Office for National Statistics

Article

# Admin-based ethnicity statistics for England, feasibility research: 2016

The Office for National Statistics (ONS) does not currently produce annual statistics by local authority on the population by ethnic group and the last official statistics were from the 2011 Census. This feasibility research combines ethnicity data from English School Census, Hospital Episode Statistics and Improving Access to Psychological Therapies to explore whether administrative data can be used to produce statistics on the population by ethnic group for 2016 at national and local authority level for England, by five-year age group and sex. These research outputs are not official statistics.

## Notice

### 7 February 2023

This is not the latest research update. For the latest update on our admin-based ethnicity statistics feasibility research, please see our [Developing admin-based ethnicity statistics for England and Wales: 2020 article](#).

# Table of contents

# 1 . Main points

- The admin-based ethnicity statistics show early promise, but further work is needed before we can produce robust estimates.

- Through combining English School Census, Hospital Episode Statistics and Improving Access to Psychological Therapies data, we are able to establish an ethnicity for 70.2% of individuals in the admin-based population estimates (ABPE) V3.0 for 2016.

- The proportion of admin-based individuals with a stated ethnicity is highest for children and lowest for males of working age; this is to be expected given the data sources used in this research.

- The proportions of the admin-based population in the Asian, Black, Mixed and White ethnic groups are broadly similar to those from the 2011 Census; we will compare with the 2021 Census in due course.

- The proportion in the Other ethnic group is higher in the admin-based ethnicity statistics than the 2011 Census; this is likely to be driven by reporting and recording differences between the data sources.

- There is under-representation in the administrative data of the Asian, Black and Mixed ethnic groups in the 20 to 24, 25 to 29 and 30 to 34 years age groups; incorporating Higher Education Statistics Agency data into the method in future should help to improve representativeness amongst these age groups.

## Disclaimer

These Research Outputs are not official statistics on the population by ethnic group, nor are they used in the underlying methods or assumptions in the production of official statistics. Rather, they are published as outputs from research into a methodology different to that currently used in the production of ethnicity statistics. These outputs should not be used for policy- or decision-making.

# 2 . About our transformation research

Ethnicity is a high priority topic for users, particularly for the analysis of inequalities, and having robust statistics on the population by ethnic group has become of increased importance during the coronavirus (COVID-19) pandemic. However, the Office for National Statistics (ONS) does not produce annual statistics by local authority on the population by ethnic group and the last official statistics available were from the 2011 Census.

Previous research focused on using Annual Population Survey (APS) data to produce population estimates by ethnic group. The first method applied ethnicity distributions from the APS to the mid-year population estimates and adjusted for communal establishments using 2011 Census data. The second method used the Generalised Structure Preserving Estimator (GSPREE) to combine APS, census and English School Census data, drawing strength from each data source. However, for both methods, small sample sizes in the APS for smaller ethnic groups affected the robustness of the resulting estimates at low geographic levels.

This research combines English School Census (ESC), Hospital Episode Statistics (HES) and Improving Access to Psychological Therapies (IAPT) data. Based on unique identifiers, these sources were linked to the admin-based population estimates (ABPE) V3.0 for 2016, which were used as the population base for the analysis. Admin-based ethnicity statistics for England at the national, regional and local authority level for 5 and 18 ethnic groups have been produced. More information on the data sources and method can be found in this accompanying article.

The research has been conducted for England only while we continue to work with the Welsh Government to acquire additional data for Wales. Scotland and Northern Ireland have devolved responsibility for producing ethnicity statistics so are not covered by this research. However, we will proactively engage with colleagues in the devolved administrations also researching this topic.

This research forms part of our population and social statistics transformation programme, which aims to provide the best insights on population, migration and society using a range of data sources. The findings will form part of the evidence base for the 2023 National Statistician's Recommendation on the future of population and social statistics.

