CITY INTELLIGENCE

Review of Ethnic Population Data Sources

September 2021

Executive Summary

There are a number of sources for estimates of the size and distribution of ethnic group populations in England. These estimates vary in quality, accuracy, timeliness, and detail; in some cases, the underlying definition of what constitutes the resident population is different. This document outlines in some detail the major sources of ethnic group information currently available at the national and regional level. It also gives a brief summary of the estimates themselves.

For analyses relating to the London resident population it is recommended that the GLA 2016-based Housing-led variant ethnic projections are used.

The sources reviewed are:

- ONS 2011 Census
- ONS Population Estimates by Ethnic Group (PEEG)
- ONS Annual Population Survey ethnic estimates (APS)
- ONS Generalised Structure Preserving Estimator Ethnic Estimates (GSPREE)
- ONS Admin-based ethnicity statistics (ABPE)
- ONS Racial Disparity Audit Aged-on Ethnic Estimates (RDA)
- GLA Ethnic Group Population Projections
- NHS Digital Management Information Ethnic Category Coverage (EGPPR/HES)
- University of Leeds ETHPOP Ethnic Group Projections

The census provides the most accurate estimate of the population across a range of characteristics. The 2011 census output data is for eighteen ethnic groups, an increase from sixteen in 2001 and ten in 1991. However, the most recent census data is for March 2011 with results of the 2021 census not due until spring 2022. For context, the total population of London at the time of the 2011 census was 8.17 million, while in 2020 (the most recent estimate) it was almost a million more at 9.03 million. This growth has impacted London's ethnic make-up creating the need for more up-to-date ethnic group estimates.

Since the 2011 census, the ONS have produced a number of ethnic group population estimate research outputs using four different approaches. These are APS, PEEG, GSPREE and RDA (see above for full names). The APS and PEEG methods both use Annual Population Survey data which is adapted from the Labour Force Survey (LFS) to extrapolate the current size of each ethnic group population from the

survey sample. The GSPREE approach uses linked administrative datasets to estimate the characteristics of the population. This approach was developed as part of an ONS project to determine the potential for using admin data as replacement for the census. Finally, the RDA estimates were produced by ONS in 2017 specifically for use as denominators in the Race Disparity Audit. The method rolls census distributions and populations forward and is the least sophisticated of the approaches.

The GLA's own ethnic group projections rely heavily on the census for both starting populations and underlying rates. Therefore, the assumptions used in the modelling become less robust over time. As a result, the decision was taken that the 2016-based projections, published in 2017, would be the last ethnic group projections until the release of new census data. The projections cover the period 2012-2050.

NHS Digital have recently created a Management Information (MI) ethnic category asset that combines data from General Practice and Hospital Episode Statistics datasets. This information aims to inform users of the quality, coverage, and distribution of ethnic category recording within this asset, and to provide a near population (England only) level view of ethnic category.

Table 1   Summary of Ethnic Population Outputs for London by Five Broad Ethnic Groups

Figures are in thousands

	Reference		Ethnic Group Population					
Source	Date	Population	White	Mixed	Asian	Black	Other	Unknown
2011 Census	27-03-2011	8,174	4,887	405	1,512	1,089	281	
ONS APS PEEG	Mid-2015	8,674	5,117	318	1,556	1,157	525	
ONS GSPREE 1	Mid-2015	8,587	5,359	260	1,332	1,119	516	
ONS GSPREE 2	Mid-2015	8,587	5,272	273	1,435	1,105	501	
ONS APS 1yr	Mid-2016	8,788	5,175	367	1,625	1,092	529	
ONS APS 3yr	Mid-2016	8,769	5,206	326	1,611	1,095	531	
ONS RDA1	Mid-2016	8,770	5,104	499	1,653	1,204	309	
ONS RDA2	Mid-2016	8,770	5,094	499	1,661	1,206	310	
EGPPR/HES	19-10-2020	10,215	5,148	389	1,926	1,160	669	923
EGPPR/HES	28-06-2021	11,352	5,882	454	2,190	1,293	756	777
GLA 2016 based	Mid-2016	8,799	5,048	484	1,750	1,179	338	

Throughout this report figures are presented in terms of the five broad ethnic groups, with the Chinese individual group included in the Asian broad group.

Table 1 summarises population outputs for London by five broad ethnic groups. For a given year there is generally good agreement in the total population between different sources, but more variation within each ethnic group. An exception is that EGPPR/HES total counts are noticeably higher, even allowing for the different time periods. This is because the EGPPR/HES outputs are not population counts and their quality is not yet mature, as explained in Section 2.4.

It is therefore more appropriate to consider the summarised outputs in terms of percentage distribution of each ethnic group (as in Table 2), and also to compare against the 2011 Census as a benchmark. The EGPPR/HES outputs indicate a much lower percentage of the White group than the other sources, particularly the 2011 Census. This may be influenced by the significant percentage of Unknown ethnic group, but these outputs are not yet mature enough to draw insight from.

Table 2 Composition Summary of Ethnic Population Outputs for London by Five Broad Ethnic Groups

Figures are percentages of total population

Source	Reference Date	Ethnic Group Percentage Composition					
		White	Mixed	Asian	Black	Other	Unknown
2011 Census	27-03-2011	59.8	5.0	18.5	13.3	3.4	
ONS APS PEEG	Mid-2015	59.0	3.7	17.9	13.3	6.1	
ONS GSPREE 1	Mid-2015	62.4	3.0	15.5	13.0	6.0	
ONS GSPREE 2	Mid-2015	61.4	3.2	16.7	12.9	5.8	
ONS APS 1yr	Mid-2016	58.9	4.2	18.5	12.4	6.0	
ONS APS 3yr	Mid-2016	59.4	3.7	18.4	12.5	6.1	
ONS RDA1	Mid-2016	58.2	5.7	18.8	13.7	3.5	
ONS RDA2	Mid-2016	58.1	5.7	18.9	13.8	3.5	
EGPPR/HES	19-10-2020	50.4	3.8	18.9	11.4	6.5	9.0
EGPPR/HES	28-06-2021	51.8	4.0	19.3	11.4	6.7	6.8
GLA 2016 based	Mid-2016	57.4	5.5	19.9	13.4	3.8	

In the following comparisons of ethnic distributions relative to the benchmark 2011 Census it is intended to illustrate the variation in uncertainty arising from the different estimation or projection methodologies, and not to suggest that one source is "better" than another. The fact that the estimates or projections are for a different point in time from the census implies that variation in uncertainty may be expected.

