Supplementary Methods

Data sources

This study used a data linkage between the Police National Computer (PNC) and the National Pupil Database (NPD). Data tables from the NPD used in this analysis were the Pupil Level Annual School Census (PLASC), Spring Census, and key stage 2 (KS2) attainment. We requested an extract from the data linkage comprising pupils born between 01/09/1990 and 31/08/1997. In this cohort, approximately 4.33 million pupils had a PLASC record, 4.36 million had a Spring Census record, 4.28 million had a KS2 attainment record, and 648,517 had a PNC record. We focused on PLASC and Spring Census records from academic years 2001/2002 to 2016/2017. Table 1 shows the age of our different birth cohorts for each of these academic years.

Table 1: Estimated age for each birth cohort at the start of included academic years.

| | **Birth cohort** | | | | | | |
Academic year	1990/1991	1991/1992	1992/1993	1993/1994	1994/1995	1995/1996	1996/1997
2001/2002	10	9	8	7	6	5	4
2002/2003	11	10	9	8	7	6	5
2003/2004	12	11	10	9	8	7	6
2004/2005	13	12	11	10	9	8	7
2005/2006	14	13	12	11	10	9	8
2006/2007	15	14	13	12	11	10	9
2007/2008	16	15	14	13	12	11	10
2008/2009	17	16	15	14	13	12	11
2009/2010	18	17	16	15	14	13	12
2010/2011	19	18	17	16	15	14	13
2011/2012	20	19	18	17	16	15	14
2012/2013	21	20	19	18	17	16	15
2013/2014	22	21	20	19	18	17	16
2014/2015	23	22	21	20	19	18	17
2015/2016	24	23	22	21	20	19	18
2016/2017	25	24	23	22	21	20	19

Note: Green cells reflect compulsory school ages, red cells reflect ages where school is no longer compulsory. This table does not contain empirical data, rather it is a visual representation of the theoretical expected age for each birth cohort at the start of included academic years.

If a pupil had multiple records within an academic year in the PLASC or Spring Census, we kept the record which the DfE had flagged as the main record. In some cases, pupils did not have a main record flagged, but had multiple 'secondary' records available. In these situations we kept the record with the lowest school unique reference number, as this uniquely identified these remaining records.

If a pupil had multiple records in the KS2 attainment table, we kept whichever record which was non-missing on the maths and English key stage 2 attainment variable, and in which they received the best result (see below for further details on this variable).

In the PNC, we dropped offences with 'implausible' offence dates and ages. Given the remit of the linkage, and of the birth cohorts we requested, offences pre-1990 (earliest birth cohort) and post-2017 (linkage end date) were considered implausible, so offences from those years were dropped. We then dropped offences where the recorded offence age was below the age of criminal responsibility for the region where it took place, as these are likely to be recording errors:

- Scotland: age <8 years
- Guernsey: age <10 years prior to 2010, and age <12 from 2010 onwards
- All other regions: <10 years

Also, given the age of the cohort, any offences committed age >27 years (the age which the oldest in the 1990 birth cohort would be by the time of 2017 linkage) were again considered implausible, and therefore dropped.

If an individual had multiple offence records following this, we kept their first offence. However, offence date often did not uniquely identify PNC records, because of multiple offences being committed on the first offence date. In this situation, we sorted records based on the remaining PNC variables of interest to this analysis, ordering first by offence category (in the order 'serious violence', 'any other offence type', 'offence type not known'), then by sex (in the order 'female', 'male', 'unknown'), then by ethnicity (in the order 'White - North European', 'White – South European', 'Black', 'Asian', 'Chinese, Japanese or South East Asian', 'Middle Eastern', 'Unknown') (see below for further details on these variables). We then kept the first record for each individual after sorting on these variables.

After cleaning and de-duplicating each dataset, we matched the PNC to the PLASC and Spring Census, only keeping individuals who successfully matched to either the PLASC or Spring Census datasets. We finally matched in KS2 attainment where available. The resulting cohort comprised n=638,228 individuals.

Gender

In the PNC, sex is available as a categorical variable ('Male', 'Female', and 'Unknown'). Sex is a mandatory field in the PNC when logging an offence. The description for the variable 'sex' is 'gender of the subject'. How the variable is interpreted can depend on the officer completing the record, and therefore it may contain either gender identity or sex assigned at birth.

