Office for National Statistics

Research report on population estimates by characteristics

Office for National Statistics research work on a simple method for producing population estimates by country of birth, nationality and ethnic group.

Contact:

Release date:
25 August 2017

Next release:
To be announced

Notice

14 January 2020

Please note that this report has been superseded by Research report on population estimates by ethnic group and religion.

Table of contents

1 . Summary

This report describes a simple method for using the Annual Population Survey to produce population estimates by country of birth, by nationality and by ethnic group. We are publishing this report to invite your views on the method and on the usefulness of these estimates.

The estimates are similar to the population estimates by country of birth and nationality already published by Office for National Statistics (ONS) but have the advantage of being consistent with the standard mid-year population estimates. Illustrative estimates for 2016 are provided in a dataset accompanying this report.

We are also looking at an alternative method for using administrative sources to produce population estimates by ethnic group. Information on these two different approaches is provided in the population estimates by ethnic group report.

2 . Background

Office for National Statistics (ONS) currently publishes annual estimates of country of birth and nationality at local authority level for England and Wales and at national level for Scotland and Northern Ireland. These estimates are derived from the Annual Population Survey (APS) and primarily relate to that part of the population living in households.

The most up-to-date formal estimates of the population by ethnic group are from the 2011 Census. Annual estimates by ethnic group were produced as Experimental Statistics from 2006 to 2011 but production was discontinued when the estimates appeared to diverge from evidence from other sources.

This work seeks to address the current situation in two ways. Firstly, we adjust the APS estimates by country of birth and nationality so they cover the entire population and are consistent with the standard mid-year population estimates. Secondly, we use a similar method to produce estimates of the population by ethnic group, with a view to better meeting the interest in more up-to-date estimates of the population by ethnic group as the 2011 Census becomes less current.

We have illustrated the use of these methods by providing estimates for 2016. These are provided in a file accompanying this report. As with the currently published APS estimates, we have produced these for countries in the UK and for local authorities in England and Wales.

We are inviting you to review and provide feedback on the proposed method and outputs, and on whether it would be useful for these estimates to be published annually in addition to the currently published estimates by country of birth and nationality.

As part of our drive to make better use of administrative data we are also looking at more complex methods for using administrative sources to produce population estimates by ethnic group. Information on these two different approaches is provided in the population estimates by ethnic group report.

3 . Data sources

The development of the system and the production of the new estimates use the three main data sources as follows.

Annual Population Survey (APS)

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

The APS has the largest sample size of any annual UK household survey and enables the generation of statistics for small geographical areas. Sampling errors are smaller compared to those using other social survey designs, because the APS has a single stage sample of addresses. More information about the quality and methodology of the APS are available from the APS Quality and Methodology Information report.

We have used the APS data for the calendar year January 2016 to December 2016 as the best available approximation for the distributions applying at mid-2016.

2011 Census

The 2011 Census provides estimates of the resident population in households and communal establishments for the UK (England, Wales, Scotland and Northern Ireland).

There is uncertainty around the 2011 Census estimates arising from both sampling and non-sampling errors as described in the 2011 Census Quality and Methodology Information report.

Mid-year population estimates (MYE)

Mid-year population estimates are based on census and are updated annually to account for estimates of population change from 1 July to 30 June. The two main contributors to population change are natural change (births minus deaths) and net migration (the difference between long-term moves into and out of the UK or local areas). The estimates cover the entire usually resident population, whether resident in households or communal establishments.

The mid-year population estimates reflect uncertainty from the 2011 Census base and from sampling and non-sampling errors in estimating the "components of change" – most notably of international and internal migration. More information on this is provided in the Population Estimates Quality and Methodology Information report.

4 . Outputs

The new estimates, for 2016 only, are provided in tables within a dataset accompanying this report. Tables A to C present the estimates in a format similar to the existing release of population estimates by country of birth and nationality to allow easy comparison of the two sets of figures. Table D provides the full set of estimates in a simpler format making it easy to analyse and reuse the figures. Tables E to I contain further analysis as described in Section 6: Findings.

