

The links between young people being imprisoned, pupil background and school quality: methodology

Methodological details for our analysis on how the quality of schools is linked with the imprisonment of young people, including the sample, logistic regression method and concepts used.

Contact:
Simeon North, Mathieu Stafford
and Holly Bathgate
integrated.data.analysis@ons.gov.uk
+44 3456 013034

Release date:
27 January 2023

Next release:
To be announced

Table of contents

1. [Data sources](#)
2. [Statistical methods](#)
3. [Concepts used](#)
4. [Sensitivity checks](#)
5. [Qualitative interviews](#)
6. [Glossary](#)
7. [Future developments](#)
8. [Acknowledgement](#)
9. [Related links](#)
10. [Cite this methodology](#)

1 . Data sources

Ministry of Justice – Department for Education data share

Our [The links between young people being imprisoned, school quality and pupil background](#) article uses Ministry of Justice (MoJ) and Department for Education (DfE) data that have been linked as part of the [Data First](#) project. The data share includes data from prisons, courts, the Police National Computer (PNC), the National Pupil Database (NPD), children looked after (CLA) and children in need (CIN). It includes demographic information as well as variables such as attainment, criminal offences, school exclusions and care experience.

There are some data limitations to be aware of as the source data was collected for administrative, rather than statistical, purposes. Overall, as described in the [MoJ-DfE Technical Note \(PDF, 459KB\)](#), the DfE assesses the match rate between the PNC and the NPD as good, with 77% of offenders matched. The data were linked using a rules-based approach, so there is some variation in the data linking accuracy. For example, it is harder to determine that records belong to one person if they have used different names and moved address often. A share of offenders are unmatched to the NPD as they did not attend education in England. The [DfE reports \(PDF, 459KB\)](#) that the matched dataset has similar characteristics to the unmatched dataset in terms of gender and age, but that there are some noted differences for ethnicity.

Further details of the linking process and data quality are available in section one of the [MoJ-DfE Technical Note \(PDF, 459KB\)](#) and in the [Administrative Data Research UK \(ADR-UK\) Data quality report \(PDF, 2.98 MB\)](#).

Ofsted school data

The MoJ-DfE data were joined to publicly available school data from Ofsted and DfE for the 2009 to 2010 academic year. These data included school-level information such as the overall effectiveness grade (see [Section 6: Glossary](#)), school size, type, and region, as well as summary statistics about the school's pupils, such as the percentage of pupils with Special Educational Needs (SEN) and the percentage of pupils eligible for free school meals.

Coverage and context

The publication looks at young people in the MoJ-DfE data share who meet all the following criteria:

- were born in the 1993 to 1994 academic year (between 1 September 1993 and 31 August 1994), and therefore typically started primary school in September 1998
- attended a state primary school in England for at least one term
- attended a state secondary school in England in January 2010
- had recorded educational attainment for key stage 1 and key stage 2

The focus of this research was the effect of mainstream and state special school quality on outcomes. Pupils registered at pupil referral units (PRUs) in January 2010 were filtered from the analysis. Some local authorities did not have PRUs but may have had special schools that focused on pupils with behavioural, emotional and social difficulties (BESD). Schools where every pupil had a primary Special Educational Need type of BESD were therefore filtered from the analysis. These school types, at which pupils were likely to have been admitted because of challenging behaviour, will be researched separately.

The total sample of the MoJ-DfE data used in this analysis is 515,226 pupils. The [first publication in this series](#) had a sample of 687,101 pupils. Of these, 100,550 were removed as they were not present in the school census data while 26,367 were removed as they were not present in the primary school data. A further 8,727 who attended PRUs or schools focused on BESD pupils were filtered out of the sample as these settings were outside the scope of this analysis. Finally, a further 36,231 individuals who were missing attainment scores or attended schools without an overall effectiveness grade were removed.

