ALSPAC and Avon & Somerset Police linked dataset: background and overview

ALSPAC Data Linkage Team

Population Health Sciences

University of Bristol

alspac-linkage@bristol.ac.uk

September 2022

Version 1.0

Summary

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a birth cohort study based at the University of Bristol. The study started with the recruitment of approximately 15,000 pregnant women living in and around the City of Bristol during 1991 and 1992. The study has collected data on all aspects of the study children's lives – including extensive data describing socio-economic, family, educational, and health circumstances. The participants' local police records have now been linked to the ALSPAC database to create a new resource for crime-related research. This report provides the background to this linkage, details of the linkage process, and a summary of the police data. The overall aim is to inform researchers of the data available, enabling them to consider research questions that could be addressed using the police data, subject to permissions from the ALSPAC Executive. Research areas of particular interest to Avon & Somerset (A&S) Police are listed in this report.

Any queries about the police data should be sent to the ALSPAC data linkage team mailbox: alspac-linkage@bristol.ac.uk.

Key points:

- ALSPAC has permission to link to criminal record data of 12,662 participants (born in 1991/92).
- As of July 2021, 1757 of these participants (14%) had an A&S police record related to a charge or out-of-court disposal.
- 73% of participants with an A&S police record were male.
- Between them the 1757 individuals have 6413 A&S police records .
- Most of the individuals with a record have a small number of records: 47% have one, and 18% have two (range 1 to >150, median 2).
- The most common offence types (as defined by the Home Office) were violence against the person, drug offences, and theft offences.
- For slightly less than half (46%) of the ALSPAC participants with an Avon & Somerset police record, the first record was before the age of 18.

Contents

1. Background to the linkage

1.1 Why link police data to a birth cohort study?

Policing in the UK increasingly seeks to take a public health approach to tackling crime, where the focus is on proactive prevention, the tackling of upstream risk factors, and on populations rather than individuals (1, 2). This approach relies on 'the skilled use and interpretation of data and the evidence base to ensure that interventions are designed, delivered and tailored to be as effective as possible'(1). However, police records of criminality (e.g. convictions and cautions) do not contain data relating to an individual's exposure to potential risk factors for perpetrating crime. In contrast, longitudinal birth cohort studies such as ALSPAC have a wealth of data on the lives of their participants, and often their families, peers, and wider contexts, across the life course. Therefore, linking police data with cohort studies has the potential to add considerable value to research on criminal behaviour, allowing the study of both the antecedents and consequences of involvement with the criminal justice system.

1.2 Avon Longitudinal Study of Parents and Children (ALSPAC)

ALSPAC is a world-leading cohort study based at the University of Bristol. Pregnant women living in a defined geographical area in and around Bristol, who had an expected due date between April 1991 and December 1992, were eligible to take part. Full details are given in the cohort profiles (3, 4). In brief, there were 14,541 pregnancies resulting in 13,988 children alive at one year of age (known as the 'core sample'). By age 18 years, an additional 718 children, who met the original study eligibility criteria, but whose mothers had not joined the study during pregnancy, had also been recruited. The mothers, their partners, and the study children have been followed-up regularly through questionnaires and clinic visits. (Note that although the ALSPAC children are now adults, we refer to them in this report as the 'study children' to distinguish them from their parents.) This has resulted in a detailed database recording information on wide-ranging aspects of participants' lives including their health, wellbeing, and family and socio-economic circumstances. It also includes measures of self-reported criminal and anti-social behaviours and involvement with the criminal justice system (these ALSPAC crime-related measures are not the focus of this report, but more detail on these is given in Appendix A). ALSPAC data are available to legitimate researchers for public benefit research purposes under controlled conditions. A

fully searchable data dictionary can be downloaded from the website (www.bristol.ac.uk/alspac/researchers/our-data/).

1.2.1 Background to record linkage in ALSPAC

Linkage of participants in cohort studies to their administrative and health data has benefits for both participants and researchers. For participants, it is a low-burden way to participate in a study compared to completing questionnaires or visiting clinics. For researchers, it provides measures that participants would not be able to self-report accurately (e.g. air pollution levels near their home, or prescription medications taken as a child).

When the ALSPAC children reached legal adulthood (age 18 years), there was a postal campaign which informed participants about ALSPAC's intention to link their ALSPAC data to their routine health and administrative records, including education, employment, earnings and benefits, and criminality records. Each participant was sent a pack that included an information booklet and consent form, which provided a clear means to opt out of ALSPAC, or to any of the proposed linkages (5, 6). This pack was designed with input from study participants and legal and ethical advisors, and underwent rigorous independent ethical review.

Due to factors related to establishing an appropriate ethico-legal basis for record linkage in ALSPAC and the negotiation of access to linked *health* records (i.e. unrelated to the criminality data linkage), the participant information materials were initially issued in two batches. Batch one sought opt-in consent, which stated that linkage would only occur with explicit participant approval, while batch two was structured as an opt-out approach and notified participants that their routine records would be linked to ALSPAC unless they specifically opted out (i.e. linkage would occur in the event of non-response). Participants who did not respond to batch one were a sent a batch two pack. This change in approach was requested and designed with input from the Information Commissioner's Office as part of a wider programme considering best practice as to how Data Protection regulations should be applied for this purpose (this programme was unrelated to ALSPAC but occurred during the time in which ALSPAC were conducting this activity). Overall, few participants have opted-out of record linkage. To date, 4% have opted-out of linkage to criminality records.

Many of the proposed data linkages have now been achieved, including to health, education, and geographic records. With regards criminality linkage, the original plan was to link the study children to records held in the Police National Computer (PNC), a large administrative database held by the Ministry of Justice (MoJ) that was started in 1974 and contains information about police cautions and court convictions held on individual offenders in England and Wales (7, 8). A pilot linkage was achieved, as detailed in the next section, but this did not progress to a full linkage.

1.2.2 Pilot linkage of ALSPAC to PNC criminality records

Full details of the pilot linkage of ALSPAC to the PNC, conducted in 2013, have been published (9). In brief, following negotiations between ALSPAC and the MoJ, it was agreed that a pilot linkage exercise would be conducted to test the feasibility of the linkage mechanism through the production of an anonymous linked extract (i.e. the extract from the PNC could not be linked to any ALSPAC data). The main conclusions of this pilot were:

(1) linkage to criminality records is acceptable to the majority of ALSPAC participants, including many who have a criminality record;

(2) there are sufficient levels of criminality in the ALSPAC cohort for it to be a resource for crime-related research;

(3) those who opted-in to criminality linkage are more socio-economically advantaged, and self-report fewer criminal behaviours, than those who were non-responders to the linkage campaign;

(4) the majority of offences (86%) were committed in the Avon and Somerset Police area, the policing area local to ALSPAC.

As a result of these findings, and in light of the full linkage to the PNC not being forthcoming, it was decided that pursuing linkage to local police data, held by Avon and Somerset Police, would be a valid and worthwhile alternative approach.

1.3 Avon and Somerset Police

Avon and Somerset Police are responsible for law enforcement in the four counties that replaced the now abolished county of Avon (Bristol, Bath and North-East Somerset, North Somerset, and South Gloucestershire), plus the county of Somerset. The A&S Police area therefore includes the ALSPAC recruitment area.

Offences committed in the Avon and Somerset area, and which come to the attention of the police, are recorded in the A&S Police database. Since September 2015, A&S Police have used the NicheRMS365 cloud platform as their record management system (10). Older records have been migrated to the Niche platform. Offences in other areas of the country, or abroad, are not recorded in their database.

The age of criminal responsibility in England is 10; children below this age cannot be arrested, charged or cautioned if they break the law. The UK has no statute of limitations for any criminal offence, meaning an offence can be reported to the police at any time. However, the police and Crown Prosecution Service cannot instigate criminal proceedings for 'summary only offences' (minor crimes that are heard in a Magistrates Court e.g. common assault, low value shoplifting, graffiti) if more than 6 months have passed since the date of the offence. For 'indictable only' offences (heard in a Crown Court) and 'triable either way offences' (can be tried as either a summary or indictable offence), there is no time limit for bringing a prosecution.]

A&S police records include the disposal outcome(s) for each crime. A description of the disposal types that ALSPAC has fair processing permission to link to is given in Appendix B. In brief, these include charges, offences 'taken into consideration' (TICs) and out-of-court disposals. The threshold of evidence needed for an individual to be charged is high (11), and the majority will go on to face trial in court. Out-of-court disposals comprise cautions, penalty notices, drugs warnings and community resolutions. To be issued with an out-of-court disposal, an individual must admit they are guilty of the offence and be eligible in terms of previous recorded offending (they are intended to deal with low level offending). Notably, and in contrast to the PNC, A&S police do not routinely record conviction data.

More details on the police data that have been linked to ALSPAC are given in Sections 3 and 4 of this report.

1.4 Priority areas for research

Data managers and researchers from the ALSPAC data linkage team collaborated with A&S Police to identify priority policing areas that could potentially be researched using a linked dataset. This led to the development of a project that began in 2021 and is investigating the association between childhood adversity and involvement in violent crime in adolescence

and young adulthood. The first part of that project involved achieving the linkage between ALSPAC and the police data, as detailed in this report. That has now been achieved, and the linked dataset is available to be used for other projects, subject to approval from the ALSPAC Executive.

Research proposals which address one or more of A&S Police's priority areas are particularly welcomed:

- What are the key risk factors to becoming involved in crime or not? Are their key 'cross roads'?
- What is the link between health/mental health and crime/crime prevention (public health system approach)?
- What universal service interventions are most likely to have the greatest impact on reducing likelihood of being involved with crime (and at what age)?

