Department for Education

Children in Need of help and protection

Data and analysis

Methodology document

March 2018

Contents

1. Introduction

The Department for Education has committed to review the support Children in Need (CIN) currently receive with the aim of understanding why their outcomes are so poor and, therefore, the additional support they might require, in and out of school. As part of the work to meet this commitment this data analysis publication brings together evidence from a range of sources to provide new insight, additional to that available through DfE's routine statistical releases on CIN outcomes in the social care and education system.

DfE: children's social care statistics

DfE official statistics provide information on many aspects of the children's social care system. These statistics help DfE and others to understand and monitor changes in the system over time, provide the basis for informed public debate and highlight areas where policy makers and local authorities may be able to make improvements.

There are three statistical first releases (SFRs) related to children's social care:

1. Characteristics of children in need in England ([Collection](#))

2. Looked after children in England ([Collection](#))

3. Children's social care workforce ([Collection](#))

Full details of the data collections and methodology behind the statistics can be found by following the links above. The statistical first releases relating to children also include accompanying statistics on educational outcomes. These are created using data from the National Pupil Database ([NPD](#)).

New evidence

This ad-hoc publication has been produced to support the call for evidence as part of the CIN review. If you require children's social care statistics for any other purpose, these should be obtained from the SFRs listed above.

The new evidence and insights presented in this data anlysis publication are based on data from either:

- The children in need census and look after children censuses
- Our bespoke dataset based on NPD data created for this publication

Where we present new information, the underlying figures can be found in the accompanying tables.

Overview of our approach

We have used the NPD to analyse the characteristics, social care experiences and educational outcomes of children in need. A more detailed description of the data sources and methodology can be found in sections 2 (social care) and 3 (education).

Our analysis is based on data collected within academic year (AY) 2015/16. This is consistent with the latest SFR information available on children's outcomes. The outcomes SFRs for AY 2016/17 will be published in March 2018.

The majority of our analysis splits children into four mutually exclusive groups. These groups are based on the level of risk and services needed from the social care system. We use data from the looked after children (LAC) and children in need (CIN) censuses to define the population of these groups. We have used a snapshot measure, based on the "looked after or in need" status on 31st March 2016, rather than a longitudinal measure, based on a child's status at any point in the year. This is consistent with the approach taken in the DfE official statistics. Table [a] sets out the definitions of the four distinct groups.

Table [a]: Groupings

Group	Details
CINP	Children who were on a child in need plan on the 31st March 2016, excluding those who were also on a child protection plan and/or looked after. We use information from the CIN census, alongside the 'CLA_31_MARCH' variable from the LAC census, to decide which children are in this group.
CPP	Children who were on a child protection plan on the 31st March 2016, excluding those who were also looked after. We use information from the CPP module of the CIN census, alongside the ' CLA_31_MARCH ' variable from the LAC census, to decide which children are in this group.
LAC	Children who were looked after on the 31st March 2016. We use the ' CLA_31_MARCH ' variable from the LAC census to decide which children are in this group.
All other children	Children who were not looked after, on a child protection plan or on a child in need plan on the 31st March 2016. This group will include children who were looked after, on a protection plan or in need at other points in the year but not at the 31st March.

This approach adds value to the current evidence base as official statistics are not grouped in this mutually exclusive way, and some children are included in both the LAC and CIN SFRs.

2. Social care experience and outcomes

Data sources

We use data from the 2015-16 CIN and LAC census as this was the latest data available at the time. Where published figures are used, we use 2015-16 figures for consistency, however 2016-17 data is now available. Users can navigate to the latest 2016-17 publications and use the same table references to get updated figures.

Analysis based on the children in need census 2015-16 and the looked after children census 2015-16 (Source: CIN and LAC censuses 2015-16) will differ from any comparable figures published in DfE statistical first releases for the reasons below.

