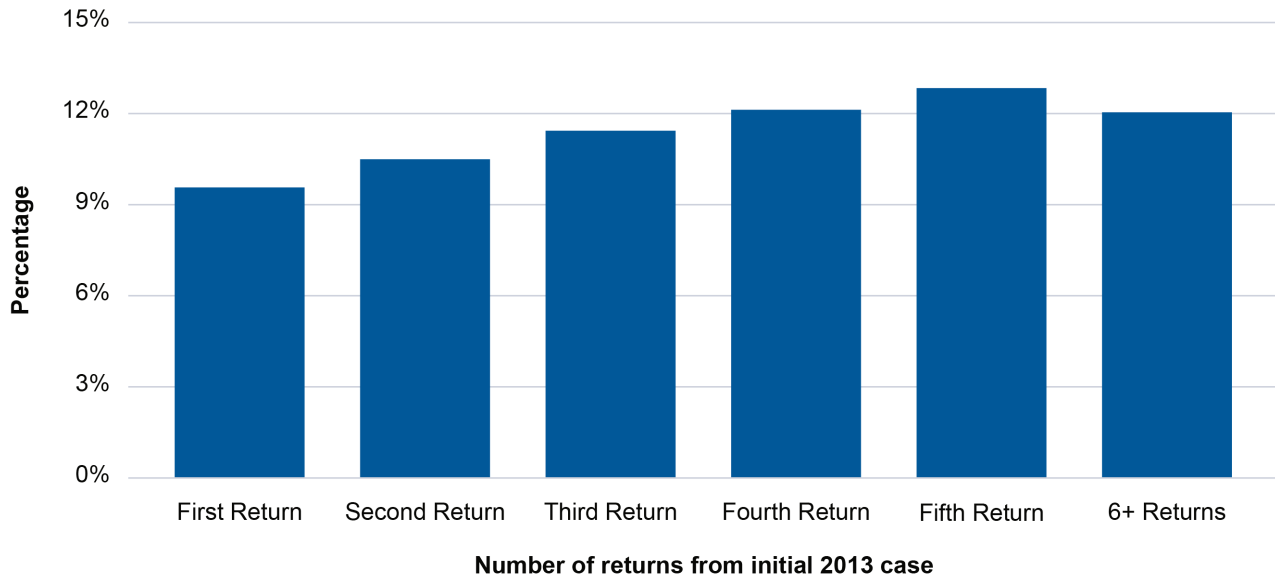


Figure 4 shows the percentage of defendants, who following an initial case in the magistrates' courts in 2013, later returned to the Crown Court. This analysis has been conducted by taking the initial methodology to identify a returning defendant (as outlined in section 3.2). The data on returning defendants was then linked to the Crown Court data to discover whether the case was sent to the Crown Court.

The pattern demonstrates that the percentage of defendants who were sent to the Crown Court increased with the number of times they returned to the magistrates' courts following their initial case. For example, 10% of defendants on their first return to the magistrates' courts were sent to the Crown Court. This increased to 13% for defendants who returned five times, until slightly decreasing to 12% of defendants who returned six or more times.

The low percentages of cases being sent to the Crown Court is expected as most proceedings within the magistrates' courts are for summary (less serious) offences. The increase of defendants being sent to the Crown Court as they returned more frequently might indicate an escalation in their offending behaviour. It also suggests an escalation in sentencing, where defendants returning multiple times for what could be similar offences, are facing escalated sentences and their cases subsequently being sent to the Crown Court.