# 3 . Population coverage

The admin-based population estimates (ABPE) V3.0 for 2016 contain around 54.2 million records. Previous work has noted that there is undercoverage in the population base and details the reasons behind this. As no single source can provide ethnicity data for the whole population between census years, we combined English School Census (ESC), Hospital Episode Statistics (HES) and Improving Access to Psychological Therapies (IAPT) data. Through doing this, we were able to establish an ethnic group for 70.2% of individuals in the ABPE. The remaining 29.8% of individuals in the ABPE consists of:

- those in the combined administrative data but with an unknown ethnic group — 4.9% of the ABPE

- those in the combined administrative data but with ethnic group refused — 9.3% of the ABPE

- those that we were unable to link to ESC, HES or IAPT — 15.6% of the ABPE

Full information on how these indicators were determined can be found in this [accompanying article](#).

Looking by age, only 66.0% of children under 1 year of age have a stated ethnicity in the combined administrative data. This is because although births are captured in HES, the ABPE has a reference date of 30 June 2016 to align with the mid-year population estimates but the HES data used cover the period 1 April 2009 to 31 March 2016. This means that those in the ABPE born between 1 April and 30 June 2016 have not been linked to HES. We are exploring the potential to use monthly HES and Birth Notifications data in future, both of which should improve coverage for those under 1 year.

The proportion with a stated ethnicity is highest for those aged 1 to 16 years, at 90.9%. This is because in addition to records from hospital visits, ethnicity data from birth records in HES have been rolled forward and ESC provides ethnicity data for most children at state-funded schools.

From the age of 17 years, there is a substantial decrease in the proportion of individuals with a stated ethnicity. This is because they are not captured in ESC after leaving school, and as younger people are less likely to need hospital services, they may also not be captured in HES. From the age of 20 years and peaking at age 35 years, a greater proportion of females than males have a stated ethnicity. As this age range corresponds to childbearing years, this divergence is likely in part related to females attending hospital for pregnancy appointments and childbirth.

The proportion with an unknown ethnicity is lowest up to the age of 16 years because of low levels of unknowns in ESC. It is also low for older ages, with only 2.0% of those aged 85 years and over having an unknown ethnic group. Males have a higher level of unknowns than females from age 18 to 60 years.
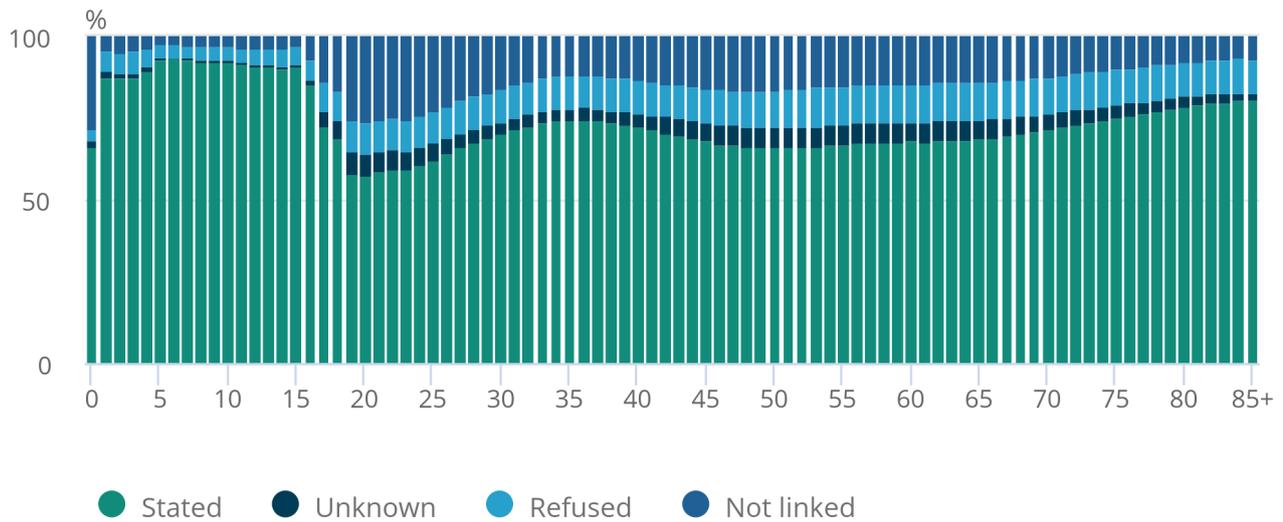
Refusal rates are lowest for children because of lower levels of refusals in ESC than in HES and IAPT. Refusal rates are similar for males and females across the age distribution.