Removing duplicates

We have removed a small number of duplicate cases from the 2015-16 CIN census, see page 4 of the following: 2015-16 CIN outcomes methodology

Merging datasets

In order to get mutually exclusive groups, we match the child protection plan (CPP) module of the CIN census to identify children on CPPs at 31st March 2016 using 'LA' and 'LAChildID' fields. Similarly, we match children from the looked after children census onto the CIN census using 'LA' and 'LAChildID' fields to identify children looked after on 31st March 2016. There were roughly 120 CPP cases and 2,000 LAC cases that couldn't be identified in the CIN census. Consequently, these cases will be present in the CINP group (see Table a).

This means that the demographics of the LAC group in particular will differ slightly from any comparable DfE SFR figures.

3. Educational outcomes and characteristics

We have used NPD data to analyse the characteristics and educational outcomes of children in need and compare these to all other pupils. This section of the methodology document sets out:

- The data sources used in this analysis;
- How we dealt with multiple records per pupil and how we brought the input data together; and
- The variables used and filters applied to produce the educational outcomes and characteristics data.

Analysis based on the National Pupil Database (NPD) will differ slightly from any comparable figures published in DfE statistical first releases for various reasons, including:

- Differences in underlying datasets – for example, using 'final' datasets where available (see Table b)
- Differences in match rates (see Figure e)
- Differences in social care classifications – for example, we use all LAC compared to LAC greater than 12 months in the DfE SFR.

Data sources

We used the schools and alternative provision (AP) censuses covering the 2015/16 academic year to derive the base pupil population and then merged information on from other, mainly NPD, data sources. Table [b] sets out all the data sources and filters applied, where relevant.

Table [b]: Data sources used to analyse educational outcomes and characteristics

Data source	Details/reason for including and filters applied
Schools census from January 2016	This provides the base pupil population and characteristics (e.g. special education needs (SEN) and whether they claim free school meals (FSM)). Schools provide this information. We include pupils where: • The link unique row number variable is null • The LA is not 702 (overseas schools)

Data source	Details/reason for including and filters applied
AP census from January 2016	This provides the base pupil population for all pupils who are educated in alternative provision. LAs are responsible for providing this information. We include pupils where: • The LA is not 702 (overseas schools)
Early years foundation stage profile (EYFSP) attainment at the end of AY 2015/16	We use this to analyse the attainment of pupils in Reception in AY 2015/16. We include results where: • The attainment results are final • The pupil has a complete profile, no 'A' flags in any of the EYFSP stages.
Key stage 2 (KS2) attainment at the end of AY 2015/16	We use this to analyse the attainment of pupils in Year 6 in AY 2015/16. We include results where: • The attainment results are final • The pupil is included in national results statistics • The pupil is at the end of key stage 2
Key stage 4 (KS4) attainment at the end of AY 2015/16	We use this to analyse the attainment of pupils in Year 11 in AY 2015/16. We include results where: • The attainment results are final • The pupil is at the end of key stage 4
Looked after children (LAC) census financial year (FY) 2015-16	We use this to flag all children who were looked after at 31st March 2016.
Children in need (CIN) census FY 2015-16	We use this to flag all children who had a child protection plan and/or were in need on 31st March 2016. Please see the 2015-16 CIN methodology document for details of the records that are included.

Data source	Details/reason for including and filters applied
Absences data for AY 2015/16 – 3 terms	We use this to analyse the number of children who were persistently absent from school in AY 2015/16.
Exclusions dataset for AY 2014/15	We use this to analyse school exclusions in AY 2014/15 for pupils in our 2015/16 dataset. We use lagged exclusions data because this is consistent with the CIN and CLA outcome SFRs that cover data from AY 2015/16.
Pupil premium allocations data for FY 2016-17.	Pupil premium allocations are based on lagged data – the FY 2016-17 pupil premium allocations are based on data from AY 2015/16. We use this data to flag the pupils who met the eligibility criteria based on AY 2015/16 data.
Ofsted data extract from March 2016	We use this to group pupils by the Ofsted rating of the schools they attend.

Removing duplicates and bringing input data sets together

There were multiple records for some pupils in almost all the data inputs. We only include each pupil once in our analysis, so we designed a process to remove multiple records.

We ensured that each input data source only had one record per pupil and did this by flagging the "main" record for each pupil, which was already available in the school census but not in the other data sources. Our process for flagging the main record in these datasets is set out in table [c] below.

