Office for National Statistics

Article

Change over time in admin-based ethnicity statistics, England: 2016 to 2020

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.

Contact:
Alison Morgan
Admin.Based.
Characteristics@ons.gov.uk
+44 1329 447187

Release date:
23 May 2022

Next release:
To be announced

Table of contents

1. Main points

2. About our transformation research

3. Population coverage

4. Ethnic breakdown

5. Incorporating the 2011 Census

6. Glossary

7. Data sources and quality

8. Future developments

9. Related links

1 . Main points

- We have produced a set of admin-based ethnicity statistics for 2016 to 2020 as part of our feasibility research.
- Across the 2016 to 2020 time series, the proportion of the admin-based population for which we were able to establish an ethnicity increased each year, from 74.7% in 2016 to 79.9% in 2020.
- Including the 2011 Census as a data source further increases the proportion of the population with a stated ethnicity, but its impact reduces over time.
- Between 2016 and 2020, the admin-based ethnicity statistics showed increases in the Asian, Black, Mixed and Other ethnic groups, and a decrease in the White ethnic group; we will use Census 2021 data when available to evaluate these trends.

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, which is different to what is currently used in the production of ethnicity statistics. These outputs should not be used for policy or decision-making.

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 and the last official statistics available were from the 2011 Census. Our Experimental Statistics that primarily used Annual Population Survey data were published in December 2021 and Census 2021 estimates will be released later this year.

In August 2021, we published findings from our initial feasibility research on producing statistics on the population by ethnic group for England from administrative data. The research was based on linking ethnicity data from Hospital Episode Statistics (HES), English School Census (ESC) and Improving Access to Psychological Therapies (IAPT) to a 2016 admin-based population base and implementing a set of rules to deal with multiple recorded ethnicities for an individual. A set of admin-based ethnicity statistics were produced for 2016 based on the proportion of people in each ethnic group; we refer to these as version 1. Full details of the previous method can be found in the methods paper published alongside the research outputs.

Since then, we have been working to improve the admin-based ethnicity statistics through incorporating three additional data sources and refining the method. Full details of the changes are available in the accompanying methodology changes article.

We have produced a new set of admin-based ethnicity statistics for 2016, version 2, and compared them with those produced in version 1. This comparison is in our accompanying article. In this article we present admin-based ethnicity statistics for 2016 to 2020, based on version 2, to explore population coverage and change in ethnicity over time.

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.