**Figure 1: The proportion of females with a stated ethnicity is highest for children, has a second peak in the mid-30s and then increases with age from age 50 years**

Proportion of females in the ABPE with ethnicity stated, unknown, and refused, and proportion not linked to ESC, HES or IAPT, by age, England

## Figure 1: The proportion of females with a stated ethnicity is highest for children, has a second peak in the mid-30s and then increases with age from age 50 years

Proportion of females in the ABPE with ethnicity stated, unknown, and refused, and proportion not linked to ESC, HES or IAPT, by age, England



Source: Office for National Statistics

Notes:

1. Stated refers to those with a stated ethnicity and no refusal on their most recent administrative data record.

2. Unknown groups together those with unknown recorded on all ethnicity records and those with multiple stated ethnicities recorded on the latest date.

3. Refused refers to those with a refusal on their most recent administrative data record.

4. Not linked refers to individuals in the ABPE who have not been linked to ESC, HES or IAPT.

5. All individuals on the ABPE have their sex recorded as either male or female.

**Figure 2: The proportion of males with a stated ethnicity is highest for children and lowest for males of working age**

**Proportion of males in the ABPE with ethnicity stated, unknown, and refused, and proportion not linked to ESC, HES or IAPT, by age, England**

Figure 2: The proportion of males with a stated ethnicity is highest for children and lowest for males of working age

Proportion of males in the ABPE with ethnicity stated, unknown, and refused, and proportion not linked to ESC, HES or IAPT, by age, England



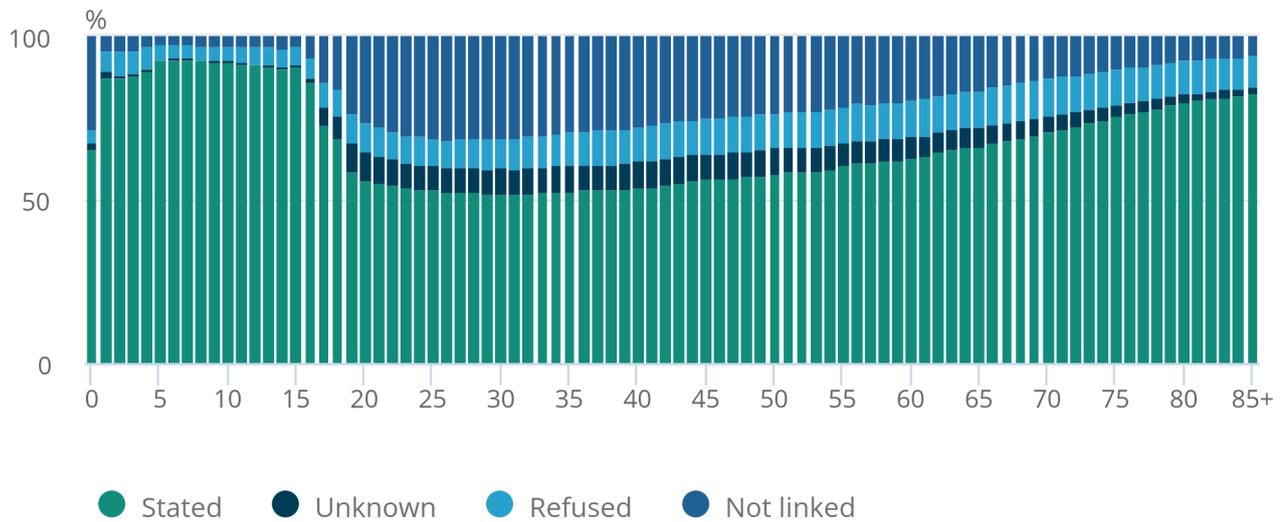**Source: Office for National Statistics**

**Notes:**

1. Stated refers to those with a stated ethnicity and no refusal on their most recent administrative data record.

2. Unknown groups together those with unknown recorded on all ethnicity records and those with multiple stated ethnicities recorded on the latest date.