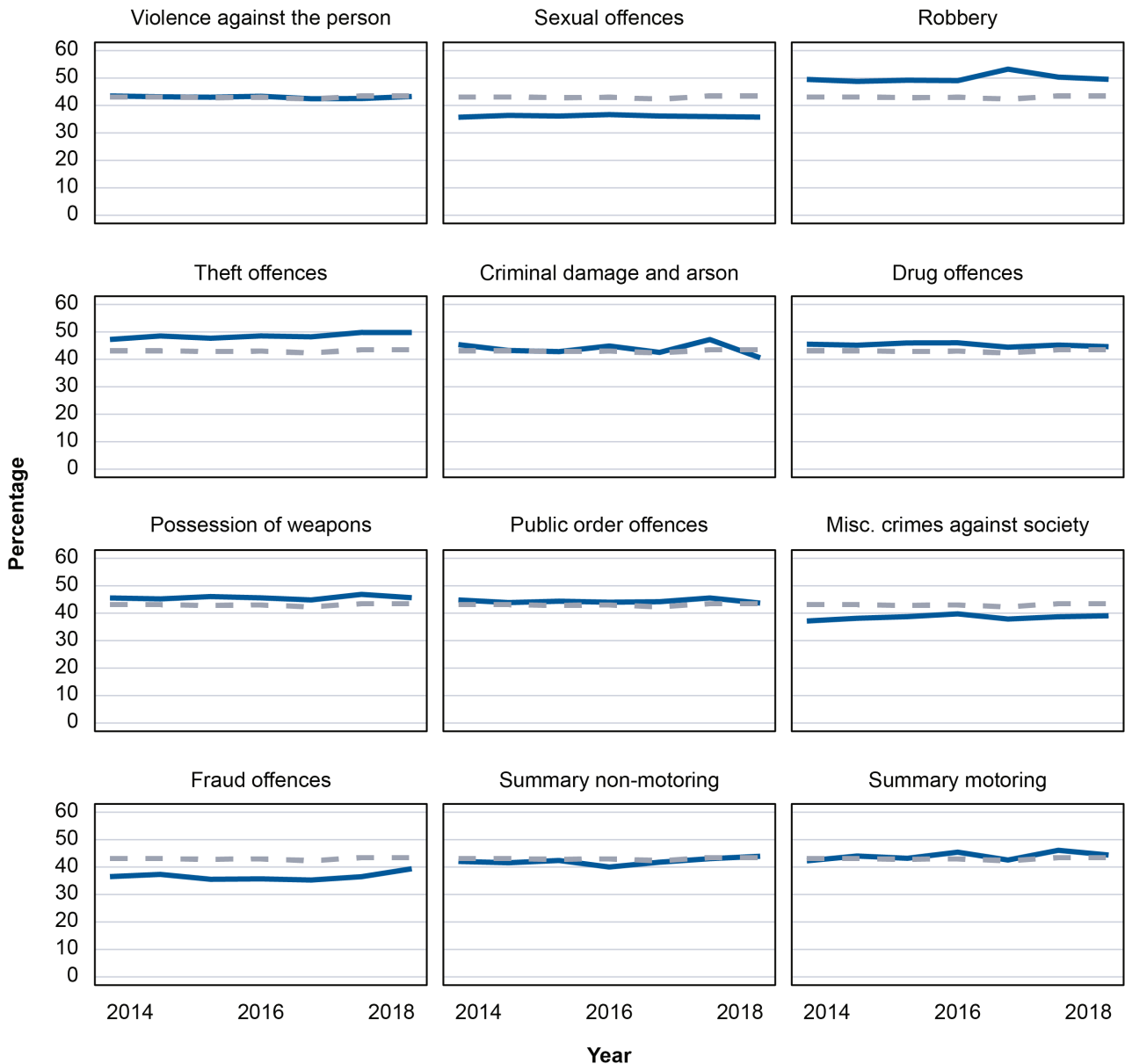
4.2 Linking Crown Court data to the Index of Multiple Deprivation

Matching defendant level records to the 2019 IMD can provide important new insights on the socio-economic context of justice system users. By matching the Lower-layer Super Output Area (LSOA)¹² information in the Crown Court dataset with the IMD, it is possible to investigate the prevalence of defendants dealt with at the Crown Court who reside in the most deprived small areas across England and Wales.

The Crown Court data covers both England and Wales, however, given that each country in the UK produces a separate version of the IMD with different indicators, the data for England and Wales cannot be combined to provide a direct comparison. A separate figure for Wales has therefore been produced (Figure 6).

¹² LSOA is a geographic area and Output Areas are clusters of adjacent unit postcodes. There is an LSOA for each postcode in England and Wales.

