

Article

# Developing admin-based ethnicity statistics for England and Wales: 2020

Research update on producing population statistics by ethnic group for England and Wales from administrative data, with comparisons with Census 2021 estimates.

Contact:  
Alison Morgan  
admin.based.  
[characteristics@ons.gov.uk](mailto:characteristics@ons.gov.uk)  
+44 1329 447187

Release date:  
7 February 2023

Next release:  
To be announced

## Table of contents

1. [Main points](#)
2. [About our transformation research](#)
3. [Population coverage](#)
4. [Ethnic breakdown](#)
5. [Developing admin-based ethnicity statistics for England and Wales data](#)
6. [Glossary](#)
7. [Data sources and quality](#)
8. [Future developments](#)
9. [Related links](#)
10. [Cite this article](#)

# 1 . Main points

- By combining ethnicity data from eight administrative data sources and the 2011 Census, we established an ethnicity for 84.9% of people in the 2020 admin-based population base for England and 88.5% of people in the 2020 admin-based population base for Wales.
- Compared with the Census 2021 estimates at national level, the 2020 admin-based ethnicity statistics for both England and Wales have a smaller proportion of people in the Asian, Black, Mixed and Other ethnic groups, and a larger proportion in the White ethnic group.
- By local authority, the 2020 admin-based ethnicity statistics broadly align with Census 2021 in showing how ethnic diversity varies across the country, but there are differences in the reported sizes of each ethnic group.
- Admin-based ethnicity statistics for Lower layer Super Output Areas (LSOAs) have been produced for the first time; these show promise but have high levels of suppression because of disclosure control.
- We have produced a set of ethnicity estimates for local authorities using the Generalised Structure Preserving Estimator method; this method does not significantly improve the statistics by local authority on the proportion of people in each ethnic group.

These 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, and they should not be used for policy- or decision-making. They are published as outputs from research into a methodology different to that currently used in the production of ethnicity statistics.

## 2 . About our transformation research

The Office for National Statistics (ONS) does not currently produce annual statistics by local authority on the population by ethnic group. The only official statistics available are from the census every 10 years, with [Census 2021](#) being the most recent.

In May 2022, we published the [latest findings from our feasibility research](#) on producing statistics on the population by ethnic group for England from administrative data. This was accompanied by an [article exploring a time series of admin-based ethnicity statistics](#) covering 2016 to 2020. Following on from these two publications, this article presents an update on our feasibility research for England, along with our first admin-based ethnicity statistics research outputs for Wales.

The previous research for England considered two approaches, which were:

- using ethnicity information from administrative data only
- using ethnicity information from administrative data and the 2011 Census

The figures presented in this article are based on the second approach, using ethnicity information from administrative data and the 2011 Census.

The method used for this article is unchanged from [that published in May 2022](#), but new for this article, we have added in [Individualised Learner Record](#) and [Welsh School Census](#) data. We have also produced admin-based ethnicity statistics for Lower layer Super Output Areas (LSOAs) for the first time.

In addition to producing admin-based ethnicity statistics using the [linked record-level data approach](#), we have produced a set of ethnicity estimates using the Generalised Structure Preserving Estimator method, which utilises both administrative and survey data. This builds on [previous work](#) using this method.

This research forms part of the [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 in England and Wales.

### 3 . Population coverage

The [2020 Statistical Population Dataset version 3.0 \(SPD V3.0\)](#), previously called the admin-based population estimates, was used as the population base for the admin-based ethnicity statistics. Ethnicity records from administrative data and the 2011 Census were linked to SPD V3.0 to create the admin-based ethnicity dataset (ABED).

We established an ethnicity for 84.9% of people in England in the ABED and 88.5% of people in Wales. The remaining people in the ABED consist of:

- those in the ethnicity data sources but with an unknown ethnic group – 0.8% of the ABED in England and 1.3% in Wales
- those in the ethnicity data sources but with ethnic group refused – 10.2% of the ABED in England and 3.0% in Wales
- those we were unable to link to the ethnicity data sources – 4.1% of the ABED in England and 7.2% in Wales

In both England and Wales, the proportion of people with a stated ethnicity is higher for females than males.

By age, the proportion with a stated ethnicity is highest for those aged 5 to 18 years. This is because most children are captured in the English School Census or Welsh School Census, with both having low levels of refusals and unknowns. This demonstrates the importance of these sources long term in the production of admin-based ethnicity statistics.

In Wales, the proportion of people with a stated ethnicity is lowest for those aged 0 to 2 years, at 70.2%. Over a quarter of children in this age group have an ethnicity record of refused, with most records coming from Birth Notifications (where refusals are recorded as "Not Stated"). In England, the proportion of children in this age group with a stated ethnicity is higher than for Wales, at 90.2%. The ethnicity records for this age group in England are mostly from Birth Notifications, Hospital Episode Statistics, or the Emergency Care Dataset. The levels of refusals in these data sources for England are lower than in the Birth Notifications data for Wales. This may be linked to differences in attitudes to providing ethnicity data but could also be caused by differences in how the "Not Stated" category is used in different hospitals. We will explore this further with NHS England and Digital Health and Care Wales.

In both England and Wales, a lower proportion of adults of working age have a stated ethnicity than children and older people. In Wales, this is because a higher proportion of people could not be linked to the ethnicity data sources. In Wales, we are heavily reliant on the 2011 Census to provide ethnicity data for adults at present, as we have not included hospital data for Wales (see [Section 7: Data sources and quality](#)). Anyone who migrated to Wales after the 2011 Census would need to have attended state-funded school or university, or used NHS hospitals or psychological services in England, for us to be able to link on an ethnicity for them. As part of our future work programme, we plan to continue to incorporate additional administrative data sources, to reduce reliance on census data and enable us to better capture the ethnicities of new entrants to the population.