We went through the process set out in Table [c] for each input data source where there were multiple records for some pupils. Please see Annex A for information on the lead variables used to choose the "main record" and the number of multiple records removed.

Figure [d] sets out the process of merging the resulting datasets together. We joined the schools and AP censuses to create our base pupil population and then merged on data from the other input data sources, using either the pupil or the school identifier to match the data together.

Table [c]: Process for choosing the main record for each pupil

Step	Details
1. Frequency count	Count the number of times each pupil appears in the data set – this is based on the unique pupil number reference.
2. Choose a "lead variable"	Choose a variable in the dataset that you will use to decide which record is the "main record" for each pupil. E.g. the claiming FSM variable for the AP dataset.
3. Record the "max lead variable"	For each pupil, create a new variable that records the maximum value for the *lead variable.* E.g. If there are two records for a pupil in the AP dataset, one record where the claiming FSM variable is 1 (pupil claims FSM) and one where the claiming FSM variable is 0 (pupil does not claim FSM), the *max lead variable* would be 1 for both records.
4. Keep main record	Where a pupil appears in a data set more than once, keep the record where the *lead variable* is equal to the max *lead variable*. E.g. In the above AP dataset example we would keep the record where the claiming FSM variable is 1 as this is equal to the max lead variable.
5. Choose first record if still multiple records	In some cases there will be multiple records per pupil where the *lead variable* is equal to the *max lead variable*. In these cases we chose the first record that appears in the dataset. E.g. in the AP dataset we group all the records for each pupil together and assign a row number. We then keep the first record for each pupil.

Figure [d]: Creating merged dataset

Step 1: Base pupil population

We joined the schools and AP censuses together. We created 'NULL' entries for AP records where variables exist on the school census but not the AP.

Result – data set with all pupil records from the school and AP census (after filters set out in Table [a] and removing duplicates in AP census). 8,037,250 records.

Step 2: Removing duplicates

There were 8,030,970 unique pupil references and 8,037,250 records in the data set after step 1. This means that 6,280 pupils are in both the schools and AP censuses. We stripped out the repeat records using the steps below:

- If pupil reference only appears once then we kept the record.

- If pupil reference appears more than once than we used the census source to decide which record to keep. We decided to keep the record from the schools census over the AP census.

After following this process we had our base pupil population. Each pupil appears only once in the base population, there are 8,030,970 records.

Termly school census
- Exclusions data collection
- Absences data collection
- Deprivation pupil premium allocations
- Pupils in schools and their characteristics

Annual AP census
- Pupils in AP and their characteristics

Annual CIN census
- Children on a CPP
- Children in need

Annual CLA census
- Looked after children
- Post LAC/LAC pupil premium allocations

Annual EYFSP return
- EYFSP attainment

Annual KS2 national curriculum assessments
- KS2 attainment

Annual KS4 assessments
- KS4 attainment

Ofsted inspections
- Ofsted rating of school

Base pupil population
(Keeping one record per pupil)

Each pupil in a school or AP provider at January 2016 with information on:

CINP, CPP, LAC status

Deprivation and SEN

Absences and exclusions

EYFSP/KS2/KS4 attainment

Pupil premium eligibility

Ofsted rating of school

Key
- **Data collections**
- **Input data sets**
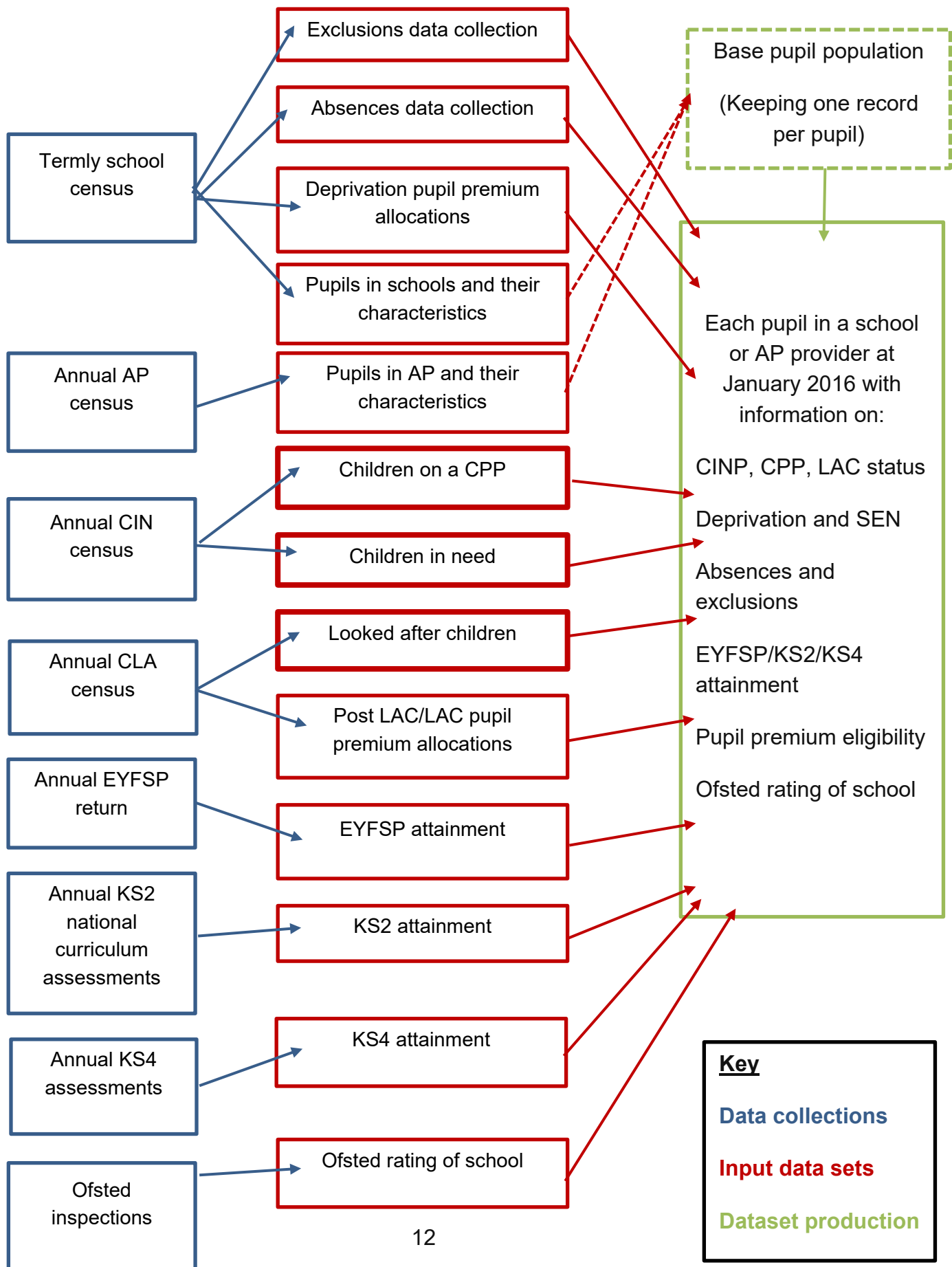- **Dataset production**

12

Figure [e]: Matching rates for CIN and LAC

	[A] Number of "school age" records in the source census	[B] Number of "school age" records in the source census with a pupil matching reference	[C] Number of "school age" children in the source census with a pupil matching reference	[D] Number of "school age" children in the NPD merged dataset
CIN at 31st	241,585 school age records (based on age at 31st March 2016) – some children will appear more than once in the data.	215,226 school age records (based on age at 31st March 2016) with a pupil matching reference – some children will appear more than once in the data. 89.1% of [A]	214,400 school age children (based on age at 31st March 2016) with a pupil matching reference – only keeping one record per child. 99.6% of [B] 88.7% of [A]	201,701 school age children (based on age at 31st March 2016) in NPD merged dataset. 94.1% of [C] 93.7% of [B] 83.5% of [A]
LAC at 31st	44,650 school age records (based on age at August 2015) – some children will appear more than once in the data.	44,466 school age records (based on age at August 2015) with a pupil matching reference - some children will appear more than once in the data. 99.6% of [A]	44,463 school age children (based on age at August 2015) with a pupil matching reference – only keeping one record per child. 100.0% of [B] 99.6% of [A]	43,347 school age children (based on age at August 2015) in NPD merged dataset. 97.5% of [C] 97.5% of [B] 97.1% of [A]