As the samples between the publications are different, the results cannot be directly compared between the publications. Readers should note that, unlike in previous releases, the model included in this publication only considers immediate custodial sentences received when individuals were aged 16 years and over. This was done to ensure that the outcome, immediate custodial sentences, and the control variables were measured at separate points in time.

2 . Statistical methods

The model used in this publication measures the association between overall effectiveness grade and imprisonment after the age of 16 years, holding the factors outlined in this section constant. The model output and data used in the article can be found in the [accompanying dataset](#).

The publication uses logistic regression, a type of generalized linear model that models the relationship between one binary (true or false) dependent variable and multiple explanatory variables. In this case, it is used to estimate the binary outcome of receiving an immediate custodial sentence from 1 September 2010, when the pupil was aged 16 years or over, until the end of 2017, controlling for observable pupil- and school-level factors. The analysis was performed in base R, using the glm function.

The likelihood of this binary outcome being true was estimated, controlling for the following observed characteristics:

- pupil demographics, including gender, ethnicity, and English as an additional language ([EAL](#)) during primary school
- pupil deprivation, including free school meal eligibility in primary school and the average of the [income deprivation affecting children index \(IDACI\)](#) score during primary school
- pupil vulnerability (during secondary school), including [child in need \(aged 14 to 16 years\)](#) and [child looked after \(aged 11 to 16 years\) \(PDF, 525 KB\)](#)
- pupil attainment, including key stage 1 and 2 attainment
- pupil Special Educational Needs (SEN)
- school information, including school type, Ofsted region of the school, urban-rural classification of the school, selective or religious school flag, gender mix of school (mixed, boys, girls), and characteristics of pupils in a school (for example, percentage of pupils with English as an additional language and percentage of pupils eligible for free school meals)

The model can be expressed as:

$$\ln \frac{p}{1 - p} = \beta_0 + \beta_1 X + \beta_2 Z$$

Where p is the estimated probability of the binary outcome being true (in this case, the probability of receiving an immediate custodial sentence in young adulthood), while

$$\frac{p}{1 - p}$$

and

$$\ln \frac{p}{1 - p}$$

are, respectively, the odds and log-odds of the binary outcome being true.

X is a vector of pupil-level characteristics relating to the individual, while Z is a vector of school-level characteristics relating to the school attended by the pupil.

Since pupils at the same school share school-level information, the standard errors of the logistic regression are clustered at the school level.

From the results, odds ratios and predicted probabilities were calculated using the variables described. When calculating predicted probabilities, reference values were used for categorical variables, and the mean average was calculated for numeric variables including key stage 1 and 2 attainment and the characteristics of pupils in a school, such as the percentage of pupils eligible for free school meals.

The analysis focuses on the contribution of schools during key stages 3 and 4. Where possible, the model controls for pupil characteristics recorded prior to key stage 3. This is so that the model does not control for characteristics that could have been the result of the school's effect.

Interpretation

The findings in our [The links between young people being imprisoned, pupil background and school quality article](#) look at the relationship between school quality and imprisonment in young adulthood. It is important to note that they do not attempt to provide evidence on causality.

Care should also be taken when interpreting the coefficients of the logistic regression. The coefficient of interest is school quality, and the pupil-level and school-level characteristics cannot be interpreted as direct associations.

For example, the regression includes a school's region and its urban-rural classification. Factors such as deprivation and ethnicity are not evenly distributed across regions or urban-rural classification. As such, the coefficient value for a region is the net of these other effects and cannot be easily interpreted as the direct effect of a school being in a particular region.

There are also factors not included in the data that may be significant. For example, the data do not include family characteristics that may contribute to the probability of imprisonment.

3 . Concepts used

Pupil-level characteristics

Pupil-level characteristics were collected during primary school, before the age of criminal responsibility (age 10 years in England). These included demographic factors such as gender and ethnicity, as well as educational attainment (in key stages 1 and 2) and socioeconomic factors such as eligibility for free school meals (FSM) and the income deprivation affecting children index (IDACI).