Regarding the population-specific sources, the ONS GSPREE methods produce a higher percentage for the White group than their APS and RDA methods, and higher than the benchmark 2011 Census. Whereas the GLA projections produce the lowest percentage for the White group, and lower than the Census.

For the Mixed group the ONS APS and GSPREE percentages are lower than the Census, whereas the ONS RDA and GLA percentages are higher.

For the Asian group the ONS APS and RDA percentages are similar compared with the Census, whereas the ONS GSPREE percentages are lower and the GLA percentage is higher.

Percentages are the most consistent between sources for the Black group. The ONS APS 1yr and 3yr percentages are slightly lower than the Census, but ONS GSPREE and RDA and GLA percentages are very similar to the Census.

To some extent, the pattern for the Other group is a reverse of the Mixed group. The ONS APS and GSPREE percentages are noticeably higher than the Census, whereas the ONS RDA and GLA percentages are similar to the Census.

It should be noted that this comparison summary is for London and five broad ethnic groups. For individual ethnic groups (where data is available) or for England, different regions or lower geographies the variation in populations and percentage compositions between sources may be more marked.

1. Introduction

This document outlines in some detail the major sources of ethnic group information currently available in England at the national and regional level. It has been produced to assist users of ethnicity information to understand their limitations, given that no up-to-date official statistics on ethnic populations are available. There are a number of sources for estimates of the size and distribution of ethnic group populations. These estimates vary in quality, accuracy, timeliness, and detail; in some cases, the underlying definition of what constitutes the resident population is different. It also gives a brief summary of the estimates themselves.

For analyses relating to the London resident population it is recommended that the GLA 2016-based Housing-led variant ethnic projections are used.

1.1 ONS (Office for National Statistics)

In England and Wales there are no fully ethnic-demographic equivalent of mid-year estimates (with breakdown by ethnicity, sex, and age) that are accredited as National Statistics.

In the period 2006 to 2011 ONS published Population Estimates by Ethnic Group (PEEG) for mid-years 2001 to 2009 at Local Authority and District level, using 2001 Census data (adjusted to mid-year) for the base year ethnic populations. These were designated as Experimental Official Statistics. PEEG for 2010 were also produced but not published. This series was discontinued when comparison with Annual Population Survey (APS) ethnicity data and 2011 Census ethnicity data raised concerns about the quality of the estimates. No replacement has been produced to date, although ONS have published periodic research outputs based on survey and administrative data. ONS have also published ad-hoc aged-on ethnic estimates for mid-years 2011 to 2018 (regions) and 2011 to 2019 (local authorities). None of these are accredited as National Statistics.

Therefore, the only ethnic population data sources that are accredited as National Statistics are censuses. The 1991 Census was the first in which an ethnicity question was included, and this continued in the 2001 and 2011 Censuses, and the recent 2021 Census. They are considered to be the most robust population data available but become more out-date over time until the next census. A census population is the population on Census Day, in England typically towards the end of March or April, whereas a mid-year population estimate (or projection) is the population on the United Nations standard reference date of 30th June.

1.2 GLA (Greater London Authority)

The Greater London Authority (GLA) have produced Ethnic Population Projections for London Boroughs since the late 1990s. These projections incorporate the latest census data when it becomes available. The latest ethnic projections are the 2016-based, published in 2017, and incorporate 2011 Census information. Thereafter, the annual ethnic projection updates have been suspended due to uncertainty about the assumptions on ethnic migration and the consequences of the 2016 Referendum. Additional uncertainty has resulted from COVID-19 Pandemic travel restrictions.

1.3 University of Leeds

The University of Leeds have produced ethnic population projections under two ETHPOP research projects funded by ESRC. The first project was *Ethnic group population trends and projections for UK local areas.* The projections were based on 2001 Census ethnic populations and were published in 2011.

The second project was *NewETHPOP - Evaluation, Revision and Extension of Ethnic Population Projections.* This re-based the projections to 2011 Census ethnic populations and publication was in 2016.

1.4   NHS Digital [MI] Ethnic Category Coverage (EGPPR/HES)

NHS Digital have recently created a Management Information ethnic category asset that combines data from General Practice (a specific dataset called GDPPR) and Hospital Episode Statistics (HES). Combining these sources increases coverage. This Management Information (MI) asset aims to inform users of the quality, coverage, and distribution of ethnic category recording within this asset. The asset is intended to provide information for coronavirus (COVID-19) planning and research.

1.5   Summary Tables

Tables summarising ethnic populations for England and London from the ONS and GLA sources are given in the Appendix.

Table 3 lists the sixteen ethnic groups used in the 2001 Census and the eighteen ethnic groups used in the 2011 Census, to show how they align at the individual ethnic group level and at the level of the five broad ethnic groups.

The White Gypsy or Irish Traveller group and the Arab group were not defined separately in the 2001 Census. The Chinese group was part of the Other broad group in the 2001 Census, but part of the Asian broad group in the 2011 Census.

Throughout this report figures are presented in terms of the five broad ethnic groups, with the Chinese individual group included in the Asian broad group.

Variations between sources in ethnic populations and their compositions (percentages) are generally greater for London (and other regions) than for England and are likely to be greater still at local authority level.

2.   Description and Location of each Data Source

An Excel summary table of the sources and output definitions described below accompanies this report.

2.1   ONS

2.1.1   Census Ethnicity Data

The 1991 England and Wales Census ethnic question presented nine ethnic groups, and outputs were in terms of ten ethnic groups, with an additional group (Other Asian) being derived from the write-in responses. In the 2001 Census the ethnic question presented sixteen ethnic groups, and the 2011 Census presented eighteen ethnic groups. Outputs were published in terms of the corresponding ethnic groups. In the 2021 Census the ethnicity question presented nineteen ethnic groups.