Researchers interested in developing a proposal in one of these areas should contact the ALSPAC Data Linkage Team in the first instance. If appropriate, we can put researchers in touch with our contact at A&S Police to obtain further details on the priority area they are interested in.

2. Linkage methodology

Data Processing Agreements for the transfer of A&S police data to ALSPAC for this project were finalised in spring 2020, and the police data were transferred to ALSPAC in summer 2021. It is anticipated this linkage will be updated every 2-3 years.

The linkage of ALSPAC to Avon and Somerset Police data took place in two stages. Stage 1 involved the ALSPAC data linkage team establishing the linkage using personal identifiers common to both the ALSPAC participant database and A&S police records. Stage 2 involved A&S Police extracting attribute data on the matching individuals, removing identifiers and securely sharing the matched records with ALSPAC. All data processing was conducted by the three Data Managers in the ALSPAC Data Linkage Team, all of whom were individually security cleared by A&S Police prior to the commencement of this project. All data processing took place within the ALSPAC Data Safe Haven, which is accredited to the ISO27001 information security standard.

2.1 Using personal identifiers to establish matches

As there is no strong, persistent personal identifier common to both ALSPAC and the A&S Police datasets (as NHS number would provide for a health data linkage for example), a number of identifiers available in both datasets were used to determine which individuals in ALSPAC had an A&S Police record. This used deterministic and probabilistic record linkage methods to maximise linkage coverage and minimise false matches. In brief: deterministic linkage compares whether identifiers in ALSPAC exactly match those found in A&S police records (e.g. first name, family name, date of birth, postcode all in agreement); probabilistic linkage makes similar comparisons but uses Bayesian statistical methods to assign a probability to the two records matching based on the level of agreement (allowing variations such as full name and short form names, or errors in postcode details) and the quality and distribution of the identifiers in the wider population (e.g. accounting for issues such as the family name 'Smith' being very common and therefore more likely to match by chance).

A&S Police sent ALSPAC the following identifiers: forename, surname, date of birth (DoB), sex and full current and historical address(es) of all individuals held in their database who were born between 1st January 1991 and 31st January 1993 (the date range in which the

ALSPAC study children were born). No information about these individuals was sent other than the identifiers. The same identifiers were extracted from the ALSPAC participant database for the 12,662 participants for whom ALSPAC had permission to link to criminal record data [comprised of those who explicitly opted-in to criminality linkage (n=5055) and those who received the opt-out linkage form and did not respond (n=7607)].

Firstly, a deterministic match was completed using forename, surname, and DoB. This yielded 1876 matches. Postcode was then used to create a match strength variable (Table 1).

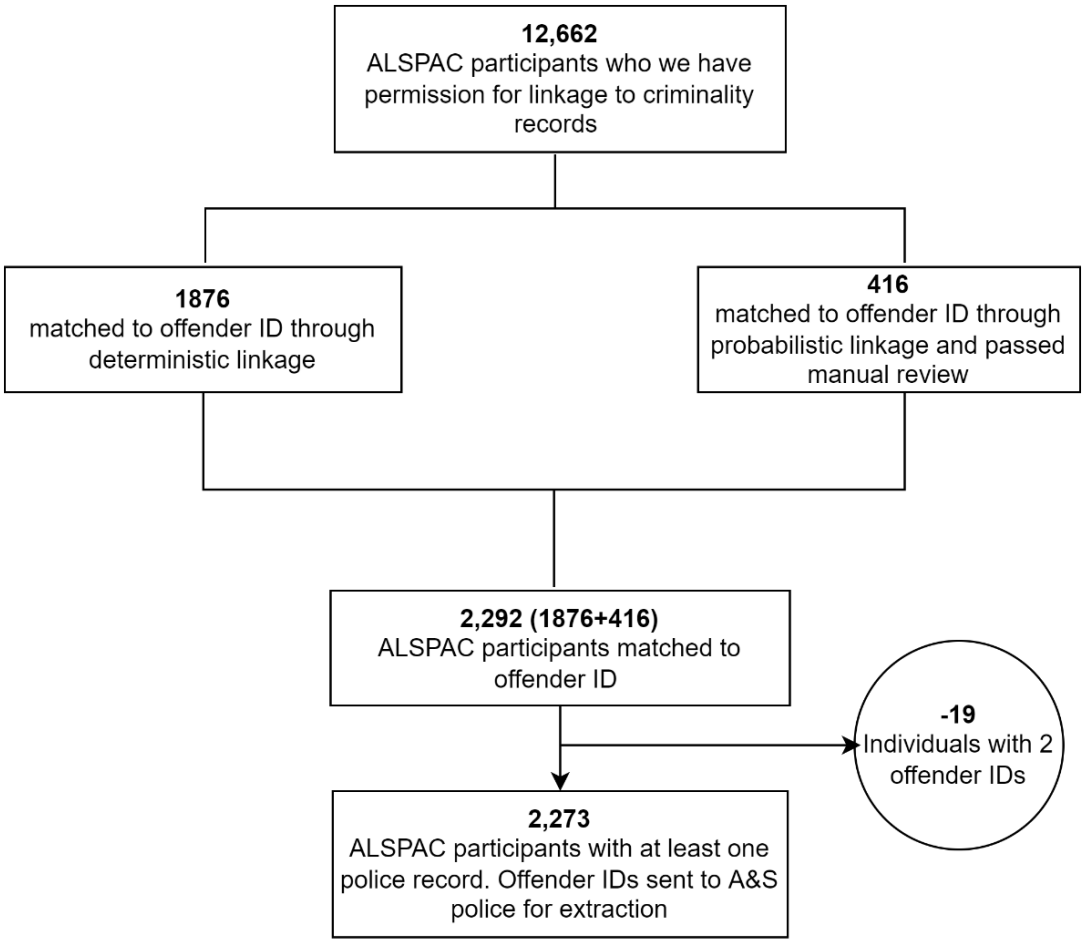
Table 1. Deterministic match criteria and number of matches

Matching Criteria	Match strength	Number of matches
Forename, surname, DoB, full postcode	1	956
Forename, surname, DoB, first half of postcode	2	403
Forename, surname, DoB	3	517
		1876 TOTAL

Secondly, a probabilistic matching procedure was carried out using the LinXmart record linkage software developed by Curtin University, Australia (12). Using this method yielded 2,292 matches. This included all 1876 linked using the deterministic linkage process plus an additional 416 matched through probabilistic linkage and who passed manual review. In final checks, it was found that 19 individuals linked to 2 *offender_ids* (police individual-level ID) i.e. the police have marked them as different people in their database, but they are the same person according to the ALSPAC database. In these cases, records belonging to both *offender_ids* were kept and linked to the same individual. Therefore, in total we identified 2,273 ALSPAC individuals who had at least one record in the A&S Police dataset (Figure 1).

At the end of this process, all personal identifiers provided by A&S Police were securely destroyed in line with ALSPAC's ISO27001 certified processes. This left an ID match variable (ALSPAC ID to A&S *offender_ids*) and linkage quality variables.

Figure 1: Flow chart of linkage of ALSPAC participants to Offender IDs



2.2 Extracting attribute data

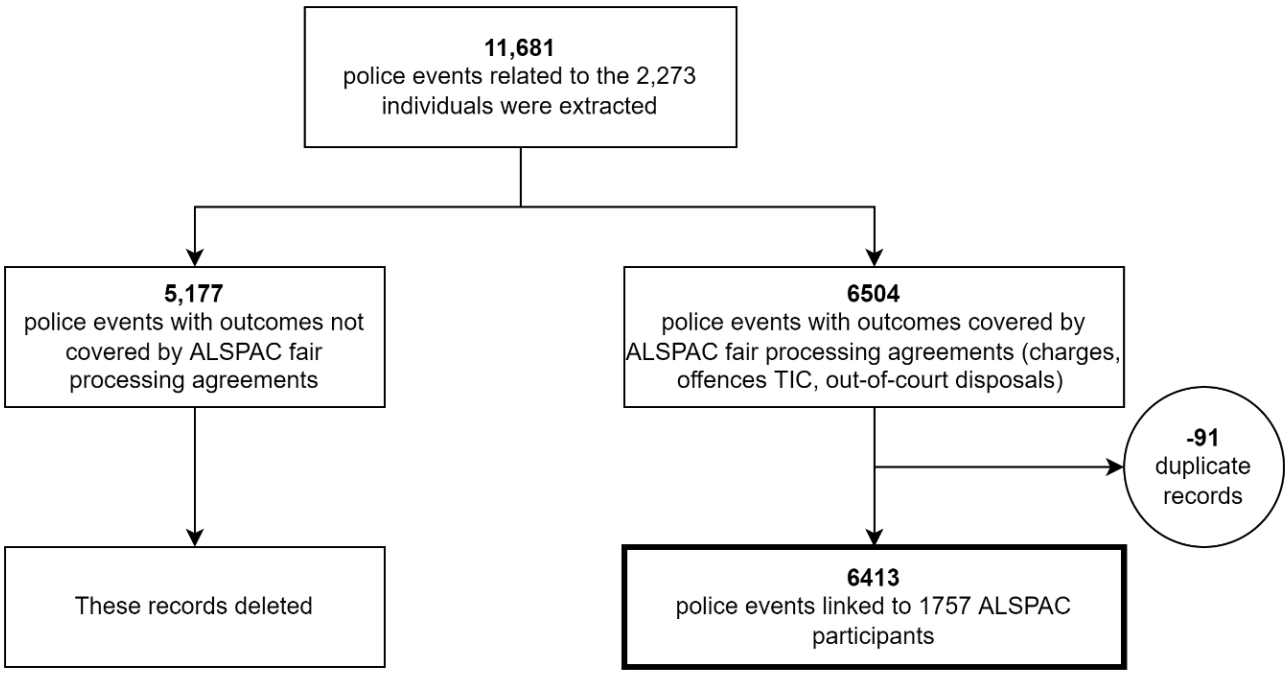
Avon and Somerset Police then extracted 11,681 police events related to the 2,273 individuals matched in stage 1 and securely transferred these to ALSPAC. In this event-based dataset, each row corresponds to an individual and a crime occurrence. The first step was to identify records with a disposal type which was in line with ALSPAC's fair processing communications with participants: charges, offences 'taken into consideration' (TICs), and out-of-court disposals (i.e. cautions, drug warnings, penalty notices, and community resolutions). As the police record many details of crimes at an offence level rather than at an individual offender level, different outcome variables had to be used to ascertain the

disposal type for offences involving one individual versus offences involving more than one person ('group crimes'). Full details of this process are given in Appendix C. In summary, the following records were linked to ALSPAC:

- **For offences involving only one person:** records were linked to ALSPAC if the main outcome variable (*currentclassificationhooutcom*) was OC1 (charged), OC2-3 (cautioned), OC4 (taken into consideration), OC6 (penalty notice for disorder), OC7 (cannabis warning), or OC8 (community resolution) (Appendix C Table 1).
- **For offences involving more than one person:** records were linked to ALSPAC if the concatenated outcome variable (*offenderclassificationconcat*) contained at least one of the terms which relate to OC 1-4 and 6-8 described above: charged, TIC, cautioned, adult conditional caution, postal requisition, reported for summons, cannabis warning, penalty notice for disorder, community resolution, prosecuted (Appendix C Table 2).

All other records were deleted (these include records where the individual had been eliminated from enquiries, or where there was insufficient evidence to proceed). This resulted in a final sample of 6413 police records related to 1757 individuals (Figure 2).

Figure 2: Flow chart of linkage of police events records to ALSPAC participants



3. The Avon & Somerset Police dataset

3.1 Variables

The dataset provided by A&S police contains 19 variables. These comprise administrative variables, date variables that specify when an offence took place and when it was reported to police, type and severity of offence variables, disposal type variables, flag variables, and variables related to Magistrate's Court appearances (Table 4). Most of the variables are complete or have very small amounts of missing data.

There are three date variables. The date an offence took place is given by *occurrencefromdate*. The other two date variables are system generated: *occurrencecreateddate* is the date the incident was reported to the police and they determined that it was a crime, and *occurrencereporteddate* is the date the crime was entered into the STORM Command and Control system, which is where the police assess 999/101 calls and allocate officers to the scene.

Initial inspection of the date variables revealed one impossible *occurrencefromdate* (date in 1930), which was set to missing. Five records had an *occurrencereporteddate* earlier than the *occurrencefromdate* (i.e. implying the offence took place after the crime was reported) - the difference between the dates for these 5 records ranges from 1 to 26 days and is likely due to data entry errors in *occurrencefromdate*.

For 50% of the records, all three date variables are identical (i.e. the crime took place on the same day that it was reported to the police and entered into the police computer systems). The two system generated date variables, which are both measures of when a crime was reported, differ in only 286 records (4.3%). In all but two of these non-matching records, *occurrencereporteddate* is the earlier date. Where there is a difference between the dates, it tends to be small: over half (54%) of the 286 records have a difference of 1 day, almost three quarters (73%) of the records have a difference of <10 days, and 85% have a difference of <1 month (31 days). (Median 1 day, mean 21 days, range 1 to 324 days).

For over half (53%) of the records *occurrencefromdate* is equal to *occurrencereporteddate* (i.e. the crime was reported on the same day it occurred). For the remaining 3040 (47%) of records, the difference between the dates ranges from 1 day to over 10 years. 45% of these records have a difference of only 1 day, and 73% have a difference of <10 days. Therefore,

the majority of crimes were reported on the day they occurred or within a few days. However, 7% of the records with a difference between the dates have a difference of over a year. Only seven records had a time difference of over 3 years.

There are four 'flag' variables which specify if a crime involved domestic abuse, knife crime, drugs, or alcohol. There is an additional variable which specifies whether an offender was using drugs and/or alcohol (*currentsubstanceusedbyoffend*) but this has very high levels of missing data as it is no longer used by the police in their reporting.

The nature of the offence is given by *currentoffencehocode* (the Home Office code for the offence), *currentoffencedescription* (a detailed categorical variable which describes these codes), and *currentoffencegroup* (a categorical variable which assigns each of the offences to one of 12 offence groups). For example, *currentoffencedescription* describes a code as 'possession of cannabis' and *currentoffencegroup* assigns that offence to the 'drug offences' category.

The variable *scorexmultiplier* indicates each offence's severity. These scores are used by the police to monitor harm as opposed to just crime volumes, and to enable them to identify the most high risk offenders and most vulnerable victims. These scores are used only by the police and not by the courts. The *scorexmultiplier* value is derived from the 'harm score' for the offence, increased if relevant by a 'multiplier'. Each Home Office offence code has a corresponding harm score, ranging from 0.01 to 100. Offences with a harm score <3 include intent to supply class A drugs (harm score of 0.8), wounding with intent to do serious bodily harm (1.45), and rape (2.9). No offences have a harm score between 3 and 8. Crimes with scores ≥8 include conspiring to traffic a person into the UK for exploitation (8), causing or inciting child pornography (10), manslaughter (30), use of noxious substance in terrorism offence (50), and murder (100). These harm scores are increased by a multiplier if the following factors are present: +30% for domestic abuse related, +50% for hate related, +5% for drug related, +10% if there is a firearm tag, and +30% if there is a safeguarding children tag. If more than one of these factors is present, the multipliers are cumulative and applied in the order listed.

The final three variables relate to Magistrates' Court appearances (available from November 2015 only): *casefileid* is the ID for that court appearance, *casefilecreateddateandtime* gives

the date of the court case, and *verdict* states whether the defendant was found guilty or not guilty (this variable has high levels of missing data).

Table 2: Variables in the dataset provided by A&S police

Variable type	Variable name	Description	% missing (100%=6413)	Available?
Administrative	occurrence_id	ID of the crime	0%	Yes[3]
	offendercount	How many offenders were involved in the crime	0%	Yes
Date	occurrencecreateddate	System generated, triggered by a 111/999 call about an occurrence that the officer later declares a crime, or similar.	0%	No (age available)
	occurrencereporteddate	Automatically entered when the crime occurrence is created (generated from STORM[1] and pushed to Niche[2]).	0%	No (age available)
	occurrencefromdate	Date of the offence, person reported via 111/999 or any other way	0.1%	No (age available)
Type/severity of offence	currentoffencegroup	12 category variable giving type of offence	0%	Yes
	currentoffencedescription	Offence description	0%	No
	currentoffencehocode	Offence Home Office code	0%	No
	scorexmultiplier	Crime severity score	0%	Yes
Disposal type	currentclassificationhooutcom	Offence-level. Home office outcome code and description	0%	No
	offenderclassificationconcat	Individual-level. String variable with up to 6 terms. This has been split into 6 separate variables.	0%	No
Flag	currentsubstanceusedbyoffend	Offender affected by: alcohol; alcohol and drugs; drugs; not affected; not known. This flag started being used in the mid-2000s but has since fallen into disuse. Not mandatory field.	95.2% (99.0% if not known category is excluded)	No
	domesticabuseindicator	Crime involved domestic abuse (no/yes)	0%	Yes
	knifecrimeindicator	Crime involved a knife (no/yes)	0%	Yes
	drugsflagged	Crime involved drugs (no/yes)	0%	Yes
	alcohol	Crime involved alcohol	98.1%	No
Magistate's Court	casefileid	ID of Magistrates' court case	87.9%	Yes[3]
	casefilecreateddateandtime	Date of court case	88.2% (3.0% of those with a casefileid)	No (age available)
	verdict	Verdict of Magistrates' court case (Not guilty; guilty)	92.0% (37.9% of those with a casefileid)	Yes

[1]STORM is the name of the Command and Control system, which is where the police assess 999/101 calls and allocate officers to scene.
[2]Niche is the crime recording system.
[3]A pseudonymised version of these variables is available.

