Methodology for creating uncertainty intervals for mid-year population estimates by single year of age and sex: July 2020

Simulations-based methods for producing more granular measures of statistical uncertainty for mid-year population estimates, by local authority, sex and single year of age.

Contact:
Louisa Blackwell
louisa.blackwell@ons.gov.uk
+44 (0)1329 444539

Release date:
27 July 2020

Next release:
To be announced

Table of contents

1 . Measuring uncertainty in mid-year estimates at local authority level

Measures of statistical uncertainty for the local authority mid-year population estimates are research statistics that aim to give users of Office for National Statistics (ONS) data information about their quality. The [uncertainty intervals for 2012 to 2016 mid-year population estimates](#) were published in 2017. These were produced for each of the 348 local authorities in England and Wales. In this article, we describe how we produce local authority uncertainty intervals by age and sex. To begin, we provide an outline of how the uncertainty intervals are produced at local authority level.

Local authority population estimates are calculated using the cohort component method. In this approach, the previous year's population is aged-on by one year and then adjusted for births, deaths, net international migration, net internal migration and special populations (such as members of the armed forces and prisoners). The data for these adjustments come from several sources:

- data on births and deaths come from the General Register Office administrative registers

- international migration estimates come from the International Passenger Survey (IPS), supplemented in the case of in-migration by a range of administrative sources

- data on asylum seekers and their dependants come from the Immigration and Nationality Directorate of the Home Office

- internal migration data are primarily based on the NHS Patient Register

- adjustments are also made for special population sub-groups that are not captured in the international and internal migration estimates, for example, members of the armed forces and prisoners

The estimation process is repeated each year, starting from the 2011 Census base and rolled forward using the cohort component method. Uncertainty from international and internal migration includes accumulated uncertainty from previous years rolled forward, plus new uncertainty for the given year. This means that the uncertainty accumulates over time. The longer the lapse since the census, the more uncertainty there will be in the estimates.

"Uncertainty" is defined here as the quantification of doubt about a measurement. The three main sources of uncertainty associated with the mid-year population estimates are the census base, international migration and internal migration (moves between local authorities). Uncertainty in the other components of change (births, deaths, asylum seekers, armed forces and prisoners) is not reflected in the methodology and is assumed for now to be zero.

For each of the three components associated with uncertainty, the estimation process that is used to produce the mid-year population estimates is replicated and the replicates are used to simulate a range of possible values that might occur. The simulated distributions for each component are combined, iteration by iteration, mirroring the standard cohort components approach that is used for the published mid-year population estimates. Therefore, in year t:

$$Population^{(t)} = \begin{array}{l} base\ population^{(t-1)} + net\ natural\ increase^{(t-1,t)} \\ + net\ international\ migration^{(t-1,t)}. \\ + net\ internal\ migration^{(t-1,t)}. \end{array}$$

The methods for producing uncertainty measures at local authority level are described in [Methodology for measuring uncertainty in ONS local authority mid-year population estimates: 2012 to 2016](#) .

2 . Methods for producing uncertainty at single year of age

To produce uncertainty intervals for the mid-year population estimates broken down by age and sex, we require simulated values for each of the three components associated with statistical uncertainty also broken down by age and sex. The production of simulated values for internal migration at local authority level uses a model of internal moves by single year of age and sex. Therefore, we already have this level of granularity for internal migration. See Step A and Step B for a summary of the approach that we take for the census base and international migration components. Estimating uncertainty at single year of age in 2012 is a two-stage process. For the first stage (Step A), we produce an initial set of 1,000 simulated values for 2012 uncertainty.

For the census base population, we use parametric bootstrapping to create 1,000 plausible values for each single year of age for each sex. The assumption is that errors are normally distributed and that the variance for each age is taken from the published variances for the respective five-year age band. [1] The simulations for international migration are more complicated, but for both immigration and emigration we mirror the methods used to create these components in the production of mid-year population estimates.

Sex and age distributions are imputed for the 1,000 simulated immigrant flows generated for mid-year population estimate uncertainty at local authority level. The method clusters local authorities based on immigrant age and sex distributions, taken from census data. Local authorities are given the mean age and sex distributions for their cluster to produce 1,000 values for single year of age for each sex within local authorities.

Sex and age distributions are imputed for the 1,000 simulated emigrant flows generated for mid-year population estimate uncertainty at local authority level. Mirroring mid-year population estimate processes, local authorities are clustered based on age, sex and citizenship (British or non-British). Three years of International Passenger Survey (IPS) data (current year plus two previous years) for emigrants are used to create age and sex distributions, separately for British and non-British. Local authorities are given the mean age, sex and citizenship distributions for their cluster.

Simulations for each component are combined in accordance with the cohort component methodology:

$$
\begin{aligned}
Population(t, x) = \\
base\ population^{(t-1,x-1)} \\
+\ net\ natural\ increase^{(t-1,t,x)} \\
+\ international\ immigration^{(t-1,t,x)} \\
-\ international\ emigration^{(t-1,t,x)} \\
+\ inflows^{(t-1,t,x)} \\
-\ outflows^{(t-1,t,x)}
\end{aligned}
$$

We age-on the base population by a year.

The second stage (step B) begins with calculating the coefficient of variation for the simulations derived in step A. This is used in parametric bootstrapping to incorporate uncertainty into the population update between Census Day (March 27) and the mid-year (June 30). The bootstrapping assumes that errors are normally distributed around the mean, which for each local authority is taken as the difference between the published census value and the mid-year population estimate. We then repeat step A, using the updated simulated values as the mid-year 2011 population base.