Census questionnaires pre-2021:
https://census.ukdataservice.ac.uk/use-data/censuses/forms.aspx

Census questionnaires in 2021:
https://www.ons.gov.uk/census/censustransformationprogramme/questiondevelopment/census2021paperquestionnaires

A variety of standard 1991, 2001, and 2011 census ethnic population outputs (by sex, age, and ethnic group) are available on the NOMIS web portal:
https://www.nomisweb.co.uk/

In addition, for 2001 and 2011 census data, ONS published Commissioned Tables, which covered specific user requests or more complex tabulations. These are published on ONS web resources but are not always easy to find.

For the 2021 Census, ONS are aiming to produce the census first release within the first year after Census Day and all other data within two years of Census Day (a release window of March 2022-March 2023). ONS are working to ensure that the outputs will retain National Statistics accreditation, and data will be available via a suite of web-based data output products:
https://www.ons.gov.uk/census/censustransformationprogramme/census2021outputs

2.1.2   ONS Population Estimates by Ethnic Group (PEEG)

The resource for the discontinued PEEG series is available on the National Archives website:
https://data.gov.uk/dataset/12fef22c-a92c-4813-9e1a-23012266c6e6/population-estimates-by-ethnic-group-experimental

The published PEEG for mid-years 2001 to 2009 at Local Authority and District level were broken down by sex (male, female, persons), broad age group (0-15, 16-64/59, 65/60+), and the sixteen 2001 Census Ethnic Groups. The mid-year 2002 to 2009 estimates were also published for the Primary Care Organisation area classification.

25 July 2013

Population Estimates by Ethnic Group (experimental), Comparison of mid-2010 Population Estimates by Ethnic Group against the 2011 Census:

https://webarchive.nationalarchives.gov.uk/20151014044651/http://www.ons.gov.uk/ons/rel/peeg/population-estimates-by-ethnic-group--experimental-/comparison-of-mid-2010-population-estimates-by-ethnic-group-against-the-2011-census/index.html

This report compares the unpublished PEEG for mid-year 2010 with 2011 Census ethnic data.

March 2014

Population estimates by ethnic group (PEEGs) - external review:

https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/populationestimatesbyethnicgrouppeegsexternalreview/2014-03-11

In December 2013, ONS commissioned the University of Manchester to conduct a review of the PEEGs. The review was produced in March 2014.

25 August 2017

Population estimates by ethnic group

New approaches for producing population estimates by ethnic group report:

https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/populationestimatesbyethnicgroup

This resource describes two alternative ONS methods considered for producing population estimates by ethnic group. These were:

1   APS-based (Annual Population Survey) population estimates by ethnic group (a "simple" method).

2   A method for estimating ethnicity from survey and administrative data, using the Generalised Structure Preserving Estimator (GSPREE) approach.

 A useful detailed summary of the main differences between these two methods is included.

More information on the APS and GSPREE research and outputs is given in the following sections.

2.1.3   ONS APS Ethnic Estimates

25 August 2017

in August 2017 ONS published a Research report on population estimates by characteristics:

https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/researchreportonpopulationestimatesbycharacteristics

This report presented a simple method for using the Annual Population Survey (APS) to produce separate population estimates by country of birth, by nationality and by ethnic group. APS data was used in combination with 2011 Census data and Mid-Year Estimates. The report was published to invite views on the method and the usefulness of the estimates. The ethnic outputs are counts of persons in the UK by countries and local authorities and higher geographies in England and Wales, as at June 2016. They are in terms of six ethnic groups, and the counts are rounded to the nearest thousand.

6 November 2017

Alternative 2015 population estimates by ethnic group produced from the Annual Population Survey (APS PEEG) were published alongside GSPREE 1 and 2 ethnic outputs. All of these were only by local

authority and district and ethnic group (six ethnic groups for GSPREE outputs, and five ethnic groups for APS outputs). See the 6 November 2017 details in the GSPREE section below.

January 2018
On 19th January 2018 ONS published an ad hoc request based only on APS data and LFS data, of ethnicity and religion outputs:
https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/labourproductivity/adhocs/00799 75tablesproducedusingtheannualpopulationsurveyandthelabourforcesurveyethnicityandreligionbroken downbycountry
The outputs are (unrounded) counts of persons by ethnic group for each UK country, for APS years Jan-Dec 2007 to Jan-Dec 2016. For Jan-Dec 2007 to Jan-Dec 2010 and UK countries excluding Northern Ireland the outputs are in terms of fifteen ethnic groups. For Jan-Dec 2007 to Jan-Dec 2010 and UK countries including Northern Ireland the outputs are in terms of six ethnic groups. For Jan-Dec 2011 to Jan-Dec 2016 and all UK countries the outputs are in terms of nine ethnic groups.

December 2019
On 4 December 2019 ONS published a Research report on population estimates by ethnic group and religion:
https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimate s/articles/researchreportonpopulationestimatesbyethnicgroupandreligion/2019-12-04
This report presented a next step in the simple method for producing population estimates by ethnic group and by religion. As for the August 2017 publication, APS data was used in combination with 2011 Census data and Mid-Year Estimates, but in this case a three-year APS pooled dataset covering January 2014 to December 2016 was used instead of a single-year APS dataset. As in the August 2017 publication, the ethnic outputs are illustrative estimates of persons in the UK by countries and local authorities and higher geographies in England and Wales, as at June 2016. They are in terms of six ethnic groups, and the counts are rounded to the nearest thousand:
https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimate s/datasets/populationcharacteristicsresearchtables

2.1.4 ONS GSPREE Ethnic Estimates

In March 2014, the National Statistician made a recommendation that the census in 2021 should be predominantly online, making increased use of administrative data and surveys to enhance the statistics from the 2021 Census. This recommendation was endorsed by the Government's formal response, which also set out its ambition that "censuses after 2021 be conducted using other sources of data… sufficiently validating the perceived feasibility of that approach".