Producing charts

There are 8,030,970 pupils in our base pupil population. To analyse the characteristics and educational outcomes of pupils we generally took a subset of this base pupil population. The specific subset changed based on the educational outcomes or characteristics that we were analysing. For transparency, we have provided details of the variables and filters we have applied to produce each output, these are set out in Annex B.

Annex A: Removing multiple records, table of lead variables for each input data set[1] and number of records removed

Data source	Step 1	Step 2/3	Step 4	Step 5
AP	22,040 records in data set. 21,940 unique pupil references.	Use "claiming FSM" as lead variable. This can be 1 or 0. Max lead variable set as max of claiming FSM.	22,010 records after only keeping records where lead variable = max lead variable.	21,940 records after keeping first record for each pupil.
EYFSP	670,870 records in data set. 670,770 unique pupil references.	Use "total EYFSP score" as lead variable. This is a number between 0 and 51. Max lead variable set as max of total EYFSP score.	670,780 records after only keeping records where lead variable = max lead variable.	670,760 records after keeping first record for each pupil.
KKS2	592,270 records in data set. 592,210 unique pupil references.	Use "expected level in reading, writing and maths" as lead variable. This can be 1 or 0. Max lead variable set as max of expected level in reading, writing and maths.	592,250 records after only keeping records where lead variable = max lead variable.	592,210 records after keeping first record for each pupil.

---

[1] After filters applied as set out in table [a].

Data source	Step 1	Step 2/3	Step 4	Step 5
KS4	630,720 records in data set. 624,500 unique pupil references.	Use "attainment 8" as lead variable. This is a number between 0 and 80. Max lead variable set as max of attainment 8.	626,290 records after only keeping records where lead variable = max lead variable.	624,500 records after keeping first record for each pupil, ordered by whether included in national results.
CLA	75,780 records in data set. 75,680 unique pupil references.	Use "looked after at 31st March" as lead variable. This can be 1 or 0. Max lead variable set as max of looked after at 31st March.	75,710 records after only keeping records where lead variable = max lead variable.	75,680 records after keeping first record for each pupil.
CIN	518,000 records in data set. 473,510 unique pupil references.	Use "CIN at 31st March" as lead variable. This can be 1 or 0. Max lead variable set as max of CIN at 31st March.	488,390 records after this step. Where children were in need at 31st March, kept record where lead variable = max lead variable. For children who were not in need at 31st March (but in need at some other point in year) kept all records.	473,510 records after keeping first record for each pupil.

We removed multiple records from the absences input data in a slightly different way. We kept all the records from AY 2015/16 and created a variable to flag whether the pupil had been persistently absent (defined as missing at least 10% of the sessions they could have attended). Some pupils were in the input data set more than once, these are pupils who attended multiple schools within the year. We kept the maximum "persistant absence" variable for each pupil – so if a pupil had been persistently absent from one school in AY 2015/16 then they are included in the persistently absent group in our analysis, even if they attended other schools in the same academic year and weren't persistently absent in these settings.

Annex B: Variables used, filters applied and pupil population for charts[2]

Chart	Variables used	Filters applied	Base population for chart/table				
			Total	LAC pupils	CPP pupils	CINP pupils	All other pupils
Total pupil count in base data set	NA	No filters applied	8,030,970	46,220	27,330	145,710	7,811,710
Table 3: Pupils split by ethnicity	LAC/CPP/CINP pupil groups, ethnicity	Only including pupils with valid ethnicity data.	7,923,100 (98.7%)	46,010 (99.5%)	27,090 (99.1%)	144,830 (99.4%)	7,705,170 (98.6%)
Table 4: Pupils split by deprivation level of residential postcode – based on IDACI score	LAC/CPP/CINP pupil groups, IDACI 2015	Only collected on the schools census, so no AP records included. Only including pupils with valid IDACI score data.	7,957,560 (99.1%)	42,420 (91.8%)	26,790 (98.0%)	141,690 (97.2%)	7,746,660 (99.2%)

---

[2] Figures rounded to the nearest 10 pupils. Percentages calculated using unrounded figures.