In the NPD, gender was derived from the PLASC (2001/2002 – 2004/2005) and Spring Census (2005/2006 – 2016-2017) as a binary variable ('Male' and 'Female'). To align with PNC coding of sex, individuals missing gender were labelled 'Unknown'. Current guidance for schools completing the school census requires that gender is "self-declared and recorded according to the wishes of the parent and/or pupil. Individuals are free to change the way their gender is recorded" (Department for Education, 2022). However, guidance on this has changed over the years: for example, the 2008 guidance document on completing the school census did not specify who should designate pupil gender. As with PNC sex, how the variable is interpreted might depend on the individual completing the record, and therefore it may contain either gender identity or sex assigned at birth.

Ethnicity

In the PNC, ethnicity is a mandatory field when logging an offence. It is described as "an indication of the ethnic appearance of a subject". Ethnicity is determined by the officer completing the record, based on their interpretation of the individual's visual appearance.

Ethnicity categories in the PNC are sometimes referred to as Phoenix ethnicity codes ('White – North European', 'White – South European', 'Black', 'Asian', 'Chinese, Japanese or South East Asian', 'Middle Eastern', 'Unknown'). Phoenix ethnicity codes can be further summarised as broader 4+1 ethnicity categories (Table 2), consistent with a previous report (Ministry of Justice, 2021).

Table 2: Categorising PNC ethnicity codes as broader ethnicity categories

Phoenix ethnicity codes	Broad ethnicity categories
White – North European	White
White – South European	
Black	Black
Asian	Asian
Chinese, Japanese or South East Asian	Other
Middle Eastern	
Unknown	Unknown

NPD ethnicity was derived from PLASC (2001/2002 – 2004/2005) and Spring Census (2005/2006 – 2016/2017) datasets. In the Spring Census, ethnicity was available as major ethnic groups (Appendix 1). In the PLASC 2002/2003 – 2004/2005, only extended categories were available, but we categorised these into major ethnic groups to match the Spring Census. In the PLASC 2001/2002, different ethnicity codes were used, and we again categorised these into major ethnic groups to approximate the Spring Census (Appendix 2). Because of small cell sizes in the extended categories, we primarily focused on major ethnic groups in this analysis. Minor ethnic groups derived from the PLASC and Spring Census were also used in one analysis, but only to further understand some of our findings relating to Mixed ethnic groups. To align with the PNC, individuals missing ethnicity, or whose ethnicity was categorised as 'Unclassified', 'Refused', 'Information not yet obtained', or any unrecognised code, were labelled 'Unknown'.

To further assist comparison with the PNC, we also created an ethnicity variable in the NPD to approximate the PNC's broader 4+1 ethnicity categories (Table 2). This was based on major ethnic groups, as shown in Table 3.

Table 3: Categorising NPD major ethnic groups as broader ethnicity categories

Major ethnic groups	Broad ethnicity categories
White	White
Black	Black
Asian	Asian
Chinese	Other
Mixed	
Other	
Unknown	Unknown

Current guidance for schools completing the school census requires that "the school must not ascribe any ethnicity to the pupil. This information must come from the parent/guardian or pupil" (Department for Education, 2022). However, as with gender, guidance on this has changed over the years: for example, the 2008 guidance document on completing the school

census did not specify who should designate pupil ethnicity. Indeed, at that time, source of pupil ethnicity was also collected during the school census, to distinguish between ethnicity provided by parents, pupils, and ascribed by schools (see below for further details on this variable). In both the NPD and PNC, interpretation of the term 'ethnicity' might depend on the individual or officer completing the record (for example, whether it primarily captures skin colour, cultural background, or nationality, among other possible interpretations).

Ethnicity source

In the NPD, a variable indicating the source of ethnicity data was available from academic years 2001/2002 to 2009/2010 (parent, child, current school, previous school or another source). We summarised this variable for the years when it was available. In the PNC, no equivalent variable is available, but it is ascribed by the officer completing the record based on their interpretation of appearance.