Figure 5: Percentage of defendants dealt with in the Crown Court from 2013 to 2019 by principal offence from the most deprived small areas in England (top 20% of the Index of Multiple Deprivation)



*Grey-dashed line represents the overall percentage of defendants dealt with across all offence groups

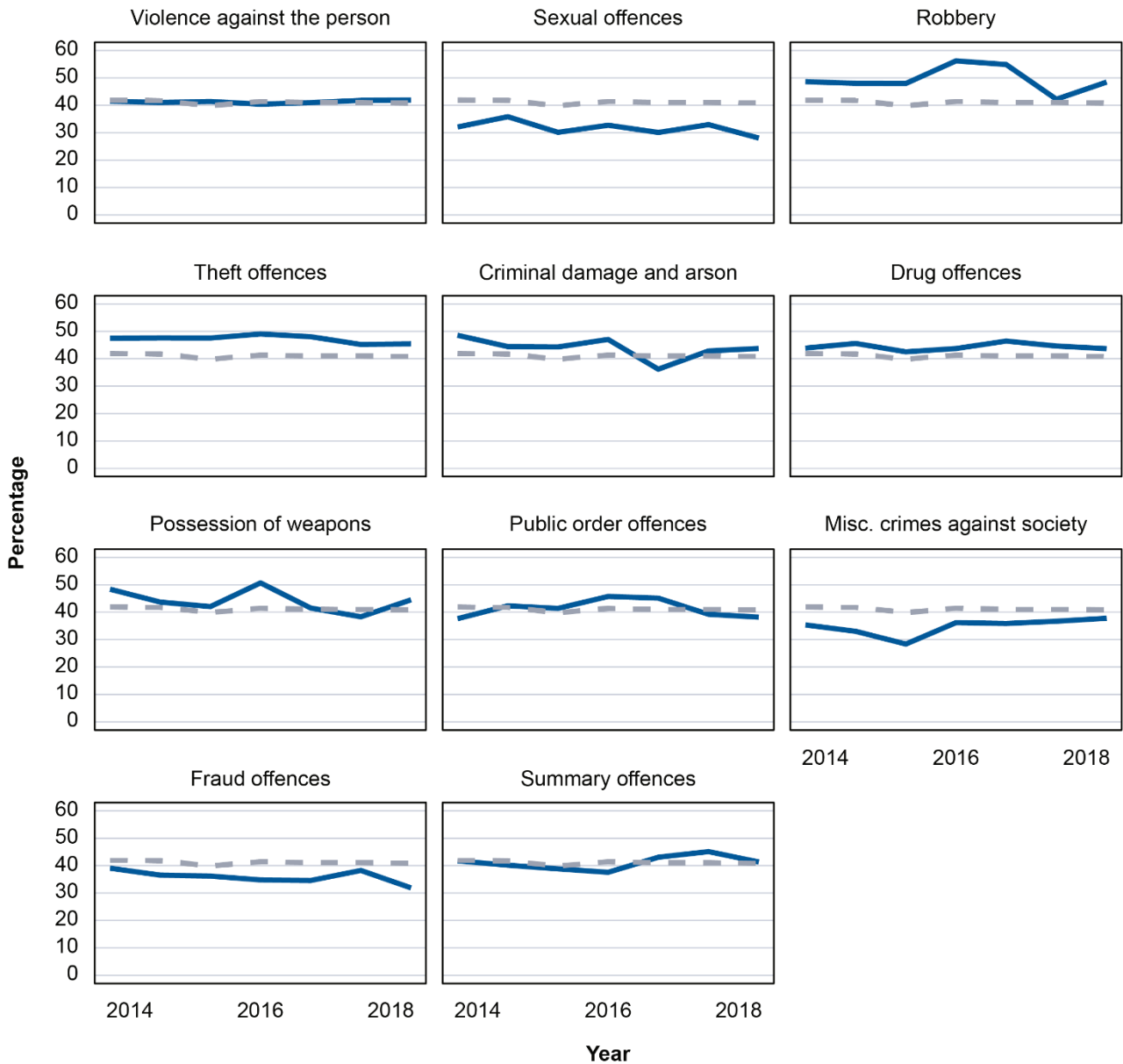
Figure 5 shows the percentages of defendants residing in the 20% most deprived small areas in England, by the principal offence dealt with at the Crown Court from 2013 to 2019. The dotted line in each sub-graph shows that overall, 42% to 44% of defendants (regardless of offence group) resided in the 20% most deprived neighbourhoods at the time of their Crown Court appearance. This is an over-representation of the general population, where according to 2011 census population data, 20% of England’s population

lived in the top quintile of deprived areas (Ministry of Housing, Communities and Local Government, 2020).

