Data Explained: Ethnic inequalities in the criminal justice system

Author: Kitty Lymeropoulou
Date: August 2022

This Data Explained summarises experiences and learning from working with the magistrates’ and Crown Court datasets in the course of producing research into ethnic disparities in the criminal justice system. This publication intends to help guide future researchers using this data and feedback into dataset development and documentation.

The data discussed in this Data Explained was made securely available through a Ministry of Justice Data First Fellowship funded by ADR UK. The data used in this research project comes from the Ministry of Justice and was accessed through the ONS Secure Research Service. The data was not originally collected for research and it is expected that there are gaps and inconsistencies in its recording, a number of which are detailed in the following.

Initial research questions

The research set out to examine the experiences and outcomes of ethnic minority groups in the court system; the relative importance of defendant, case, and court factors in explaining ethnic differentials at different stages of the CJS; and make recommendations about ways of effectively addressing ethnic inequalities. The initial research questions were:

1. How do the experiences and outcomes of defendants in the magistrates’ court and Crown Court differ by ethnicity?
2. Do ethnic disparities persist after controlling for defendant, case and court characteristics?
3. What are the main drivers of any differences by ethnicity in court outcomes, and how much of the gap is attributed to defendant, case and court factors?

Research methodology

The research examined disproportionality using the Relative Rate Index (RRI) to document the extent of disparities in court experiences and outcomes between detailed self-identified ethnic minority groups and the White British. Multilevel modelling was then used to examine the effect of defendant (age, gender, ethnicity, deprivation), case (offence type, offence severity, previous convictions, pre-sentencing remand, plea, no of defendants in cases) and court (court case workload and conviction rate, ethnic density) characteristics on remand and sentencing outcomes, and to determine whether ethnic disparities persist after controlling for these characteristics. Separate models by ethnic group were estimated to examine variations in remand and sentencing outcomes within and between ethnic groups. Ethnic differences in sentencing outcomes were also examined using the Blinder-Oaxaca decomposition method.
Datasets and variables used

The research draws on administrative data on defendant appearances in magistrates’ (2011-2020) and Crown courts (2013-2020) and the linking datasets developed through the Data First programme. Key variables examined include:

<table>
<thead>
<tr>
<th>Individual defendant, case and linking (between magistrates' and Crown Court) identifiers.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethnicity (self-identified and police-defined)</td>
</tr>
<tr>
<td>Age (at committal/offence)</td>
</tr>
<tr>
<td>Sex</td>
</tr>
<tr>
<td>LSOA of residence (matched to the 2015 IMD)</td>
</tr>
<tr>
<td>Police force area</td>
</tr>
<tr>
<td>Court (id/code)</td>
</tr>
<tr>
<td>Offence (most serious disposal/final offence)</td>
</tr>
<tr>
<td>Home Office disposal code</td>
</tr>
<tr>
<td>Home Office offence code</td>
</tr>
<tr>
<td>Plea</td>
</tr>
<tr>
<td>Remand status (on committal to Crown Court, Remand by the Police, Remanded in custody, Remand during magistrates’ court proceeding)</td>
</tr>
<tr>
<td>Total defendants in a case</td>
</tr>
<tr>
<td>Convicted rank</td>
</tr>
<tr>
<td>Case type</td>
</tr>
<tr>
<td>Disposal/final result date</td>
</tr>
<tr>
<td>Unit and length of disposal</td>
</tr>
<tr>
<td>Jsas result group</td>
</tr>
<tr>
<td>Result code and qualifier</td>
</tr>
<tr>
<td>Mode of trial</td>
</tr>
</tbody>
</table>

Additional datasets and variables used:

Indices of Deprivation 2019 Income and Employment domains combined for England and Wales: Income deprivation score for Lower Layer Super Output Areas (LSOAs)

2011 Census: Local Authority population by ethnicity

Cambridge Harm index 2021: High severity offences (with a starting point sentence of three years)
Data Limitation encountered

There are issues with several remand variables, namely, remand_status_a (remand status during proceedings), remand_status_c (remand status on committal to Crown Court) and remand_cust (denoting if the defendant was remanded in custody at any point during their case) included in the magistrates’ courts dataset. Cross-tabulations between these variables and other outcome variables suggest these variables are unreliable. For example, cross-tabulations show that remand_a_status classifies incorrectly defendants in the ‘NA’ category and defendants remanded by the Police and underestimates custody rates. Remand_status_c seems to overestimate custody rates. Cross-tabulations with other variables indicate that defendants in cases for minor offences, and in cases that were not sent to Crown Court for trial or sentencing, have been classified incorrectly as being remanded in custody on committal to Crown Court.

The convicted_rank_dc variable (the overall convicted rank for the defendant’s case) does not capture convictions accurately. A comparison of convicted rank and disposal code shows there are instances where a defendant had a disposal code ‘acquitted or dismissed’ but the convicted rank is ‘convicted’. Similarly, the convicted rank may indicate non-conviction i.e., ‘other or acquitted’ while the disposal code indicates a custodial sentence. For this reason, it is recommended that disposal codes are used to identify defendants with convictions.