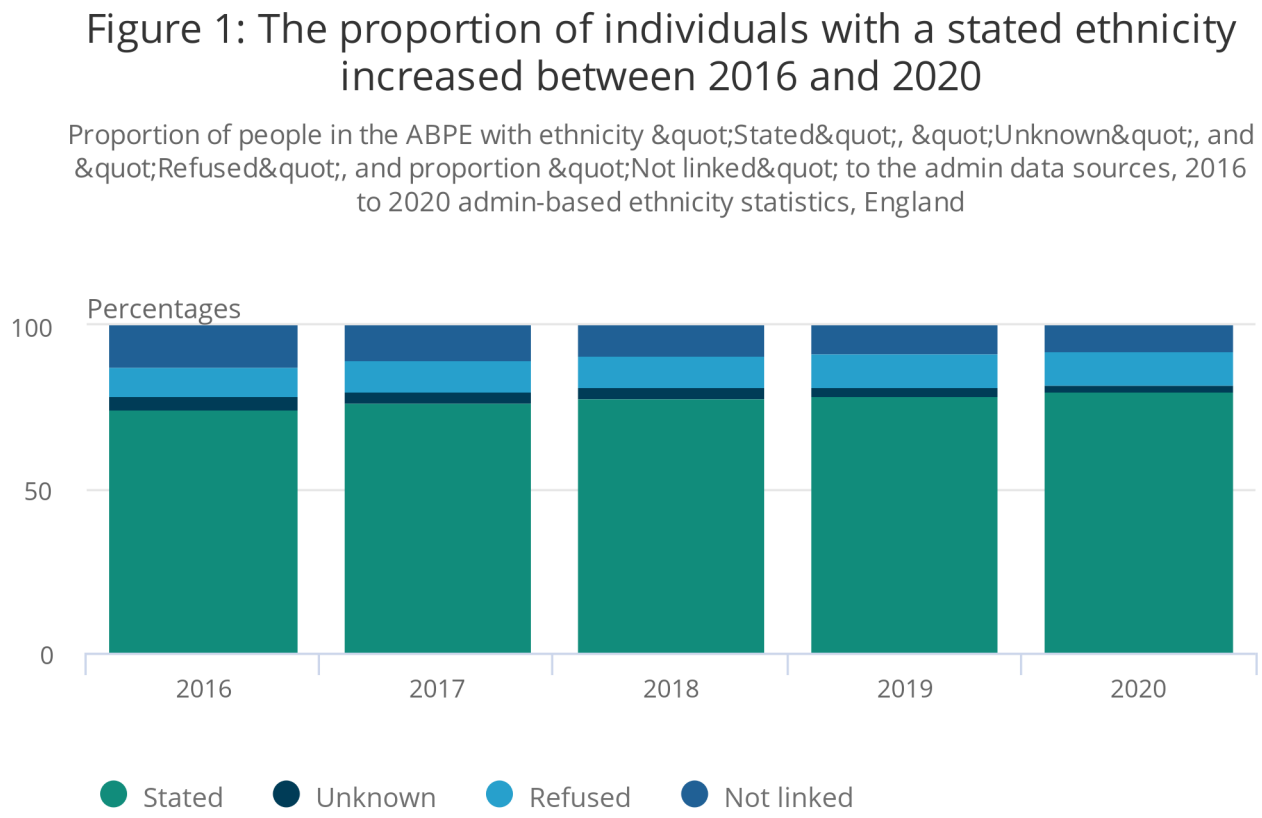
3 . Population coverage

The admin-based population estimates (ABPE) v3.0 datasets for each year from 2016 to 2020 were used as the population bases for the admin-based ethnicity statistics. The ABPEs aim to approximate the usually resident population as at 30 June of the reference year.

The proportion of individuals in the ABPEs for which we have a stated ethnicity has increased each year over the time series, from 74.7% in 2016 to 79.9% in 2020. This is because of an increase in the proportion that could be linked to an ethnicity data source. The proportion of people with an unknown ethnicity has decreased slightly, from 4.0% in 2016 to 2.4% in 2020, but the level of refusals has increased slightly, from 9.0% in 2016 to 10.5% in 2020.

Figure 1: The proportion of individuals with a stated ethnicity increased between 2016 and 2020

Proportion of people in the ABPE with ethnicity "Stated", "Unknown", and "Refused", and proportion "Not linked" to the admin data sources, 2016 to 2020 admin-based ethnicity statistics, 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.