3.2 Changes made to the A&S police data to prevent disclosure

The ALSPAC Data Linkage managers have made the following changes to the police data to prevent disclosure of ALSPAC participants' identities during research use:

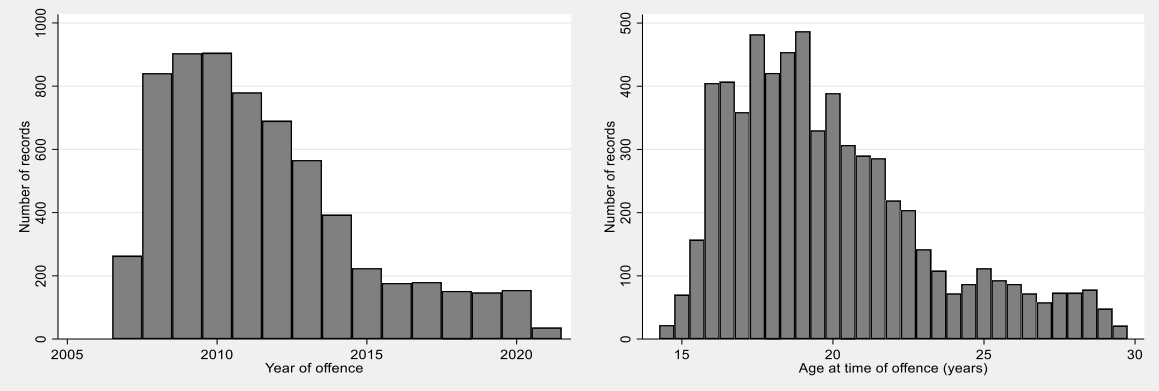
- The *occurrence_id* and *casefileid* variables have been pseudonymised but retain equivalent functionality.
- The date variables will not be released. Instead, the age of the participant on each of these dates has been calculated (in months) using their date of birth. Month and year of offence will be available.
- The original outcomes variables (*currentclassificationhooutcom* and *offenderclassificationconcat*) will not be released. A binary variable has been derived (participant has a police record, yes or no). This binary variable ensures all ALSPAC participants with a record are treated equally (as details on type of disposal are not available for individuals involved in group crimes prior to September 2015, as described above).
- The variables that describe the nature of the offence in detail (*currentoffencehocode* and *currentoffencedescription)* will not be released to researchers in their original format as they have many categories with small cell counts. The offence group variable (*currentoffencegroup*) will be available. If required, researchers can discuss with the ALSPAC Data Linkage Team options for grouping the Home Office codes in a different way to that available in *currentoffencegroup* variable.
- The *scorexmultiplier* variable will be aggregated at the upper end due to a small number of records with a high score.
- Variables with high levels of missing data will not be released.
- All variables will be assessed for small cell counts before being released to researchers.

4. Summary of A&S police data linked to ALSPAC

4.1 Overall summary of the records

The 6413 A&S police records cover the time period from when the participants were in their early teens through to their late 20s. The years 2009-2010 saw the largest number of offences (when the participants would have been aged in their late teens) (Figure 3, Table 3).

Figure 3: Distribution of records by year of offence and age at offence



Due to small numbers of offences from 2002-2006, the 2007 year includes all offences committed up to and including 2007. 2021 only includes records up until July 2021 (when data were extracted).
Due to small numbers of offences at the youngest and oldest ages, any offences below the age of 14.5 are included in the 14.5 group, and any offences over the age of 29.5 are included in the 29.5 group.

The police records linked to ALSPAC cover a wide range of offence groups. The most common groups are violence against the person (22% of records), drug offences (19%), theft (17%) and public order offences (11%) (Table 5). More details (from the *currentoffencedescription* variable) on the subgroups of each of the offence groups are provided in Table 6 for subgroups with >80 records.

Almost a third of records have a flag, most commonly for drugs (21% of records) (Table 3). Over three quarters of records are for a crime involving only one person. In terms of crime severity, the *scorexmultiplier* variable has a range of 0.01 to over 100, with most records having a relatively low score (74% of records have a severity score of ≤0.2). There were 55 records with a *scorexmultiplier* value between 2 and 3, and 11 with a score>3.

Table 3: Summary of A&S police records linked to ALSPAC

		N	% (6413=100%)
Offence group	Arson and criminal damage	807	12.7
	Burglary	466	7.3
	Drug offences	1237	19.3
	Fraud	44	0.7
	Miscellaneous crimes against society	157	2.5
	Possession of weapons	85	1.3
	Public order offences	683	10.7
	Robbery	102	1.6
	Sexual offences	45	0.7
	Theft	1077	16.8
	Vehicle offences	277	4.3
	Violence against the person	1433	22.4
Flag	Alcohol	118	1.8
	Drugs	1334	20.8
	Knife crime	109	1.7
	Domestic abuse	631	9.8
	Any flag	2068	32.3
Group crime	No	4954	77.3
Year offence committed	2002-2006	15	0.2
	2007-2008	1089	17.0
	2009-2010	1809	28.2
	2011-2012	1470	22.9
	2013-2014	959	15.0
	2015-2016	401	6.3
	2017-2018	332	5.2
	2019-2021	338	5.3

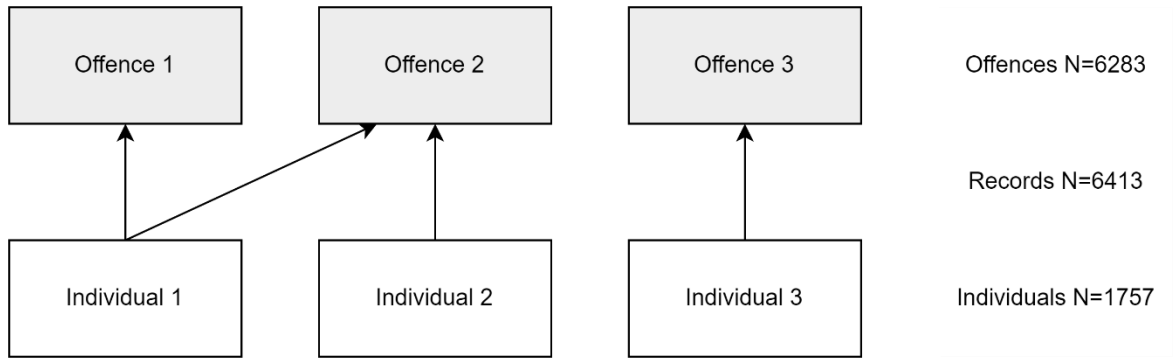
Table 4: Summary of offences and subgroups for A&S police records linked to ALSPAC

Offence Group (Total number of records)	**Subgroups with >80 records**		**N**	**%** (of total records in that offence group)
	Home Office Code	Offence description		
Violence against the person (N=1433)	008/06	Assault occasioning actual bodily harm	707	49.3
	105/01	Common assault and battery	373	26.0
Drug offences (N=1237)	092/66	Possess cannabis - class C (recordable)	849	68.6
	092/50	Having possession of a controlled drug - Cocaine	91	7.4
	092/61	Having possession of a controlled drug - Cannabis	83	6.7
Theft (N=1077)	046/00	Theft from shops and stalls	759	70.5
	049/10	Theft if not classified elsewhere	139	12.9
Arson and criminal damage (N=807)	149/00B	Other Criminal Damage to a Building Other than a Dwelling - valued under £5000	259	32.1
	149/00C	Other Criminal Damage to a Vehicle - valued under £5000	193	23.9
	149/00A	Other Criminal Damage to a Dwelling - valued under £5000	170	21.1
	149/00D	Other Criminal Damage, Other - valued under £5000	161	20.0
Public Order Offences (N=683)	125/12	Harassment, alarm or distress	315	46.1
	125/11	Fear or provocation of violence	94	13.8
	125/09	Causing intentional harassment, alarm or distress	88	12.9
Burglary (N=466)	028/01	Burglary/Attempted Burglary/Conspire to commit burglary - dwelling	277	59.4
	030/01	Burglary/Attempted Burglary/Conspire to commit burglary - other than a dwelling	140	30.0
Vehicle offences (N=270)	045/10	Theft from a motor vehicle	136	50.4

4.2 Summary of number of offences and number of individuals

The 6413 records linked to 1757 individuals, and to 6283 separate offences. The relationship between number of records, number of individuals, and number of offences is summarised in Figure 4. Each arrow in the figure can be thought of as a record in the dataset. In this example, Individual 1 has two records, one for Offence 1 and one for Offence 2. Individuals 2 and 3 have one record each. Individual 2's record is for offence 2, meaning this is a 'group crime' as more than one individual was involved.

Figure 4: Relationship between individuals, records and offences



Therefore, of the 12,662 ALSPAC participants who we had permission to link to police data, 13.9% (n=1757) have at least one A&S police record related to a charge, offence TIC, caution, or other out-of-court disposal. The percentage of participants not in the ALSPAC core sample (i.e. not recruited during their mother's pregnancy) is similar for those with (6.8%) and without (6.2%) a police record. Most of the individuals with a record have a small number of records: 47% have one, and 18% have two (range 1 to >150, median 2) (Table 5).

Table 5: Summary of number of records

	Number of participants (100%=12,662)
Any record	1757 (13.9%)
Number of records	
1	832 (6.6%)
2	316 (2.5%)
3-10	493 (3.9%)
11+	116 (0.9%)

6164 of the 6283 offences (98%) appear in the dataset only once (i.e. these are non-group crimes where only one individual was involved, or are group crimes where only one of the group is in the ALSPAC sample). The highest number of individuals in the dataset related to one offence is 4.

Of the 925 individuals with more than one police record, 45 had offences all recorded as being committed on the same day. For the other 880, the time span covered by the offences (calculated as the time between first and last offence) ranged from a few days to over 10 years, with a median of 3.9 years and a mean of 4.5 years (Table 6).

Table 6: Summary of time difference between first and last offences

Time difference	N of individuals (100%=925[1])
0 (offences all on same day)	45 (4.9%)
<6 months	87 (9.4%)
6 months to <1 year	51 (5.5%)
1 to <3 years	231 (25.0%)
3 to <5 years	189 (20.4%)
5 to <7 years	109 (11.8%)
7 to <9 years	82 (8.9%)
9 years +	131 (14.2%)

[1]The number of individuals in ALSPAC with more than one police record

Age at first (or only) A&S police recorded offence covers a wide range, from early teens to late 20s (mean 18.7 years, median 18.2 years). For slightly less than half (46%) of the ALSPAC participants with an A&S police record, the first record is before the age of 18. A quarter were aged 20 or over at first offence (Table 7).