The analysis provides a basis for exploring any over- or under- representation of defendants living in deprived neighbourhoods across specific offence groups within the defendant population. The respective proportions of defendants dealt with for several offence groups (violence against the person, public order offences, criminal damage and arson, summary motoring and non-motoring) who lived in the most deprived areas were similar to the average (dotted line) across all defendants.

However, there were substantial differences for the remaining offences. In 2019, half (50%) of defendants dealt with in the Crown Court for a robbery offence resided in the top quintile of deprived small areas in England. The percentage of defendants dealt with for a theft offence has gradually increased from 47% in 2013 to 50% in 2019. Defendants dealt with for drug offences (45% in 2019) and possession of weapons (46% in 2019) were also over-represented in deprived neighbourhoods compared to all Crown Court defendants. In contrast, there were fewer defendants from the most deprived areas dealt with in the Crown Court for principal offences, such as sexual offences (36% in 2019), fraud offences (39% in 2019 but as low as 35% in 2017), and miscellaneous crimes against society (39% in 2019) compared to the overall percentage of defendants in the 20% most deprived neighbourhoods.

Figure 6: Percentage of defendants dealt with in the Crown Court from 2013 to 2019 by principal offence from the most deprived small areas in Wales (top 20% of the Index of Multiple Deprivation)¹³



*Grey-dashed line represents the overall percentage of defendants dealt with across all offence groups

Figure 6 shows similarities to England for the Welsh defendant population. From 2013 to 2019, between 40% to 42% of defendants resided at the time of their appearance before the Wales Crown Court in the 20% most deprived Welsh neighbourhoods (dotted grey

¹³ Summary and summary non-motoring figures have been merged into summary offences due to low counts.

line). Similar patterns to those in England were identified, where the percentage of defendants dealt with for a sexual offence in 2019 who reside in the most deprived small areas in Wales, is the lowest of all offence types (28%).¹⁴ Robbery (49%), theft offences (46%) and possession of weapons (45%) remain the offence groups with the highest percentage of defendants dealt with who reside in the most deprived small areas in Wales in 2019.

The over-representation of defendants in England and Wales residing in the most deprived areas, as in the case of robbery and theft, suggest that these acquisitive offences are driven by financial need. In contrast, the relative under-representation of defendants in England and Wales residing in the most deprived areas for sexual and fraud offences, suggest that these are more likely to be individually motivated than financially, though this would warrant further investigation into other IMD quintiles.

¹⁴ The findings on sex offenders' lower than average English and Welsh neighbourhood deprivation disagree with previous evidence on USA registered sex offenders' residential area profile (Mustaine and Tewksbury, 2005; Mustaine, Tewksbury and Stengel, 2006).

5. Implications of the findings

This report draws on analysis of the magistrates' courts and Crown Court datasets shared under Data First. From 2011 to 2019, over half (56%) of defendants returned to the magistrates' courts at least once, mostly for theft, robbery and drug offences (as shown in Figure 1 and Figure 2). Furthermore, defendants who returned to the court multiple times were more likely to have their case sent to the Crown Court, suggesting an escalation in returning defendants' offending behaviour and subsequent sentences. These findings suggest that the sentencing or interventions imposed, particularly for offence groups such as theft, robbery and drug offences, have not stopped most defendants from returning to the criminal courts.

A disproportionate number of defendants returned for the same offence. Previous research has shown that offenders serving a sentence for an acquisitive offence was an important predictor for re-offending (Brunton-Smith and Hopkins, 2013). The findings in this report also show high proportions of defendants returning after a theft offence, but additionally high proportions of defendants returning for theft offences after possession of weapons or miscellaneous crimes against society offences. Theft offences include a wide array of specific crimes, such as burglary, vehicle crime, shop-theft and a suite of other unauthorised taking. Therefore, expanding the analysis in this report to examine returning defendants for more narrow offence definitions is one avenue of future work.

