



Data Explained: An investigation into racial bias in court case outcomes in England and Wales

Author: Dr Angela Sorsby

Date: May 2022

This Data Explained summarises experiences and learning from working with the Ministry of Justice Data First Crown Court and magistrates' court defendant case level datasets. It intends to help guide future users of this data and feedback in dataset development and documentation.

The administrative data used by Data First to produce these datasets was originally collected for the purposes of administering the justice system rather than for statistical research purposes. Production of the Data First datasets required data from different systems and organisations to be merged and processed. Inconsistencies in, for example, definitions and data formats between different information management systems and over time mean that bringing together data from different systems is a highly complex process for which best practice is still being developed.

Making the data available to the research community will result in increased understanding of the source data and the data's strengths and weaknesses for research purposes. However, as the data was not originally intended for statistical research, it is to be expected that there will be various challenges, a number of which are detailed in the following document.

Project details

In this project the datasets were used to investigate ethnic and gender disparities in experiences of the criminal justice system. Previous research has found that people of minority ethnic backgrounds are overrepresented within the criminal justice system in England and Wales (Phillips and Bowling, 2017: 196). However, prior to the Data First programme, detailed data on individual court cases which could be used in ethnicity research was not generally available for academic research purposes. As a result, available statistics on ethnic and gender disparities in court case outcomes don't generally control for other factors such as age and type of offence. The de-identified magistrates' and Crown Court datasets give access to case level administrative data on criminal court use. They include information on things such as age, type of offence, whether there are co-defendants, plea and mode of trial.



This project uses regression analysis to establish whether ethnic disparities remain after taking account of things such as age, plea, offence type and offence severity. In terms of court case outcomes, I look at whether the case results in a conviction and, where convicted, the severity of the sentence. I also examine whether patterns are the same or different for men and women.

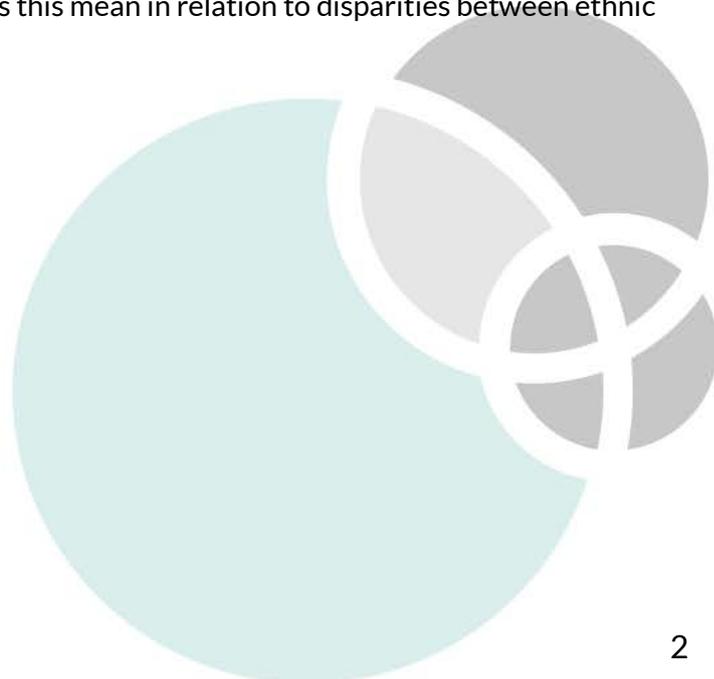
As there are indications from previous research that defendants from ethnic minority backgrounds are more likely to elect to be tried in the Crown Court (MOJ, 2019: 38), I examine whether there are differences between the courts in terms of outcomes, providing more information on the consequences of electing for trial in the Crown Court.

I also look at patterns over time, such as whether people from particular ethnic groups are repeatedly acquitted and the number of magistrates' court cases in which people appear before there is a Crown Court case.