Within the tables we have categorised the population characteristics of interest into broad groupings as follows.

Ethnic group

Six groups used: White British, All Other White, Mixed, Asian, Black, and Other. This grouping is derived from the harmonised classification of ethnic groups where All Other White is the aggregation of the standard groups of White Other, White Irish, and Gypsy or Irish Traveller. Chinese is included within the Asian group as set out in the standard classification.

Country of birth

Two groups used: UK Born and non-UK Born.

Nationality

Two groups used: UK Nationality and non-UK Nationality.

The existing release of estimates by country of birth and nationality provides a much more detailed disaggregation of those topics. We have not prepared similarly detailed estimates as part of this work, but invite comments on the usefulness of these groupings and any alternatives.

5 . Method adopted

The proposed method is based on the assumptions that:

- the proportion of the total population living in households remains constant at the level measured in the 2011 Census
- the communal establishment population will have different characteristics to the household population but these characteristics will have changed since the 2011 Census in a similar way to those of the household population

A detailed explanation of the method is provided in Annex A and a simple worked example is provided in Annex B. In summary, the method adopted is as follows:

- the mid-year population estimates are split into estimated household and estimated communal establishment populations by assuming that the proportion of people living in households has remained unchanged since the 2011 Census
- the Annual Population Survey (APS) distributions (what proportion of people belong to a particular ethnic group, say) are applied to the estimated household population
- in order to estimate the distribution for the communal establishment population a combination of 2011 Census and APS data is used; we assume that the distribution in communal establishments has changed since the 2011 Census in a similar way to how the distribution for the household population has changed
- the estimated communal establishment distributions are applied to the estimated communal establishment population
- the estimates for households and communal establishments are brought together to arrive at the final population estimates by population characteristics

6 . Findings

Figures used in this section are provided in Tables E to I in the accompanying dataset.

Effect of adjusting to mid-year estimates (MYEs)

The difference between the new estimates and the existing Annual Population Survey (APS) estimates of the total population within each area is set out in Table E. These differences are purely the result of differences between the APS estimates – which are derived from a combination of the mid-year population estimates and the subnational population projections and adjusted to remove the communal establishment population – and the standard mid-year population estimates rather than the result of the methods previously described.

The great majority of differences (330 out of the 394 areas (excluding Isles of Scilly and City of London)) are in the range 1% to 3%, which is intuitively reasonable as an adjustment for the communal establishment population. At the top of the range, Rutland (11% difference) and Richmondshire (10%) are understood as reflecting large communal establishments in the area (a prison and armed forces establishments respectively).

Counter-intuitively, nine local authorities show negative differences (that is, an APS estimate higher than the mid-year estimate). This reflects the use of the subnational population projection in the derivation of the APS estimate (needed as the 2016 population estimate is not available until after the APS results are processed). These negative differences indicate that the change of the population measured in the mid-year estimate for 2016 was lower than the change projected for that year.

Practical impact on estimated shares of the population

Here we look at the difference between the new estimates and the existing APS estimates of the proportion of the population of each area that are non-UK born or of non-British nationality. By looking at proportions we avoid the complications of the effects of scaling to the MYEs (described previously).

Table F shows the estimated proportions from the APS estimates and from the new estimates, together with the percentage point difference between the proportions (for example, if the APS estimated that 10% of the (household) population of an area was born outside the UK and the new estimates were that 11% of the population fell in that category, the percentage point difference would be 1 percentage point).

The table shows that the impact of the new methods on estimated shares for these variables is fairly limited.

The correspondence of the two sets of estimates tells us nothing about the accuracy of the new estimates but means that the conclusions we would be likely to reach if using the APS estimates would be unlikely to change if using the new estimates.

