ADR UK

Data Explained

—

Ministry of Justice – Department for Education linked dataset

Exploring the intersections between ethnicity and care experience and youth justice involvement
Author: Dr Katie Hunter
Date: October 2022

This Data Explained summarises experiences and learning from working with the Ministry of Justice-Department for Education (MoJ-DfE) linked datasets in the course of producing research into youth justice involvement and how it might vary by ethnicity and care experience. This publication is intended to help guide future researchers using this data and to provide feedback into future dataset development and documentation.

The administrative data discussed in this Data Explained was made securely available through Data First and ADR UK. The data used in this research project comes from the Ministry of Justice and Department for Education and was accessed through the ONS Secure Research Service (SRS). 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.

Project details

In England and Wales, youth justice is characterised by longstanding over-representation of children from Black and ethnic minority backgrounds and those with experience of the care system (Lammy, 2017; Youth Justice Board (YJB) & MoJ, 2022). Furthermore, there is considerable overlap between these groups, with the Laming Review estimating that 44% of looked after children in custody come from an ethnic minority background (Prison Reform Trust, 2016). Despite this, there was previously a lack of quantitative data on the intersections between ethnicity, care experience and justice systems contact (Hunter, 2019; Fitzpatrick et al 2022). The newly linked MoJ-DfE datasets have made it possible to explore these intersections for the first time.

My Fellowship project is directly building upon my Economic and Social Research Council (ESRC)-funded PhD research, which utilised mixed methods to investigate the over-representation of care-experienced children and Black and ethnic minority children in the youth justice system. My thesis determined that Black and ethnic minority children in care experience a 'double whammy' of disadvantage that influences their trajectories into the youth justice system (Hunter, 2022). However, it found that we needed to improve the quantitative evidence base to better understand the nature and extent of these disproportionalities (Hunter, 2019; Hunter, 2022).

Initial research questions

1. What proportion of individuals have experience of the looked after system and what are the details of offending?

2. Does offending profile vary according to ethnicity and/or legal status and placement type?

3. How do care-experienced and non-care-experienced individuals' sentence lengths compare for three offence types (actual bodily harm, robbery and possession of an article with blade or point), and do these relationships vary by ethnicity?

4. How do factors (offender characteristics, legal status, placement type) impact frequency of offending for care-experienced individuals?

Research methodology

Individuals with experience of the looked after care system will be identified across four birth cohorts born between 1996 and 1999 (see below). Descriptive statistics will enable the offending profiles (offence type, frequency and disposal) of care-experienced and non-care-experienced individuals to be produced and disaggregated by ethnicity. To compare sentencing for three offence types (actual bodily harm, robbery and possession of an article with blade or point), log-transformed custodial sentence length for each offence type will be considered using logistic regression.

Ethnicity and care experience will be included as key variables in these models, while also controlling for potential confounders. Where appropriate, multilevel models will be used to account for individuals who have multiple disposals within the offence types. Linear regression will be used to measure average frequency of proven offences for care-experienced and non-care-experienced individuals over a specified time frame, accounting for exposure to periods of imprisonment. Multilinear regression models will then be used to explore the extent to which ethnicity and legal status impact upon average frequency of proven offences for care-experienced individuals.

Key variables

Type of data	Datasets used	Variables used
Educational census data	• SpringCensus2006 • SpringCensus2007 • SpringCensus2008 • Spring Census2009 • PRUCensus • APCensus	• *YearOfBirth* • *Gender* • *Ethnicity,* *EthnicGroupMinor* • *EthnicGroupMajor*
Information about care experience	*CLA_05-06_to_16-17* (looked after dataset	• *CLA_ACADYR* • *CLA_PROCESSING_YEAR* • *CLA_POC_START* • *CLA_POC_LENGTH* • *CLA_LEGAL_STATUS* • *CLA_PLACEMENT* • *CLA_CLA_PP_1_DAY*
Information about youth justice involvement	*PNC_00_to_17* (Police National Computer (PNC) dataset)	• *CaseID* • *OffenceID* • *CourtCautionDate* • *CourtCode* • *CautionType* • *DisposalID* • *HODisposalCode* • *DisposalAmount* • *DisposalDuration* • *HOOffenceCode* • *OffenceStartDate* • *AdjudicationCode* • *IsPrimaryOffence* • *OffenceStartAge*

Data is being matched using the following ID Variables: *PupilMatchingReferenceAnonymous* and *MoJUID* (where applicable).