This understanding of returning defendants, including by specific offence groups, can help inform targeted interventions to better meet the needs of criminal court users and reduce reoffending (as indicated by the number of offenders who are drawn back to the criminal courts). For example, exploring whether and, if so, how different sentencing outcomes of previous burglary convictions may impact on the likelihood of returning to the criminal courts can provide policy insights to the evidence base of national crime-specific initiatives to tackle burglary. More detailed analysis of returning defendant characteristics, including how age, sex, ethnicity and locality impact on the nature of a return to the criminal courts, will further enhance these insights.

Further analysis found that higher than average proportions of defendants dealt with in the Crown Court for theft and robbery lived in the 20% most deprived small areas of England and Wales. Such implications suggest that area-based policy tools are likely to be appropriate for offender rehabilitation; for example, initiatives to improve residents' health, education and employment outcomes. Furthermore, where acquisitive offences are committed close to an offenders' home, target hardening¹⁵ to reduce crime opportunities adds to the armour of policy tools which existing literature suggests has a likely immediate effect (Hirschfield, Newton and Rogerson, 2010; Brantingham and Brantingham, 1993).

By contrast, the relative under-representation of defendants dealt with for sexual and fraud offences living in the most deprived areas suggest that individual-based rehabilitation initiatives may be more appropriate.

The findings support criminological theory and previous evidence on the relationship between neighbourhood deprivation and criminality, to the extent that offences were committed in close proximity to a defendant's residence (Brantingham and Brantingham, 1993) and that crime occurs in socially and economically disadvantaged neighbourhoods (Shaw and McKay, 1942; Sampson, 2006; Cole, 2019).

These findings are also aligned with established evidence on the links between neighbourhood deprivation and crime type-specific offending and offender characteristics (Kearns, Livingston, Galster and Bannister, 2019; Dorminey, Fleming, Kranacher and Riley, 2012; Bernasco, 2010; Steffensmeier and Haynie, 2000). For example, acquisitive offence and assault offenders operate in areas they know well, i.e. their present and past neighbourhoods (Bernasco, 2010). There is also evidence of intra-neighbourhood transmission of offending whereby a high concentration of past offenders in a neighbourhood encourages the uptake of criminal behaviour by 'newly active' offenders in the area (Kearns, et al., 2019). It suggests therefore, that addressing area deprivation is one way of reducing crime.

Importantly, the findings here invite further detailed investigation of defendants' over- or under-representation across IMD quintiles and/or ONS area classifications (ONS, 2011)

¹⁵ The process of making a potential crime target less attractive, more difficult and/or riskier to attack in order to discourage offending and thus reduce crime.

for specific offence groups, accounting for offender characteristics, specialisation and previous convictions. These insights could inform further theoretical developments and provide policy evidence for reducing (re-)offending, a strategic MoJ Area of Research Interest (MoJ, 2020).

References

ADR UK, Data First: Harnessing the potential of linked administrative data for the justice system, (2020). < <https://www.adruk.org/our-work/browse-all-projects/data-first-harnessing-the-potential-of-linked-administrative-data-for-the-justice-system-169/> > date accessed 6th September 2021.

Bernasco, W. (2010). A sentimental journey to crime: effects of residential history on crime location choice. *Criminology*, 48(2), 389-416. <https://doi.org/10.1111/j.1745-9125.2010.00190.x>

Brantingham, P. and Brantingham, P. (1993). Environment, Routine, and Situation: Toward a Pattern Theory of Crime. In R. V. Clarke and M. Felson (Eds.) *Routine Activity and Rational Choice*. Brunswick, NJ: Transaction Publications.
<https://doi.org/10.4324/9781315128788-12>

Brunton-Smith, I. and Hopkins, K. (2013). The factors associated with proven re-offending following release from prison: findings from Waves 1 to 3 of SPCR. *Ministry of Justice Analytical Series*. Retrieved from:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/491119/re-offending-release-waves-1-3-spcr-findings.pdf

