Data Explained: Understanding the nature, extent and outcomes of serious and organised crime cases heard before the Crown Court in England and Wales

Author: Dr Tim McSweeney

Date: August 2022

This Data Explained output summarises experiences and learning from working with the Ministry of Justice Data First Crown Court and magistrates' courts defendant case level and prisoner custodial journey level linked datasets in the course of producing research into the nature, extent and outcomes of serious and organised crime cases. This publication is intended to help guide future researchers using this data and to provide feedback into dataset development and documentation.

The data discussed in this Data Explained was made securely available via the Data First programme: a ground-breaking data linkage initiative, led by the Ministry of Justice (MoJ) and funded by ADR UK. The data used in this research project comes from the magistrates' courts defendant case level dataset, the Crown Court defendant case level dataset and the prisoner custodial journey level data and was accessed through the Secure Research Service (SRS) hosted by the Office for National Statistics (ONS). The data was not originally collected for research

Initial research questions

Drawing on these resources as part of an inaugural [Data First Fellowship](#), the project aimed to answer the following research questions (RQ):
- What is the nature and extent of serious and organised crime (SOC) heard before the Crown Court in England and Wales between 2013 and 2020?
- How much cumulative crime harm did these SOC cases account for in the higher courts' caseload during this period?
- Was this crime harm equally distributed across (i) offence types, (ii) the different groups involved in SOC, and (iii) different locations?
- Was there an association between cases involving SOC and the likelihood of Crown Court proceedings being discontinued, dismissed or a defendant being acquitted?
- Among SOC cases, which factors (linked to defendant characteristics, group size, main offence, and location) were predictive of Crown Court proceedings being discontinued, dismissed or a defendant being acquitted?
- Was there an association between involvement in SOC and repeat appearances before the criminal courts?

Datasets and variables used

The study drew upon a range of Data First resources, including the:
- magistrates' court defendant case level dataset
- Crown Court defendant case level dataset
- prisoner custodial journey level dataset.

Details of the key variables considered and examined when answering each of the project's main RQs, by source, are set out in the table below.

Research question (RQ)	Dataset used	Variable name	This variable was used to…
RQ1	Crown Court	*'estimated_defendant_ids'*	Calculate the number of defendants in a case, in order to identify and describe the prevalence and incidence of SOC defendants and cases.
		'disposal_code'	Identify custodial sentences imposed, in an effort to identify SOC defendants using the definition developed by Francis and colleagues (2013).
		'duration1', 'units1', 'duration2' and 'units2'	Determine the length of any disposals imposed (in years), as part of the process to identify SOC defendants.
		'offence_ho_code_mso'	Isolate the Home Office code of the offence that was flagged as the most serious at commital, to identify and describe any offences associated with SOC.
		'age_at_committal', 'sex' and 'ethnicity_self_defined_group'	Describe the demographic characteristics of SOC and non-SOC defendants committed to the Crown Court.
		'remand_on_committal'	Describe the defendant's remand status when committed to the Crown Court and any differences between SOC and non-SOC defendants.
		'plea_rank_desc_dc'	Determine the proportion of defendants entering a guilty plea and the extent to which this varied by SOC status.
		'first_hearing_date', 'last_hearing_date' and 'case_total_hearings'	Calculate the duration of the trial and total number of hearings associated with it, and assess any differences between SOC and non-SOC defendants and cases.
		'convicted_rank_desc_dc'	Identify the proportion of SOC and non-SOC defendants convicted at Crown Court.
RQ2	Crown Court	*'offence_ho_code_mso'*	Recode the relevant Home Office code for the offence that was flagged as the most serious at commital to its corresponding ONS Crime Severity Score (CSS).