Error due to methodological assumptions

As noted previously, the method is based on two assumptions:

- the proportion of the total population living in households remains constant at the level measured in the 2011 Census
- the communal establishment population will have different characteristics to the household population but these characteristics will have changed since the 2011 Census in a similar way to those of the household population

Whilst we cannot directly measure the effect of those assumptions in these estimates we can use the results of the 2001 and 2011 Censuses to assess how making those assumptions would have led to inaccuracies in estimates for 2011 based on the 2001 Census.

Table G presents an indicative analysis of whether born outside the UK using data from the 2001 and 2011 Censuses for England and Wales as a whole.

The table shows that the proportion of the population living in communal establishments stayed stable at 1.8% of the population between 2001 and 2011. The proportion of the household population that was born overseas rose from 8.8% to 13.3% – a much faster growth rate than the corresponding proportion for the communal establishment population, which rose from 16.8% to 19.2%.

The analysis then proceeds in three steps.

The first is to take the 2011 Census population as given and to calculate the difference in the estimated communal establishment population if the proportion observed in 2001 is applied rather than the 2011 Census proportion. The effect is a very small increase in the estimated CE population of about 1,000 and a corresponding decrease in the household population.

The second is to take the 2011 Census estimate of the proportion of the household population that is born overseas as given (this is broadly comparable to our having the APS estimate for the household population on which to base our analysis) and to use this (using the change in the odds ratio as described in Annex A) to estimate the corresponding proportion for communal establishments. The effect of this is to increase the estimated proportion of communal establishment residents who are born outside the UK from 19.2% to 24.3%.

Thirdly, and finally, we apply that newly-estimated proportion to the new estimate of the communal establishment population and combine that with the estimate for the household part of the population to derive the net effect of the assumptions.

This process results in the total estimate of the non-UK born population in England and Wales rising by about 52,000 (0.7% of the actual 2011 Census estimate), or an average of about 5,000 each year between 2001 and 2011.

Caution must be taken with the results of this analysis, which is based on just two data points in 2001 and 2011. The assumptions underlying the method may prove to be less, or more, accurate for years since 2011 and any divergence from the assumptions will not proceed at a precisely constant rate over time. Relative errors will inevitably be higher for smaller geographical areas and for smaller population groups (in particular, smaller ethnic groups). The analysis does not take account of all sources of additional uncertainty or inaccuracy introduced by the new method (for example, the indirect effect of the APS sampling error on the estimated distributions within communal establishments or the inclusion of some students in halls of residence in the APS).

We can, however, use this indicative analysis to suggest that the error in the estimate of the non-UK born population of England and Wales due to the methodological assumptions described here may be of a similar scale, by the end of the inter-censal period, to the expected error (as measured by the standard error of approximately 73,000) in the corresponding APS estimate.

Estimates not available

A limitation of using the APS data in producing these estimates is that estimates for some combinations of area and category are not available because the APS sample for that combination was too small.

Table H provides a summary of the numbers of not available estimates. For the largest category in each topic (that is, ethnic group is White British; country of birth is UK, nationality is British) the only estimates not available are those for the very small local authorities of the Isles of Scilly and the City of London. Though small numbers of estimates for non-UK born and non-British nationality are not directly available, a comparison of the estimate of the total population and the corresponding UK-born or British nationality estimate for those areas provides reassurance that the unavailable estimate is likely to be very small.

Higher numbers of not available estimates are seen in the estimates by ethnic group with each group other than White British and Other White having more than 20% of estimates not available, and the Black and the Other groups having more than 40% of estimates not available. Despite these relatively high levels of not available estimates for individual groups, an estimate of the non-“White British” (that is, the population excluding the White British group) or the non-White population could be derived for the very great majority of areas by subtracting the White British, or both White groups, estimate from the total population.

Reliability of the estimates by ethnic group

The preceding analysis has concentrated on comparing the new estimates with the existing statistics on country of birth and nationality. To some extent, the conclusions of that analysis can be taken to be likely to apply similarly to the estimates by ethnic group. However, there are some differences that should be borne in mind.