Consequently, ONS is aiming to produce the type of information that is collected by a ten-yearly census from an Administrative Data Census. The approach to date has been application of Generalised Structure Preserving Estimation (GSPREE) to produce population estimates by local authority in England for various variables such as ethnic group or housing tenure. The GSPREE approach combines more timely survey estimates with more detailed, but outdated, census distributions and also timely information available in administrative sources for subsections (but not all) of the population. The GSPREE procedures improve the quality of estimates for cells in a contingency table.

Explaining the Generalised Structure Preserving Estimator (GSPREE)
https://www.ons.gov.uk/peoplepopulationandcommunity/housing/articles/researchoutputssubnational
dwellingstockbytenureestimatesengland2012to2015/2018

The Generalised Structure Preserving Estimator (GSPREE) method can be used to estimate a cross tabulation of the population between two or more categorical variables, for example local authority and ethnic group or local authority and housing tenure. Estimates are produced by combining information from different sources including administrative, survey and census data and drawing on the strengths of each. The method provides a framework for estimating population characteristics for specific geographies in non-census years, where such comprehensive data are not available.

As a minimum for GSPREE, there must be two sources of input data: one from a survey and the other from a proxy source. The GSPREE methodology treats the ethnic group or housing tenure category definitions in the survey data as the target definitions, even if the proxy data are defined slightly differently. A proxy data source contains detailed information for the same cross-tabulated set of areas and categories of the target table, but it does not exactly measure it. This could be because of a number of factors such as differing time periods, differences in the target population and/or inconsistent variable category definitions.

The GSPREE method obtains the best possible estimates by combining the survey and proxy data through a statistical model. Having obtained estimates of the cross tabulation, the method then benchmarks the estimates to row and column margins, using iterative proportional fitting. It is vital that these row and column margins are of high quality and have undergone a rigorous quality assurance process, such as National Statistics that have been certified as compliant with the Code of Practice for Statistics. Benchmarking estimates ensures that they are correctly scaled to represent the population and are consistent across different geographic levels.

There may be several sources of information relating to the target categorical variable (e.g., ethnic group or housing tenure), but no one source alone provides the exact information required.

The following are a brief summary of GSPREE ethnic population research and outputs published by ONS.

November 2016
In November 2016 ONS published the report GSS Methodology Series No 42, Assessing the Generalised Structure Preserving Estimator (GSPREE) for Local Authority Population Estimates by Ethnic Group in England:
https://www.ons.gov.uk/file?uri=/methodology/methodologicalpublications/generalmethodology/curre
ntmethodologyarticles/assessingthegeneralisedstructurepreservingestimatorgspreeforlocalauthoritypop
ulationestimatesbyethnicgroupinengland.pdf
This report describes the GSPREE methodology in detail and assess the experimental outputs, including:

Application: 2014 Population Estimates by Local Authority and Ethnic Group in England

Validation Study: Comparing 2011 Census to 2011 GSPREE Population Estimates by Local Authority and Ethnic Group in England

The data sources used for the GSPREE approach were the 2011 Census, the 2014 English School Census (ESC), the 2014 Annual Population Survey (APS), and the 2014 Mid-Year Estimates. Three GSPREE

models were developed, Model 1, Model 2, and Model 3 each in turn implementing more elements such as age breakdown from the data sources to improve the robustness of the ethnic estimates.

Other than the comparative ethnic population data presented as tables and charts in this report, no ethnic populations were published.

25 August 2017
In August 2017 ONS published a web resource summarising GSPREE Models 1 and 2 for 2011 comparison:
Research Outputs: An approach for estimating ethnicity from survey and administrative data, 2011.
The research investigates the use of the Generalised Structure Preserving Estimator and survey data to enhance administrative data ethnicity estimates at local authority level in England, 2011. These Research Outputs are not official statistics:
https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusresearchoutputs/populationcharacteristics/researchoutputsanapproachforestimatingethnicityfromsurveyandadministrativedata2011
This methodology used the 2010/11 Annual Population Survey (APS), the 2011 English Schools Census (ESC), and the 2011 Census as inputs to the models. No GSPREE output tables were published.

6 November 2017
Research Outputs: Ethnicity estimates from survey and administrative data, 2015.
Research to investigate the ability to produce estimates of ethnicity at local authority level for 2015 using the Generalised Structure Preserving Estimator approach:
https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusresearchoutputs/populationcharacteristics/researchoutputsethnicityestimatesfromsurveyandadministrativedata2015
The data sources used for the GSPREE approach in this publication were the 2011 Census, the 2015 English School Census (ESC), the 2015 Annual Population Survey (APS), and the 2015 Statistical Population Dataset V2.0 (SPD V2.0). The SPD V2.0 is an experimental non-ethnic mid-year estimate dataset generated only from administrative sources, intended as a future method of producing mid-year estimates without requiring input from a preceding census, and in fact to replace the decennial questionnaire-based census.

For GSPREE Model 1 (APS, ESC, and SPD V2.0) and Model 2 (APS, ESC, 2011 Census, and SPD V2.0) using these input datasets, experimental ethnic population outputs for 2015 by local authority and district and six ethnic groups were published (no sex or age breakdown).

This publication also included outputs of alternative 2015 population estimates by ethnic group produced from the Annual Population Survey (APS PEEG), for comparison with the GSPREE 1 and 2 outputs. These were by local authority and district and five ethnic groups.

2.1.5   ONS Admin-based ethnicity statistics (ABPE)

6 August 2021
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.
https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/articles/adminbasedethnicitystatisticsforenglandfeasibilityresearch/2016

Previous research (described in Sections 2.1.3 and 2.1.4 above) 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 were produced.

On 6 August 2021 ONS also published information on:
Producing admin-based ethnicity statistics for England: methods, data and quality
https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/articles/producingadminbasedethnicitystatisticsforenglandmethodsdataandquality/2021-08-06

Since the previous work by ONS using the Statistical Population Dataset V2.0 (SPD V2.0) as the base population for GSPREE ethnicity estimates the SPD has been superseded by the Admin-based population estimates (ABPE) V3.0 dataset.