3. Refused refers to those with a refusal on their most recent administrative data record.

4. Not linked refers to individuals in the ABPE who have not been linked to ESC, HES or IAPT.

5. All individuals on the ABPE have their sex recorded as either male or female.

# Region

The proportion of individuals with a stated ethnicity is lowest in London (63.4%) and highest in the North West (75.1%). The low figure for London is largely because 21.7% of individuals on the ABPE in London could not be linked to ESC, HES or IAPT. This may be because of London having a younger population than the other regions, and younger people being less likely to attend hospital.

The proportion with an unknown ethnic group is similar across the regions and the proportion of refusals ranges from 7.2% in Yorkshire and The Humber to 13.3% in the South East.

Table 1: Proportion of individuals on the ABPE in each region of England with ethnicity stated, unknown, and refused and proportion not linked to ESC, HES or IAPT

|  | Stated | Unknown | Refused | Not linked |
|---|---|---|---|---|
|  | Percentage | | | |
| **North East** | 74.2 | 4.0 | 11.1 | 10.8 |
| **North West** | 75.1 | 4.5 | 7.7 | 12.7 |
| **Yorkshire and The Humber** | 73.9 | 5.2 | 7.2 | 13.7 |
| **East Midlands** | 71.8 | 4.3 | 8.6 | 15.4 |
| **West Midlands** | 73.1 | 5.1 | 7.8 | 14.0 |
| **East of England** | 70.1 | 6.2 | 7.7 | 16.0 |
| **London** | 63.4 | 4.5 | 10.4 | 21.7 |
| **South East** | 65.5 | 4.9 | 13.3 | 16.3 |
| **South West** | 72.3 | 5.0 | 8.7 | 13.9 |

Source: Office for National Statistics

Notes

1. Proportions may not sum to 100% because of rounding.

2. Stated refers to those with a stated ethnicity and no refusal on their most recent administrative data record.

3. Unknown groups together those with unknown recorded on all ethnicity records and those with multiple stated ethnicities recorded on the latest date.

4. Refused refers to those with a refusal on their most recent administrative data record.

5. Not linked refers to individuals in the ABPE who have not been linked to ESC, HES or IAPT.

## Local authority

The proportion of individuals in the ABPE with a stated ethnicity ranges from 37.2% in the City of London to 87.0% in Knowsley.

The proportion of individuals with an unknown ethnicity ranges from 1.7% in Halton to 10.6% in Cambridge. Of the 10 local authorities with the highest rates of unknowns, eight are in the East of England.

The proportion of individuals in the ABPE with ethnicity refused ranges from 1.9% in North East Lincolnshire to 33.8% in Gosport. There is a high proportion of refusals in several local authorities along the south coast, including Havant (33.1%), Fareham (32.2%), Hastings (29.4%), Rother (29.2%) and Portsmouth (29.0%). Another cluster can be seen in the West Midlands, with Tamworth and Lichfield having refusal rates of 29.3% and 27.5% respectively.

The proportion of individuals in the ABPE not linked to ESC, HES or IAPT ranges from 6.8% in Knowsley to 50.8% in the City of London. The highest proportions are in London local authorities and the university cities of Oxford and Cambridge. These differences are largely because of differing levels of hospital usage with a number of factors likely to be underlying this. As part of our future development, we plan to incorporate Higher Education Statistics Agency data into the method. This should help to reduce the proportion of individuals in the ABPE in the unlinked group within university towns and cities.

## Figure 3: There are large variations across the country in the proportion with a stated ethnicity in the admin-based ethnicity statistics

**Proportion of individuals in the 2016 ABPE V3.0 with ethnicity stated, unknown and refused, and proportion not linked to ESC, HES or IAPT, by local authority, England**

**Notes:**

1. Stated refers to those with a stated ethnicity and no refusal on their most recent administrative data record.

2. Unknown groups together those with unknown recorded on all ethnicity records and those with multiple stated ethnicities recorded on the latest date.