Table 7: Summary of age at first offence

Age at first offence	N of individuals (100%=1757)
<16 years	212 (12.1%)
16 to <18 years	599 (34.1%)
18 to <20 years	498 (28.3%)
20+ years	448 (25.5%)

4.3 Summary of the records by sex

Of the 1757 individuals in the sample with a record, 73% (n=1279) are male and 27% (n=478) female. Therefore, one in five males in the ALSPAC sample has a police record compared to less than 8% of the females (Table 8). The males linked to a total of 5255

police records (82% of the total) and the females to 1158 (18% of the total). The mean age at first offence is similar for both sexes.

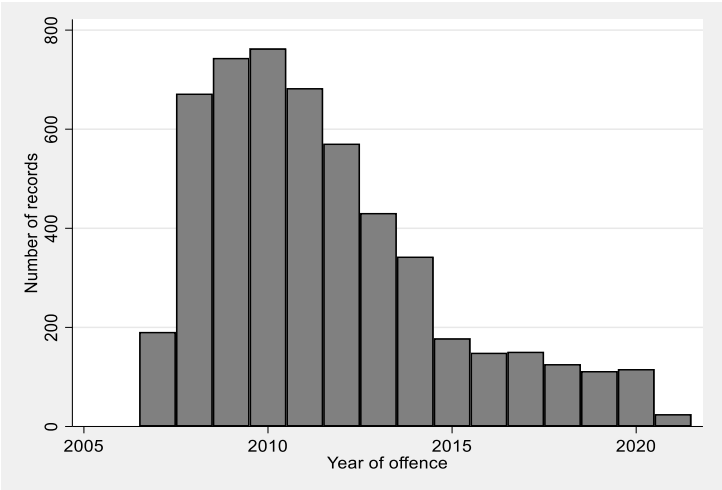
Table 8: Summary of police records by sex

		Males N=6368	Females N=6294
Any record	Yes	1279 (20.1%)	478 (7.6%)
Of those with records:			
Number of records	1	541 (8.5%)	291 (4.6%)
	2	229 (3.6%)	87 (1.4%)
	3-10	408 (6.4%)	85 (1.4%)
	11+	101 (1.6%)	15 (0.2%)
Age at first offence (years)	Mean (95% CI)	18.6 (18.5-18.8)	18.8 (18.5-19.1)
	<16 years	142 (11.1%)	70 (14.6%)
	16 to <18 years	434 (33.9%)	165 (34.5%)
	18 to <20 years	385 (30.1%)	113 (23.6%)
	20+ years	318 (24.9%)	130 (27.2%)

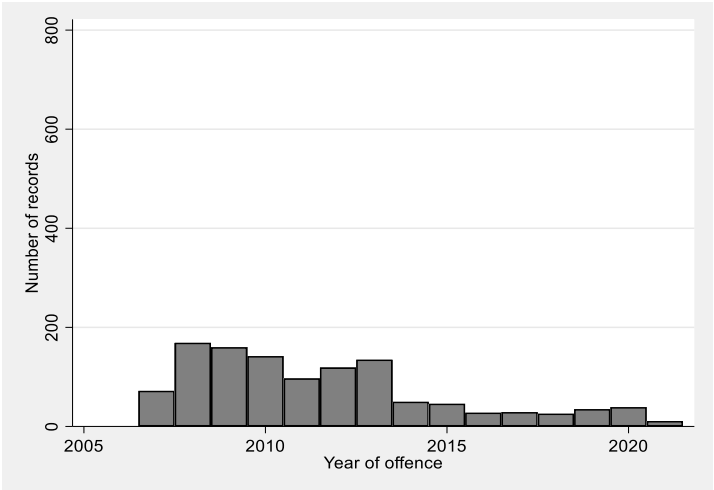
In addition to sex differences in levels of offending, there are differences in when it took place. The overall peak in number of records around the years 2009-2010 is driven largely by the boys' offending (Figures 5a and 5b). The girls have considerably fewer records, and the distribution of their records by year of offence is flatter.

Figure 5a: Distribution of year of offence by sex

(i) Males



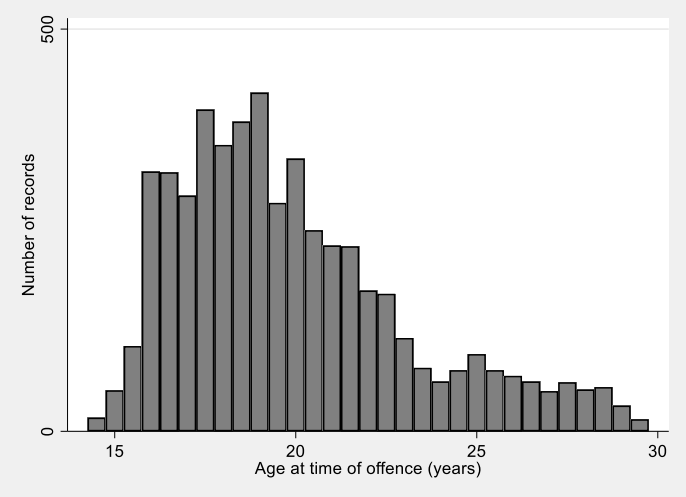
(ii) Females



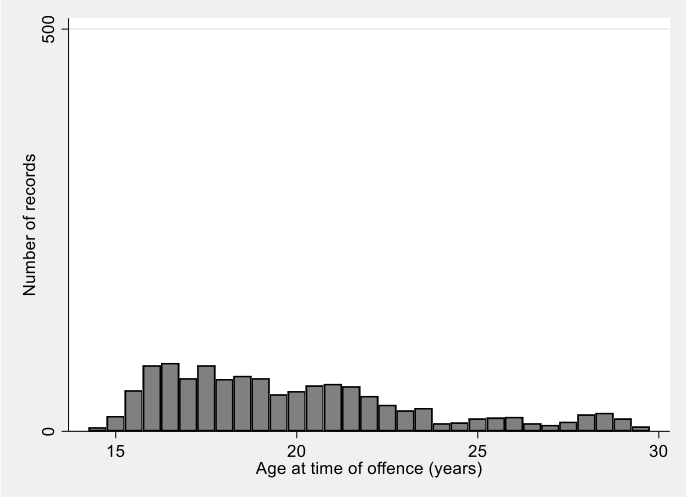
[1]Due to small numbers of offences from 2002-2006, the 2007 year includes all offences committed up to and including 2007. 2021 only includes records up until July 2021 (when data were extracted).

Figure 5b: Distribution of age at offence by sex

(i) Males



(ii) Females



[1]Due to small numbers of offences at the youngest and oldest ages, any offences below the age of 14.5 are included in the 14.5 group, and any offences over the age of 29.5 are included in the 29.5 group.

Violence against the person and drug offences are the most common offences for males; theft and violence against the person for females (Table 9). All offences were more common in males than females. The differences by sex were particularly large for sexual offences, vehicle offences and burglary (for these three offence types, >96% of records belonged to males). Differences by sex were smallest for theft (58% of records belonged to males).

Table 9: Summary of number of police records, by offence group and sex

Home Office offence group	Males N records (100%=5255)	Females N records (100%=1158)
Arson and criminal damage	732 (13.9)	75 (6.5)
Burglary	451 (8.6)	15 (1.3)
Drug offences	1095 (20.8)	142 (12.3)
Fraud[1]	-	-
Miscellaneous crimes against society	131 (2.5)	26 (2.3)
Possession of weapons	77 (1.5)	8 (0.7)
Public order offences	572 (10.9)	111 (9.6)
Robbery	88 (1.7)	14 (1.2)
Sexual offences[1]	-	n<5
Theft	622 (11.8)	455 (39.3)
Vehicle offences	268 (5.1)	9 (0.8)
Violence against the person	1141 (21.7)	292 (25.2)

[1] Cell counts suppressed to prevent calculation of the small cell count for sexual offence records for females.

5. Summary of ALSPAC data available for those with a crime record

5.1 ALSPAC questionnaire and clinic data

Those with an A&S police record have lower questionnaire response rates at all ages for most questionnaire types – maternal, partner (these are the mothers' partners – usually the study child's father), child-based (completed by the mother about the child), and child-completed (Figures 6a and 6b). Full details on the ages covered by these questionnaires can be found in the ALSPAC documentation, but as a guide, the mother and partner questionnaires shown in Figure 6a cover a period from the pregnancy with the study child up until the child was aged 12, and the child-based questionnaires shown in Figure 6b cover from 4 weeks to 19 years, and the child clinic from ages 7 to 24 years. An exception is seen for the teacher-completed questionnaires (for school years 3-8) where no difference in response is seen between children with and without a police record. This reflects the fact that, for most children, the teacher questionnaires were sent directly to the schools for completion, so completion (by the teacher) was not associated with any characteristics of the individual child or their family. Attendance rates at the child clinics (Figure 6b) are also lower for those with a police record.

Note that many participants complete some questionnaires, or attend some clinics, but not others. A lack of response to any given questionnaire, or non-attendance at a clinic, does not mean that the individual had dropped out of the study. Very few of the ALSPAC children have withdrawn their consent for their data to be used (n=28 as of March 2022) and it is estimated that around 10,000 of the original families are still involved with the study in some way (even though they don't all respond to every questionnaire or attend every clinic).