Chart	Variables used	Filters applied	Base population for chart/table				
			Total	LAC pupils	CPP pupils	CINP pupils	All other pupils
Table 9: Pupils claiming FSM	LAC/CPP/CINP pupil groups, FSM eligible	Only including pupils with valid data on claiming FSM.	7,991,820 (99.5%)	45,230 (97.9%)	27,030 (98.9%)	144,240 (99.0%)	7,775,320 (99.5%)
Table 10: Pupils eligible for different pupil premium funding grants	LAC/CPP/CINP pupil groups, pupil premium allocations for 2016/17 (based on 2015-16 data)	Only including pupils who are aged between 4 and 15 at the start of the academic year and included in the schools census. This fits with the pupil premium eligibility criteria.	7,201,910 (89.7%)	41,470 (89.7%)	26,250 (96.0%)	134,540 (92.3%)	6,999,660 (89.6%)
Table 11: Pupils with a SEN recorded in census	LAC/CPP/CINP pupil groups, SEN provision	Only including pupils with valid data on SEN.	7,991,820 (99.5%)	45,230 (97.9%)	27,030 (98.9%)	144,240 (99.0%)	7,775,320 (99.5%)
Table 12: Pupils with SEN split across SEN type	LAC/CPP/CINP pupil groups, SEN provision	Pupils who have a SEN recorded on the schools or AP censuses.	1,164,450	24,980	10,630	69,410	1,059,430

Chart	Variables used	Filters applied	Base population for chart/table				
			Total	LAC pupils	CPP pupils	CINP pupils	All other pupils
	LAC/CPP/CINP pupil groups, SEN provision, Primary SEN type	Pupils with detailed SEN information – only captured on the schools census.	1,150,920 (98.8%)	22,710 (90.9%)	10,530 (99.0%)	67,430 (97.2%)	1,050,260 (99.1%)
Table 13:	LAC/CPP/CINP pupil groups, FSM eligible, pupil premium allocations for 2016/17 (based on 2015-16 data), SEN provision	Figures for FSM, SEN and PP as in Tables 9, 10 and 11. When looking at the overlap between groups, only include pupils with valid FSM and SEN data who are aged between 4 and 15 at the start of the academic year and included in the schools census.	7,167,330 (89.2%)	40,670 (88.0%)	25,960 (95.0%)	133,280 (91.5%)	6,967,420 (89.2%)
Table 14: Pupils achieving a good level of development at	LAC/CPP/CINP pupil groups, GLD	Pupils with valid data for the EYFSP good level of development measure.	657,730	1,710	1,900	7,500	646,620

Chart	Variables used	Filters applied	Base population for chart/table				
			Total	LAC pupils	CPP pupils	CINP pupils	All other pupils
EYFSP, split by SEN and FSM	LAC/CPP/CINP pupil groups, GLD, FSM eligible, SEN provision	Pupils who also have valid FSM and SEN data.	656,860 (99.9%)	1,700 (99.1%)	1,900 (99.8%)	7,480 (99.7%)	645,790 (99.9%)
Table 15: Pupils achieving the expected level of development at KS2, split by SEN and FSM	LAC/CPP/CINP pupil groups, VALRWM, RWMEXP	Pupils who are included in the valid pupil number for reading, writing and maths.	584,870	3,380	2,230	11,070	568,190
	LAC/CPP/CINP pupil groups, VALRWM, RWMEXP, FSM eligible, SEN provision	Pupils who also have valid FSM and SEN data.	584,580 (99.9%)	3,370 (99.8%)	2,230 (99.7%)	11,050 (99.8%)	567,930 (100.0%)
Tables 16, 17: Pupils achieving an A*-C in English and Maths and the	LAC/CPP/CINP pupil groups, NATRES, L2BASICSLLPTQEE, ATT8	Pupils who are included in the national KS4 results.	553,620	5,900	1,500	12,070	534,150