3. Refused refers to those with a refusal on their most recent administrative data record.

4. Not linked refers to individuals in the ABPE who have not been linked to ESC, HES or IAPT.

Download the data

# 4 . Ethnicity comparisons

The Office for National Statistics (ONS) does not currently produce annual estimates of the population by ethnic group at local authority level. We have used the latest official estimates from the 2011 Census as a comparator and will use 2021 Census data when available to continue to evaluate the admin-based ethnicity statistics. As we have not been able to assign an ethnicity to everyone in the admin-based population estimates (ABPE), comparisons have been made based on the proportion of people in each ethnic group, rather than the absolute number. This allows us to explore how representative those that we have assigned an ethnicity to are of the underlying population.

When comparing the admin-based ethnicity statistics with the 2011 Census it is important to bear in mind that differences could be related to any of the following:

- population change between 2011 and 2016

- differences in reporting, recording and mode of collection

- lack of representativeness of the administrative data used to assign an ethnicity

- differences in response options, particularly the lack of the Gypsy or Irish Traveller response options in Hospital Episode Statistics (HES) and Improving Access to Psychological Therapies (IAPT), and the Arab response option in all three administrative sources

- bias in the admin-based population estimates

At the national level, we have also compared the admin-based ethnicity statistics with estimates from the year ending June 2016 Annual Population Survey (APS) data. This removes the issue of population change but introduces sampling error. Additionally, the APS only covers people living in private households.

Of those with a stated ethnicity in the administrative data, the proportions in each of the Asian, Black, Mixed and White ethnic groups are broadly similar to the 2011 Census (Table 2). The proportion in the Other ethnic group, however, is double the proportion from the 2011 Census (2.1% compared with 1.0%).

For 18 ethnic groups, the proportions in most ethnic groups are also broadly comparable with the 2011 Census and 2016 APS. However, there are notable differences for some ethnic groups.

The proportions in four of the five other ethnic groups (Black Other, Mixed Other, White Other and Any Other ethnic group) are higher in the admin-based ethnicity statistics than in the 2011 Census statistics. This could reflect changes in the population over time, but research by the Nuffield Trust on ethnicity coding in English health services datasets suggests that some individuals are being incorrectly coded into the other ethnic groups.

The proportion of people in the Irish ethnic group in the admin-based ethnicity statistics is half the size of the proportion in this ethnic group in the 2011 Census (0.5% compared with 1.0%). This may be caused by mis-recording, with research by the Nuffield Trust identifying incorrect coding of people from the White Irish ethnic group as White British. Record-level comparisons of HES data with the 2011 Census showed similar findings. However, the Good Friday Agreement states that those from Northern Ireland can "identify themselves and be accepted as Irish or British, or both, as they may so choose", so this may also be a factor.

The cause of incorrect coding of some ethnic groups is unknown. NHS guidance is that ethnicity should be self-reported. However, the Nuffield Trust report outlines some of the potential challenges of this. These include patients lacking the capacity to provide this information, staff being unaware of the required procedure for capturing ethnicity information and inconsistent response options across organisations and care settings.

Table 2: Proportion of the population in each ethnic group by data source, England

| Ethnic group | 2016 admin-based ethnicity statistics | Annual Population Survey, year ending June 2016 | 2011 Census |
|---|---|---|---|
| | Percentage | | |
| **Asian** | 7.6 | 8.0 | 7.8 |
| Bangladeshi | 0.9 | 0.9 | 0.8 |
| Chinese | 0.4 | 0.6 | 0.7 |
| Indian | 2.4 | 2.9 | 2.6 |
| Pakistani | 2.5 | 2.3 | 2.1 |
| Asian Other | 1.5 | 1.3 | 1.5 |
| **Black** | 3.7 | 3.6 | 3.5 |
| African | 1.9 | 2.3 | 1.8 |
| Caribbean | 1.1 | 1.2 | 1.1 |
| Black Other | 0.7 | 0.2 | 0.5 |
| **Mixed** | 2.2 | 1.8 | 2.3 |
| White and Asian | 0.5 | 0.5 | 0.6 |
| White and Black African | 0.3 | 0.2 | 0.3 |
| White and Black Caribbean | 0.6 | 0.6 | 0.8 |
| Mixed Other | 0.9 | 0.5 | 0.5 |
| **White** | 84.3 | 84.8 | 85.4 |
| English/Welsh/Scottish/Northern Irish/British | 78.2 | 78.4 | 79.8 |
| Gypsy or Irish Traveller | 0.1 | 0.0 | 0.1 |
| Gypsy/Roma | 0.1 | x | x |
| Irish Traveller | 0.0 | x | x |
| Irish | 0.5 | 0.6 | 1.0 |
| White Other | 5.5 | 5.8 | 4.6 |
| **Other** | 2.1 | 1.8 | 1.0 |
| Arab | x | 0.4 | 0.4 |
| Any other ethnic group | 2.1 | 1.4 | 0.6 |