In England, the lower proportion of people with a stated ethnicity among adults of working age is caused by both a higher proportion of people not linking to the ethnicity data sources than in other age groups and a refusal rate of 12.0% across adult ages. Those not linking to the ethnicity data sources are again likely to mainly be people who have migrated to England since the 2011 Census. The high level of refusals is a reflection of the refusal rate in the Hospital Episode Statistics data, which are the main source of ethnicity data for adults in England.

**Figure 1: The proportion of people with a stated ethnicity in the 2020 admin-based ethnicity dataset is lowest for children aged 0 to 2 years in Wales, and low for adults of working age**

Proportion of people in the 2020 admin-based ethnicity dataset with a stated ethnicity by age and sex, England and Wales

Figure 1: The proportion of people with a stated ethnicity in the 2020 admin-based ethnicity dataset is lowest for children aged 0 to 2 years in Wales, and low for adults of working age

Proportion of people in the 2020 admin-based ethnicity dataset with a stated ethnicity by age and sex, England and Wales



Source: Office for National Statistics

Notes:

1. "Stated" refers to people with a stated ethnicity and no refusal on their most recent administrative data record.
2. All individuals in the Statistical Population Dataset (SPD) have their sex recorded as either male or female.
3. Age is as at 30 June 2020.

The proportion of people in the ABED with a stated ethnicity varies across the local authorities in England, ranging from 45.4% in City of London to 94.2% in East Devon. In England, 80.6% of local authorities have a stated ethnicity for at least 80% of people in the ABED.

In Wales, the proportion of people with a stated ethnicity does not vary much across the local authorities, ranging from 86.2% in Newport to 90.8% in Caerphilly. There is greater variability, however, within the local authorities.

Figure 2 and the [accompanying dataset](#) provide coverage rates by Lower layer Super Output Area (LSOA).

**Figure 2: There are large variations across England and Wales in the proportion of people with a stated ethnicity in the 2020 admin-based ethnicity dataset**

Proportion of people in the 2020 admin-based ethnicity dataset with a stated ethnicity, by Lower layer Super Output Area (LSOA), England and Wales

Notes:

1. “Stated” refers to people with a stated ethnicity and no refusal on their most recent administrative data record.
2. LSOA boundaries are as of 2021. The admin-based ethnicity dataset is based on the 2011 Output Areas (OAs) so a best fit lookup between the 2011 and 2021 OA codes was used to attach 2021 OA and LSOA codes to the dataset. However, as this is a best-fit approach, the residents of some LSOAs have been assigned to neighbouring LSOAs. This means that data are not available for a small number of LSOAs.

## Download the data

.xlsx

## 4 . Ethnic breakdown

The admin-based ethnicity statistics are based on the proportion of people in each ethnic group, of those with a stated ethnicity in the admin-based ethnicity dataset. To assess the accuracy of the admin-based ethnicity statistics, as there are no official statistics on the population by ethnic group for 2020, we have conducted aggregate-level comparisons with the Census 2021 estimates. Through the comparisons, we aim to understand how representative those with a stated ethnicity are of the underlying population.

When comparing the admin-based ethnicity statistics with Census 2021 estimates, it is important to bear in mind that the data sources cover different time periods. The 2020 admin-based ethnicity statistics are based on data up to 30 June 2020 while Census 2021 is a snapshot of data as at 21 March 2021. This means that:

- age has been calculated at different time periods; we have not aged-on people in the admin-based ethnicity statistics to their age at Census 2021
- between the two time periods there may have been population changes, particularly because of the coronavirus (COVID-19) pandemic, both in terms of who is in the population and where they live
- people may have changed the ethnicity they identify as in the period between their ethnicity being recorded in administrative data and completing their Census 2021 form

It is also important to note that there are differences in the reporting, recording and mode of collection in the administrative data sources compared with Census 2021. There are also differences in the response options, as detailed in the [accompanying dataset](#).

There may be bias in the [Statistical Population Dataset version 3.0 \(SPD V3.0\)](#). For the next iteration of the research, we will be switching to SPD V4.0 as the population base, so we should see a reduction in bias.

There is also uncertainty in the census estimates, as outlined in the [census quality measures](#).

As part of our future work programme, we plan to produce admin-based ethnicity statistics for 2021 and repeat the aggregate-level comparisons. We also plan to conduct record-level comparisons of linked administrative and census data to explore the issues outlined in this article and better understand the causes of differences seen between the admin-based ethnicity statistics and Census 2021 estimates. We will use these insights to make further improvements to the data sources and method.

## Comparisons of ethnicity statistics

Compared with Census 2021 estimates, the 2020 admin-based ethnicity statistics for both England and Wales have a smaller proportion of people in the Asian, Black, Mixed and Other ethnic groups, and a larger proportion in the White ethnic group.

In England, despite a smaller proportion of people being recorded in the Asian, Black, and Mixed ethnic groups in the admin-based ethnicity statistics than in the Census 2021 estimates, the proportions in the Asian Other, Black Other, and Mixed Other ethnic groups are larger. This is the case across most age groups and most local authorities. This suggests that there may be overuse of these categories in the administrative data collections, as has been previously [reported by the Nuffield Trust](#) in their analysis of English health datasets.