The SPD V2.0 was created using these admin sources:

• the NHS Patient Register (PR)

• the Department for Work and Pensions (DWP) Customer Information System (CIS)

• data from the Higher Education Statistics Agency (HESA)

• School Census data

The ABPE was created by combining additional administrative data sources compared to those used for used for the SPD, including:

• Benefits and Income Datasets

• NHS Patient Register and Personal Demographic Service

• Higher Education Statistics Agency

• English School Census

• Welsh School Census

• Births Registrations

To produce the admin-based ethnicity statistics, the ABPE dataset was used as the population base and ethnicity data was linked on from the Hospital Episode Statistics (HES), English School Census (ESC) and Improving Access to Psychological Therapies (IAPT) administrative data sources.

Although Admin-based ethnicity statistics for England at the national, regional and local authority level for five and eighteen ethnic groups were produced, these are summarised only on the web resource. The only figures that appear to have been published are England statistics for mid-year 2016 as percentages by five ethnic groups and 5-year age band to 85+, and local authority statistics for mid-year 2016 as percentages by five ethnic groups only. The percentages were calculated out of those with a stated ethnicity in the linked information.

The latest information on the ABPE V3.0 base populations was published by ONS on 21 June 2019:

Developing our approach for producing admin-based population estimates, England and Wales: 2011 and 2016.
Research into developing a new methodology to create population estimates from administrative data. These estimates are not official statistics on the population:
https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/developingourapproachforproducingadminbasedpopulationestimatesenglandandwales2011and2016/2019-06-21

2.1.6 ONS RDA (Racial Disparity Audit) Aged-on Ethnic Estimates

In 2018 ONS published two datasets of ethnic populations for use as denominators in the Race Disparity Audit. These denominators are neither National Statistics nor standard published experimental statistics and were not produced using methods which have undergone formal Quality Assurance. The method was essentially a process of aging-on the distributions of ethnic group populations in the 2011 Census by one year for each successive mid-year.

Dataset ONS RDA 1
These were initially published on 27 July 2018 and were for mid-years 2011 to 2017. Mid-year 2018 was added on 6 September 2019.
Population denominators by ethnic group, regions and countries: England and Wales, 2011 to 2018:
https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/adhocs/008780populationdenominatorsbyethnicgroupregionsandcountriesenglandandwales2011to2017
The breakdown was by the eighteen 2011 Census ethnic groups, for regions in England, England as a whole, Wales, and England and Wales as a whole, and by single year of age and sex.

Dataset ONS RDA 2
These were initially published on 27 July 2018 and were also for mid-years 2011 to 2017. Mid-year 2018 was added on 4 August 2020 and mid-year 2019 was added on 26 August 2020.
Population denominators by broad ethnic group and for White British, local authorities in England and Wales: 2011 to 2019:
https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/adhocs/008781populationdenominatorsbybroadethnicgroupandforwhitebritishlocalauthoritiesinenglandandwales2011to2017

The breakdown was by local authority in England and Wales, the standard five broad ethnic group aggregation (White, Mixed, Asian, Black, Other) with an additional split for White British, and by single year of age and sex.

The method used by ONS for ethnic populations for the Race Disparity Audit is as follows:

1. The distribution by ethnic group (by LA, sex and age) is taken from the 2011 Census. This distribution is taken to apply at mid-2011.

2. Estimates for 2011 were produced by applying the distribution calculated in 1. to the mid-year population estimate for 2011.

To derive estimates for the next year:
3. The Census distribution was 'aged on' by one year - so the distribution relating to 0-year olds in 2011 related to 1-year olds in 2012. It was assumed that the ethnic distribution for cohorts not born in 2011 was the same as the distribution of 0-year olds in 2011. Cohorts aged 89 were absorbed in the 90+ group after ageing on.

4. The 'aged-on' distribution was applied to the mid-year population estimate for 2012.

3 and 4 were repeated to produce estimates for 2013-2018/2019.

The overall populations are wholly consistent with mid-year estimates. However, the RDA aged-on ethnic estimates may not reflect dynamic effects due to ethnic fertility, mortality, and migration parameters, or their changes over time. Since higher migration rates typically occur in younger ages the aged-on estimates for these ages may be less robust than those for older ages.

2.2   GLA Ethnic Group Population Projections

GLA ethnic populations have only been produced at local authority level for London boroughs. Along with non-ethnic GLA population projections and household projections, they are freely available from the Demography homepage on the GLA London Datastore:
https://data.london.gov.uk/demography/

Up to 2012 GLA ethnic projections were based on ethnic populations and ethnic migration patterns derived from 2001 Census data. They were published for the period covering 2001 to 2031. The ethnic breakdown was in terms of ten ethnic groups aggregated from the sixteen 2001 Census ethnic categories, and further breakdown was by sex and five-fear age band.

From 2013, once 2011 Census data became available these were incorporated into the GLA ethnic projections. These have been published for the period covering 2011 to 2051. The ethnic breakdown is in terms of seventeen ethnic groups aggregated from the eighteen 2011 Census ethnic groups , and further breakdown is by sex and single-year-of-age.

The last GLA ethnic projections were published in 2017 and are referred to as GLA 2016-based. Normally GLA ethnic projections are updated annually using the latest GLA main projection non-ethnic outputs as constraints, but this is on hold due to uncertainty about how patterns of migration by ethnic group have changed since 2011 as a consequence of the 2016 Referendum result and the COVID-19 pandemic. Therefore, the GLA 2016-based ethnic projections are not consistent with subsequent GLA non-ethnic projections, although the overall populations are likely to be similar.

Since 2016, ONS long-term international migration data indicates that net migration from EU countries has declined substantially, whereas net migration from other parts of the world (e.g., India) has increased. This means that any ethnic projection model using ethnic parameters based on the 2011 census is likely to become less representative as time moves further beyond 2016. Uncertainty in ethnic migration patterns has also increased as a result of COVID-19 pandemic travel restrictions.