Source: Office for National Statistics

Notes

1. Ethnicities have been presented alphabetically at the 5-category level (except for Other) and then alphabetically within each 5-category ethnic group.

2. APS estimates have been weighted to the household population.

3. Proportions have been calculated out of those with a stated ethnicity.

4. Proportions may not sum to totals because of rounding.

5. The HES and IAPT data sources do not include the Gypsy or Irish Traveller response option. ESC has separate categories for Traveller of Irish heritage and Gypsy/Roma. The 2016 APS and 2011 Census only include a combined Gypsy or Irish Traveller response option.

6. The Arab ethnic group is not a response option for any of the three administrative data sources used to produce the admin-based ethnicity statistics.

The Chinese ethnic group appears to be under-represented in the administrative data, with 0.4% of the admin-based population recorded as Chinese compared with 0.7% and 0.6% in the 2011 Census and 2016 APS respectively. Data from the 2011 Census showed that a third of the Chinese population aged 16 years and over in England were students. Looking at the admin-based ethnicity statistics by age (Figure 4), within the 20 to 24, 25 to 29 and 30 to 34 years age groups, the Asian ethnic group are under-represented in the administrative data. This holds across all Asian sub-groups and is largest for the Chinese ethnic group.

The Black and Mixed ethnic groups also appear to be under-represented in the 20 to 24, 25 to 29 and 30 to 34 years age groups in the admin-based ethnicity statistics. Looking at data by 18 ethnic groups, there is over-representation of the Black Other and Mixed Other ethnic groups but under-representation of the other sub-groups, particularly the Mixed White and Asian ethnic group. For the White ethnic group within these ages, there is over-representation of the White British ethnic group but under-representation of the other sub-groups.

One of the planned next steps in the research is to incorporate the Higher Education Statistics Agency data into the method. This should improve the representativeness of the admin-based ethnicity statistics for 20- to 34-year-olds. However, people aged 20 to 34 years not at university are still likely to be a challenging group to obtain ethnicity data for.

## Figure 4: The Other ethnic group is over-represented in the admin-based ethnicity statistics across all age groups

**Proportion of people in each ethnic group by age and data source, England**

**Source: Office for National Statistics**

**Notes:**

1. Proportions have been calculated out of those with a stated ethnicity.

Download the data

# Region

In each region, the proportion of the population in each ethnic group in the admin-based ethnicity statistics is broadly similar to the 2011 Census, except for the Other ethnic group. The national trend of a higher proportion of the population in the Other ethnic group in the admin-based ethnicity statistics compared with the 2011 Census is seen across all regions.

Table 3: Proportion of the population in each ethnic group in the 2016 admin-based ethnicity statistics and 2011 Census, by English regions

