

Data Explained

Ministry of Justice & Department for Education linked dataset - England

How do different types of social care involvement affect children's education and offending outcomes?

Author: Dr Anna Leyland

Date: July 2024

This Data Explained summarises experiences and learning from working with the Ministry of Justice and the Department for Education (MoJ-DfE) linked dataset in the course of producing research into the impact of social care involvement in childhood on education and offending outcomes. 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 ADR UK (Administrative Data Research UK), an Economic and Social Research Council (ESRC) investment (part of UK Research and Innovation). [Grant number: [171554-1](#)]. The data used in this research project comes from the Department for Education and Ministry of Justice and was accessed through the Office for National Statistics 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.

Project details

Little is known about how different degrees and durations of involvement from child social services may affect outcomes in two key areas: a child's education, and their involvement with criminal justice services. To date, it has been challenging for researchers to unpick the dynamic risk or protective effect of formal child social care services on these later outcomes.

The Ministry of Justice and the Department for Education linked dataset is a large, de-identified dataset that connects information on social care, education and crime. Researchers can use this dataset to answer some of these important and complex questions.

This project used this dataset to explore how involvement from formal social care services in childhood impacts on attainment and engagement in education and contact with criminal justice services in childhood and early adulthood. Comparisons were made between those children with different types and forms of formal child social care services involvement (e.g., a child in need, a child protection order, a looked after child) to assess the potential differences between group risk or protective effects on later outcomes.

Initial research questions

1. What are the effects of formal child social care intervention in childhood on education attainment and engagement, and contact with criminal justice services?

2. Are the education and offending outcomes for children who engage with formal child social care in childhood equal across local authorities?

Research methodology

Research question one explores the effects of different modes of child welfare service interventions (none, referral only, child in need, child protection plan, child looked after) on:

- education attainment (level 4 outcomes: GCSE or equivalent)
- education engagement (number of unauthorised absences in secondary school; sessions missed through temporary exclusion; permanent school exclusion)
- criminal justice service outcomes (convictions: none, one, two or more; severity of disposal: caution, non-custodial, custodial).

The degree to which social care services intervene with families is based on clinical and in some cases legal judgments about the needs and safety of the child and the family. As such, those who have the most severe type of intervention, placement in an out of home setting as a looked after child, will also most likely have the most challenging and potentially harmful or unsafe home environments. It is therefore difficult to disentangle the role of social care referral or interventions on children's outcomes, with the impact of neglect or abuse on child development. In addition, the reason that social care services intervene with children and families can include the child's criminal or antisocial behaviour. In these instances, the criminal justice system formal contact may come ahead of the social care intervention, and as such drawing a causal link from social care intervention to criminal justice systems formal contact would be inaccurate. Although in the majority of cases social care intervention precedes the criminal justice formal contact by one year or more.

The social care data was used to identify the highest intervention ever received by the individual across childhood. This approach makes the data appropriate for later descriptive and inferential analyses, but it does have some limitations. The linked dataset used in the present case contains only the most recent period or episode of care within each reporting year. As such, if there were multiple periods or episodes of care within a reporting year, then the details of any change of care level in these earlier episodes/ periods are not captured – which may lead to misclassification at a lower level of intervention.

The second research question made use of the same data sources and knowledge from research question one, but the effects were placed in the geographical context of the local authority. Specifically, for this question, other community level indicators were added as predictors in the model (e.g., community crime rates) and spatial mapping was used for descriptive purposes. In this analysis, the comparison is made between local authorities with reference to the national average, to compare the likelihood of a criminal caution or conviction for those in receipt or a social care referral or at each level of social care intervention, adjusting for the level of offending in the non-social care population in that local authority and the local authority poverty rate.

Key variables

This Data Explained document focuses on the data items from the child in need (CIN_08-09_to_16-17) and child looked after (CLA_2006_to_2018) datasets used in the current research which form two of the available datasets within the MoJ-DfE linked dataset. Rstudio was used for all data preparation and analysis and so the discussion gives examples of how the data was manipulated using the R coding language.