Cole, S. (2019). Social and physical neighbourhood effects and crime: Bringing domains together through collective efficacy theory. *Social Sciences*, 8(147), 1-18.
<https://doi.org/10.3390/socsci8050147>

Department of Communities and Local Government. (2015). The English Indices of Deprivation 2015. Retrieved from:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/465791/English_Indices_of_Deprivation_2015_-_Statistical_Release.pdf

Dorminey, J., Fleming, A., Kranacher, M. and Riley, R. (2012). The Evolution of Fraud Theory. *Issues In Accounting Education*, 27(2), 555-579. <https://doi.org/10.2308/iace-50131>

Fellegi, I.P. and Sunter, A.B. (1969). A Theory for Record Linkage. *Journal of the American Statistical Association*, 64(328), 1183-1210.

<https://doi.org/10.1080/01621459.1969.10501049>

Francis, B., Soothill, K. and Fligelstone, R. (2004). Identifying Patterns and Pathways of Offending Behaviour: A New Approach to Typologies of Crime. *European Journal of Criminology*, 1(1), 48-87. <https://doi.org/10.1177/1477370804038707>

Gov.uk, Ministry of Justice: Data First, (2021). <<https://www.gov.uk/guidance/ministry-of-justice-data-first>> date accessed 4th September 2021.

Government Analysis Function, Splink: MoJ's open source library for probabilistic record linkage at scale, (2020). <<https://www.gov.uk/government/publications/joined-up-data-in-government-the-future-of-data-linking-methods/splink-moj-s-open-source-library-for-probabilistic-record-linkage-at-scale>> date accessed 4th September 2021.

Hirschfield, A., Newton, A. and Rogerson, M. (2010). Linking Burglary and Target Hardening at the Property Level: New Insights Into Victimization and Burglary Protection, *Criminal Justice Policy Review*, 21(3), 319–337.

<https://doi.org/10.1177/0887403409356965>

Hunter, J., Garius, L., Hamilton, P. & Wahidin, A. (2018). Who steals from shops, and why? A case study of prolific shop theft offenders in an English Core City. In: V. CECCATO and R. ARMITAGE (eds). *Retail Crime: International Evidence and Prevention*. London: Palgrave MacMillan. 71-97. <https://doi.org/10.1007/978-3-319-73065-3>

Kearns, A., Livingston, M., Galster, G. and Bannister, J. (2019). The effects of neighbourhood offender concentrations on the number, type, and location of crimes committed by resident offenders. *British Journal of Criminology*, 59(3), 653-673.

<https://doi.org/10.1093/bjc/azy065>

Ministry of Housing, Communities and Local Government. (2019). The English Indices of Deprivation 2019. Retrieved from:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/835115/loD2019_Statistical_Release.pdf

Ministry of Housing, Communities and Local Government. (2020). People living in deprived neighbourhoods. Retrieved from:

<https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/demographics/people-living-in-deprived-neighbourhoods/latest/downloads/imd-and-domains-of-deprivation.csv>

Ministry of Justice. (2020). Areas of Research Interest. Retrieved from:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/944000/areas-of-research-interest.pdf

Mustaine, E.E. and Tewksbury, R. (2005). Social Disorganization and Residential Locations of Registered Sex Offenders: Is This a Collateral Consequence? *Deviant Behavior* 27(3), 329-350. <https://doi.org/10.1080/01639620600605606>

Mustaine, E.E., Tewksbury, R. and Stengel, K. (2006). Residential location and mobility of registered sex offenders. *American Journal of Criminal Justice*, 30(2), 177-192.

<https://doi.org/10.1007/BF02885890>

Office for National Statistics, 2011 residential-based area classifications, (2011)

<<https://www.ons.gov.uk/methodology/geography/geographicalproducts/areaclassification/s/2011areaclassifications>> data accessed 4th September 2021.

Sampson, R.J. (2006). Collective efficacy theory: Lessons learned and directions for future inquiry. In F.T. Cullen, J.P. Wright, & K. Blevins (Eds.) Taking stock: The status of criminological theory. *Advances in criminological theory*, 15, 149-167.