Ethnicity data is collected by the police. There are two methods of recording ethnicity: officer identified ethnicity and self-identified ethnicity. Officer identified ethnicity is recorded by a police officer or by administrative staff and is based on visual appearance. Self-identified ethnicity is collected during a custody interview. The defendant is asked to identify their own ethnicity from a list of 16 options. The information is input into the police administrative system and passed to Her Majesty's Courts and Tribunal Service to record in the court administrative system.

Initial research questions

- Are there disparities between ethnic groups in:
 - the outcomes of cases after controlling for other factors?
 - the number of times individual people are prosecuted over a specified time period?
 - the proportion of convictions, compared to other outcomes, that a defendant receives?
 - the number of cases in which a defendant appears in the magistrates' court prior to appearing in a crown court case?
- Are disparities the same or different for men and women?
- Are the outcomes of 'triable either way' cases in the magistrates' and Crown Court different, after taking account of other factors, and what does this mean in relation to disparities between ethnic groups?



Datasets and variables used

Data First Crown Court defendant case level dataset variables used:			
ID variables	estimated_defendant_id	case_id_hash	defendant_in_case_id_hash
Personal characteristics of defendant	ethnicity_police_defined_group	ethnicity_self_defined	ethnicity_self_defined_group
	age_at_committal	sex	LSOA_residence
Court information	court_id		
Offence variables	offence_code_msd	offence_ho_code_msd	offence_ho_group_msd
	offence_code_mso	offence_ho_code_mso	offence_ho_group_mso
Proceedings variables	case_type	receipt_type	receipt_date
	bail_status_after_last_hearing	bail_status_before_first_hearing	comm_bc_status
	either_way_type_desc		
Outcome variables - principle offence	disp_title	disposal_ho_code	disposal_date
	disposal_ho_code_desc	disposal_type	disposal_ho_group_desc
	duration1	units1	monetary
Data First magistrates' court defendant case level dataset variables used:			
ID variables	estimated_defendant_id	case_id_hash	defendant_in_case_id_hash
Personal characteristics of defendant	sex	age_at_offence	ethnic_appearance
	ethnicity_police_defined_group	ethnicity_self_defined_group	LSOA_residence
	ethnic_assessment		
Offence variables - principle offence	offence_date	cjs_offence_code	final_offence_ho_code
	offence_ho_group		
Offence variables - most serious offence	cjs_offence_code_mso	final_offence_ho_code_mso	offence_ho_group_mso
Proceedings variables	court_code	final_result_date	initiation_date
	proceeding_type	mode_of_trial_code	plea_code_desc
	remand_a_status_desc	remand_c_status_desc	remand_p_status_desc
	remanded_cust		
Outcome variables relating to - principle offence	result_desc_full	final_disposal_ho_code	primary_disposal_days
	disposal_monetary_value_pounds	jsas_result_group_desc	

MoJ Data First criminal courts and prisons linking dataset variables used:

All variables used

Data limitations encountered

There is limited ethnicity data in the magistrates' court dataset. Ethnicity is 'not stated' for many (indeed the majority) of cases in all the data fields relating to ethnicity in the magistrates' court dataset. This is particularly the case for summary offences. As ethnicity data is collected by the police this is to be expected because for many summary offences the defendant may not be seen by the police or appear in court.

Ethnicity is missing for the majority of appeal cases in the Crown Court dataset. This was not an issue for this project as there was no intention to study appeal cases. It would however be an issue for a project which intended to focus on or include appeal cases.

Ethnicity is available for a much greater proportion of trial and sentencing cases in the Crown Court dataset. Nevertheless, although ethnicity is stated for the majority of these cases, it is still 'not stated' for a substantial number of cases.

Ethnicity is more likely to be 'not stated' for some Crown Courts compared to others.

Ethnicity is more likely to be 'not stated' for certain types of offence such as fraud.

The datasets include only cases from 2011 onwards for the magistrates' court dataset and 2013 onwards for the Crown Court dataset. There is therefore no available data for defendants' previous convictions prior to these dates. The linked magistrates' and Crown Court dataset therefore only contains all convictions from 2013 onwards.