Table 1: Proportion of people in each ethnic group by data source, England and Wales

**Percentages**

<b>Ethnic Group</b>	<b>England</b>		<b>Wales</b>	
	2020 admin-based ethnicity statistics	Census 2021 estimates	2020 admin-based ethnicity statistics	Census 2021 estimates
<b>Asian</b>	8.9	9.6	2.4	2.9
Bangladeshi	0.9	1.1	0.4	0.5
Chinese	0.8	0.8	0.5	0.5
Indian	2.7	3.3	0.5	0.7
Pakistani	2.5	2.8	0.5	0.6
Asian other	2.0	1.7	0.6	0.7
<b>Black</b>	4.0	4.2	0.7	0.9
African	2.3	2.6	0.5	0.6
Caribbean	1.0	1.1	0.1	0.1
Black other	0.7	0.5	0.1	0.1
<b>Mixed</b>	2.7	3.0	1.5	1.6
White and Asian	0.6	0.8	0.4	0.5
White and Black African	0.3	0.4	0.2	0.3
White and Black Caribbean	0.6	0.9	0.4	0.4
Mixed other	1.1	0.8	0.6	0.4
<b>White</b>	83.4	81.0	94.9	93.8
British	75.4	73.5	90.9	90.6
Gypsy, Roma or Irish Traveller	0.1	0.3	0.1	0.2
Irish	0.5	0.9	0.3	0.4
White other	6.6	6.3	2.2	2.7
White not specified	0.7	[z]	1.4	[z]
<b>Other</b>	1.1	2.2	0.5	0.9
Arab	0.1	0.6	0.2	0.4
Any other ethnic group	1.0	1.6	0.3	0.5

Source: Office for National Statistics

Notes

1. The admin-based ethnicity proportions have been calculated out of those with a stated ethnicity.
2. Proportions may not sum to totals because of rounding.
3. The exact wording used for each ethnic group differs across data sources, as detailed in the accompanying dataset.
4. The Gypsy, Roma and Irish Traveller ethnic groups have been aggregated because of differences in response options across data sources, meaning that it is not possible to separate them.
5. Hospital Episode Statistics (HES), Emergency Care Data Set (ECDS) and Improving Access to Psychological Therapies (IAPT) do not include any Gypsy, Roma or Irish Traveller response options.
6. Higher Education Statistics Agency (HESA) data for England and Wales only have categories for White and Gypsy or Traveller within the higher-level White ethnic group. Those with a sub-category ethnicity of White in HESA were re-coded as White not specified.
7. Birth Notifications, HES, ECDS, IAPT and the English School Census (ESC) main codes do not include an Arab response option.
8. Welsh School Census (WSC) and HESA collections in England and Wales do not have a White Irish response option.
9. [z] denotes not applicable, as the category was not used in the data collection.
10. Confidence intervals for the Census 2021 estimates are available as part of the Census quality measures.

The admin-based ethnicity statistics for local authorities and Lower layer Super Output Areas (LSOAs) broadly align with Census 2021 in showing how ethnic diversity varies across England and Wales. However, there are differences compared with Census 2021 in the reported sizes of each ethnic group. These are largely in line with differences seen at the national level but with additional localised differences that are explored further in this article.

## **Asian ethnic group**

Overall, there is a smaller proportion of people in the Asian ethnic group in the 2020 admin-based ethnicity statistics than the Census 2021 estimates, at 8.9% compared with 9.6% in England, and 2.4% compared with 2.9% in Wales. It is for the 5-year age bands between 30 and 49 years where the 2020 admin-based ethnicity statistics are the furthest below the Census 2021 estimates, with differences of up to 2.4 percentage points in England and 1.4 percentage points in Wales. This may be because of migrants entering the population since 2011 and their ethnicity not being identified in the administrative data.