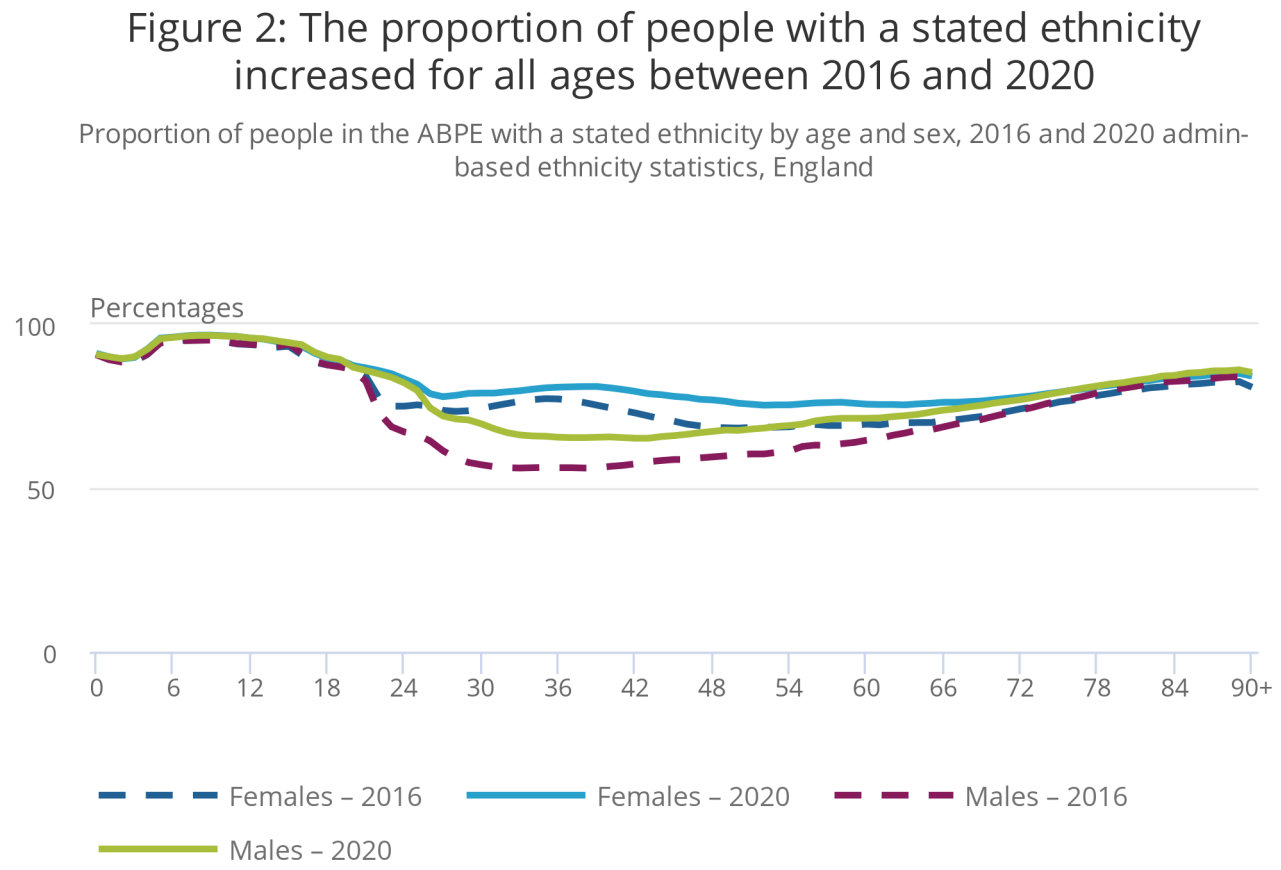
3. "Refused" refers to those with a refusal on their most recent administrative data record.

4. "Not linked" refers to individuals in the ABPE who have not been linked to any admin data source.

Between 2016 and 2020, the proportion of people with a stated ethnicity increased across all ages, with the biggest increases being for males in their 20s and 30s. The difference in coverage between males and females has reduced, as have the differences in coverage by age.

Figure 2: The proportion of people with a stated ethnicity increased for all ages between 2016 and 2020

Proportion of people in the ABPE with a stated ethnicity by age and sex, 2016 and 2020 admin-based ethnicity statistics, 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. All individuals in the ABPE have their sex recorded as either male or female.

3. Age is at 30 June of the year of interest.

When looking at the data by local authority, the proportion of the ABPE population with a stated ethnicity generally increased over time, with the majority of local authorities having a higher proportion of people with a stated ethnicity in 2020 compared with 2016. St Helens is the local authority with the highest proportion of the ABPE population with a stated ethnicity in 2020, at 90.9%. The local authorities with the biggest increases were:

- Gosport at 19.6 percentage points
- Isle of Wight at 19.5 percentage points
- Havant at 18.2 percentage points
- Fareham at 17.7 percentage points

The increases in these four local authorities were the result of a decrease in the proportion of individuals with ethnicity refused along with a decrease in the proportion of individuals that could not be linked to one of the ethnicity data sources.

Throughout the years, the proportion of individuals with a stated ethnicity in the City of London has remained consistently low and a decrease of 3.5 percentage points can be seen from 2016 to 2020. The other local authorities with a decrease in the proportion of individuals with a stated ethnicity were:

- East Staffordshire at 0.4 percentage points
- Eastbourne at 0.2 percentage points
- Stockton-on-Tees at 0.1 percentage points

Figure 3: Coverage has generally improved by local authority between 2016 and 2020

Proportion of people in the ABPE with a stated ethnicity in the 2016 and 2020 admin-based ethnicity statistics, 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. Local authority boundaries are as of 2021.