Firstly, as noted previously, the relatively small size of some ethnic groups leads to a much higher proportion of estimates not available.

Secondly, the household-based cluster design of the APS means that estimates by ethnic group are likely to have somewhat higher sampling errors, other things being equal, than estimates by country of birth or nationality.

Thirdly, the practical impact of the new methods will be expected to be greater for groups that form a larger part of the communal establishment population. Table I shows that the minority groups (non-“White British” ethnic groups, born outside the UK, and holding a non-British passport (the proxy for nationality available through the 2011 Census)) all form a larger part of the communal establishment population than the household population with the single exception of the “Other” ethnic group. However, the difference from the household population is smaller for the non-“White British” ethnic groups than for non-UK born or non-UK passport held.

7 . Conclusion

It is possible to use existing National Statistics and a very simple method to produce population estimates by country of birth, by nationality and by ethnic group, which are consistent with the standard mid-year population estimates.

The new estimates for country of birth and nationality are similar enough to the existing Annual Population Survey (APS) estimates that they would be unlikely to change the conclusions of any analysis using the data. However, they have the advantage over the APS estimates of being consistent with the standard mid-year population estimates.

The initial primary source of uncertainty in the estimates will be the sampling error inherent in the APS estimates. The proposed method introduces additional uncertainty through the estimation of the size and distribution of the communal establishment population as we move from the 2011 Census.

This uncertainty is difficult to quantify but an indicative analysis based on census data for the non-UK born population suggests that inaccuracies through these assumptions might grow over the inter-censal period to become broadly similar in scale to those due to sampling error in the APS estimates.

The estimates have the following strengths and weaknesses.

Strengths

- they are consistent with the standard mid-year population estimates and are based on standard classifications, which make it easy to use the estimates in conjunction with other datasets
- they are based on existing National Statistics, which have supporting quality information
- they reflect differences in the characteristics of the household and the communal establishment populations
- estimates by ethnic group are more up-to-date than the standard source of the 2011 Census

Weaknesses

- sampling error and other sources of uncertainty make the estimates less reliable than the 2011 Census results were at that time
- these sources of uncertainty have a particular impact for smaller groups (such as ethnic groups other than White British) or when measuring change over time
- while information on sampling error in the APS is available, uncertainty through the new methods, and the assumptions underlying them, described in this report, cannot be easily quantified
- the estimates produced do not use the level of detailed classifications in the current population estimates by country of birth and nationality release
- estimates cannot be produced for areas or small population groups where the APS sample is too small

8 . Providing feedback

You are invited to provide feedback on these research outputs. In particular, we are interested in your thoughts on:

- the method we have used
- whether the groupings of country of birth, nationality and ethnic group used in the outputs are appropriate
- whether you would find an annual release of the estimates useful

Please provide your comments via email to pop.info@ons.gov.uk.

9 . Annex A: Methods

Note: in the following equations * represents multiplied by and / represents divided by.

1. The proposed method starts by assuming that the proportion of the people living in households has remained unchanged since the 2011 Census. Given that, we split the mid-year estimates (MYEs) into household (HH) and communal establishment (CE) population as follows:

   $P_{Census\ HH} = N_{Census\ HH} / N_{total}$

   where:

   - $N_{Census\ HH}$ is the number of people living in households as at the 2011 Census
   - N_{total} is the total population as at the 2011 Census
   - $P_{Census\ HH}$ is the proportion of people living in households as at the 2011 Census

   The estimated (EST) household population is calculated by applying this proportion to the MYEs:

   $N_{EST\ HH} = MYE_{2016} * P_{Census\ HH}$

   where:

   - MYE_{2016} is the 2016 mid-year population estimate
   - $P_{Census\ HH}$ is the proportion of people living in households as at the 2011 Census
   - $N_{EST\ HH}$ is the estimated household population

   whilst the estimated communal establishment population is given by:

   $N_{EST\ CE} = MYE_{2016} – N_{EST\ HH}$

2. The next step is to apply the Annual Population Survey (APS) ethnic distribution to the estimated household population:

   $N_{EST\ HH}(A) = N_{APS}(A) / N_{Total\ APS} * N_{EST\ HH}$

   where:

   - $N_{APS}(A)$ is the number of people in the ethnic group A in the area according to the APS
   - $N_{Total\ APS}$ is the total number of people living in the local authority
   - $N_{EST\ HH}$ is the estimated household population
   - $N_{EST\ HH}(A)$ is the estimated household population of the group A

3. The estimated ethnic group distribution of the communal establishment population is derived using a combination of APS and census data.

   The proportion of each ethnic group in APS and census is calculated using the following formula:

   $P(A) = n(A) / n(S)$

   where:

- n(A) is the APS estimate of the number of people in the ethnic group A in the area (for example, the number of White British)
- n(S) is the APS figure for the total number of people living in the area
- P(A) is the proportion in the ethnic group A

The odds for each ethnic group in the 2011 Census HH, 2011 Census CE and 2016 APS HH are calculated as follows:

ODDS(A) = P(A) / (1 – P(A))

where:

- P(A) is the proportion that ethnic group A makes up of the relevant household or communal establishment
- ODDS(A) is the odds of the ethnic group A

The odds of somebody in the CE population being from a given ethnic group are estimated using the following:

$ODDS_{Census\ HH}$ (A) / ODD_{APS} (A) = $ODDS_{Census\ CE}$ (A) / $ODDS_{EST\ CE}$ (A) (1)

therefore:

$ODDS_{EST\ CE}$ (A) = ($ODDS_{APS\ HH}$ (A) * $ODDS_{Census\ CE}$ (A)) / $ODDS_{Census\ HH}$ (A)

where:

- $ODDS_{Census\ HH}$ (A) is the odds of somebody from the household population being in ethnic group A in 2011 according to the census
- $ODDS_{APS}$ (A) is the odds of somebody from the household population being in ethnic group A in 2016 according to the APS
- $ODDS_{Census\ CE}$ (A) is the odds of somebody from the communal establishment population being in ethnic group A in 2011 according to the census
- $ODDS_{EST\ CE}$ (A) is the odds of somebody from the communal establishment population being in ethnic group A in 2016

The odds here are used so that we can avoid facing issues associated with probabilities that are greater than 1. If we used probabilities directly we may need to deal with the following:

$P_{Census\ HH}$ (A) / P_{APS} (A) = $P_{Census\ CE}$ (A) / $P_{EST\ CE}$ (A)

For example, if $P_{Census\ HH}$ (A) = 0.6, $P_{Census\ CE}$ (A) = 0.91 and P_{APS} (A) = 0.8 then:
$P_{ESTCE}(A) \approx 1.2$

Then we transform the odds to the proportions of each ethnic group for the estimated population in communal establishments as follows:

$P_{EST\ CE}$ (A) = $ODDS_{EST\ CE}$ (A) / (1 + $ODDS_{EST\ CE}$ (A))

where:

- $P_{EST\ CE}$ (A) is the estimated proportion of the communal establishment population who belong to ethnic group A in 2016
- $ODDS_{EST\ CE}$ (A) is the estimated odds of somebody from the communal establishment population belonging to ethnic group A in 2016

Note: transforming odds to proportions in this way can result in proportions that do not sum precisely to 1 across the ethnic groups within a local authority. To overcome this and ensure that these proportions sum to 1, we calculate and apply a calibration factor.

The calibration factor x is calculated as follows:

If $P_{EST\ CE}$ (White British) + $P_{EST\ CE}$ (Other White) +…+ $P_{EST\ CE}$ (Other) = Y

The calibration factor is x = 1 / Y

We then multiply the proportion in each ethnic group by this calibration factor to derive proportions that sum to 1.