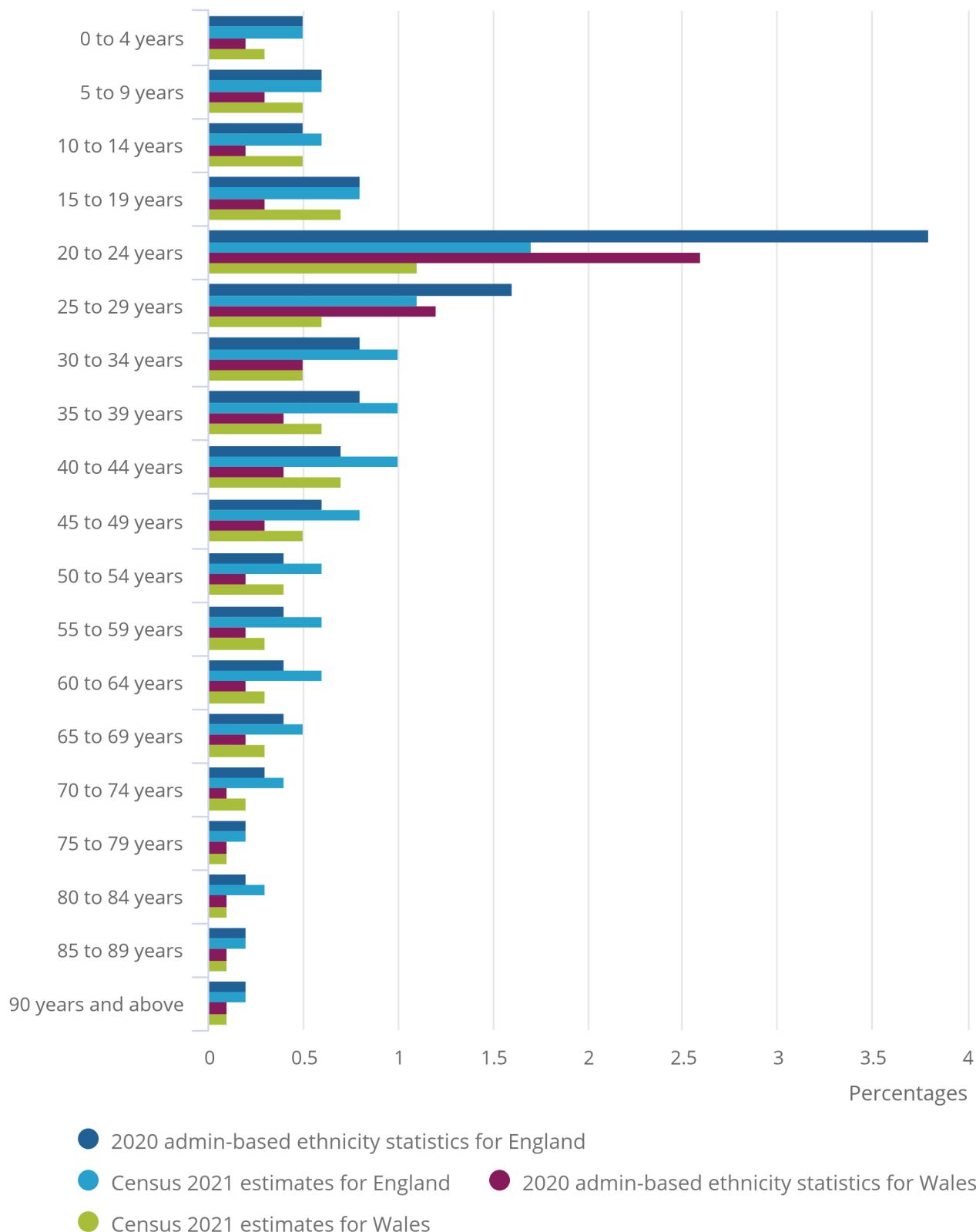
Contrary to the overall picture, for the 20 to 24 years age group in both England and Wales, the 2020 admin-based ethnicity statistics for the Asian ethnic group are higher than the Census 2021 estimates. This is driven by the Chinese ethnic group, with the proportion of people in the Chinese ethnic group in the 2020 admin-based ethnicity statistics over double that in the Census 2021 estimates. For this age group, the ethnicity record has been taken from the Higher Education Statistics Agency (HESA) data for 92.2% of people in the Chinese ethnic group in England and 98.6% in Wales. The differences between the 2020 admin-based ethnicity statistics and Census 2021 estimates may therefore, at least in part, be caused by changes in migration for study because of the COVID-19 pandemic. In other words, this is because students were included at their term-time address as recorded in the HESA data in the 2020 SPD V3.0 but not resident in England and Wales as at Census 2021 because of deferring their studies or doing distance learning. Local authorities containing large universities are also affected by this.

**Figure 3: The proportion of people aged 20 to 24 years in the Chinese ethnic group in the 2020 admin-based ethnicity statistics is over double the proportion in the Census 2021 estimates**

Proportion of people in the Chinese ethnic group by age and data source, England and Wales

**Figure 3: The proportion of people aged 20 to 24 years in the Chinese ethnic group in the 2020 admin-based ethnicity statistics is over double the proportion in the Census 2021 estimates**

Proportion of people in the Chinese ethnic group by age and data source, England and Wales



**Source: Office for National Statistics**

**Notes:**

1. The admin-based ethnicity proportions have been calculated out of those with a stated ethnicity.

## **Black ethnic group**

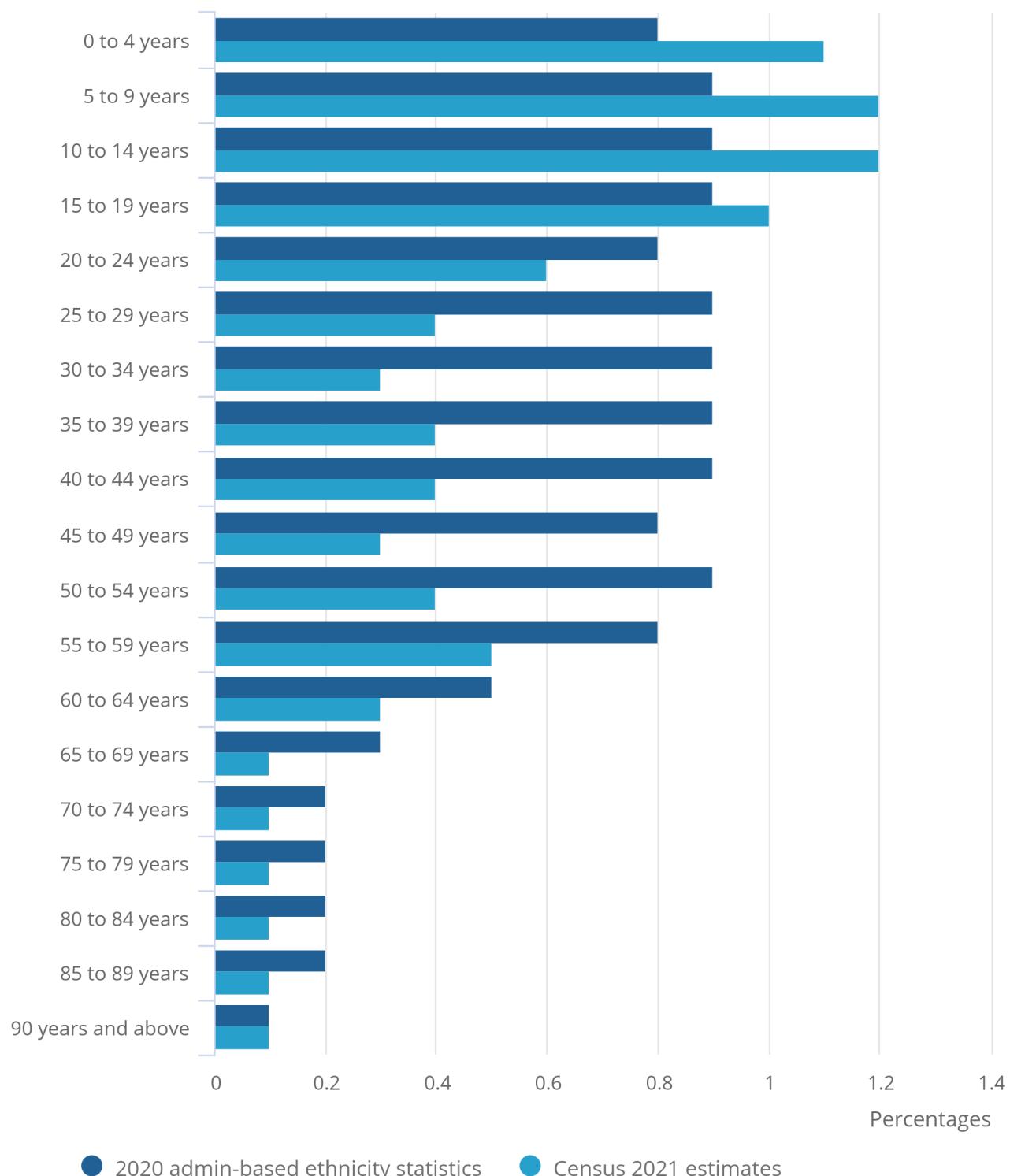
In England, the 2020 admin-based ethnicity statistics have a slightly lower proportion of people recorded as Black African and Black Caribbean than the Census 2021 estimates and a slightly higher proportion recorded as Black Other. This finding for the Black Other ethnic group does not hold across all ages though, with the census proportion higher than the admin proportion for those aged under 18 years (Figure 4).