Dataset name		
CLA_2006_to_2018 <i>example of code names if imported as they appear in original dataset</i>	MoJUID	CLA_POC_LENGTH
	CLA_ACADYR	CLA_CAT_NEED
	CLA_PupilMatchingRefAnonymous (usedplyr::recodetorenameas pupil_id)	CLA_DATE_EPI_COMM
	CLA_CLA_LA	CLA_DATE_EPI_CEASED
	CLA_PROCESSING_YEAR	CLA_LEGAL_STATUS
	CLA_POC_START	CLA_PLACEMENT
		CLA_REC
CIN_08-09_to_16-17 <i>example of code names if imported with janitor::clean.names(., "snake")</i>	mo_juid	CLA_CLA_31_MARCH
	cin_acadyr	cin_latest_category_of_abuse
	cin_pupil_matching_ref_anonymous (usedplyr::recodetorenameas pupil_id)	cin_number_of_previous_cpp
	cin_cin_at31march	cin_primary_need_code
	cin_category_of_abuse	cin_referral_nfa
	cin_initial_category_of_abuse	cin_reason_for_closure
		cin_age_start_of_cin_period
	cin_age_end_of_cin_period	

Note: Access to several other data items from CLA_2006_to_2018 were requested but are not discussed here (CLA_CLA_6_MONTHS; CLA_CLA_12_MONTHS; CLA_CLA_PP_6_MONTHS; CLA_CLA_PP_1_DAY; CLA_REASON_PLACE_CHANGE).

When the datasets are first imported into Rstudio, in most cases the missing items are left as blank spaces (" ", " ") rather than NA (not applicable) or another identifier. Attempting to recode the missing items as part of the import code can be slow when working in the Office for National Statistics (ONS) Secure Research Service.

Most data items are recognised by Rstudio as character variables, including dates (e.g., *cin_acadyr*) whereas other factor or logical variables are recognised as numeric (e.g., *cin_cin_at31march*).

Compared to other datasets within the MoJ-DfE linked dataset, the coverage of the child in need and child looked after datasets is good. One exception is the reason the care placement has changed (*cla_reason_place_change*), where only 2.5% of the data is complete as this variable has only been collected since 2016.

Three variables are key identifiers to linking across datasets and when moving between long and wide format: academic year, ministry of justice ID and the pupil matching reference. In each

dataset from the MoJ-DfE dataset these variables are coded differently, such that the prefix or suffix will contain an identifier for the dataset e.g., *KS4_PupilMatchingRefAnonymous*. As such, it is helpful to consistently use a common code name for these variables across the datasets, for example *moj_uid*, *pupil_id*, and *academic_year*.

Summary of comments on specific variables

The child looked after data (*CLA_2006_to_2018*) can be used to calculate care episode duration for the most recent care episodes of each academic year. This data can be used to calculate duration in care and changes in care placement types, but does not give full coverage for care episodes. To calculate care duration, the variables relating to the date the care episode started (*cla_date_epi_comm*) and ended (*cla_date_epi_ceased*) can be used in conjunction with academic year (*cla_acadyr*). Academic year (e.g. 2005/2006) can be split into two columns to create the start and end year, which makes it easier to use in other transformations (`tidyr::separate`).

The same variables can be used to calculate the first and last date of social care contact for the child looked after dataset, (`min(cla_date_epi_comm)` and `max(cla_date_epi_ceased)`) for each pupil (`dplyr::group_by(pupil_id)`).

It is also possible to see which types of care children have had, for example, foster care, kinship care, using the child looked after placements variable (*cla_placements*). Again, this is complicated if the goal is to see how long children have spent in each type of care and is limited as the data refers to only the most recent episode of care in each reporting year. A look-up [table can be created using the MoJ-DfE metadata](#) spreadsheet, which can be ingested by the [ONS Secure Research Service \(SRS\) customer support team](#) to the users' workspace on the SRS. Look up files can also be created for the reason the care episode has ended (*cla_rec*), the legal status of the care placement (*cla_legal_status*) and the category of need (*cla_cat_need*).

The child in need (CIN) data (*CIN_08-09_to_16-17*) has similar information regarding the start and end date of CIN episodes. When using CIN data, the completeness of data from 2008/09 until 2011/12 needs to be considered (please see DfE, 2023 for more details). It is also possible to use the academic year item (*cin_acadyr*) and the indicator for whether the child was a child in need on 31 March (*cin_cin_at31march*). A low precision indicator of years as child in need can then be calculated by totaling the occasions across multiple entries that the child was recorded as being a child in need (`dplyr::group_by(pupil_id)`). Updated versions of the CIN and CLA data in the MoJ-DfE data linkage will include all episodes of care, allowing for a more precise estimate of time in care and care trajectories.