Chart	Variables used	Filters applied	Base population for chart/table				
			Total	LAC pupils	CPP pupils	CINP pupils	All other pupils
average attainment 8 score at KS4, split by SEN and FSM	LAC/CPP/CINP pupil groups, NATRES, L2BASICSLLPTQEE, ATT8, FSM eligible, SEN provision	Pupils who also have valid FSM and SEN data.	551,080 (99.5%)	5,710 (96.8%)	1,460 (97.3%)	11,880 (98.4%)	532,040 (99.6%)
Table 18: Average progress 8 score with confidence intervals at KS4, split by SEN and FSM	LAC/CPP/CINP pupil groups, INP8CALC, P8SCORE	Pupils included in the progress 8 calculations.	519,900	4,800	1,390	11,070	502,640
	LAC/CPP/CINP pupil groups, INP8CALC, P8SCORE, FSM eligible, SEN provision	Pupils who also have valid FSM and SEN data.	519,440 (99.9%)	4,750 (99.0%)	1,380 (99.3%)	11,020 (99.6%)	502,280 (99.9%)
Table 20: Pupils split by the type of school they attend and schools split by type of pupils that attend	LAC/CPP/CINP pupil groups, PHASE	Pupils with valid data on the phase of education they attend.	8,030,970 (100.0%)	46,220 (100.0%)	27,330 (100.0%)	145,710 (100.0%)	7,811,710 (100.0%)

Chart	Variables used	Filters applied	Base population for chart/table				
			Total	LAC pupils	CPP pupils	CINP pupils	All other pupils
Table 21: Pupils split by the Ofsted rating of the school they attend – state-funded primary schools	LAC/CPP/CINP pupil groups, PHASE	Pupils attending primary, middle-deemed primary, early years or nursery schools.	4,677,670 (58.2%)	18,010 (39.0%)	16,850 (61.7%)	68,540 (47.0%)	4,574,270 (58.6%)
Table 21 Pupils split by the Ofsted rating of the school they attend – state-funded secondary schools	LAC/CPP/CINP pupil groups, PHASE, Overall Effectiveness	Pupils attending secondary, middle-deemed secondary or all-through schools.	3,212,200 (40.0%)	19,010 (41.1%)	8,600 (31.5%)	47,110 (32.3%)	3,137,480 (40.2%)
Table 22: Distribution of CIN pupils across schools	LAC/CPP/CINP pupil groups, CENSUS SOURCE	LAC, CPP or CINP pupils with a record on the schools census – no school information available on the AP census.	NA	43,710 (94.6%)	27,210 (99.5%)	143,600 (98.6%)	NA

Chart	Variables used	Filters applied	Base population for chart/table				
			Total	LAC pupils	CPP pupils	CINP pupils	All other pupils
Tables 23, 24: Pupils who are persistently absent, have a fixed term exclusion and have a permanent exclusion	LAC/CPP/CINP pupil groups, CENSUS SOURCE, PHASE, age at start of academic year, persistent absence, total fixed exclusions, permanent exclusion count	Pupils aged 4 to 15 at the start of the academic year, who have their main registration in a primary, secondary or special school. Does not include pupils on the AP census or in PRUs.	7,150,950 (89.0%)	39,510 (85.5%)	25,380 (92.8%)	130,890 (89.8%)	6,955,180 (89.0%)
Table 25: Pupils who joined their current school outside of August or Sepetmeber and proportion of pupils in Y11 who joined school in Y9 or later	LAC/CPP/CINP pupil groups, EntryDate, NCYearActual	Pupils aged 4 to 15 at the start of the academic year, who have valid entry date information.	7,201,910 (89.7%)	41,470 (89.7%)	26,250 (96.0%)	134,540 (92.3%)	6,999,660 (89.6%)

Follow us on Twitter:
@educationgovuk

Like us on Facebook:
facebook.com/educationgovuk