**Figure 4: In England, the proportion of people in the Black Other ethnic group in the 2020 admin-based ethnicity statistics is lower than the Census 2021 estimate for children but higher for adults**

Proportion of people in the Black Other ethnic group by age and data source, England

Figure 4: In England, the proportion of people in the Black Other ethnic group in the 2020 admin-based ethnicity statistics is lower than the Census 2021 estimate for children but higher for adults

Proportion of people in the Black Other ethnic group by age and data source, England



● 2020 admin-based ethnicity statistics

● Census 2021 estimates

**Notes:**

1. The admin-based ethnicity proportions have been calculated out of those with a stated ethnicity.

In Wales, a similar proportion of people are recorded as Black Caribbean and Black Other in the 2020 admin-based ethnicity statistics compared with the Census 2021 estimates, whereas a smaller proportion of people are recorded as Black African.

Contrary to the overall picture for the Black African ethnic group, looking at the data by local authority, in Herefordshire and Powys there is a much higher proportion of people recorded as Black African in the 2020 admin-based ethnicity statistics than in the Census 2021 estimates. We have identified a potential ethnicity data quality issue within the hospital data for some areas in England that is affecting the reporting of people in the Black African ethnic group. As Powys neighbours Herefordshire, one of the affected areas, and 49.0% of ethnicity records in Powys have come from the hospital data, Powys is also affected by this potential issue. We plan to explore this further with NHS England.

## Mixed ethnic group

In both England and Wales, the proportion of people in the White and Asian, White and Black African, and White and Black Caribbean ethnic groups is lower in the 2020 admin-based ethnicity statistics than the Census 2021 estimates, whereas the proportion in the Mixed Other ethnic group is higher. These findings hold across all age groups and most local authorities. Two notable exceptions are Herefordshire and the neighbouring Welsh local authority of Powys. In Herefordshire, the proportion of people recorded as White and Asian is 1.4% in the admin-based ethnicity statistics but only 0.4% in the Census 2021 estimates. These figures are 0.4% and 0.3%, respectively, for Powys. This appears to be caused by mis-recording of ethnicity in the hospital data, with the finding for Powys reflecting use of hospitals in England.