Attainment during key stage 1 and key stage 2 is calculated in accordance with [the Department for Education's technical guide](#) for measuring attainment. Attainment in key stage 1 uses component scores in reading, writing, and mathematics, while key stage 2 uses component scores in English, mathematics, and science.

Some pupil-level information was not available during primary school, such as social care background or the type of Special Educational Needs (SEN) recorded. The type of Special Educational Need was collected instead from the first spring school census of secondary school, the first year that SEN type was recorded. Since there are a variety of SEN types, SEN type was separated into three categories: not being recorded with SEN, being recorded with the behavioural, emotional and social difficulties (BESD) SEN type, and being recorded with a SEN type other than BESD. Our [previous article](#) identified that BESD was the primary SEN type most recorded in the imprisoned group.

Social care information about the pupils included whether a pupil was registered as a child in need (CIN) [or a child looked after \(CLA\)](#). The previous publication found that a larger share of these groups received immediate custodial sentences. Information was only available from the 2005 to 2006 academic year for children looked after; that is, from when pupils were aged 11 or 12 years. Children in need were only identified in the year 2008 to 2009, when pupils in our cohort would have been aged 14 or 15 years. To capture this, two variables were included in the model to indicate whether a pupil was present in either of the two social care groups at any point before the 2009 to 2010 academic year.

School-level characteristics

Pupils were assigned to a school based on their registration in the spring school census of the 2009 to 2010 academic year, when the pupils would have been aged 15 or 16 years. School-level characteristics were collected using publicly available DfE and Ofsted data from the 2009 to 2010 academic year for the school assigned to each pupil.

The school-level characteristics included the Ofsted region in which the school was located and its urban-rural designation using Census 2011 classifications. Other characteristics included the school type, whether it was a mixed or single-sex school, its admissions policy (selective or not), and whether it was religiously affiliated. They also included the school size and the proportion of Special Education Needs students with and without a statement, free school meals, White British pupils, and pupils with English as an additional language.

4 . Sensitivity checks

To test the robustness of the findings from the main model used in our [The links between young people being imprisoned, pupil background and school quality article](#), we ran a series of sensitivity models, using the same logistic regression method outlined previously.

When restricting the sample to males, the model showed the association between overall effectiveness grade and imprisonment was stronger compared with the full sample and remained statistically significant. A model including only females found the association was not statistically significant. However, because of the small number of females in the sample who were imprisoned, the model had limited statistical power.

As moving school means pupils can attend schools with different overall effectiveness grades, we ran two models to account for such moves. The first flagged whether a student moved schools between 2006 and 2010. When pupils had attended schools that merged, split, or became academies, these were not counted as school moves. All other school moves, including for behavioural or personal reasons, or moving from middle school to upper school where applicable, were included. The model found the association between overall effectiveness grade and imprisonment was weaker but remained statistically significant. A second model allocated pupils according to the school they were enrolled in during 2005 to 2006. The model found the association between overall effectiveness grade remained statistically significant. This shows that the association with school quality is not solely because of pupil movement during school years 7 to 11.

We ran a model to flag people whose first imprisonment was during the 2011 riots in England (6 to 11 August 2011) as there was a higher number of imprisonments, including first imprisonments, during this concentrated period compared with other weeks. The riots were also geographically concentrated with there being a possibility that imprisonment was concentrated in a limited number of schools. The results from this model found that the association between overall effectiveness grade and imprisonment was stronger compared with the full sample and remained statistically significant.

Building on the descriptive [analysis of looked-after children](#) recently published by the Office for National Statistics (ONS), we ran an additional model on the association between school quality and imprisonment for looked-after children. The model found a statistically significant association, with looked-after children in satisfactory or inadequate schools having a 28% increase in the odds of being imprisoned compared with looked-after children in good or outstanding schools.