Figure 6a: Maternal and partner questionnaire response rates by police record status

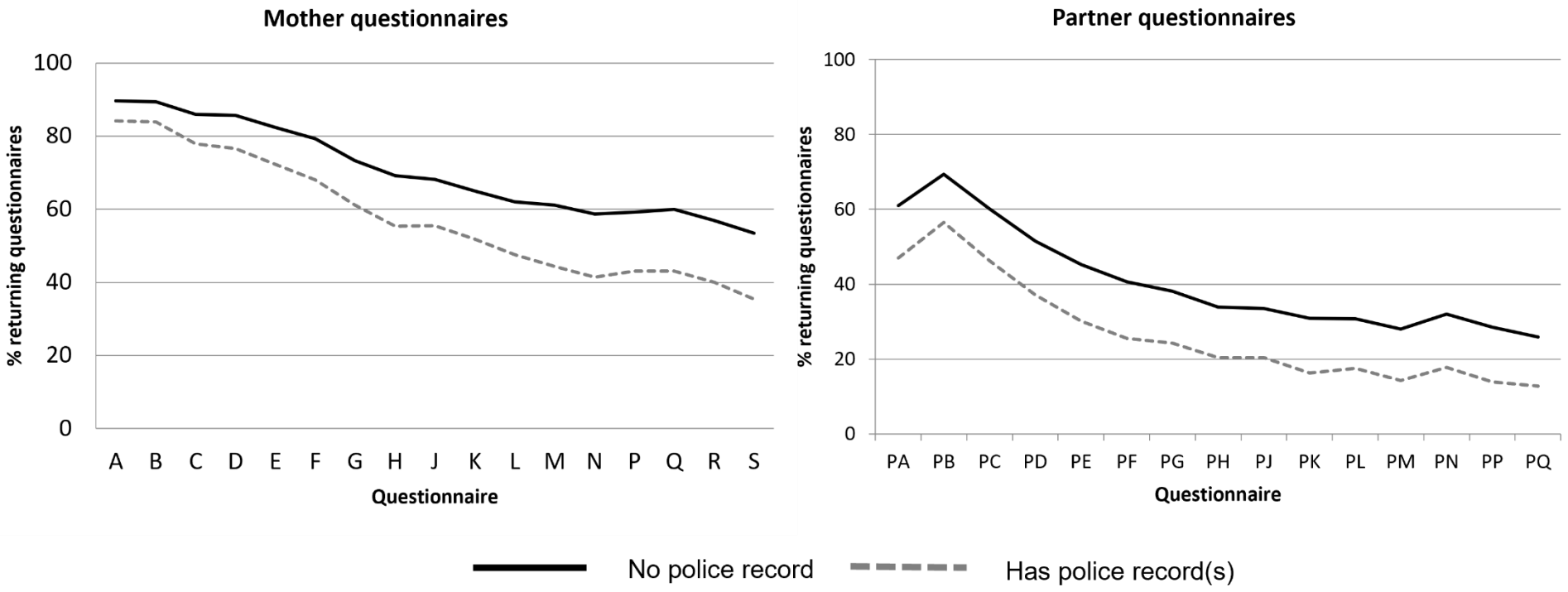
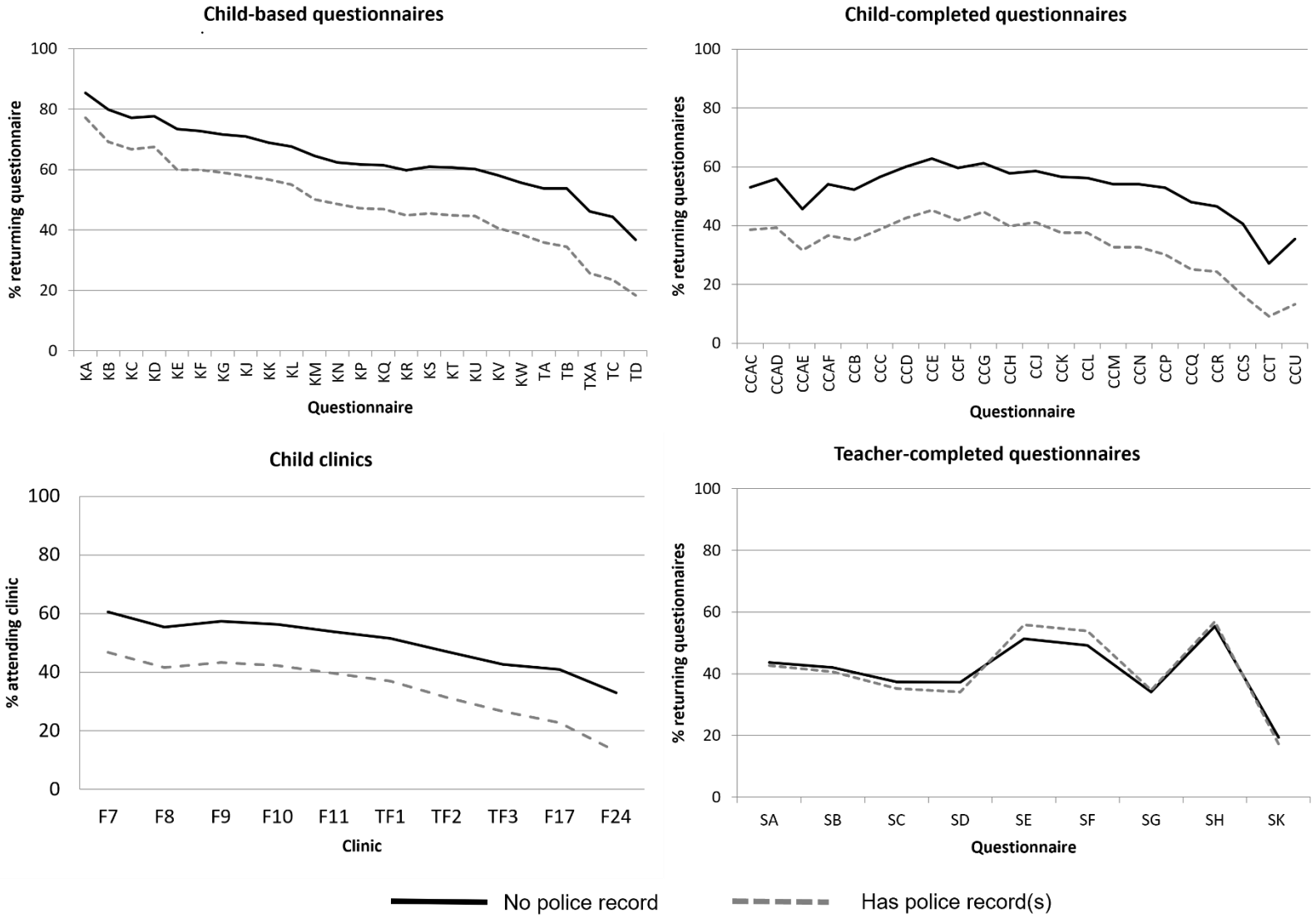


Figure 6b: Child questionnaire and clinic, and teacher questionnaire, response rates by police record status



5.2 Other linkage data

ALSPAC has already achieved linkage to primary and secondary health care data, and to education data, for many participants. Of those participants for whom we have permission to link to their crime data (N=12,662), for the vast majority we also have permission to link to their health and education records (Table 10). However, those with a crime record are much less likely to have actively consented, and much more likely to be non-responders. This illustrates the importance of including non-responders in any analyses using linkage data where possible.

Table 10: Consent status for other linkages by criminal record status

		Does not have crime record (N=10,905)	Has crime record (N=1757)
Health	Dissented	82 (0.8%)	n<5
	Consented	4701 (43.1%)	>300 (>17%)[1]
	Non-response	6122 (56.1%)	1439 (81.9%)
Education	Dissented	26 (0.2%)	n<5
	Consented	4758 (43.6%)	>300 (>17%)[1]
	Non-response	6121 (56.1%)	1439 (81.9%)

[1]Exact numbers not shown to prevent calculation of small cell counts for dissented group

5.3 Using area of residence to define denominator

As the A&S police data only cover crimes committed in A&S, it is important to be able to identify who was living in this area so that an appropriate denominator can be defined. Flags have been derived that denote whether an individual was living in A&S on each of their birthdays. [This is based on the contact address ALSPAC held for that child's family at each time point and may not be completely accurate]. At age 10 (the youngest age someone can have a police record), almost 90% of the ALSPAC sample for whom we have consent to link to crime data had an address in A&S (Table 11). This proportion declined only slightly through adolescence but then dropped to 76% by age 24 and 66% by age 28. Overall, over 60% of the sample had a contact address in A&S for every birthday from age 10 through to 28 years. Only 7% had an address in another area on every birthday during that time. Note that over time the number of children for whom their address is unknown increases, with a particularly large jump between ages 20 and 24, and again to age 28 (i.e. likely coinciding with moving out of their family home, and the responsibility for updating

their contact details with ALSPAC shifting from the mothers to their children). Very few of the sample (just over 1%) had no known contact address between ages 10 and 28.

Table 11: Percentage of ALSPAC sample living in Avon and Somerset by age

Time point[1]	In area	Out of area	Unknown address/not in ALSPAC at that age
10th birthday	11,125 (87.9%)	1288 (10.2%)	249 (2.0%)
16th birthday	10,938 (86.4%)	1411 (11.1%)	313 (2.5%)
20th birthday	10,688 (84.4%)	1491 (11.8%)	483 (3.8%)
24th birthday	9605 (75.9%)	1824 (14.4%)	1233 (9.7%)
28th birthday	8308 (65.6%)	2121 (16.8%)	2233 (17.6%)
All birthdays from 10 to 28[2]	7831 (61.8%)	890 (7.0%)	162 (1.3%)

[1]Flags are available for every birthday, but for brevity only five time-points are presented here.
[2]The percentages in the 'all birthdays' row do not total 100% as individuals who moved between the in area/out of area/unknown address categories over time are excluded.