<http://doi.org/10.4135/9781412959193.n220>

Shaw, C.R. and McKay, M.D. (1942). Juvenile Delinquency and Urban Areas. *Chicago: Chicago University Press*. <http://doi.org/10.4135/9781412959193.n228>

Steffensmeier, D. and Haynie, D. (2000). Gender, structural disadvantage, and urban crime: do macrosocial variables also explain female offending rates? *Criminology* 38(2), 403-438. <https://doi.org/10.1111/j.1745-9125.2000.tb00895.x>

Appendix A

Quality of deduplication and linking

All references to the number of times a defendant has returned to the court are based on probabilistic data linkage rather than directly from the recorded data. This estimation has been calculated through a process of deduplication in both the magistrates' courts and Crown Court datasets, and linkage between the two using the [Fellegi-Sunter model](#). The model looks at pairs of records as an input, and outputs a score between 0 and 1 depending on the likelihood the two records are a match. The result of the deduplication process is a dataset which provides an estimate of all matching records belonging to the same person. This information is what allows estimates on the number of times a defendant has returned to the courts to be made.

The quality of the deduplication can only be as good as the quality of the raw data collected by the courts and police. Inconsistent recording of personal details, and a lack of truly distinct identifiers collected in the court system, makes identifying records belonging to the same person more challenging. Even when the data are good quality and two records look convincingly like a match, there is no other possible way beyond what is in the data to confirm that these are same person.

The criminal courts do not assign unique and consistent IDs to all court users in the course of processing cases, so some estimate through linkage must be used. Even where administrative systems do attempt to track individuals it is important to remember there will always be sources of error through mis-recording and duplicate or conflated IDs being issued; and sources of bias, with some individuals systematically less likely to be correctly identified with their previous appearances by the services they interact with (for example, not having a national insurance number or having changed their name or address).

A quality assurance process has been carried out on the deduplication to assess the accuracy of the matches. The process takes a random sample of records from the original dataset, and for each record a manual search to find every single record which would be the same is carried out. This is a fundamentally different process and methodology which is much more human intensive than Splink. When a record is thought to be the same, a

fractional score between 0 and 1 is given which gives an idea on the certainty of the human guess. A key difference between the methodology of the manual quality assurance process and Splink is that the human manual labelling can look at more than just pairs of records at the same time. In contrast, Splink initially looks for pairs of matching records, and at a later point accounts for transitivity of record comparisons. This comparison of clerically reviewed records compared with records matched by Splink allows an estimate on the accuracy of the deduplication to be made.

Figure 7: Receiver operating characteristics curve, Splink compared with clerical and rules-based matching¹⁶