| | Asian | | Black | | Mixed | | White | | Other | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 2016 admin-based ethnicity statistics | 2011 Census | 2016 admin-based ethnicity statistics | 2011 Census | 2016 admin-based ethnicity statistics | 2011 Census | 2016 admin-based ethnicity statistics | 2011 Census | 2016 admin-based ethnicity statistics | 2011 Census |
| | Percentage | | | | | | | | | |
| **North East** | 2.4 | 2.9 | 0.5 | 0.5 | 0.7 | 0.9 | 95.5 | 95.3 | 0.9 | 0.4 |
| **North West** | 6.2 | 6.2 | 1.6 | 1.4 | 1.5 | 1.6 | 89.6 | 90.2 | 1.2 | 0.6 |
| **Yorkshire and The Humber** | 7.6 | 7.3 | 1.5 | 1.5 | 1.6 | 1.6 | 88.1 | 88.8 | 1.3 | 0.8 |
| **East Midlands** | 6.2 | 6.5 | 1.9 | 1.8 | 1.9 | 1.9 | 88.7 | 89.3 | 1.3 | 0.6 |
| **West Midlands** | 11.7 | 10.8 | 3.8 | 3.3 | 2.6 | 2.4 | 80.2 | 82.7 | 1.7 | 0.9 |
| **East of England** | 4.6 | 4.8 | 2.2 | 2.0 | 2.1 | 1.9 | 89.9 | 90.8 | 1.2 | 0.5 |
| **London** | 18.1 | 18.5 | 14.8 | 13.3 | 4.5 | 5.0 | 55.4 | 59.8 | 7.3 | 3.4 |
| **South East** | 5.0 | 5.2 | 1.7 | 1.6 | 2.2 | 1.9 | 89.5 | 90.7 | 1.6 | 0.6 |
| **South West** | 1.6 | 2.0 | 0.9 | 0.9 | 1.3 | 1.4 | 95.5 | 95.4 | 0.7 | 0.3 |

Source: Office for National Statistics

Notes

1. Proportions have been calculated out of those with a stated ethnicity.

2. Proportions may not sum to 100% because of rounding.

# Local authority

Figure 5 shows the 2016 admin-based ethnicity statistics against the 2011 Census. Each point represents one of the 326 local authorities in England. The dashed diagonal line shows perfect agreement between the two measures. If the point is below the line, the admin-based figure is lower than the 2011 Census figure. If the point is above the line, the admin-based figure is higher than the 2011 Census figure.

**Figure 5: The admin-based ethnicity statistics are broadly similar to the 2011 Census estimates at local authority level, except for the Other ethnic group**

**Proportion of the population of each local authority in England in each ethnic group in the 2016 admin-based ethnicity statistics and the 2011 Census**

**Notes:**

1. Proportions have been calculated out of those with a stated ethnicity.

[Download the data](#)

For the Asian, Black and Mixed ethnic groups, the proportions from the admin-based ethnicity statistics are broadly similar to those from the 2011 Census. This can be seen in Figure 5 by them "following the diagonal".

The over-representation of the Other ethnic group at the national and regional levels is seen for all local authorities except three (Gravesham, South Tyneside and North East Lincolnshire).

The majority of local authorities have a lower proportion of the population recorded as White in the admin-based ethnicity statistics compared with the 2011 Census. This is perhaps to be expected, if the decrease in the proportion in the White ethnic group seen between 1991 and 2011 continued between 2011 and 2016. The biggest differences are in London local authorities.

The admin-based ethnicity statistics show early promise, but further work is needed before we can produce robust estimates. The two main challenges to be addressed in order to produce representative and high-quality admin-based ethnicity statistics are coverage and mis-recording. Section 7 outlines the planned next steps for the research.

# 5 . Glossary

## Ethnic group

The self-reported ethnic group of the individual, according to their own perceived ethnic group and cultural background.

## Ethnicity refused

In the English School Census (ESC), if a parent/guardian or pupil has declined to provide ethnicity data, this is recorded as "refused". In Hospital Episode Statistics (HES) and Improving Access to Psychological Therapies (IAPT), where a patient chooses not to state their ethnicity, the code "Z - Not Stated" is recorded.

## Ethnicity stated

Ethnicity stated refers to the ethnicity being recorded as a specific ethnic group and not refused or unknown.

## Ethnicity unknown

In ESC, where the ethnicity has not yet been collected, this is recorded as "NOBT" (information not yet obtained). In HES and IAPT, the default code "99 Not Known" is used where the person's ethnicity is unknown.

In this article, the unknown category also includes individuals with multiple recorded ethnicities where the rules did not lead to a final ethnicity being selected. These have been termed "ethnicity unresolved".

## Ethnicity unresolved

Where multiple ethnicities were recorded on the latest date (and for HES, a dataset hierarchy of Admitted Patient Care, Accident and Emergency, Outpatients didn't resolve the conflict), these have been coded as "unresolved" and grouped into the "unknown" category for the analysis in this article.