5 . Qualitative interviews

Our [The links between young people being imprisoned, pupil background and school quality](#) article also includes quotes from interviews with Marcus Isman-Egal from EPIC, a youth crime prevention organisation based in Doncaster, and James Hadley from the Haverstock School in Camden, London. Both organisations were selected for interview because of their experience of, and expertise in, working with young people. Interviewees could illustrate some of the challenges of working with young people who were perceived to be at risk of involvement with the criminal justice system.

A data journalist at the Office for National Statistics (ONS) conducted the interviews through video calls. The quotes are not intended to be representative of the diverse lived experiences of young people or the adults that work with them, nor does their inclusion imply best practice that should be replicated. The interview also does not mean that EPIC or the Haverstock School endorse the findings from the article.

6 . Glossary

Behavioural, emotional and social difficulties (BESD)

This is a type of Special Educational Need. Children and young people with BESD experience challenges with personal, social, and emotional development. This can have an impact on behaviour and emotional responses, and their relationships with self. Since 2014, social, emotional, and mental health (SEMH) has replaced the term BESD and emphasises that behaviour is a symptom of the needs that underlie it.

English as an additional language (EAL)

A pupil is recorded as having English as an additional language where English is not recorded as their first language, and they are exposed to a language at home that is known or believed to be other than English. This includes children who were born in the UK and whose parents were also born in the UK.

Overall effectiveness grade

The judgement Ofsted reached at their most recent inspection. Schools could be graded as outstanding, good, satisfactory, or inadequate. In 2012 to 2013, the satisfactory category was renamed “requires improvement”.

School type

School type was chosen to reflect the range of state-funded schools when the cohort of interest were in secondary education (2005 to 2010). There are some differences to the current composition of schools. The following types of school are included in this analysis:

- academy
- City Technology College
- community
- community special
- foundation
- foundation special
- non-maintained special
- voluntary aided
- voluntary controlled

Urban-rural classification

[Urban and rural](#) settings are classified into one of six categories based on the dwelling density of the area. Rural areas are separated into whether they are in a sparse setting, which takes into account the dwelling density of up to 30 kilometres beyond the area, or not. Of the urban classifications, only city and town can also be classified as in a sparse setting.

7 . Future developments

The authors welcome feedback and suggestions for future developments.

This release is the third in a series investigating how the probability of imprisonment varies by prior education and social provision, and how that impact might differ according to young people’s backgrounds.

The next phase of the work will focus on areas of interest that have emerged over the course of this analysis, for example, school moves and the support offered by pupil referral units (PRUs). The Office for National Statistics (ONS) also plans to collaborate with a charity partner to explore these themes with children and young people.

8 . Acknowledgement

The team would like to thank colleagues, including Philip Noden, Paul Moore, and John Jerrim, at the Office for Standards in Education, Children's Services and Skills (Ofsted) for their valuable data and analytical contributions to this piece as well as their feedback and comments throughout the process.

9 . Related links

[The education and social care background of young people who interact with the criminal justice system: May 2022](#)

Article | Released 11 May 2022

Examining educational attainment and provision, social care provision and demographics of young people educated in England who subsequently received a custodial sentence, and comparing with their peers who did not.

[The education background of looked-after children who interact with the criminal justice system: December 2022](#)

Article | Released 5 December 2022

Educational attainment and provision, social care background and demographics of looked-after children educated in England who subsequently received a custodial sentence.

[Ministry of Justice – Department for Education linked dataset: Feasibility of evaluating early interventions for violence prevention \(PDF, 2.98 MB\)](#)

Report | Released 25 July 2022

Quality report looking at the reliability of the Ministry of Justice and Department for Education (MoJ-DfE) linked dataset. This describes the data sources, the linkage process, and a range of quality assessments.

[The links between young people being imprisoned, school quality and pupil background: data](#)

Dataset | Released 27 January 2023

Data on how background characteristics and school quality are linked with the criminal justice system interactions of young people in England, including the sample, logistic regression method, and analysis conducted.

10 . Cite this methodology

Office for National Statistics (ONS), released 27 January 2023, ONS website, methodology, [The links between young people being imprisoned, pupil background and school quality: methodology](#)