The model incorporates ethnic-specific parameters for fertility and migration, though mortality rates are common across ethnic groups. Ethnic-specific fertility rates are based on Hospital Episode Statistics (HES) analysis of delivery episodes. Base ethnic populations and ethnic migration patterns are based on 2011 census data.

The census indicates that an infant's (child under 1 year old) recorded ethnicity can be different from that of the father and/or mother. This feature of transgenerational ethnicity of births is accounted for in GLA ethnic projections alongside the HES-based ethnic fertility rates.

In England and Wales ethnicity has never been collected on Death Certificates. Due to the resultant lack of information regarding ethnic mortality, this is not incorporated in the GLA ethnic projection model.

Summary of variants

There are two variants of the GLA 2016-based Ethnic Group Population Projections:

1   2016-based Central Trend
    This variant is constrained to the main GLA 2016-based central trend borough projection, which estimates the future trajectory of London's population at local authority (borough) level. The main

central trend projection assumes that future population growth will be accommodated by natural growth in dwellings wherever necessary.

2   2016-based Housing-led
    This variant is constrained to the main GLA 2016-based housing-led borough projection. The GLA housing-led model attempts to reconcile two competing trends that influence the density at which housing is occupied, these being the changing demographic profile of the population and local authority constraints in housing availability. It then constrains the sum of resulting local authority-level populations and components to be consistent with the equivalent trend-based projection for the region as a whole. Therefore, at London level, ethnic and non-ethnic 2016-based projections are the same but differ at borough level.

2.3   University of Leeds ETHPOP Ethnic Group Projections 2008, 2016

The University of Leeds ETHPOP research project is described on a dedicated website. Outputs from the updated NewETHPOP 2011 census-based work can be accessed there (registration is required): https://www.ethpop.org/

Both the ETHPOP and NewETHPOP outputs were produced for Local Authorities and Districts in England, Unitary Authorities in Wales, Councils in Scotland, and District Councils in Northern Ireland. The first projections were based on the 2001 Census ethnic populations and parameters and were published in 2010. The projections were updated to 2011 Census based ethnic populations and parameters and were published in 2016. The ethnic breakdown was in terms of a consistent twelve ethnic groups aggregated from the slightly different census ethnic classifications used in the constituent countries of the United Kingdom. Further breakdown was by sex and single-year-of-age.

Outputs from the ETHPOP 2001 census-based projections were for mid-years 2001 to 2051, and those from the NewETHPOP 2011 census-based projections were for mid-years 2011 to 2061.

Summary of variants

2.3.1   ETHPOP 2001 census-based

October 2010

(a) Uptap ER/EF: Emigration rates (ER), Emigration flow (EF), mortality improvement 2% per annum

(b) Trend EF: Aligned to the 2008-based national population projections

(c) Benchmark ER/EF: Constant components from 2001 onwards

28.9.2011
Major update on all projections by improving assumptions on internal migration, mortality in older ages and fertility schedules in the youngest and oldest ages for women in the fertile age range.

Two more projections available one assuming constant demographic intensities since 2001 (BENCH) and another excluding international migration and simulating natural increase.

2.3.2   NewETHPOP 2011 census-based

August 2016
Updated the ETHPOP database with new 2011 based population projections.

There are 3 projections available:

(d) LEEDS 1 projection is based on a BREXIT scenario

(e) LEEDS 2 is aligned to ONS assumptions.

(f) LEEDS 2b is constrained to ONS population projections (2014-based)

ETHPOP 2001 census-based Methodology:
A Working Paper published in July 2010 gives a detailed description of the ETHPOP 2001 census-based methodology:
ETHNIC POPULATION PROJECTIONS FOR THE UK AND LOCAL AREAS, 2001-2051
http://doc.ukdataservice.ac.uk/doc/6777/mrdoc/pdf/6777_workingpaper.pdf

(or
https://www.researchgate.net/publication/225337613_Ethnic_population_projections_for_the_UK_2001-2051)

A brief summary is as follows.

All base populations and basic demographic parameters (fertility, mortality, and migration) are incorporated on an ethnic basis, using census data. All variants are not aligned with ONS mid-year estimates.

Although all demographic parameters are incorporated on an ethnic basis, the census data used is often modelled proxy data as no direct ethnic information relating to fertility or mortality was collected in the census.

Ethnic fertility was estimated using census data, birth statistics and survey tables.

For ethnic mortality: census data on limiting long-term illness was used to predict ethnic mortality levels and regression equations between local Standardized Illness and Mortality Ratios for all ethnicities were assumed to apply to individual ethnic groups.

Estimates of internal migration by ethnic group were constructed from census and register migration data.

A databank of administrative variables related to international migration for local areas was constructed in order to develop estimates for long-term international immigration and emigration.

NewETHPOP 2011 census-based methodology:
A paper published in 2016 gives some detailed description of the NewETHPOP 2011 census-based methodology:
THE FUTURE IS DIVERSITY: NEW FORECASTS FOR THE UK'S ETHNIC GROUPS
Paper for the European Population Conference, 31 August to 3 September, 2016, Mainz, Germany:
https://epc2016.princeton.edu/papers/161123

The NewETHPOP methodology is similar to the original ETHPOP methodology, with some refinements. An important one is a change in modelling ethnic mortality. In the original ETHPOP work two ways of estimating ethnic mortality were developed: a method based on the relationship between limiting long-term illness (LLTI) and mortality (Standardized Illness and Mortality Ratios), and a method that used Local Authority mortality rates weighted by the different geographical profiles of ethnic groups (Geographically Distributed Method). Whereas ETHPOP used the Standardized Illness and Mortality Ratios method, subsequent research showed that LLTI and mortality were only partially correlated, therefore NewETHPOP used the Geographically Distributed Method.

2.4   NHS Digital [MI] Ethnic Category Coverage (EGPPR/HES)

NHS Digital have recently created a Management Information (MI) ethnic category asset that combines data from General Practice (GDPPR) and Hospital Episode Statistics (HES) datasets. Combining these sources increases coverage. This information aims to inform users of the quality, coverage, and distribution of ethnic category recording within this asset, and to provide a near population (England only) level view of ethnic category.