Offence type

Offence type was derived from the PNC to indicate whether a serious violence offence was committed. Home office offence codes were flagged as 'serious violence' if they were in the list of such codes used in a previous publication (Department for Education and Ministry of Justice, 2022). We additionally flagged as 'serious violence' any home office offence codes which more broadly indicated violence against the person offences, robbery offences, or possession of weapons offences committed in Guernsey, Jersey, Isle of Man, Northern Ireland or Scotland (more granular offence codes were not available for offences committed in these locations). All remaining valid offence codes, including breach offences, were defined as 'any other offence type'. Any invalid, missing, or 'not known' offence codes were labelled 'unknown'.

KS2 attainment

KS2 attainment data was derived from KS2 attainment tables in the NPD. Maths and English assessment national curriculum levels were converted to a binary indicator of whether each pupil achieved a Level 4, the expected attainment threshold at KS2 (Table 4). We then created a binary indicator of whether each pupil achieved a Level 4 in both maths and English KS2 assessments (Table 5).

If pupils were missing the maths and English Level 4 indicator after this procedure, we filled it in using the pre-made (but less complete) indicator supplied by the DfE.

Table 4: Converting national curriculum levels at KS2 to a binary indicator of achieving expected attainment thresholds.

National Curriculum Level	Level 4 achieved
Disapplied from national curriculum (D)	-
Invalid (IN)	-
Left school before test (L)	-
Missing (M)	-
Took test in past (P)	-
Lost scripts (X)	-
Not eligible for tests (Y/Z)	-
Absent (A)	No
Working below the level of the test (B)	No
Did not achieve level (N)	No
Maladministration (Q)	No
Unable to access tests (T)	No
2	No
3	No
4	Yes
5	Yes
6	Yes

Table 5: Creating a binary indicator of achieving Level 4 in both maths and English KS2 assessments.

Level 4 maths	Level 4 English	Level 4 maths and English
Yes	Yes	Yes
Yes	No	No
No	Yes	No
No	No	No
No	Missing	No
Missing	No	No
Yes	Missing	Missing
Missing	Yes	Missing
Missing	Missing	Missing

Statistical analysis

Analyses focused on gender/sex/ethnicity as recorded at each individual's first offence in the PNC, and gender/sex/ethnicity as recorded in the most recently available school census record in the NPD. To identify the most recently availably school census record, we first identified the 'academic year' in which the offence took place. For example, an offence committed 31st August 2005 took place in the academic year 2004/2005, whereas an offence committed 1st September 2005 took place in the academic year 2005/2006. We then preferentially took gender/sex/ethnicity from the school census record for that individual in that same academic year. If they did not have a school census record that year, or if gender/sex/ethnicity was recorded as 'Unknown', we took gender/sex/ethnicity from the school census conducted the previous academic year, and so on. Using this method ensured that, in most cases, gender/sex/ethnicity were usually being compared across the

datasets from reasonably proximate timepoints (Table 6). However, using this method also meant that for the small number of individuals who did not have a school census record from the same academic year as the offence, or from any previous year, they did not have NPD gender/ethnicity data available for analysis.

Table 6: Proximity between most school census record with available gender or ethnicity data and first offence record used in analysis, n=638,228

Proximity of most recent school census record to first offence record	Gender, n (%)	Ethnicity n (%)
Same academic year	372,812 (58.4%)	364,328 (57.1%)
1 academic year	68,019 (10.7%)	68,610 (10.8%)
2 academic years	53,716 (8.4%)	54,583 (8.6%)
3 academic years	43,350 (6.8%)	44,482 (7.0%)
4 academic years	33,121 (5.2%)	34,154 (5.4%)
≥5 academic years	64,995 (10.2%)	68,997 (10.8%)
No prior school census record available	2,215 (0.4%)	3,074 (0.5%)

We summarised the availability of gender/sex/ethnicity, and drew comparisons between the NPD and PNC, using descriptive statistics (frequencies and percentages).

Appendix 1: Mappings between extended categories, minor ethnic groups, and major ethnic groups in the NPD.