Summary of comments on datasets and specific variables

Educational Census Datasets

There are limitations in the educational census data (*SC_Pupil_01-02_to_17-18_SUM; PRU_Census_09-10_to_12-13; Alt_Provision_07-08_to_17-18*) with regards to years covered. The main school census data is available from 2001 whereas Alternative Provision (AP) census and Pupil Referral Unit (PRU) census data are only available from 2007 and 2009 respectively. Therefore, individuals educated in PRUs or AP prior to these census years may be missing from the data. Nonetheless, the birth cohort approach taken (see below) mitigates some of this risk since most children would be in a mainstream school at age 10 (the point at which the data has been extracted).

Looked After Dataset

There are some key limitations with the looked after dataset (*CLA_05-06_to_16-17*). For example, researchers cannot know whether an individual is care-experienced if they ceased to be looked after before the first data collection period in 2005/06. This information is not recorded and such individuals cannot be included as 'care-experienced' in analyses. There is also a risk that some individuals may be missed from analysis if they were looked after by more than one local authority, because they will be assigned multiple IDs, preventing linkage of care records (see McGrath-Lone et al 2016). This may mean that some of the most disadvantaged individuals are not captured by these analyses. Individuals who were adopted after being in care would also be assigned a new ID, so these individuals would appear as two different people in the datasets.

This dataset is significantly limited by the fact that it only includes information about an individual's most recent episode in care. Therefore analyses cannot consider the number of care episodes, previous legal statuses and placement types, placement stability and most significantly, onset of youth justice involvement in relation to entry to care. All of these are vital to understanding the relationship between care experience and youth justice involvement. The existence of the *CLA_EPI_IDX* variable, denotes the unique episode of care index, offers some hope that future data shares will include more detailed care histories.

PNC Dataset

There are issues with some variables contained in the PNC dataset (*PNC_00_to_2017*). *AddressStationCode* has a high proportion of missing data and so it has been excluded from my analyses. In *CourtCautionDate* a small number of cases are incorrectly recorded as occurring before the offence date listed in *OffenceStartDate*. There is also a small number of cases which obviously contain errors (e.g. offences listed as occurring in the year 2040). In *OffenceStartAge,*

there are some cases where information appears to be incorrectly recorded (for example, age at offence listed as - 17). All erroneous records have been excluded from analysis.

There are some cases where *OffenceStartAge* is listed as below 10 (with the vast majority recorded as age 9). Given that the minimum age of criminal responsibility is 10 in England and Wales, these cases have also been excluded from the analyses. In *HOOffenceCode* there is a very small number of records which contain codes for offences committed outside of England and Wales (i.e. in Scotland). Since these records pertain to cases in other youth justice jurisdictions, they were excluded from the analysis.

Further exploration of the PNC dataset is required to determine the suitability of *CautionType*, *DisposalAmount* and *DisposalDuration* for the current research. Nonetheless, a broader limitation of the PNC data dataset is the lack of contextual information provided about an offence; there may be other mitigating and/or aggravating factors that influence disposals not evidenced in the data.

How you dealt with data limitations

After initial exploration of the available datasets, a birth cohort approach was developed with my mentor, Professor Brian Francis. My original research design involved analyses of both youth and adult criminal justice systems involvement. However, the time periods covered by the different datasets meant exploration of adult justice systems involvement would be limited. Therefore, we took the decision to compare youth justice involvement for four cohorts of individuals born between 1996 and 1999. Snapshot demographic information for individuals born in each birth year were extracted from the respective 2006 to 2009 spring censuses. To maximise the number of individuals taken from AP and PRU census data, demographic information for all those born between 1996 and 1999 were taken from these censuses.

The cohort approach allows us to account for varying levels of exposure to youth justice involvement since we can explore PNC records for each individual when they were aged between 10 and 17-years-old (see Table 1). As such, each person has an equal chance of having a PNC record (i.e. equal exposure in the data). Using snapshot information for individuals aged 10 (rather than collating data from across multiple censuses) not only saves time, but also minimises the risk of including individuals missing from later datasets (i.e. because they have died or left the country). This could reduce the proportion of those with youth justice contact.