The main purpose of all data or information collection in the NHS is to help managers deliver health services. In this context data or information collections are regarded as management information assets. In the context of Information Governance data or information collections are required to be recorded on asset registers. Hence the published EGPPR/HES ethnicity coverage outputs are referred to as a Management Information (MI) asset.

HES (Hospital Episode Statistics) is a database of all publicly funded hospital activity in England.

GDPPR (GPES Data for Pandemic Planning and Research) is a specific dataset extracted by the General Practice Extraction Service (GPES) into NHS Digital, and is intended to provide information for coronavirus (COVID-19) planning and research.

It should be noted that the EGPPR/HES outputs are not population counts. They are counts of ethnic group of persons as currently identified from the two datasets. General Practice List Sizes of their patients are typically higher than comparable ONS Mid-Year Estimates. It is also likely that there is duplication of persons between the two datasets. Therefore, the EGPPR/HES total counts published to date are noticeably higher than comparable ONS Mid-Year Estimates or comparable General Practice List Sizes routinely published by NHS Digital. As the EGPPR/HES outputs and their quality mature over time one would expect the total counts for England and CCGs to resemble those of routinely published General Practice List Sizes.

The published ethnicity coverage datasets should therefore be used as "supporting information" on the quality and maturity of ethnicity coverage in the primary and secondary care systems of the NHS rather than definitive population counts. Interpretation of the information in the datasets should take into consideration the inflated nature of the figures compared to GP List Sizes or demographic population estimates.

Summary counts of patients by ethnic category are available here:
https://digital.nhs.uk/data-and-information/publications/statistical/mi-ethnic-category-coverage/current

19 November 2020
The first publication of ethnic count information was on 19 November 2020, for the Reporting Period end date 19 October 2020. The GPES extract is run on a fortnightly basis, and a file incorporating ethnic counts for the latest reporting period is added to the published resource, also at fortnightly intervals.

The counts are of patients by ethnic category for Clinical Commissioning Groups (CCGs) of registration in England. Eighteen ethnic groups (as in the 2011 Census) are used, plus an Unknown group.

Only patients who have interacted with GP practices and/or hospitals in England are included within this MI asset.

Useful link on GDPPR

https://digital.nhs.uk/coronavirus/gpes-data-for-pandemic-planning-and-research/guide-for-analysts-and-users-of-the-data

APPENDIX – SUMMARY TABLES

Table 3
Grouping of the Eighteen 2011 Census Ethnic Groups and the Sixteen 2001 Census Ethnic Groups into the Five Broad Ethnic Groups (England and Wales)

2011 Census Ethnic Group (18)	2001 Census Ethnic Group (16)	Broad Ethnic Group
White British	White British	
White Irish	White Irish	
White Gypsy or Irish Traveller[1]		White
Other White	Other White	
White and Black Caribbean	White and Black Caribbean	
White and Black African	White and Black African	
White and Asian	White and Asian	Mixed
Other Mixed	Other Mixed	
Indian	Indian	
Pakistani	Pakistani	
Bangladeshi	Bangladeshi	Asian
Chinese[2]		
Other Asian	Other Asian	
Black African	Black African	
Black Caribbean	Black Caribbean	Black
Other Black	Other Black	
Arab[1]		
	Chinese[2]	Other
Any other ethnic group	Any other ethnic group	

[1] *The White Gypsy or Irish Traveller group and the Arab group were not defined separately in the 2001 Census.*
[2] *The Chinese group was part of the Other broad group in the 2001 Census, but part of the Asian broad group in the 2011 Census.*

Table 4
2011 Census Ethnic Populations for England and London
by Census Ethnic Group and Broad Ethnic Group
(These ethnic populations are National Statistics)

2011 Census Ethnic Group	England Population	%	London Population	%
All usual residents	53,012,456	100.0	8,173,941	100.0
White British	42,279,236	79.8	3,669,284	44.9
White Irish	517,001	1.0	175,974	2.2
White Gypsy or Irish Traveller	54,895	0.1	8,196	0.1
Other White	2,430,010	4.6	1,033,981	12.6
White and Black Caribbean	415,616	0.8	119,425	1.5
White and Black African	161,550	0.3	65,479	0.8
White and Asian	332,708	0.6	101,500	1.2
Other Mixed	283,005	0.5	118,875	1.5
Indian	1,395,702	2.6	542,857	6.6
Pakistani	1,112,282	2.1	223,797	2.7
Bangladeshi	436,514	0.8	222,127	2.7
Chinese	379,503	0.7	124,250	1.5
Other Asian	819,402	1.5	398,515	4.9
Black African	977,741	1.8	573,931	7.0
Black Caribbean	591,016	1.1	344,597	4.2
Other Black	277,857	0.5	170,112	2.1
Arab	220,985	0.4	106,020	1.3
Any other ethnic group	327,433	0.6	175,021	2.1

Broad Ethnic Group				
White	45,281,142	85.4	4,887,435	59.8
Mixed	1,192,879	2.3	405,279	5.0
Asian	4,143,403	7.8	1,511,546	18.5
Black	1,846,614	3.5	1,088,640	13.3
Other	548,418	1.0	281,041	3.4