The disposals recorded as 'Custody for life', 'Detained during her majesty's pleasure', 'Detention for life under s.226', and 'Life imprisonment' in the field *disp_title* in the Crown Court dataset never have a duration. This is to be expected, as these sentences are of indeterminate duration but it is therefore not possible to include these sentences in establishing for example average sentence length for a particular type of offence, if one uses only the data contained in the dataset.

Incarceration periods for the magistrates' and Crown Court datasets are for the periods given by the court rather than the period which is actually served. More information on the sentence served is now available in the Data First Prisoner Custodial Journey Dataset. That dataset was not available when I formulated and started my project, so I did not have the time resources to use it.

In both the magistrates' and Crown Court datasets it is not possible to compare the severity or intensity of different types of community sentence. In relation to some types of community sentence (e.g. mental health treatment) the duration seems to relate to the time period over which the order

extends rather than the amount of input required from the defendant; while for other types of order, such as unpaid work, the duration is the time input required from the defendant. In addition, for some types of community sentence (e.g. in the Crown Court: 'Supervision Requirement' and 'Programmes aimed at changing offender behaviour') duration is never provided. This may be because the requirements are to meet up with an offender manager on a number of occasions or to complete a set of specified courses, in which case a defined duration may not be appropriate. It should be noted that comparing the severity of community sentences is not just limited by the available data. Comparing the severity of community sentences is always going to be problematic considering such sentences take so many different forms and for some forms the aims are more about punishment while for others the aims are more about rehabilitation.

In addition, in relation to community sentences, it is not possible to specifically identify 'Unpaid work' in the magistrates' court dataset. The qualifier code BK (Perform Unpaid Work for a number of hours - Minimum 40 maximum 100) is rarely or never used and the result code 1155 (Community Payback Order) in the field *result_qualified* never appears to be used.

Necessary modifications to initial research questions or research design

The time period for the Crown Court dataset is 1 January 2013 to 31 December 2020. Hence, Crown Court convictions are only available for cases received from 1 January 2013. This means that the combined magistrates' and Crown Court datasets only have information on all convictions from 2013 onwards (even though the magistrates' court dataset has cases initiated from 2011 onwards). In view of this, in relation to research questions where it was desirable to control for the number of previous convictions (are there disparities between ethnic groups in the outcomes of cases after controlling for other factors?), or where it was necessary to count the number of convictions (the proportion of convictions, compared to other outcomes, that a defendant receives), analyses were undertaken using only cases from 2013 onwards and only defendants who were born in 1995 or later (calculated using *receipt_date* and *age_at_committal* for the Crown Court dataset and using *offence_date* and *age_at_offence* for the magistrates' court dataset). These defendants were under 18 on the day before the start of the time period for which the Crown Court data relates. The datasets therefore include all adult convictions for these defendants. A consequence of this is that these analyses relate only to defendants aged 10 to 25. The analyses do not however relate to all defendants in this age band as there are defendants who are 25 or under at the start of the time period to which the data relates but were born prior to 1995.

In order to include all convictions (youth and adult) separate analyses were undertaken using only defendants who were born in 2003 or later and were therefore under 10 on the day before the start of the time period for which the Crown Court data relates. A consequence of this is that these analyses relate only to defendants aged 10 to 17. The analyses do not however relate to all defendants in this age band as there are defendants who are 17 or under at the start of the time period to which the data relates but were born prior to 2003.

In relation to the research question - are there disparities between ethnic groups in the number of cases in which a defendant appears in the magistrates' court prior to appearing in a Crown Court case? - it is possible to identify from the 2011 and 2012 magistrates' court data whether cases are sent to the Crown Court using the field *jsas_result_group_desc*. However, in view of the limited ethnicity data in the magistrates' court dataset and because the outcome of cases is not known for

2011 to 2012 cases sent to the Crown Court, I decided to focus on cases from 2013 onwards in relation to this research question.

In calculating offence severity (see below) and in examining differences between ethnic groups and between men and women in sentence length it is not possible to include the disposals 'Custody for life', 'Detained during her majesty's pleasure', 'Detention for life under s.226', and 'Life imprisonment' as these sentences never have a duration.