Download the data

.xlsx

4 . Ethnic breakdown

The admin-based ethnicity statistics show increases in the proportions of the population in the Asian, Black, Mixed and Other ethnic groups between 2016 and 2020 and a decrease in the proportion in the White ethnic group. These trends were also seen in the majority of local authorities. Comparing our 2019 Experimental Statistics on the population by ethnic group with the 2011 Census showed similar trends, except for the Mixed ethnic group where the 2019 Experimental Statistics were lower than the 2011 Census figures. When the Census 2021 estimates of the population by ethnic group are released, we will use them to evaluate the reliability of the trends shown in the admin-based ethnicity statistics and the experimental Annual Population Survey (APS)-based estimates.

Changes in the proportion of people in each ethnic group in the admin-based ethnicity statistics are driven by a number of factors.

Population change

Identified by those only in the admin-based population estimates (ABPE) in either year one or year two. It should be noted that changes in the ABPE population could represent real changes in the population owing to births, deaths and migration, or could be because of the availability of the activity data, which are used to determine if someone is in the population. More information on the ABPEs can be found in Population and migration statistics system transformation – recent updates: evaluating coverage and quality in the admin-based population estimates.

Change in the availability of ethnicity data

Either being able to link on an ethnicity record from the admin data when we could not previously, or people moving in or out of the unknown and refused groups

Change in the ethnicity recorded for an individual

An individual being recorded as a different ethnicity in year two compared with year one.

It is the net impact of each factor and the relative change in the size of each ethnic group that affects the proportion of the population that they make up.

Looking in more detail at the White ethnic group, Table 1 shows a decrease in the proportion of people in the White British ethnic group between 2016 and 2020. The main driver behind this was population change. Across all pairs of consecutive years (2016 to 2017, 2017 to 2018, 2018 to 2019), the number of people recorded as White British who left the ABPE was greater than the number of people recorded as White British who entered.

For the White Other ethnic group, the opposite trend was seen, with an increase in the proportion of people recorded as White Other between 2016 and 2020. However, rather than being caused by population change, this was mainly driven by changes in the availability of ethnicity data. This may reflect a time lag in identifying the ethnicity of migrants, but this is something that we need to explore through further research.

The decrease in the White not specified ethnic group in Table 1 is driven by change in recorded ethnicity. The additional step in the ethnicity selection process for the White not specified ethnic group is a factor. If an individual was recorded in another admin data source and then as White not specified in HESA, they would be assigned to their previous ethnicity.

The additional step also happens for Any other ethnic group, but the proportion of people recorded as Any other ethnic group has still increased over time. This is because of population change and availability of ethnicity data.

Population change is also an important driver of changes in the Chinese and Mixed ethnic groups. Changes in the size of the Chinese ethnic group are likely to be driven by changes in patterns of immigration to England for study, and emigration on completion of studies. For the Mixed ethnic group however, births are likely to be an important factor, with birth statistics showing an increasing proportion of babies being of Mixed ethnicity.

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

Ethnic Group	Admin-based ethnicity statistics					2011 Census	2019 APS-based Experimental Statistics
	2016	2017	2018	2019	2020		
Asian	8.4	8.5	8.6	8.8	8.9	7.8	8.3
Bangladeshi	0.9	0.9	0.9	0.9	0.9	0.8	1.1
Chinese	0.7	0.7	0.7	0.8	0.8	0.7	0.6
Indian	2.5	2.5	2.6	2.6	2.7	2.6	2.9
Pakistani	2.5	2.5	2.5	2.6	2.5	2.1	2.4
Asian Other	1.8	1.8	1.9	1.9	2.0	1.5	1.3
Black	3.9	4.0	4.0	4.1	4.1	3.5	3.7
African	2.1	2.2	2.2	2.3	2.3	1.8	2.4
Caribbean	1.1	1.1	1.0	1.0	1.0	1.1	1.1
Black Other	0.7	0.7	0.7	0.8	0.8	0.5	0.2
Mixed	2.3	2.4	2.5	2.6	2.7	2.3	1.9
White and Asian	0.5	0.5	0.5	0.5	0.6	0.6	0.5
White and Black African	0.3	0.3	0.3	0.3	0.3	0.3	0.3
White and Black Caribbean	0.6	0.6	0.6	0.6	0.6	0.8	0.5
Mixed Other	1.0	1.0	1.1	1.1	1.2	0.5	0.5
White	84.2	83.9	83.6	83.2	82.9	85.4	84.2
British	76.4	75.8	75.4	74.8	74.3	79.8	77.6
Gypsy, Roma or Irish Traveller	0.1	0.1	0.1	0.1	0.1	0.1	[low]
Irish	0.5	0.5	0.5	0.5	0.5	1.0	0.6
White Other	5.7	6.0	6.3	6.6	6.7	4.6	6.0
White not specified	1.5	1.4	1.3	1.3	1.3	[z]	[z]
Other	1.2	1.2	1.2	1.3	1.4	1.0	1.9
Arab	[low]	0.1	0.1	0.1	0.1	0.4	0.5
Any other ethnic group	1.2	1.2	1.2	1.2	1.3	0.6	1.5