## Not linked

This refers to individuals who are in the admin-based population estimates (ABPE) V3.0 for 2016 but have not been linked to ESC, HES or IAPT.

# 6 . Data sources and quality

## Admin-based ethnicity statistics

The admin-based ethnicity statistics were produced from three administrative datasets:

- English School Census (ESC), 2011 to 2016: this is a statutory data collection about pupils in state-funded schools in England

- Hospital Episode Statistics (HES), 2009 to 2016: this is a database containing details of all attendances at NHS hospitals in England; it is made up of three sub-datasets: Admitted Patient Care (APC), Accident and Emergency (AE) and Outpatients (OP)

- Improving Access to Psychological Therapies (IAPT), 2012 to 2016: this is a dataset containing individuals accessing NHS psychological therapies in England

Most individuals have the same ethnicity on all records in the data. However, some individuals have multiple recorded ethnicities. We implemented a method to select a final ethnicity per person, firstly for each individual dataset and then looking across the three datasets.

The general approach was to take the ethnicity from the most recent record. If an individual refused to provide an ethnicity on the most recent date, their ethnicity was coded as "refused". If the ethnicity on the most recent date was unknown, the last stated ethnicity or refusal was selected where available, otherwise their ethnicity was coded as "unknown". Where multiple ethnicities were recorded on the latest date (and for HES, a dataset hierarchy of APC, AE, OP did not resolve the conflict), these were coded as "unresolved" and grouped into the "unknown" category for the analysis in this article. Of those in the ABPE, 0.03% of individuals had a final ethnicity of "unresolved".

After selecting one ethnicity per person within each administrative dataset, the data were linked to the admin-based population estimates (ABPE) V3.0 for 2016 based on a unique identifier. Records that did not link to the ABPE were dropped. A final ethnicity was selected using a similar process as previous, with a hierarchy of IAPT, ESC, HES used when individuals had different ethnicities recorded in different data sources in the same year.

Records where the final ethnicity was unknown or refused have been excluded when calculating the proportion of people in each ethnic group.

## Population base

The 2016 ABPE V3.0 was used as the population base for the admin-based ethnicity statistics. The quality of the population base will impact on the quality of the admin-based ethnicity statistics. More information about the coverage of the population base can be found in a previous report.

## Annual Population Survey (APS)

The APS is a continuous household survey, comprising the Labour Force Survey (LFS) supplemented by sample boosts in England, Wales and Scotland to ensure small areas are sufficiently sampled. The APS does not include most people living in communal establishments (such as care homes or prisons) or anyone else living outside private households. Information on some students living in halls of residence is collected where the students' parents live in a sampled household.

Further information on the methods and data sources can be found in the accompanying article.

# 7 . Future developments

This research shows promise for the ability to produce ethnicity statistics down to local authority level from administrative data. This would be an improvement on using survey data, where estimates can be unreliable at lower geographic levels because of small sample sizes. We will continue to explore how we can further improve upon the admin-based ethnicity statistics through:

- trialling alternative methods for handling multiple recorded ethnicities and refusals

- incorporating additional data sources to improve the population coverage

- combining the administrative data with survey data using the Generalised Structure Preserving Estimator (GSPREE), building on previous work using this method

- continuing research into producing survey-based ethnicity statistics, to provide a more robust comparator and an improved survey source to feed into GSPREE

- producing admin-based ethnicity statistics for Wales

- producing admin-based ethnicity statistics for other years

- exploring the potential to produce multivariate statistics on ethnicity by other characteristics

- engaging with existing efforts within the health sector to improve data collection practices

- collaborating with external experts and peer organisations conducting research in this area

## Feedback

We welcome feedback on the method used to produce the admin-based ethnicity statistics and the planned future developments. Please email your feedback to Admin.Based.Characteristics@ons.gov.uk.

# 8 . Related links

Producing admin-based ethnicity statistics for England: methods, data and quality
Article | Released 6 August 2021
An overview of methods, data sources and data quality for the feasibility research on producing statistics on the population by ethnic group from Hospital Episode Statistics, English School Census and Improving Access to Psychological Therapies data.