6. Strengths, limitations and bias

The linkage of ALSPAC to A&S police records has created a valuable resource for researchers interested in the causes and consequences of interaction with the criminal justice system. Strengths include the sufficient levels of criminality in the ALSPAC sample for research on both males and females, and records that cover from early adolescence through to the current day (when they were aged almost 30). However, there are a number of limitations to these data that any researcher considering using them should consider.

Importantly, this linkage to local police records does not provide any data on crimes committed outwith Avon and Somerset. While we know from our pilot linkage to the PNC – which is national - that the majority of offences committed by ALSPAC participants occurred in this area (9), there were some committed in other areas of England and in Wales (and possibly in other countries too, but we could not ascertain that from the PNC). A lack of police record in our linked dataset therefore does not mean a participant has no criminality records. The likelihood of us missing records probably increases over time as the cohort ages and becomes less geographically clustered. It is therefore important to consider restricting the analysis sample to those believed to be living in A&S during the time period of interest.

A further limitation is that longitudinal cohorts with general population samples such as ALSPAC are not best placed to study rare outcomes as the number of participants who experience them is small. With regards the police data, this means that ALSPAC cannot be used to study risk factors for rare crimes, which will include the most serious crimes such as rape or murder. Even for more common, lower-level crimes, numbers will often be too low in ALSPAC for these crimes to be considered individually at the level recorded in the *currentoffencedescription* variable, and it is more likely that aggregate outcomes that combine several offence types will need to be used (e.g. 'violent crime').

Finally, it is important for researchers using police records data to be aware that there are several sources of bias. These include bias in terms of whose criminal behaviour is detected by the police, and the disposal type they are given. For example, the disproportionate use of Stop and Search on Black, Asian and Minority Ethnic communities (13). Bias may also be introduced through the data linkage process due to participants with a criminal record being less likely to be active in ALSPAC, resulting in their identifier information (e.g. current name and address) held by the study being out of date. This may introduce bias if propensity to drop out of study follow-up is associated with criminal activity. It is also known that linkage error can be patterned by socio-demographic circumstances (e.g. non-traditional UK names may be entered into official records incorrectly) and may be associated with criminality (e.g. using 'fake' identifiers).

Funding

Avon and Somerset Police provided funding for the linkage of their data to ALSPAC and the writing of this summary report. The analysis of the data and writing of the manuscript has been conducted independently by University of Bristol staff with full academic freedoms. The UK Medical Research Council and the Wellcome Trust [217065] and the University of Bristol provide core support for ALSPAC. AT and RC have been awarded funding by the ESRC through their Secondary Data Analysis Initiative for the exemplar project described in this report, which will examine the relationship between childhood adversity and violent crime (Grant No ES/T014393/1). JM is partly funded by the National Institute for Health Research Applied Research Collaboration West (NIHR ARC West)

Author roles & affiliations

The authors of this report are Alison Teyhan, Rosie Cornish, Andy Boyd, Richard Thomas, Mark Mumme, Amy Dillon, and John Macleod. AB, AT & JM negotiated access to the A&S data and with A&S staff designed the data sharing framework and project plan. Authors MM & RT implemented the linkage and MM oversaw the processing of the data with AT, RC and AD. AT drafted this report and all other authors contributed. JM takes responsibility for this research.

AT is a Research Fellow in the ALSPAC Data Linkage team. RC is a Senior Research Fellow in the ALSPAC Data Linkage team and in the Integrative Epidemiology Unit at the University of Bristol. MM is Linkage Data Manager of the ALSPAC Data Linkage team. JM is a Professor of Clinical Epidemiology, the academic lead of the ALSPAC Data Linkage team, and the Director of The National Institute for Health Research Applied Research Collaboration West (NIHR ARC West) at University Hospitals Bristol and Weston NHS Foundation Trust. RT and AB worked for the ALSPAC Data Linkage team at the time of this linkage, but now both work for the UK Longitudinal Linkage Collaboration (UK LLC): RT is a Senior Data Manager and AB is the Director. AD was a Research Associate in the ALSPAC Data Linkage Team at the time of the linkage and is now a Research Project Manager in Bristol Medical School.

Acknowledgements

Achieving this linkage between A&S police data and ALSPAC has been a team effort. The authors would like to thank Jacqui Jenson, former Director of Children's Services at Bristol City Council, Steven Cullen, former Assistant Chief Constable at A&S Police, and Jonathan Dowey, Head of Performance and Insight at A&S Police, for their support and enthusiasm for this work, which was instrumental in the data linkage between ALSPAC and A&S Police data being achieved.

We thank Dr. Adrian Brown and Dr. Anna Ferrante of the Curtin School of Public Health at Curtin University in Perth, Australia who developed the LinXmart linkage system and guided the ALSPAC data managers in its use.

We thank the following A&S police staff: Kate Britton, Data Protection Officer and Michelle Radcliffe, Information Sharing Co-ordinator for helping us to complete the DPIA; and Alan Salmon, Technical Specialist in Business Intelligence, and Kevin Britton, Data Officer, for preparing the police data set for us and for answering our many queries.

We are extremely grateful to all the families who take part in ALSPAC, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses.

References

1. Christmas H, Srivastava J. Public health approaches in policing: a discussion paper. Public Health England, College of Policing. 2019.
2. National Police Chiefs' Council. Policing, Health and Social Care consensus: working together to protect and prevent harm to vulnerable people. 2018.
3. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, et al. Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. Int J Epidemiol. 2013;42(1):111-27.
4. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, et al. Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. Int J Epidemiol. 2013;42(1):97-110.
5. ALSPAC. Linkage to routine health and social records [Available from: http://www.bristol.ac.uk/alspac/researchers/our-data/linkage/.
6. ALSPAC. Using your official records in Children of the 90s research [Available from: www.bristol.ac.uk/alspac/participants/using-your-records/.
7. Ministry of Justice. Data First: An Introductory User Guide. Harnessing the potential of linked administrative data for the justice system.; 2021.
8. Health and Safety Executive. Recording prosecutions on the Police National Computer [Available from: https://www.hse.gov.uk/foi/internalops/ocs/100-199/168_15.htm.
9. Boyd A, Teyhan A, Cornish R, Croft J, Thomas R, Brennan I, et al. The potential for linking cohort participants to official criminal records: a pilot study using the Avon Longitudinal Study of Parents and Children (ALSPAC) [version 2; peer review: 2 approved with reservations]. Wellcome Open Research. 2022;5(271).
10. Niche. NicheRMS365: The World's Leading RMS in the Cloud 2021 [Available from: https://nicherms.com/nicherms-365/.
11. Crown Prosecution Service. The Code for Crown Prosecutors. 2018.
12. Boyd JH, Randall S, Brown AP, Maller M, Botes D, Gillies M, et al. Population Data Centre Profiles: Centre for Data Linkage. International Journal of Population Data Science. 2020;4(2).
13. Avon and Somerset Criminal Justice Board. Identifying disproportionality in the Avon and Somerset Criminal Justice System. 2022.
14. Home Office. Police Recorded Crime and Outcomes: Open Data Tables User Guide. 2016.

Appendix A: Brief overview of ALSPAC questionnaire and clinic measures of anti-social and criminal behaviours

ALSPAC has several measures of the study children's criminal, anti-social and troublesome behaviours. Most are self-reported by the study children themselves, but there are also mother and teacher reported measures. Below is a brief overview of the measures available, with the questionnaire or clinic these measures come from specified in brackets. Full details on all these measures can be found in the ALSPAC Data Dictionary, which can be downloaded from the study's website: www.bristol.ac.uk/alspac/researchers/our-data/.

The ALSPAC children have been asked about various criminal and anti-social behaviours at several time points from childhood through to young adulthood: during clinics when aged approximately 8.5 (F@8), 10.5 (F@10), and 12.5 (TF1) and via postal questionnaire when aged approximately 14 (CCQ), 16 (TC), 18.5 (CCT) and 21 (CCU). There are also later measures. The questions vary somewhat by age but examples of behaviours they were asked about include stealing, deliberately setting fire to a car or property, carrying or using a weapon or knife, using illegal drugs, physical violence, graffiti, and animal cruelty. The study children have also reported on their contact with the police and criminal justice system at some of these time points. It is worth noting that the children were asked about their behaviours in the past 12 months, but the time between the measurement points is two years or more, meaning there are gaps in coverage.

The study mothers were first asked about their child's 'troublesome' behaviours when they were aged 7.6 years via postal questionnaire (KR). These included starting fights, bullying or threatening, running away and playing truant. When their child was aged 16.5 years (TC) they were asked in a postal questionnaire about a wider range of behaviours including stealing, selling drugs, physical violence, breaking and entering, and setting fire to things on purpose.

Teachers reported via questionnaire on troublesome behaviours when the study children were in school years 3 (SABC) and 6 (SEFG), aged around 7 and 10 respectively. These measures included using weapons, bullying, physically cruel, stealing, and being in trouble with the law.

The table below gives an example of the repeated measures available in ALSPAC for one particular criminal behaviour (knife/weapon carrying and use). It can be seen that while this behaviour has been asked about on multiple occasions, both the question wording/content and the respondent change over time. This is true for many of the other criminal and antisocial behaviour measures.