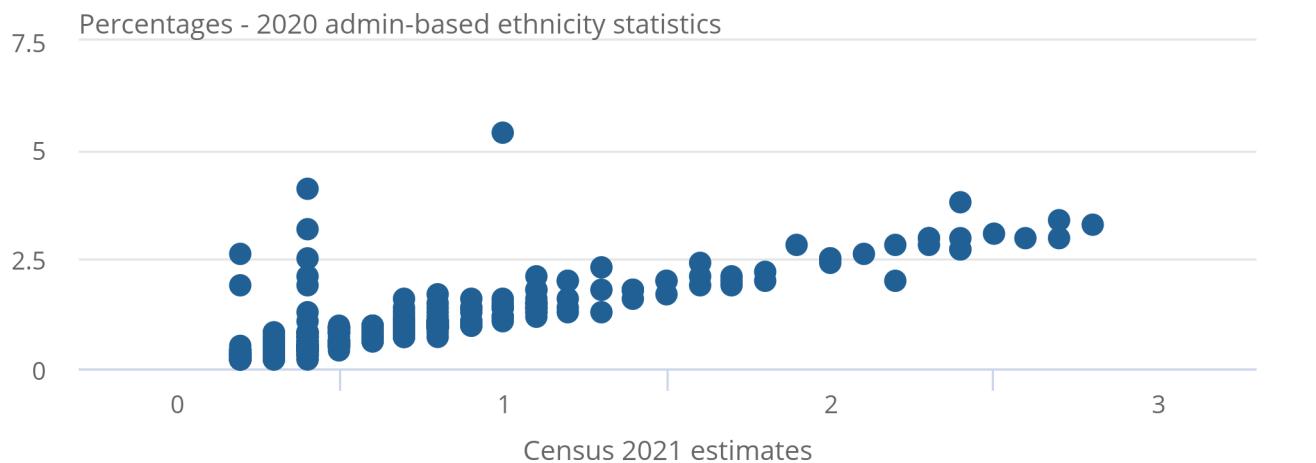
There are two clusters of local authorities in England where the proportion in the Mixed Other ethnic group in the 2020 admin-based ethnicity statistics is between 3 and 12 times the Census 2021 proportion; one is West Lindsey, South Holland, East Lindsey, and Boston and the other is East Suffolk, Babergh, Mid Suffolk, and Ipswich (Figure 5). Additionally, in Wirral, the proportion of people recorded as Mixed Other in the 2020 admin-based ethnicity statistics is 3.2% whereas the Census 2021 proportion is only 0.4%. These differences appear to also be caused by mis-recording of ethnicity in the hospital data. In East Lindsey, for example, 91.2% of Mixed Other records have come from either the Accident and Emergency part of the Hospital Episode Statistics (HES A&E) data or the Emergency Care Data Set (ECDS) (compared with 21.5% of all ethnicity records in East Lindsey). Where we were able to link an ethnicity from the 2011 Census to the admin-based ethnicity dataset (ABED), of those with a Mixed Other record taken from HES A&E or ECDS, 97.5% were recorded as White British in the 2011 Census and only 0.4% were recorded as Mixed Other. Potential sources of error in ethnicity data collection in hospitals are explored in our [Methods and systems used to collect ethnicity information in health administrative data sources, England: 2022 article](#).

**Figure 5: For several local authorities in England, the 2020 admin-based ethnicity statistics for the Mixed Other ethnic group are much higher than the Census 2021 estimates**

Proportion of people in the Mixed Other ethnic group in each local authority in the 2020 admin-based ethnicity statistics versus Census 2021 estimates, England and Wales

**Figure 5: For several local authorities in England, the 2020 admin-based ethnicity statistics for the Mixed Other ethnic group are much higher than the Census 2021 estimates**

Proportion of people in the Mixed Other ethnic group in each local authority in the 2020 admin-based ethnicity statistics versus Census 2021 estimates, England and Wales



**Source:** Office for National Statistics

**Notes:**

1. The 2020 admin-based ethnicity proportions have been calculated out of those with a stated ethnicity.
2. Local authority boundaries are as of 2021.

## **White ethnic group**

Overall, the proportion of people in the five-category White ethnic group is higher in the 2020 admin-based ethnicity statistics than the Census 2021 estimates, at 83.4% compared with 81.0% in England, and 94.9% compared with 93.8% in Wales. This holds across most local authorities and most age groups.

Currently, it is not possible to produce an accurate breakdown of the White ethnic group for three reasons. Firstly, Gypsy, Roma and Irish Traveller are not response options on all administrative data collections. Secondly, White Irish is not a response option in the Welsh School Census. Thirdly, the HESA data collections for universities in England and Wales only contain response options of White and Gypsy or Traveller within the higher-level White ethnic group. There are not response options for White British, White Irish, or White Other. Where possible, a previous ethnicity was taken from another data source for those with a most recent ethnicity in HESA of White. It was not possible to take a previous response for 0.7% of people in England and 1.4% of people in Wales. However, by single year of age, these figures peak at 4.5% for those aged 21 years in England and 10.3% for those aged 29 years in Wales.

It is our ambition to be able to produce more detailed statistics on the White ethnic group in future, but changes to data collections are required to enable this. Progress is being made though. The [ethnicity categories used in HESA data collections were changed in August 2022](#) to align with the latest census data collections in each country of the UK. A review is underway of the ethnicity categories used in NHS data collections in England and the [harmonised standard for ethnicity data collection is under review](#).

## **Other ethnic group**

It is not possible to produce accurate statistics from administrative data on the size of the Arab ethnic group because it is not a response option across all administrative data sources. As for the White ethnic group, it is our ambition to be able to do this in future, and progress is being made for some data sources to add an Arab response option.

For Any other ethnic group, the proportion in the 2020 admin-based ethnicity statistics is much lower than Census 2021, at 1.0% compared with 1.6% for England, and 0.3% compared with 0.5% for Wales. This is likely because of [the rule we implemented for the Any other ethnic group](#) when constructing the admin-based ethnicity dataset to correct for over-recording of people into this category. It appears that this rule has overcorrected the Any other ethnic group.

However, it is worth noting that between the 2011 Census and Census 2021, there was an increase in the proportion of people who chose to identify their ethnicity through the Any other ethnic group write-in option (0.6% in 2011 to 1.6% in 2021 in England; 0.3% in 2011 to 0.5% in 2021 in Wales). There was a new search-as-you-type function introduced for the 2021 online census questionnaire and we believe that [the new functionality encouraged individuals to identify their ethnic group in this way](#). We need to explore this further though to understand the extent to which the increase in the proportion of people identifying as Any other ethnic group was driven by population change versus change in ethnic identity. We will also review our ethnicity selection rules through record-level comparisons of the admin-based ethnicity dataset and Census 2021 data. If needed, we will then consider amending the rules to ensure that we can accurately report on the size of this ethnic group using administrative data.