RQ3	Crown Court	*'offence_ho_code_desc_mso'*, *'case_id_hash'* and *'LSOA_residence'*/*'LA_residence'*	Determine the extent to which cumulative crime harms – as measured using ONS CSS – varied according to: (i) SOC offence types, based on the Home Office code that was flagged as the most serious at committal; (ii) the presence of a SOC element within a case, using a unique identifier for the case which can be shared between multiple defendants; and, (iii) different geographical locations (including associations with indices of deprivation), as determined using data based on the local authority and area of residence for the defendant on committal to court, at the lower super output area (LSOA) level.
RQ4	Crown Court	*'disp_title'*, *'disposal_ho_code'* *'convicted_rank_desc_dc'*, and *'disposal_ho_code_desc'*	Describe the main types of disposal imposed by the Crown Court and to test for associations between SOC status and these outcomes, including discontinuation, dismissal and acquittal.
RQ5	Crown Court	*'age_at_committal'*, *'sex'* *'ethnicity_self_defined_group'*, *'case_id_hash'*, *'offence_ho_code_mso'*, *'LSOA_residence'*/*'LA_residence'*, *'disp_title'*, *'disposal_ho_code'*, *'convicted_rank_desc_dc'*, and *'disposal_ho_code_desc'*	Identify those factors (linked to defendant characteristics, group size, main offence, and location) which were predictive of Crown Court proceedings being discontinued, dismissed or a defendant being acquitted.
RQ6	Crown Court	*'estimated_defendant_id'*, *'receipt_date'*, *'first_hearing_date'*, *'convicted_rank_desc_dc'* and *'outcome_date'*	Test for any association between involvement in SOC and repeat appearances before the Crown Court.
	Magistrates' court	*'estimated_defendant_id'*, *'offence_date'*, *'arrest_date'*, *'initiation_date'*, *'first_hearing_date'*, and *'final_disposal_ho_code'*	Test for any association between involvement in SOC and repeat appearances before the magistrates' courts.
	Prisoner journey	*'row_id_hash'*, *'estimated_mc_cc_ps'*, *'offender_id_hash'*, *'first_movement_date'*, *'first_sentenced'*, *'imprisonment_status_category'*, and *'effective_release_date'*	Accurately assess the 'time at risk', post-release, of those sentenced by the Crown Court to a period of imprisonment between 2013 and 2020.

Data limitations encountered

Given that the Data First linked criminal courts data is derived from case management systems, there are inevitably some limitations encountered when undertaking secondary analyses for research purposes. The main limitations encountered by the project – and discussed in more detail below – were largely due to the absence of:

- a dedicated SOC flag or marker within the datasets
- information on the full range of offences being prosecuted before the court
- any data on complainants and victims
- details of aggravating and mitigating factors which may have been relevant to sentencing decisions.

An important data limitation encountered by the project is that there are no designated flags or markers to identify or highlight SOC defendants and cases within the respective case management systems currently used by the magistrates' courts (LIBRA) and Crown (XHIBIT) courts, and upon which the Data First criminal courts' datasets are derived. This means that such a flag had to be constructed using existing fields within the datasets. Importantly, the absence of a relevant SOC marker means that individuals who may have been involved in SOC-related offences, but were prosecuted in isolation from other group members during this period, will not be captured by the current study. This in part reflects the current study replicating an existing definition of organised crime set out in previous Home Office funded research.

The linked criminal courts' datasets provide information on only the most serious offence at the point of committal to court and sentencing. As such, they underestimate the full range of offending that may be attributable to defendants (either acting alone or as part of a wider group) and the scale of the crime harms attributable to them. The absence of this information also restricts the ability of researchers to explore any associations and correlations between different types of offences within the cases being heard before the courts (e.g., where drug supply or importation offences also involve an element of violence, exploitation, or the presence of weapons).

Data on defendants' area of residence (at the point of committal) is a valuable asset which enables the geographic distribution of offences and harms to be quantified and mapped (and for any associations with indices of deprivation, for instance, to be assessed), while preseving the anonymity of court users' data. However, it is not possible to determine where these harms are being directly experienced or felt, or whether some individuals, groups or communities are disproportionately affected by them, as information relating to complainants and victims are not captured or available through these Data First assets.

The datasets also lack any information on the range of aggravating and mitigating factors which may be relevant to a case or individual defendant, and which should be considered within sentencing.

Finally, and more broadly, as a project reliant on administrative data, it is important to acknowledge that the offending which comes to official attention through the courts will be influenced and subject to a range of biases associated with the way in which different actors within the criminal justice process identify, report, record and respond to crime. Any such biases will therefore be reflected in the data and findings produced using this and other linked data sources.

Necessary modifications to initial research questions or research design

One amendment to the originally proposed design of the study involved accessing and analysing records from the prisoner custodial journey dataset in order to allow for a more accurate identification of custodial entry and release dates for those sentenced to imprisonment by the Crown Court between 2013 and 2020. Without this information, the study would be required to estimate time spent in custody in a less robust and rigorous way (i.e., by assuming release would automatically be triggered for all prisoners at the midway point of their custodial sentence). The prisoner journey dataset was not available for use when the project was initially accredited.