Table 5a
Summary of ONS Ethnic Population Outputs for England by the Five Broad Ethnic Groups
ONS Census figures are National Statistics
Remaining ONS figures are Experimental or Illustrative Statistics

| England | Reference | | Ethnic Group Population | | | | | | | |
Source	Date	Population	White	Mixed	Asian	Black	Other	Unknown	White British	All Other White
2011 Census	27-03-2011	53,012,456	45,281,142	1,192,879	4,143,403	1,846,614	548,418		42,279,236	3,001,906
ONS APS PEEG	Mid-2015	54,786,336	46,627,496	924,716	4,240,731	1,986,886	1,006,507			
ONS GSPREE 1	Mid-2015	55,081,017	47,609,893	792,584	3,887,640	1,835,003	955,897			
ONS GSPREE 2	Mid-2015	55,081,028	47,609,900	792,584	3,887,637	1,835,010	955,897			
ONS APS 1yr	Mid-2016	55,268,000	46,728,000	1,001,000	4,490,000	1,999,000	1,050,000		43,066,000	3,662,000
ONS APS 3yr	Mid-2016	55,268,000	46,926,000	968,000	4,396,000	1,946,000	1,032,000		43,519,000	3,407,000
ONS RDA1	Mid-2016	55,268,067	46,622,623	1,452,626	4,548,822	2,038,236	605,761		43,443,664	3,178,958
ONS RDA2	Mid-2016	55,268,067	46,629,001	1,457,579	4,536,290	2,040,165	605,032		43,472,792	3,156,209
EGPPR/HES	19-10-2020	59,374,890	42,813,470	1,188,090	5,328,610	2,087,045	1,478,955	6,478,720	37,600,255	5,213,215
EGPPR/HES	28-06-2021	64,382,370	48,170,300	1,389,115	6,094,430	2,425,810	1,682,290	4,620,425	41,682,035	6,488,265

Table 5b
Composition Summary of ONS Ethnic Population Outputs for England by the Five Broad Ethnic Groups
ONS Census data are National Statistics
Remaining ONS data are Experimental or Illustrative Statistics
Figures are percentages of total population

England	Reference	Ethnic Group Percentage Composition							
Source	**Date**	**White**	**Mixed**	**Asian**	**Black**	**Other**	**Unknown**	White British	All Other White
2011 Census	27-03-2011	85.4	2.3	7.8	3.5	1.0		79.8	5.7
ONS APS PEEG	Mid-2015	85.1	1.7	7.7	3.6	1.8			
ONS GSPREE 1	Mid-2015	86.4	1.4	7.1	3.3	1.7			
ONS GSPREE 2	Mid-2015	86.4	1.4	7.1	3.3	1.7			
ONS APS 1yr	Mid-2016	84.5	1.8	8.1	3.6	1.9		77.9	6.6
ONS APS 3yr	Mid-2016	84.9	1.8	8.0	3.5	1.9		78.7	6.2
ONS RDA1	Mid-2016	84.4	2.6	8.2	3.7	1.1		78.6	5.8
ONS RDA2	Mid-2016	84.4	2.6	8.2	3.7	1.1		78.7	5.7
EGPPR/HES	19-10-2020	72.1	2.0	9.0	3.5	2.5	10.9	63.3	8.8
EGPPR/HES	28-06-2021	74.8	2.2	9.5	3.8	2.6	7.2	64.7	10.1

Table 6a
Summary of ONS Ethnic Population Outputs for London by the Five Broad Ethnic Groups
and GLA Outputs for London
ONS Census figures are National Statistics
Remaining ONS figures are Experimental or Illustrative Statistics

London	Reference		Ethnic Group Population							
Source	Date	Population	White	Mixed	Asian	Black	Other	Unknown	White British	All Other White
2011 Census	27-03-2011	8,173,941	4,887,435	405,279	1,511,546	1,088,640	281,041		3,669,284	1,218,151
ONS APS PEEG	Mid-2015	8,673,715	5,116,643	318,381	1,555,873	1,157,355	525,463			
ONS GSPREE 1	Mid-2015	8,586,712	5,358,885	259,540	1,332,414	1,119,397	516,476			
ONS GSPREE 2	Mid-2015	8,586,719	5,271,957	273,093	1,435,134	1,105,442	501,093			
ONS APS 1yr	Mid-2016	8,788,000	5,175,000	367,000	1,625,000	1,092,000	529,000		3,753,000	1,422,000
ONS APS 3yr	Mid-2016	8,769,000	5,206,000	326,000	1,611,000	1,095,000	531,000		3,841,000	1,365,000
ONS RDA1	Mid-2016	8,769,659	5,103,788	499,320	1,653,029	1,204,109	309,413		3,816,985	1,286,803
ONS RDA2	Mid-2016	8,769,659	5,093,548	499,147	1,661,222	1,205,934	309,808		3,816,168	1,277,380
EGPPR/HES	19-10-2020	10,215,120	5,147,860	389,175	1,926,230	1,159,565	669,030	923,260	3,152,280	1,995,580
EGPPR/HES	28-06-2021	11,351,645	5,881,830	453,900	2,189,825	1,292,605	756,230	777,255	3,481,445	2,400,385
GLA 2016 based	Mid-2016	8,798,957	5,048,316	484,481	1,749,765	1,178,743	337,651		3,554,644	1,493,672

Table 6b
Composition Summary of ONS Ethnic Population Outputs for London by the Five Broad Ethnic Groups and GLA Outputs for London
ONS Census data are National Statistics
Remaining ONS data are Experimental or Illustrative Statistics
Figures are percentages of total population

London Source	Reference Date	Ethnic Group Percentage Composition							
		White	Mixed	Asian	Black	Other	Unknown	White British	All Other White
2011 Census	27-03-2011	59.8	5.0	18.5	13.3	3.4		44.9	14.9
ONS APS PEEG	Mid-2015	59.0	3.7	17.9	13.3	6.1			
ONS GSPREE 1	Mid-2015	62.4	3.0	15.5	13.0	6.0			
ONS GSPREE 2	Mid-2015	61.4	3.2	16.7	12.9	5.8			
ONS APS 1yr	Mid-2016	58.9	4.2	18.5	12.4	6.0		42.7	16.2
ONS APS 3yr	Mid-2016	59.4	3.7	18.4	12.5	6.1		43.8	15.6
ONS RDA1	Mid-2016	58.2	5.7	18.8	13.7	3.5		43.5	14.7
ONS RDA2	Mid-2016	58.1	5.7	18.9	13.8	3.5		43.5	14.6
EGPPR/HES	19-10-2020	50.4	3.8	18.9	11.4	6.5	9.0	30.9	19.5
EGPPR/HES	28-06-2021	51.8	4.0	19.3	11.4	6.7	6.8	30.7	21.1
GLA 2016 based	Mid-2016	57.4	5.5	19.9	13.4	3.8		40.4	17.0