In examining discrepancies between ethnic groups and gender disparities the project aims to control for offence severity. Currently available offence severity scales such as the Cambridge Crime Harm Index and the ONS Crime Severity Score do not cover all offences. I therefore decided to adapt the method used by the ONS Crime Severity Score to produce my own scale using the data in the magistrates' and Crown Court datasets to produce a score for each of the Home Office Offence codes which appear in the datasets.

The ONS method produces a score or weight for each of the offences that it covers (notifiable offences) by using a combination of the proportion of disposals for the offence that are dealt with by way of custody, community sentences and fines and the mean length or value of each of these. In relation to custody the method uses mean length in days but in relation to fines and community sentences it converts the sentence length to a custodial equivalent. The data used by the ONS to produce the ONS Crime Severity Score did not have the length of community orders and community orders were not broken down by type. The ONS therefore used sentencing guidelines and based the weight on the midpoint of unpaid work for the appropriate level of sentence (high, medium or low) that the offence would receive. This method assumes that all community sentences for the same crime are of equivalent severity.

In calculating my own severity score I had initially thought to use the actual length of all community sentences but realised that doesn't work for a number of community sentences where the duration given seems to relate to the time period over which the sentence extends rather than the amount of input from the defendant (e.g. mental health treatment). In addition, in relation to some types of community sentence there is no duration in the datasets, as set out above.

I therefore decided that in calculating my own offence severity score it would probably be best to base the length of community sentences just on unpaid work and assume that all community sentences for the same crime are of equivalent severity (as the ONS method does). However, while it is possible to identify sentences of unpaid work in the Crown Court dataset, it is not possible to specifically identify 'Unpaid work' in the magistrates' court dataset, as set out above. It has therefore been necessary to only use sentences of unpaid work in the Crown Court to calculate the severity of community sentences. This means that it is not possible to establish an average order length for the few summary only offences that receive community sentences. At the time of writing this work was still ongoing but it was intended to assume that summary only offences which received community sentences received the lowest level of community sentence.

Necessary modifications to the data

In the Crown Court dataset there are some cases with receipt dates prior to 2013. I was advised that these cases should be excluded from any analysis as the data does not provide a full picture of cases prior to 2013 for the Crown Court dataset. These cases were therefore removed prior to analysis.

I was advised that cases which are recorded as 'transferred_out' in the field disposal_type are duplicate cases. These cases were therefore removed prior to analysis.

In the Crown Court dataset there are quite a number of defendants with more than one row of data with the same disposal date. These rows have the same *estimated_defendant_id* but the *case_id_hash* is different for the different rows. It is not entirely clear whether these should be treated as separate cases. In relation to some of these defendants the outcomes and offences are quite different for the different rows while in other cases the outcomes and/ or offences are the same for the different rows. In a large proportion of cases where there are multiple rows with the same *estimated_defendant_id* and disposal date, the multiple rows link to only one magistrates' court case. Where that is the case, in my analysis, I have used only the main Crown Court case which links to the one magistrates' court case. I defined the main Crown Court case as the case with the most serious disposal which reached the match threshold, as indicated by the *match_threshold_flag*. Where the multiple rows of Crown Court data match to more than one magistrates' court case I have treated these as separate cases.

In the Crown Court dataset, the fields *plea_rank_dc* and *plea_rank_desc_dc* cannot be used to determine whether or not the defendant pleaded guilty to the offences that are recorded in the offence variable fields relating to the most serious disposal or the offence variable fields relating to the most serious offence. A guilty plea will be recorded in *plea_rank_desc_dc* if the defendant has pleaded guilty to a different offence in the case. Hence plea was obtained from the field *plea_code_desc* for the corresponding case in the magistrates' court dataset.

There is a problem in relation to the field *remand_c_status_desc* in the magistrates' court dataset. The vast majority of cases are recorded as '04 Mixed (custody at first hearing, bail later)' in this field. It is not possible that this number of defendants have been remanded in custody at some point. In addition, what is recorded in the field *remand_c_status_desc* is inconsistent with data in other fields. Large numbers of defendants coded as '04 Mixed (custody at first hearing, bail later)' in the field *remand_c_status_desc* are recorded as '09 Not applicable' or '01 On bail throughout period' in the field *remand_a_status_desc*. The field *remand_c_status_desc* should not be used.