Source: Office for National Statistics

Notes

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

2. The Gypsy, Roma and Irish Traveller ethnic groups have been aggregated due to differences in response options across data sources meaning that it is not possible to separate them. The HES, IAPT and ECDS data sources do not include any Gypsy, Roma or Irish Traveller response options.

3. Out of the administrative data sources, the Arab ethnic group is only a response option in the HESA data.

4. The 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.

5. [z] denotes not applicable, as the category was not used in the data collection.

6. [low] denotes that the proportion is less than 0.05.

In addition to exploring the net impact of each component of change, linking the data across time gives insights into the level of stability of ethnicity recording within the admin data.

Out of the people in the admin-based population estimates (ABPE) v3.0 between 2016 and 2019, for 5.1%, the final ethnicity we selected had changed at least once during the time series. The proportion that had not changed will include people who only appeared in one year of the time series and those where we have taken the same administrative data record each year. When we look at those who at some point had the opportunity to provide an update to their ethnicity, meaning that they had an admin data record in a later year, this figure increases to 7.0%. This proportion is higher than we would expect given that research published on the stability of ethnic identity in England and Wales 2001 to 2011 found that 4.0% of people had a different ethnic group recorded in the 2011 Census compared with the 2001 Census. However, there may have been greater changes in ethnic identity in more recent years than between 2001 and 2011; this can be explored further once Census 2021 data are available.

For all pairs of years (2016 to 2017, 2017 to 2018, 2018 to 2019), the highest level of stability in ethnicity recording was in the White British ethnic group, followed by the Chinese, Bangladeshi, Pakistani and Indian ethnic groups. The lowest level of stability was in the Black Other ethnic group, with less than 70% of people recorded as Black Other in year one also recorded as Black Other in year two. However, they were still likely to be recorded as an ethnicity within the five-category Black ethnic group, with just under 25% recorded as either Black African or Black Caribbean. For the Mixed and Other ethnic groups, where the ethnicity was different in year two, it was more common to be recorded as an ethnicity within a different five-category ethnic group, than the five-category ethnic group from year one.

Figure 4: The level of change in recorded ethnicity varied by ethnic group

Proportion of records where the ethnicity was the same in the 2017 admin-based ethnicity statistics as in 2016, by ethnic group as per 2016

Source: Office for National Statistics

Notes:

1. Proportions are based on: out of those in the ethnic group in the 2016 admin-based ethnicity statistics, what ethnic group they had in the 2017 admin-based ethnicity statistics.

2. Only those with a stated ethnicity in both years and who had an admin data record in 2017 (rather than an earlier record being rolled forward) have been included.

Notes for: Ethnic breakdown

1. We were unable to include 2020 in the record-level analysis due to a change in the underlying linkage meaning there was not a common identifier to link the datasets together.

5 . Incorporating the 2011 Census

In addition to producing the admin-based ethnicity statistics using administrative data only, as described in our accompanying article, we have produced a set of figures based on incorporating 2011 Census as an additional data source. This is because we want to make the best use of all available data sources and the 2011 Census is the most complete source of ethnicity data as at Census Day. It also demonstrates what may be possible in future using Census 2021 data to ensure we maximise the utility of this rich data source.

Incorporating the 2011 Census increased the proportion of individuals with a stated ethnicity in 2016 from 74.7% to 84.7%. Section 5 of our accompanying publication provides further data on the impact of incorporating 2011 Census into our 2016 admin-based ethnicity statistics.