The CIN data can also be used to identify those children where no further action was taken following assessment (e.g., “referral only”) and where children also had a child protection plan:

1. Referral only: Those children without any incidence of being a child in need on 31 March but whose data is in the CIN dataset, can be coded as children who were “referral only” (`cin_cin_at31march == 0`). The accuracy of this coding can be compared with when there was no further assessment (`cin_referral_nfa == >= 1`)

2. Child protection Plan (CPP): There are two methods to assess whether a child in need has a child protection plan and these can then be cross referenced to check for discrepancies.

a) Category of abuse items: Due to coverage the *cin_category_of_abuse* and *cin_abuse_cat_first* items can be merged using the `dplyr::coalesce` function. Then all missing (NA) entries can be recoded as “no CPP” and those that are not missing (NA) can be coded as “CPP”.

b) Previous child protection plan: A new item can be created by summing the counts for previous child protection plans across academic years grouping by pupil id (`dplyr::sum(cin_prev_cpp)`) across academic years for each child. All entries that have a value of 1 or more can be recoded as “CPP” and 0 as “no CPP”.

The following describes one possible approach for working with the child welfare system data. These identifiers derived from the child looked after and child in need dataset can be joined together to have a single ordered factor item for whether the child was: child looked after > child protection plan > child in need > referral only. Some children appear on both the child looked after and child in need datasets and as such coding needs to take account of occasions where a child was both a “child looked after” and on a “child protection plan” or a “child in need”. For the purposes of the present study the child was coded as being the highest of the ordered classifications, for example if they were a “child in need” and a “child looked after” they were coded as “child looked after”.

When the child looked after and child in need data is joined to the other data within the MoJ-DfE linked dataset, then there are a majority of children who will have had no contact with child social care services, and as such do not appear in the child looked after and child in need datasets. For these children, any child looked after or child in need items will need to be accurately coded as missing data. A separate category of “no social care contact” should be made available.

Suggested improvements recommended to data owners

The newest release of the linked data will have every care episode and period, allowing for a complete assessment of the time spent receiving each level of social care intervention. This will allow for a more accurate assessment of aspects of care trajectories including the duration and timing of interventions, and for those in out of home placements, details of time spent in different types of placements and numbers of placement changes or returns to the birth family.

In the update to the MoJ-DfE data linkage in child looked after dataset, the reasons identified at assessment will be included and available to researchers. This will give details on the factors at assessment that resulted in the social welfare intervention, such as domestic violence or parent drug misuse. This would allow analyses to take consideration of the adversity and exposures experienced by the child and family.

In the child in need dataset it would be advantageous to know the intervention that was offered to the child and family by formal social care or other services. This could include those cases where there was no further action taken following assessment and where these families may have been signposted to other support services.

Additional data which would help to further develop the research

It would be most beneficial for the MoJ-DfE linked dataset to be linked to details of the parents' socioeconomic status (e.g., income, employment status, education level) and health data (e.g., hospital episode statistics, primary care) for the child and parents. The inclusion of such data would present a comprehensive picture of child and family level factors that may be greatly influential for children's outcomes. In particular, this includes children's education engagement and attainment and later offending, but also determinants of social care support and in school support for special educational needs.

References

DfE (2023) Methodology: Children in Need, available online (accessed on 18/06/2024)
<https://explore-education-statistics.service.gov.uk/methodology/characteristics-of-children-in-need-methodology>

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 Secure Research Service or data owners in relation to the interpretation or analysis. 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

This work is supported by ADRUK (Administrative Data Research UK). ADRUK is a partnership transforming the way researchers access the UK's wealth of public sector data, to enable better informed policy decisions that improve people's lives. ADRUK is an Economic and Social Research Council (ESRC) investment (part of UK Research and Innovation). [Grant number: 171554-1]. This work is also supported by colleagues from The University of Sheffield, including Dr Calum Webb of the [Sheffield Methods Institute](#), Prof. Matthew Bennett of the [Centre for Care](#) and Prof. Nathan Hughes of the Department for [Sociological Studies](#).

Contact

Name: Dr Anna Leyland
Email: anna.f.leyland@sheffield.ac.uk
[Follow Dr Anna Leyland's research on Research Gate.](#)