Approved extended categories	Minor ethnic group	Major ethnic group
White – British	White - British	White
White – Cornish		
White – English		
White – Scottish		
White – Welsh		
Other White British		
White – Irish	White - Irish	
Traveller of Irish heritage		
Any other White background	Any other White background	
Albanian		
Bosnian-Herzegovinian		
Croatian		
Greek/Greek Cypriot		
Greek		
Greek Cypriot		
Italian		
Kosovan		
Portuguese		
Serbian		
Turkish/Turkish Cypriot		
Turkish		
Turkish Cypriot		
White European		
White Eastern European		
White Western European		
White other		
Gypsy/Roma	Gypsy/Roma	
Gypsy		
Roma		
Other Gypsy/Roma		
White and Black Caribbean	White and Black Caribbean	Mixed/Dual background
White and Black African	White and Black African	
White and Asian	White and Asian	
White and Pakistani		
White and Indian		
White and any other Asian background		
Any other Mixed background	Any other Mixed background	
Asian and any other ethnic group		
Asian and Black		
Asian and Chinese		
Black and any other ethnic group		

Approved extended categories	Minor ethnic group	Major ethnic group
Black and Chinese		
Chinese and any other ethnic group		
White and any other ethnic group		
White and Chinese		
Other Mixed background		
Indian	Indian	Asian or Asian British
Pakistani	Pakistani	
Mirpuri Pakistani		
Kashmiri Pakistani		
Other Pakistani		
Bangladeshi	Bangladeshi	
Any other Asian background	Any other Asian background	
African Asian		
Kashmiri other		
Nepali		
Sri Lankan Sinhalese		
Sri Lankan Tamil		
Sri Lankan other		
Other Asian		
Black Caribbean	Black - Caribbean	Black or Black British
Black - African	Black - African	
Black - Angolan		
Black - Congolese		
Black – Ghanaian		
Black – Nigerian		
Black - Sierra Leonean		
Black – Somali		
Black – Sudanese		
Other Black African		
Any other Black background	Any other Black background	
Black European		
Black North American		
Other Black		
Chinese	Chinese	Chinese
Hong Kong Chinese		
Malaysian Chinese		
Singaporean Chinese		
Taiwanese		
Other Chinese		
Any other ethnic group	Any other ethnic group	Any other ethnic group
Afghan		
Arab other		
Egyptian		
Filipino		

Approved extended categories	Minor ethnic group	Major ethnic group
Iranian		
Iraqi		
Japanese		
Korean		
Kurdish		
Latin/South/Central American		
Lebanese		
Libyan		
Malay		
Moroccan		
Polynesian		
Thai		
Vietnamese		
Yemeni		
Other ethnic group		
Refused	Refused	Refused
Information not yet obtained	Information not yet obtained	Information not yet obtained

Appendix 2: Mappings between PLASC 2001/2002 ethnic groups, minor ethnic groups, and major ethnic groups in the NPD.

PLASC 2001/2002 ethnic groups	Minor ethnic group	Major ethnic group
White, UK heritage	White - British	White
White - British		
White - Irish	White - Irish	
White, European	Any other White background	
White, other (known)		
White - any other White background		
White, type not known		
Black, Caribbean heritage	Black - Caribbean	Black or Black British
Black or Black British - Caribbean		
Black, African heritage	Black - African	
Black or Black British - African		
Black, other	Any other Black background	
Black or Black British - any other Black background		
Indian	Indian	Asian or Asian British
Asian or Asian British - Indian		
Pakistani	Pakistani	
Asian or Asian British - Pakistani		
Bangladeshi	Bangladeshi	
Asian or Asian British - Bangladeshi		
Asian or Asian British - any other Asian background	Any other Asian background	
Chinese	Chinese	Chinese
Mixed - White and Black Caribbean	White and Black Caribbean	Mixed/Dual background
Mixed - White and Black African	White and Black African	
Mixed - White and Asian	White and Asian	
Mixed - any other mixed background	Any other Mixed background	
Other (known)	Any other ethnic group	Any other ethnic group
Any other ethnic group		
Parent/pupil preferred not to say	Refused	Refused
Ethnic group information not sought	Information not yet obtained	Information not yet obtained
Information not obtained		

References

Department for Education. (2022). Complete the school census. Retrieved from
https://www.gov.uk/guidance/complete-the-school-census/data-items-2021-to-2022

Department for Education and Ministry of Justice. (2022). Education, children's social care and
offending. Retrieved from https://www.gov.uk/government/publications/education-
childrens-social-care-and-offending