## GSFREE

In addition to producing ethnicity statistics directly from the ABED, we have explored producing statistics for local authorities using the Generalised Structure Preserving Estimator (GSFREE) method. This is a small area estimation technique that, through a statistical model, aims to produce the best possible estimates by combining and drawing strength from multiple data sources. In our case, this involves combining administrative and survey data.

Our research builds on [earlier work](#) using this technique by using the ABED as the administrative data input instead of using the English School Census. Additionally, SPD V3.0 population totals were used instead of SPD V2.0. As was done previously, weighted ethnic group distributions from the Annual Population Survey (APS) were used for the ethnic group totals for England and Wales. APS data on ethnic group by local authority were also fed into the model.

We have produced GSPREE estimates for the five high-level ethnic groups. To assess the GSPREE approach, we have compared both the GSPREE estimates and the admin-based ethnicity statistics with the Census 2021 estimates (Figure 6). Overall, the GSPREE method does not significantly improve the statistics by local authority on the proportion of people in each ethnic group.

In England, for the Asian and Black ethnic groups, a similar proportion of local authorities are closer to the Census 2021 estimates for the GSPREE method and admin-based ethnicity statistics. The admin-based ethnicity statistics are closer to the Census 2021 estimates for the majority of local authorities for the Mixed and White ethnic groups.

For local authorities in Wales, the GSPREE method has made very little difference to the proportion of people in the Asian, Black, and White ethnic groups compared with calculating the proportions directly from the ABED. For the Mixed ethnic group, similar to England, the 2020 admin-based ethnicity statistics are closer to the Census 2021 estimates than the GSPREE estimates are for all local authorities in Wales.

For the Other ethnic group, for most local authorities in England and Wales, the GSPREE estimates are closer to the Census 2021 estimates than the admin-based ethnicity statistics are. The Other ethnic group in the admin-based ethnicity statistics is affected by the implementation of the rule to correct for over-use of the Any other ethnic group category in administrative data collections. For most local authorities, GSPREE increases the proportion in the Other ethnic group, bringing it closer to the Census 2021 proportion.

### **Figure 6: The Generalised Structure Preserving Estimator (GSPREE) method does not significantly improve the statistics by local authority on the proportion of people in each ethnic group**

Proportion of people in each ethnic group in each local authority from the 2020 GSPREE method and the 2020 admin-based ethnicity statistics, versus Census 2021 estimates, England and Wales

#### Notes:

1. The 2020 admin-based ethnicity statistics proportions have been calculated out of those with a stated ethnicity.
2. Local authority boundaries are as of 2021.
3. Data do not include Isles of Scilly.
4. The dashed diagonal line shows perfect agreement between the Census 2021 estimates and the 2020 research outputs. If the point is below the line, the 2020 research output is lower than the Census 2021 estimate. If the point is above the line, the 2020 research output is higher than the Census 2021 estimate. If the admin-based ethnicity statistics point is closer to the line than the GSPREE point, the admin-based ethnicity statistic is closer to the Census 2021 estimate than the GSPREE estimate is.

## Download the data

[.xlsx](#)

The GSPREE method relies on having robust and recent ethnicity data at least at national level. Table 2 shows the year ending June 2020 APS proportions for England and Wales, which were used in the GSPREE model. For the non-White ethnic groups, the APS estimates are lower than the Census 2021 estimates. These differences are influencing the results we can attain from the GSPREE method. Without more robust survey estimates at national level, the GSPREE method cannot produce more accurate estimates of the population by ethnic group at local authority level than using the admin-based ethnicity dataset alone.

Table 2: Proportion of people in each ethnic group by data source, England and Wales

Ethnic Group	2020 admin-based ethnicity statistics	Year ending June 2020 APS estimates	Census 2021 estimates
Asian	8.5	8.0	9.3
Black	3.9	3.7	4.0
Mixed	2.6	2.0	2.9
White	84.0	84.5	81.7
Other	1.0	1.9	2.1

Source: Office for National Statistics

## Notes

1. The 2020 admin-based ethnicity proportions have been calculated out of those with a stated ethnicity.
2. APS estimates have been weighted to the household population.
3. Generalised Structure Preserving Estimator (GSPREE) estimates are available at the local authority level only, not the national level; GSPREE uses the national-level APS ethnic group distributions within the model.

## 5 . Developing admin-based ethnicity statistics for England and Wales data

[Developing admin-based ethnicity statistics for England and Wales](#)

Dataset | Released 7 February 2023

Data on population coverage and ethnic breakdowns from the admin-based ethnicity dataset for England and Wales.

## 6 . Glossary

## Ethnic group

The Office for National Statistics notes that there is no consensus on [what constitutes an ethnic group](#). For this article, we define ethnic group as the self-reported ethnic group of the individual, according to their own perceived ethnic group and cultural background.

However, when providing their ethnic group in the administrative and census data collections, respondents must choose from the response options presented to them or decide not to respond. Response options differ across the different data collections. For some administrative data collections, a more detailed set of response options was used. We have aggregated these into the ethnic groups used in this article, as outlined in the [accompanying dataset](#).

A [harmonised standard](#) is available for the collection of data on ethnicity. [The standard is currently under review](#) to ensure that it continues to meet user needs and reflects the diversity of the population.