There is also a problem in relation to the field *remanded_cust* in the magistrates' court dataset. This is the flag denoting if the defendant was remanded in custody at any point during their case. There are a number of cases where *remanded_cust* is recorded as false but the case is recorded as '02 In custody throughout period' in the field *remand_a_status_desc* (this field is for the variable: Remand status during proceedings, e.g. on bail throughout period, in custody throughout period). The field *remanded_cust* should not be used.

There are problems in relation to the field *convicted_rank_dc* in the Crown Court dataset. This field should not be trusted as a source of information and should not be used. I used *disposal_ho_code* instead. Where disposal *disposal_ho_code* is recorded as 999 then it was necessary to use *disposal_ho_code_desc* to obtain the specific disposal.

There are problems in relation to the field *total_defendants_in_case* in the Crown Court dataset. In view of this, the number of defendants involved in the case was obtained by counting the number of different *estimated_defendant_id*'s with the same *case_id_hash*.

Recommendations to data owners

It would be useful to provide information in relation to when different offences are counted as separate cases for a defendant given that there are rows of data with the same disposal date which have the same *estimated_defendant_id* but the *case_id_hash* is different for the different rows. It would be useful to know whether these should be treated as separate cases or the same case.

It would be useful to provide warnings in relation to data fields where issues have been found or even to remove these fields from the datasets to prevent them being inadvertently misused.

Additional data which would help to further develop the research.

It would be very useful for the court data to be linked to probation data. This would provide more information on case outcomes, particularly in relation to community orders. Probation data may enable comparisons to be made between groups of people in relation to the type and intensity of community orders received. This would allow for a fuller examination of differences in court case outcomes for different groups of people. Probation data may also allow for comparisons to be made between different groups of people in relation to recalls and licence conditions. In addition, data on compliance may allow for compliance with previous orders to be included as a control variable in comparing the severity of subsequent sentences.

It would be useful to be able to link to the Offender Assessment System (OASys) data. This would enable better control for offending history using Offender Group Reconviction Scale (OGRS) scores, OASys General Predictor (OGP) Scores and OASys Violent Predictor (OVP) scores. OASys data would also allow for the control of additional variables, such as substance misuse, in making comparisons in court case outcomes between different groups of people.

As an alternative to using offending history data from the Offender Assessment System it would be useful to be able to link to Police National Computer (PNC) data to control for offending history. This would allow for better control of previous offending history than was possible in the present study which only had data on convictions from 2013 onwards. This of course assumes that the PNC data would have full offending history, as opposed to only convictions since a certain time period. PNC data would also allow for comparisons to be made in relation to the administration of cautions as well as using cautions as control variables in other analyses.

If available, it would be useful to have data on the average amount of time served for disposals recorded as 'Custody for life', 'Detained during her majesty's pleasure', 'Detention for life under s.226', and 'Life imprisonment' as well as information on the amount of variability in time served for these sentences.

Acknowledgements

This work was produced using administrative data accessed through the ONS Secure Research Service. The use of the ONS data in this work does not imply the endorsement of the ONS or data owners (e.g. MoJ and HM Courts and Tribunals Service) in relation to the interpretation or analysis of the statistical data.

This work uses research datasets which may not exactly reproduce National Statistics aggregates. National statistics follow consistent statistical conventions over time and cannot be compared to Data First linked datasets.

References

1. Ministry of Justice, (2019). *Statistics on Race and the Criminal Justice system 2018*. London: MOJ. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/849200/statistics-on-race-and-the-cjs-2018.pdf.
2. Phillips, C., and Bowling, B., (2017). *Ethnicities, racism, crime and criminal justice*
3. Liebling, A., Maruna, S., and McAra, L., (eds.). (2017). *The Oxford Handbook of Criminology*. Oxford: Oxford University Press.

Contact

Dr Angela Sorsby

Email: a.sorsby@sheffield.ac.uk