The specific birth years were chosen to account for variation in time periods covered by the respective datasets. The School Census covers 2001/02 to 2017/18 whereas the AP census covers 2007/08 to 2017/18. The PRU census covers 2009/10 to 2012/13 with PRU pupils being included in the mainstream School Census from 2013/14. The looked after children dataset includes records from 2005/06 to 2016/17 and the PNC dataset covers 2000 to 2017, with some limited information about offences committed before the millennium. Therefore, the earliest birth year was chosen because looked after data is only available from 2005/06 and individuals born in

1996 would reach the age of criminal responsibility in that record year. The latest birth year was selected because the PNC data runs until 2017 and so individuals born in 1999 would reach the age of 18 in that year. The censuses used and the years included are detailed in Table 1 below.

Table 1: Census type, birth year and PNC years used

Census used	Birth year(s) used	PNC years used
School Spring Census 2006	1996	2006 to 2014
School Spring Census 2007	1997	2007 to 2015
School Spring Census 2008	1998	2008 to 2016
School Spring Census 2009	1999	2009 to 2017
PRU Census (2010-2013)	1996 1997 1998 1999	2006 to 2014 2007 to 2015 2008 to 2016 2009 to 2017
AP Census (2007-2018)	1996 1997 1998 1999	2006 to 2014 2007 to 2015 2008 to 2016 2009 to 2017

Suggested improvements recommended to data owners

It would be useful for data owners to communicate the precise form of datasets to prospective data users. When accessing the datasets for the first time, I was surprised to find that there were three mainstream School Censuses per year and that these all came in separate datasets. As such, I was faced with 36 School Census datasets, as well as the AP census and PRU census datasets. Having this knowledge prior to accessing the data would have allowed more time to develop a data cleaning plan while waiting for access. Data owners should also include comprehensive metadata files and information about how administrative data is collected (i.e. guidance on how to fill in the School Census) in the SRS as standard. This will prevent unnecessary time and effort being spent on ingestion requests when researchers start to access the datasets.

Additional data which would help to further develop the research

It is vital that the data owners work towards providing all episodes in care to ensure that researchers can better understand the relationship between care experience and justice systems

involvement. Data owners will have to develop solutions to carefully mitigate the risk of identification which is no easy task. As a starting point, data owners could provide more information about care histories such as number of care episodes and the number and type of care placements. However, the most pressing question concerns onset and the relationship between entry to care and first justice system involvement. This question cannot be addressed without precise information about when an individual was in care. It is currently not possible with the existing datasets and must be addressed as a matter of urgency.

References

Fitzpatrick, C., Hunter, K., Shaw, J. & Staines, J. (2022) *Final Report: Disrupting the Routes between Care and Custody for Girls and Women*. Available at: http://wp.lancs.ac.uk/care-custody/files/2022/05/Final-Care-Custody-Report-May-2022.pdf

Hunter, K. (2019) *Institutionalised Criminalisation: Black and Minority Ethnic Children and Looked After Children in the Youth Justice System in England and Wales* (unpublished thesis). University of Liverpool.

Hunter, K. (2022) 'Out of place': The criminalisation of Black and minority ethnic looked after children in England and Wales. *Prison Service Journal*. No. 258. Available at: https://www.crimeandjustice.org.uk/sites/crimeandjustice.org.uk/files/PSJ%20258%20January%202022.pdf

Lammy Review (2017) *An Independent Review into the Treatment of, and Outcomes for, Black, Asian and Minority Ethnic individuals in the Criminal Justice System*. Available at: https://www.gov.uk/government/publications/lammy-review-final-report

McGrath-Lone, L., Harron, K., Dearden, L., Nasim, B., & Gilbert, R. (2016) Data resource profile: Children Looked After Return (CLA). *International Journal of Epidemiology*, 45(3): 716–717f.

Prison Reform Trust (2016) *In Care, Out Of Trouble: An Independent Review Chaired by Lord Laming*. Available at: http://www.Prisonreformtrust.org.uk/Portals/0/Documents/In%20care%20out%20of%20trouble%20summary.pdf

Youth Justice Board & Ministry of Justice (2022) *Youth Justice Statistics 2020/2021*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1054236/Youth_Justice_Statistics_2020-21.pdf

Disclaimer

Acknowledgements

Contact

Name: Dr Katie Hunter
Email: k.hunter@mmu.ac.uk