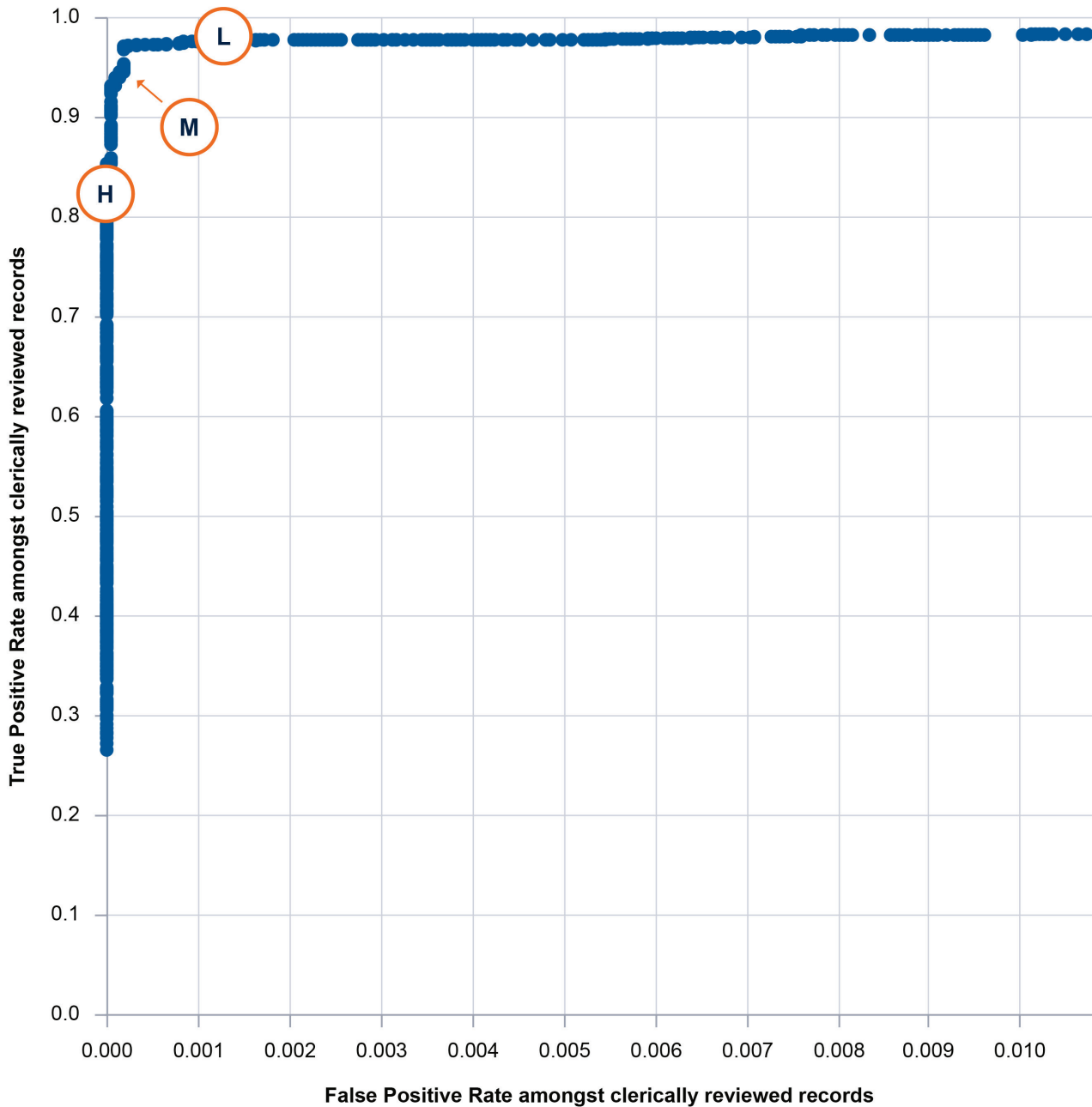


Figure 7 displays accuracy statistics from the labelled magistrates' courts records using a receiver operating characteristic (ROC) curve. Three thresholds have been chosen using the ROC curve to illustrate the trade-off between false positives and false negatives. A higher threshold results in fewer false positives, but also fewer true positives, whereas

¹⁶ Labels L, M and H on Figure 7 relate to the different thresholds as shown in Table 1 below. 'L' stands for low, 'M' for medium and 'H' for high.

lower thresholds allow more true positives to be captured at the expense of false positives creeping in.

The first threshold ('H' for 'high') has a true positive rate amongst clerically reviewed records at 85%, with no false positives detected in the sampled records. This is the high threshold which estimates around 6.7 million distinct defendants in the magistrates' courts data from 2011 to 2020. The low ('L') threshold shown in Figure 7 has a higher true positive rate, but consequently higher false positive rate of around 1 false positive for every 250 true positives. There are an estimated 6.2 million distinct defendants using the low threshold. The medium ('M') threshold has a relatively high true positive rate, with a reasonably low false positive rate of around one false positive for every 1000 true positives, which in total gives an estimated 6.3 million distinct defendants.

The medium threshold is the most reasonable estimate and has subsequently been used for this report and in the shared ONS SRS data. The range of estimated distinct defendants between the low and high thresholds is because of the quality of administrative magistrates' courts data. Similarly, the range of between around 620,000 to around 640,000 distinct defendants is a result of ambiguities in some records in the Crown Court data. This is shown in Table 1.

Table 1: Estimated number of distinct defendants

Threshold	Magistrates' courts distinct defendants	Crown Court distinct defendants
Low	6,177,352	619,998
Medium	6,320,831	624,413
High	6,680,727	639,007
Total records	13,357,982	1,000,827