Given the contested and nebulous nature of existing definitions surrounding SOC, and in the absence of a specific marker for this within the criminal courts' datasets, discussions around these definitional issues evolved during the course of the project. Though not resulting in a direct modification to the research questions or the overall design of the study, these discussions were informed through direct stakeholder engagement and dialogue which was facilitated by ADR UK and the MoJ as an integral component of the Data First Fellowship. These forms of engagement indicated that a sharper focus on the definition of SOC enshrined in recent legislation may serve to enhance the potential utility and policy relevance of the research. An important strength of Data First is that the datasets allow for the testing and application of alternative definitions of SOC.

Similarly, debates around measuring and quantifying 'crime harms' occupy an equally contested space. Two prominent indicators were considered as suitable proxy measures: ONS Crime Severity Scores (CSS) and the Cambridge Crime Harm Index (CCHI). The limited comparative research that has been undertaken to date involving these two measures suggests that while both have their relative strengths, they also have different weaknesses and can produce substantially different estimates of crime harm or severity. For the current project the CSS measure was adopted as the preferred proxy measure of severity largely on practical grounds: it offered far more coverage than the CCHI in terms of the number of offences it accommodates. The CCHI is derived from starting points outlined in sentencing guidelines, but these do not exist for all offences. Consequently, there was not a corresponding CCHI score for almost half the Home Offence codes contained within the Crown Court dataset.

Necessary modifications to the data

Following discussion with the Data First team, records relating to hearings transferred out from the Crown Court, cases involving an appeal and those with missing data on the most serious offence were excluded from the analyses. Observations relating to both police and defendants' self-defined ethnicity, for example, had disproportionately high rates of missing data.

Recommendations to data owners

The key recommendations to the data owners are to consider the scope that may exist for addressing the four main limitations outlined above. The most pertinent of these, given the focus of the current project, would be for the data owners – in consultation with other relevant stakeholders like the Home Office - to consider the feasibility of developing and testing a dedicated SOC flag or marker for use within the datasets. It is likely that demand for other markers or flags (e.g., linked to identifying involvement in 'gangs', 'county lines', joint enterprise offences, etc.) may also emerge from this work. The various Data First resources would seem to offer a valuable and unique test bed for trialling, adapting and developing such activity.

Developing a 'time spent in custody' or 'time at liberty' concept is likely to be an important consideration for future studies drawing on the Data First linked criminal courts data which feature court reappearances as an outcome. This seems especially important should any of these studies adopt a quasi-experimental or comparative design.

The current project, for example, sought to compare the rate and frequency of court reappearances by SOC-related and other defendants over a defined period of time. In order to do this, it was important to be confident that both groups had equivalent 'time at liberty', post-release from any custodial sentence imposed, in order for them to be able to reoffend and reappear at court.

A failure to adequately measure and control for this could lead us to incorrectly conclude that a lower prevalence or incidence of reappearances among one group was a proxy for lower rates of reoffending, when in fact this may be due to some of these individuals still being in custody and therefore at reduced risk of reoffending and reappearing at court.

The prisoner journey dataset is a valuable resource which enables accredited researchers to measure and control for this. If this resource cannot be used to determine prison release dates at an individual level, then some less reliable assumptions (i.e., that imprisoned offenders were eligible for release after serving half their sentence) may have to be made using information on sentence dates and length available within the criminal courts data as an alternative measure.

Accredited researchers accessing only the Crown Court dataset are likely to benefit from access to variables relating to dates during which offences were committed ('offence_date') contained within the magistrates' court defendant case level dataset. The integration or importation of these variables as a standard feature of the Crown Court dataset, for instance, would be particularly useful when dealing with and adjusting for 'pseudo-convictions', whereby sanctions are imposed or acquired for historical offences.

Additional data which would help to further develop the research.

On a practical level, a regularly updated and approved directory deposited within the ONS Secure Research Service could provide a useful resource for accredited researchers looking to source and integrate a wide range of relevant lookups, code or data (e.g., relating to things like official population estimates, indices of multiple deprivation, crime severity scores, police force area boundaries, etc.).

In the longer term, Data First may create opportunities for new data linkages involving court users which can in turn fill important knowledge gaps and help address a range of government research priorities. This includes potential for integrating data from sources like the Police National Computer (PNC) and the prisons and probation Offender Assessment System (OASys) in order to better understand the demographic characteristics of defendants (including their nationality for instance), their social circumstances, known criminal histories and any offending-related needs (for example, as they relate to substance misuse, thinking, attitudes, lifestyle or relationships), and improve our understanding of risk factors for involvement in different types of offending, including SOC.

Acknowledgements

Contact

Dr Tim McSweeney

Email: t.mcsweeney@herts.ac.uk