Over time, the impact of including 2011 Census data reduced and by 2020, the proportion of people with a stated ethnicity was only 4.7 percentage points higher when the 2011 Census was included than when using admin data only. This reduced impact is a result of population change and the increased proportion of people linking to at least one of the admin sources providing the ethnicity data.

6 . 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), it is recorded as “refused” if a parent or guardian, or pupil has declined to provide ethnicity data. In Hospital Episode Statistics (HES), the Emergency Care Dataset (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.

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, ECDS, IAPT and Birth Notifications, the default code “99 Not known” is used where the person's ethnicity is unknown. All blank and null ethnicity values in Birth Notifications were also treated as unknown. In HESA, “90 Not known” is used.

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 admin-based population estimates (ABPE) v3.0 but have not been linked to any sources of ethnicity data.

Usually resident

As defined in our latest ABPE publication, we are currently adopting the 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).

Version 1

Version 1 refers to the admin-based ethnicity statistics produced using HES, IAPT and ESC data and published in August 2021.

Version 2

Version 2 refers to the admin-based ethnicity statistics produced using HES, ECDS, IAPT, ESC, HESA and Birth Notifications data and with the new ethnicity selection rules.

7 . Data sources and quality

The admin-based ethnicity statistics were produced using administrative data sources. These are:

- English School Census (ESC), 2011 to 2020: a statutory data collection about pupils in state-funded schools in England.
- Hospital Episode Statistics (HES), 2009 to 2020: 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). AE data are not included in the 2020 HES data and instead are replaced by the Emergency Care Data Set (ECDS).
- Emergency Care Data Set (ECDS), 2020: a dataset containing information about people who have attended emergency departments in England. It has replaced the AE part of the HES dataset.
- Improving Access to Psychological Therapies (IAPT), 2012 to 2018: a dataset containing information about individuals who have accessed NHS psychological therapies in England.
- Birth Notifications, 2006 to 2020: a database containing details of babies born in England, Wales and the Isle of Man.
- Higher Education Statistics Agency (HESA), 2010 to 2020: a dataset containing information about students at publicly funded higher education institutions plus the University of Buckingham.

Ethnicity records from these admin data sources were linked to the admin-based population estimates (ABPE) v3.0 for the relevant time period based on a unique identifier. Records that did not link to the ABPE were dropped. See our accompanying methodology changes article for more information. Of those in the 2016 ABPE who could be linked to at least one of the ethnicity data sources, 77.7% of individuals had the same ethnicity on all records in the data and 13.4% had multiple recorded ethnicities within and across datasets. The remaining 8.9% of individuals only had "Unknown" or "Refused" on all ethnicity records. A method to select a final ethnicity per person was implemented, as described in the accompanying article.

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 ABPE v3.0 datasets for each year from 2016 to 2020 were used as the population bases for the admin-based ethnicity statistics. The ABPEs aim to approximate the usually resident population as at 30 June of the reference year. The quality of the population base will have an 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.

2019 Experimental Statistics

The 2019 Annual Population Survey (APS)-based estimates are Experimental Statistics that were produced from the three-year-pooled APS, Mid-Year Population Estimates and the 2011 Census. Data from the 2011 Census were incorporated into the methodology to capture the population living in both households and communal establishments. Further information on the method used to produce these estimates can be found in Population estimates by ethnic group and religion, England and Wales: 2019.

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

8 . Future developments

The research presented in this and our accompanying article continues to show 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:

- incorporating additional data sources to improve the population coverage for England and expand coverage to Wales
- exploring the potential to produce multivariate statistics on ethnicity by other characteristics
- exploring methods to adjust for missingness in the admin data
- exploring the potential to produce admin-based ethnicity statistics for smaller geographic areas
- engaging with data suppliers to better understand and improve data collection practices
- combining the administrative data with survey data using the Generalised Structure Preserving Estimator (GSPREE), building on previous work using this method
- conducting public acceptability testing on ethnicity selection methods
- using Census 2021 data to further assess the quality of the admin-based ethnicity statistics

Feedback

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

9 . Related links

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.

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.

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.

Population estimates by ethnic group and religion, England and Wales: 2019
Article | Released 16 December 2021
Experimental Statistics of population by ethnic group and religion by age and sex at the national and regional level for England and Wales. This builds upon previous research and user feedback. We are publishing these estimates for the first time to provide a more timely picture of the population between censuses and allow for user feedback.