Overview of weapon/knife measures available for the ALSPAC children

Age of child (years)	**7**	**8.5**	**10**	**12.5**	**13.9**	**15.5**	**16**	**17**	**18.7**	**21**
Questionnaire/Clinic	**SABC[1]**	**F@8[2]**	**SEFG[1]**	**TF1[2]**	**CCQ[4]**	**TF3[3]**	**TC[5]**	**TF4[3]**	**CCT[4]**	**CCU[4]**
Carried a weapon/knife					X	X	X	X	X	X
- Frequency of carrying					X	X	X	X	X	X
Type of weapon/knife								X		
Use of a weapon/knife	X	X	X	X				X		
- Frequency of use	X		X					X		
Injury caused								X		

X = question asked at that time point

[1]Questionnaire completed by child's teacher

[2]Question completed by child as part of an interview with a survey assistant at an ALPAC clinic

[3]Question completed by child as part of a computer session interview at an ALSPAC clinic

[4] Postal questionnaire completed by child

[5]Postal questionnaire completed by mother

Appendix B: Description of disposal types that, through fair processing, we have permission to link to ALSPAC

Disposal type	Description[1]
Prosecutions	
Charge	Individuals charged with a crime will usually go to court for a trial. The police will decide if the individual can go home with awaiting their first court hearing ('on bail') or if they will be remanded into police custody.
Reported for summons	A summons is a written order to attend court to answer an allegation. All prosecutions start by an individual being either summonsed to attend court or arrested then charged with an offence.
Postal requisition	A postal requisition is a legal document notifying an individual that a decision has been made to prosecute their offence(s) at court. It is a type of summons. It effectively means the individual is being charged with a criminal offence by post.
Taken into consideration (TIC)	TICs are crimes taken into consideration at the time of sentencing for another crime. The individual may volunteer these offences, or they may be asked by the police if they accept them. Either way, the individual must formally admit their guilt to the additional crime(s) while under caution. An individual may admit to TICs in order to 'wipe the slate clean' and demonstrate remorse; the sentence is likely to be shorter than if the offences were heard at a separate trial.
Out of court disposals	
Caution/ Conditional caution	A caution is a 'warning' given for minor crimes. Those aged 10 and over can receive a caution in England. The individual has to admit an offence and agree to be cautioned. They can be arrested and charged if they don't agree. Individuals who receive a conditional caution have to agree to rules and restrictions e.g. fixing damage caused, or getting treatment for drug abuse. They could be charged with a crime if they don't stick to the conditions. A caution is not a criminal conviction, but it could be used as evidence of bad character if they go to court for another crime.
Penalty notice for disorder (PND)	PND are given for offences like shoplifting, possessing cannabis, being drunk and disorderly in public, and minor criminal damage. They are only given to individuals aged 18 and over. Individuals given a PND do not get a criminal conviction if they pay the penalty fine. If they disagree with the PND they can ask for a trial.
Cannabis/khat warning	A cannabis or khat warning may be given where the offender is found in possession of a small amount of cannabis or khat consistent with personal use and the offender admits the offence. The drug is confiscated, and a record of the warning will be made on local systems. If there are aggravating features (e.g. quantity of drug, related anti-social behaviour, persistent offending) the individual may receive a PND or be charged instead.
Community resolution	These are aimed at first time offenders for less serious offences. A community resolution is not a caution or conviction, does not constitute a criminal record, and they are not currently recorded on the Police National Computer (PNC). They are however recorded on police information systems and can be accessed for intelligence purposes, and a previous Community Resolution will be taken into consideration if further offences are committed. It is a victim-focused outcome, allowing the victim to be both part of the decision making process and involved in the resolution of the crime. This could include a simple apology, compensation, or a promise to clear up any graffiti or criminal damage.

[1]The information in this table was compiled from several freely available Government, legal and policing websites.

Appendix C: Determining disposal outcome filtering rules for the ALSPAC extract

In order to determine which records had disposal outcomes that we had permission to link to ALSPAC we needed to assess each outcome vis-à-vis the permitted purpose for research as set out to participants in ALSPAC's *fair processing* materials. We used two outcome variables and a further variable which states how many individuals were involved in each crime (this allowed us to identify group crimes i.e. a single recorded act of crime that involved more than one person).

The main outcome variable, *currentclassificationhooutcom,* is a 22 category variable that gives the Home Office outcome code for each offence (14), as summarised in Table 1.

Table 1: Description of recorded crime outcomes (main outcome variable)

Outcome	Description	Permission to link to ALSPAC?
1	Charged/Summonsed	YES
2	Caution – youths	YES
3	Caution – adults	YES
4	Taken into consideration (TIC)	YES
5	Offender died	NO
6	Penalty Notices for Disorder(PND)	YES
7	Cannabis/Khat warning	YES
8	Community Resolution	YES
9	Not in public interest (CPS)	NO
10	Not in public interest (Police)	NO
11	Prosecution prevented – suspect under age	NO
12	Prosecution prevented: suspect too ill	NO
13	Prosecution prevented: victim/key witness dead/too ill	NO
14	Evidential difficulties: suspect not identified; victim does not support further action	NO
15	Evidential difficulties: suspect identified; victim supports action	NO
16	Evidential difficulties: suspect identified; victim does not support further action	NO
17	Prosecution time limit expired	NO
18	Investigation complete no suspect identified	NO
19	National Fraud Intelligence Bureau filed (NFIB only)	NO
20	Action undertaken by another body/agency	NO
21	Further investigation to support formal action not in the public interest	NO
22	Diversionary, educational, intervention – no further action	NO

This is an offence-level variable: this means that for offences involving more than one offender ('group crimes'), everyone with a record for that offence is assigned the same outcome, the most serious outcome for the group. For example, if there are three people

involved in an offence and two are cautioned and one charged, all three will have 'charged' as their *currentclassificationhooutcom* outcome. For offences that involve only one person, *currentclassificationhooutcom* is in effect an individual-level variable. Therefore, for offences involving one person, it can be used to identify which records can be linked to ALSPAC. However, for offences involving more than one person, this variable cannot be used as we do not know if all individuals involved in that offence had that outcome.

The second outcome variable *offenderclassificationconcat* is an individual-level, concatenated variable which lists several terms (up to 6) for each record (e.g. 'suspect; arrested; charged' or 'possible suspect; involved party; eliminated') (each of the individual terms is listed in Table 3). Note that prior to September 2015 (when police recording software was changed), the term 'prosecuted' was used in the concatenated variable to cover TICs and all out of court disposals (cautions, penalty notices, drug warnings, and community resolutions). We used this *offenderclassificationconcat* variable to determine the outcomes for individuals who had been involved in group crime. For any given record, if *offenderclassificationconcat* included at least one of the terms shaded in Table 2, then that record was linked to ALSPAC.

Table 2: Summary of terms included in concatenated outcome variable

Term in concatenated variable	Permission to link to ALSPAC?
Charged	Yes
Reported for summons	Yes
Postal requisition	Yes
Taken into consideration (TIC)	Yes
Cautioned[1]	Yes
Adult conditional caution[1]	Yes
Penalty notice for disorder[1]	Yes
Cannabis warning[1]	Yes
Community resolution[1]	Yes
Prosecuted	Yes
ABE Interview	No
Arrested	No
Breach of Injunction	No
Driver	No
Drugs education program	No
Education	No
Eliminated	No
Insufficient to proceed	No
Involved party	No
Mentioned	No
No further action	No
Not charged police	No
Other	No

Possible suspect	No
Restorative justice	No
Stop and search	No
Suspect	No
Voluntary attendee	No
Witness	No

[1]Prior to September 2015, all of these terms are included in the term 'prosecuted' in the concatenated variable

Table 3 shows simplified police data for 5 imaginary individuals. Individuals 1, 2, 3 and 4 were involved in one crime occurrence each, and individual 5 was involved in two crime occurrences. Individuals 1 and 5 are the only individuals involved in their respective crimes (*offendercount*=1): *currentclassificationhooutcom* can therefore be used as their individual-level outcome. Individuals 3 and 4 are both linked to the same crime (crime 3). The variable *currentclassificationhooutcom* tells us that at least one of them has been charged with this crime. We have to look at the concatenated variable to determine the individual-level outcome. This shows that individual 3 was charged, and so this record can be linked to ALSPAC. Individual 4 was not prosecuted and so this record will be deleted. Individual 2 is the only person in the dataset linked to crime 2. However, the *offendercount* variable has a value of 4 – this means that three others, who are either not in ALSPAC or who are in ALSPAC but have opted out of crime linkage, were also involved in this crime. We therefore use the concatenated variable to determine the outcome for this individual: they were prosecuted and so we can link this record to ALSPAC. As this crime occurred prior to 2015, the concatenated variable does not specify the precise disposal type, so although *currentclassificationhooutcom* tells us that one of the group was charged (OC1), we do not know if that was individual 2.

Table 3: Example of police data

Individual ID	Crime occurence ID	Number of offenders	Offence date	Offence group	Home office outcome variable[1]	Concetanated outcome variable[2]	Link to ALSPAC?
1	1	1	30/8/2016	Drug	OC3	Cautioned	Yes
2	2	4	22/12/2012	Theft	OC1	Suspect; Prosecuted	Yes
3	3	2	1/4/2017	Arson	OC1	Arrested; Charged	Yes
4	3	2	1/4/2017	Arson	OC1	Suspect; No further action	No
5	4	1	16/9/2018	Burglary	OC3	Cautioned	Yes
5	5	1	1/2/2019	Burglary	OC1	Suspect; Charged	Yes

[1] Variable is *currentclassificationhooutcom*; [2]Variable is *offenderclassificationconcat*