4. Then the estimated ethnic group distribution for communal establishments is applied to the estimated communal establishment population:

   $N_{EST\ CE}$ (A) = $P_{EST\ CE}$ (A) * $N_{EST\ CE}$

   where:

   - $P_{EST\ CE}$ (A) is the estimated proportion of people living in a communal establishment in 2016 who belong to ethnic group A
   - $N_{EST\ CE}$ is the estimated communal establishment population (for all ethnic groups)
   - $N_{EST\ CE}$ (A) is the estimated communal establishment population of the group A

5. The final estimate is calculated by adding the estimated household and communal establishment population:

   N_{EST} (A) = $N_{EST\ HH}$ (A) + $N_{EST\ CE}$ (A)

   where:

   - $N_{EST\ HH}$ (A) is the estimated household population of ethnic group A
   - $N_{EST\ CE}$ (A) is the estimated communal establishment population of ethnic group A
   - N_{EST} (A) is the estimated total population of the ethnic group A

10 . Annex B: Worked example

Note: in the following equations * represents multiplied by and / represents divided by.

This annex provides a simple worked example of how the methods would be used to produce an estimate of the non-UK born population in a fictional local authority (LA).

1. In 2016 LA Everytown had a mid-year population estimate of 100,000. The 2011 Census had estimated that 98% of the population of Everytown lived in households.

   The first step is to split the 2016 population into the household part and the communal establishment part. We assume that the proportions are unchanged from the 2011 Census so:

   Household population = 98,000
   Communal establishment = 2,000

2. The 2016 Annual Population Survey (APS) estimated that 90% of the household population of Everytown was born in the UK. We therefore estimate that the household population of Everytown is made up of:

   Born in the UK = 98,000 * 0.9 = 88,200
   Born outside the UK = 98,000 * 0.1 = 9,800

3. The 2011 Census estimated that 95% of the household population and 70% of the communal establishment population was born in the UK. We calculate odds ratios for being born in the UK.

   If p is the probability of an event happening, then the odds ratio is defined as r = p / (1 – p)

   So the 2011 Census odds ratio for a member of the household population being born outside the UK is 0.95 / (1 – 0.95) = 19.

   The 2016 APS odds ratio for a member of the household population being born outside the UK is 0.90 / (1 – 0.90) = 9.

   The ratio of these two ratios is 9 / 19 = 0.47. That is, the odds ratio for being born outside the UK has changed by a ratio of 0.47 between the 2011 Census and the 2016 APS.

   We can now calculate the 2016 odds ratio for the communal establishment population.

   First, using the same approach as previously described, the 2011 Census odds ratio for being born outside the UK for the communal establishment population is calculated as 0.70 / (1 – 0.70) = 2.3.

   We now apply the change ratio seen in the household odds ratio to the communal establishment odds ratio.

   So, the 2016 odds ratio for a member of the communal establishment population being born outside the UK is 2.3 * 0.47 = 1.08.

   Now we have the odds ratio for the communal establishment population we can convert this to a probability (or proportion) by using the formula p = r / (1 + r).

   So the estimated proportion of the communal establishment population in Everytown that was born outside the UK is p = 1.08 / 2.08 = 0.52.

   (Similarly, the estimated proportion that a member of the communal establishment population in Everytown was born inside the UK can be calculated, using the same application of odds ratios as previous, as 0.48. Since the proportions of being born inside the UK (0.48) and outside the UK (0.52) sum to 1 we don't need to scale the proportions as we generally need to do when producing estimates for a number of different groups).

   Applying these proportions to the communal establishment population we calculated in step 1:

   Born in the UK = 2,000 * 0.48 = 960
   Born outside the UK = 2,000 * 0.52 = 1,040

4. Finally, we add the household and communal establishment populations together to produce our estimate for the population by country of birth of Everytown.

   Born in the UK = 88,200 + 960 = 89,160
   Born outside the UK = 9,800 + 1,040 = 10,840