## Ethnicity refused

In the English School Census (ESC) and Welsh School Census (WSC), if a parent or guardian, or pupil, has declined to provide ethnicity data, this is recorded as "Refused". In Hospital Episode Statistics (HES), the Emergency Care Data Set (ECDS), Birth Notifications and Improving Access to Psychological Therapies (IAPT), where a patient chooses not to state their ethnicity, the code "Z – Not Stated" is recorded. In the Higher Education Statistics Agency (HESA) data, the code "98 Information Refused" is recorded. In the Individualised Learner Record (ILR) data, the code "99 Not Provided" 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 and WSC, where the ethnicity has not yet been collected, this is recorded as "NOBT" (information not yet obtained). In HES, ECDS, IAPT and Birth Notifications, where the person's ethnicity is unknown, the default code "99 Not known" is used. In HESA, "90 Not known" is used. ILR does not have a specific code for unknown.

Across the other datasets, blank and null ethnicity values were treated as unknown, as were codes not included in the relevant dataset documentation, unless stated otherwise in the [accompanying dataset](#).

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, 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 Statistical Population Dataset version 3.0 (SPD V3.0) but have not been linked to any sources of ethnicity data.

## Usually resident

As defined in [our latest admin-based population estimates \(ABPE\) publication](#), inclusion in SPD V3.0 is based on the United Nations (UN) definition of "usually resident". This is the place at which a person has lived continuously for at least 12 months, not including temporary absences for holidays or work assignments, or intends to live for at least 12 months (United Nations, 2008).

# 7 . Data sources and quality

The admin-based ethnicity statistics were produced using the following administrative data sources:

- [English School Census \(ESC\)](#), 2011 to 2020: a statutory data collection about pupils in state-funded schools in England
- [Welsh School Census \(WSC\)](#), 2011 to 2020: a statutory data collection about pupils in state-funded schools in Wales
- [Individualised Learner Record \(ILR\)](#), 2008 to 2020: a dataset containing information about individuals who attend training from providers in the Further Education (FE) and Skills sector in England
- [Higher Education Statistics Agency \(HESA\)](#), 2010 to 2020: a dataset containing information about students at publicly funded higher education institutions in the United Kingdom plus the University of Buckingham
- [Hospital Episode Statistics \(HES\)](#), 2009 to 2020: a database containing details of all attendances at NHS hospitals in England
- [Emergency Care Data Set \(ECDS\)](#), 2020: a dataset containing information about people who have attended emergency departments in England, and which has replaced the Accident and Emergency part of the HES dataset
- [Birth Notifications](#), 2006 to 2020: a database containing details of babies born in England, Wales, and the Isle of Man
- [Improving Access to Psychological Therapies \(IAPT\)](#), 2012 to 2018: a dataset containing information about individuals who have accessed NHS psychological therapies in England

In addition to the administrative data sources listed, we also linked in ethnicity data from the 2011 Census.

We considered adding in Patient Episode Data for Wales (PEDW) and the Emergency Department Data Set (EDDS) for Wales but have not done so because of the high proportion of records in the data with an ethnicity code of "Z – Not Stated". We are in discussions with Digital Health and Care Wales about these datasets and will revisit them in the next phase of the research.

Ethnicity records from the data sources provided in this section were linked to the Statistical Population Dataset version 3.0 (SPD V3.0) based on a unique identifier. Records that did not link to SPD V3.0 were dropped. A [rules-based approach](#) was used to select a final ethnicity per person, largely based on taking the most recently recorded ethnicity but with some additional steps for those with a most recently recorded ethnicity of Any other ethnic group or White not specified. Of those with a stated ethnicity recorded in the admin-based ethnicity dataset, the ethnicity was taken from the 2011 Census for 6.8% of people in England and 55.9% of people in Wales.

The admin-based ethnicity statistics are based on the proportion of people in each ethnic group. Records where the final ethnicity was unknown or refused were excluded when calculating the proportions.

## 8 . Future developments

The research presented in this article continues to show potential for producing ethnicity statistics down to local levels from administrative data. We will continue to explore how we can further improve upon the admin-based ethnicity statistics through:

- incorporating additional data sources to improve the population coverage
- reviewing and refining the rules to deal with multiple ethnicity records for an individual
- producing admin-based ethnicity statistics for 2021, using the Statistical Population Dataset version 4.0 (SPD V4.0) as the population base, and conducting both aggregate-level and record-level comparisons with Census 2021
- further exploring methods to adjust for missingness in the administrative data
- engaging with data suppliers to better understand and improve data collection practices

## Feedback

We welcome feedback on the admin-based ethnicity statistics presented in this article and the [accompanying dataset](#), particularly whether the level of accuracy is sufficient to meet user needs. We also welcome feedback on the planned future developments. Please email your feedback to [Admin.Based.Characteristics@ons.gov.uk](mailto:Admin.Based.Characteristics@ons.gov.uk).

## 9 . Related links

### [Producing admin-based ethnicity statistics for England: changes to data and methods](#)

Article | Released 23 May 2022

An overview of the changes to data sources and methods in the feasibility research on producing statistics on the population by ethnic group from administrative data.

### [Developing admin-based ethnicity statistics for England: 2016](#)

Article | Released 23 May 2022

Update on feasibility research on producing statistics on the population by ethnic group for England from administrative data, following the inclusion of additional data sources and method changes. These research outputs are not official statistics.

### [Change over time in admin-based ethnicity statistics, England: 2016 to 2020](#)

Article | Released 23 May 2022

Feasibility research on producing statistics on the population by ethnic group for England for 2016 to 2020 from administrative data, exploring change over time. This builds on work to improve the 2016 admin-based ethnicity statistics. These research outputs are not official statistics.

### [Admin-based ethnicity statistics for England, feasibility research: 2016](#)

Article | Released 6 August 2021

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.

### [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.

## 10 . Cite this article

Office for National Statistics (ONS), released 7 February 2023, ONS website, article, [Developing admin-based ethnicity statistics for England and Wales: 2020](#)