Other issues with the data more briefly include:

- Total_defendants_in_case (total number of defendants in the case) doesn’t match the total number of defendants estimated using the defendant ID and case ID.
- There are duplicate cases included in the Crown court datasets relating to cases transferred out.
- The age of defendant variable in the Crown Court datasets includes some extreme (unrealistic) age values.
- The disposal code (offence_ho_group_msd) in the Crown Court datasets includes values not shown (e.g. 0) in the metadata files.
- The offence_ho_group variable in the magistrates’ dataset includes code 13 (not captured in the metadata files).

A comparison of the self-identified and police-defined ethnicity of defendants appearing in magistrates’ and Crown Courts showing how defendants in each self-identified category have been ascribed ethnicity by the police based on visual appearance reveals a high degree of consistency between the aggregated self-identified and police-defined ethnicity classification for White, Black and Asian groups. In contrast, people with a mixed ethnic background and those belonging to ‘other’ groups (other than the White, Black, and Asian groups) are not recorded consistently by the Police. Police-defined ethnicity may, therefore, give a different picture of the experiences of ethnic minority groups in courts than self-identified ethnicity.
Defendants’ ethnicity has a high proportion of missingness, particularly in the magistrates’ courts dataset. There is substantial variation in ethnicity recording across the years covered in the Data First datasets with a marked decline in ethnicity recording over more recent years. Analysis of missing ethnicity reveals that ethnicity coverage varies across courts and is less complete for females compared to males, and for older than younger defendants. There is more substantial variation in the coverage of ethnicity of defendants by type of offence. In the magistrates’ court dataset, the highest proportion of missingness of ethnicity relates to defendants in cases for summary offences. Summary offences make up the majority of cases in the magistrates’ courts which explains the high proportion of defendants with missing ethnicity in the magistrates’ courts dataset since information on ethnicity has primarily been collected for defendants charged by the police.

The linking dataset can be used to join up cases across the magistrates’ and Crown Court datasets to identify defendants with previous convictions, as this can indicate higher culpability and affect sentencing decisions. Since convictions outside the dataset years (2011-2020) cannot be determined, the linking dataset does not capture adequately the defendants’ criminal history.

**Necessary modifications to initial research questions or research design**

The issues with the remand variables in the magistrates’ courts dataset (described above) meant that it was not possible to examine remand during court proceedings. Instead, the analysis focused on remand status before the case starts derived from information on remand by the Police.

It was not possible to examine ethnic disparities in sentencing outcomes for less serious offences because there was a high proportion of defendants with missing ethnicity for these offences. As the majority of cases in magistrates’ courts relate to summary offences and defendants’ ethnicity for summary offences had a high proportion of missingness, the analysis on sentencing outcomes drew only on Crown Court defendants.

**Necessary modifications to the data**

Magistrates’ and Crown Court datasets were not combined because of issues with the reliability and comparability of variables (e.g., remand).

Following feasibility and missing ethnicity analyses which showed differences in ethnicity coverage by offence and large under-recording of the ethnicity of defendants appearing in courts for less serious offences, the analyses excluded summary offences.

Multiple years of the data were combined to examine ethnic disparities in remand and sentencing outcomes to overcome issues relating to the small sample size of some ethnic minority groups in the 16+1 self-identified ethnicity classification.
Recommendations to data owners

- Improve the recording of ethnicity, particularly in the magistrates’ courts dataset and collect information on additional self-identified ethnic groups (e.g., by adopting the 18+1 ethnicity classification which includes Gypsy or Irish Traveller and Arab groups). This is pertinent for understanding the experiences of ethnic minority defendants, given that ethnicity outcomes based on aggregated groups (White’, ‘Mixed’, ‘Asian’, ‘Black’, and ‘Chinese or Other’) mask substantial variation in outcomes between the constituent groups comprising the aggregated ethnic categories.
- Include comparable information (e.g., remand, age at offence) in Crown Court and magistrates’ court variables to enable the two datasets to be combined seamlessly and in a meaningful way.
- The metadata files need improving. The description of the variables in many cases is not sufficient to determine what the variables capture, how they are derived and how they might relate to each other.

Additional data which would help to further develop the research

- Information on defendants’ criminal history beyond the years captured in the datasets.
- More detailed information on defendants’ personal circumstances such as socio-economic status, education, employment, and illness/disability.
- More detailed information on aggravating and mitigating factors including victim harm, offence seriousness and conviction type (e.g., Joint Enterprise)
- Whether a pre-sentence report (PSR) was requested by the court, and information on the risk and needs assessment of the defendant, the offence(s) committed and sentence proposal.
- Information on legal representation in cases, and the judiciary in courts including the number of judges and ethnicity of judges in each court or CPS area.
- Police force area where the defendant is charged.
Disclaimer

This work was produced using administrative data accessed through the ONS Secure Research Service. The use of the 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 analyses 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.

Acknowledgements

Thanks to Thomas Jackson for his advice and support with the magistrate’s and Crown Court datasets and members of the Data First team for their feedback on the Data Explained report.

Contact

Dr Kitty Lymperopoulou
k.lymperopoulou@mmu.ac.uk