Process for calculating the mid-year 2012 simulations by single year of age, sex and local authority

Step A: Calculate preliminary mid-year 2012 simulations from the March 2011 census simulations

Step A1: Census populations March 2011 simulations by single year of age, sex and local authority

These use published variances for corresponding five-year age group to derive variances by single year of age, as follows:

$$SD_{SYOA} = \frac{census_{SYOA}}{census_{group}} SD_{group}$$

We then use parametric bootstrapping from the normal distribution ~ N (census $_{SYOA}$, SD$_{SYOA}$) to create 1,000 simulations for the census component for each local authority by single year of age and sex.

Step A2: Mid-2012 natural changes (births minus deaths plus minor adjustments) internal inflow simulations

These are already by single year of age, sex and local authority.

Step A3: Mid-2012 internal inflow simulations

These are already by single year of age, sex and local authority.

Step A4: Mid-2012 internal outflow simulations

These are already by single year of age, sex and local authority.

Step A5: Mid-2012 international immigration simulations

These mirror the methodology used by the Population Estimates Unit to calculate international immigration estimates by age and sex. 2011 Census data on immigrants are used to cluster local authorities with similar age and sex profiles. Sex and age within the international in-migration component for each local authority are imputed based on the mean distributions within the cluster that the local authority has been assigned to.

Step A6: Mid-2012 international emigration simulations

These mirror the methodology used by the Population Estimates Unit to calculate international emigration estimates by age and sex. The 2011 Census is used to cluster local authorities based on sex, age and citizenship (British or non-British). Within each cluster, we use International Passenger Survey (IPS) data to create age, sex and citizenship distributions. British and non-British emigrants are assumed to have different age structures.

Three years of IPS data (current year and previous two years) provide a smoothed (centred average) single year of age distribution by citizenship and sex for each cluster. Sex and age are then imputed for each local authority's emigration simulations, based on the distribution in cluster that local authority was assigned to.

Step A7: Combine simulations from Steps A1 to A6 to derive preliminary mid-2012 simulations by single year of age, sex and local authority

Step B: Derive 2011 mid-year simulations, then calculate final mid-year 2012 simulations

Step B1: Derive 2011 mid-year population simulations by single year of age, sex and local authority

Update the March 2011 Census simulations to the 2011 mid-year population, by single year of age, sex and local authority, as follows:

i. calculate the coefficients of variation of the mid-2012 simulations, by single year of age, sex and local authority, generated in step A

ii. using parametric bootstrapping, simulate values for the three-month census to mid-2011; this is done by generating values from the normal distribution, using the coefficients of variation calculated in the previous step, while the mean is taken as the difference between the published 2011 mid-year estimate and the 2011 Census estimate by single year of age, sex and local authority (this incorporates uncertainty around these updates)

iii. add the values generated in the previous point (ii) to the census simulations

Step B2: Step A processes A2 to A6, as before

Step B3: Combine simulations from Steps B1 and B2, to produce final mid-2012 simulations by single year of age, sex and local authority

Notes for: Methods for producing uncertainty at single year of age

1. This assumes that the coefficient of variation for a single year of age is the same as for their respective five-year age group. Empirical testing suggests this may understate variance for 15-year-olds.
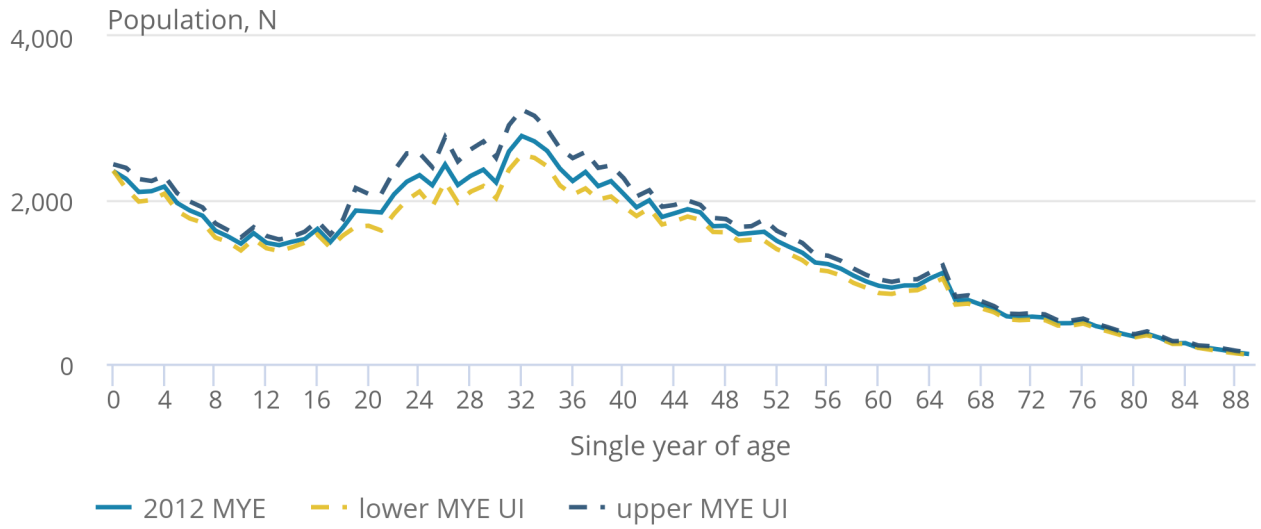
3 . Uncertainty intervals

The uncertainty intervals generated by this method reflect known patterns of doubt about population estimation: they are generally wider for men than for women and they are wider at student and young working ages.

Figures 1 and 2 provide examples.

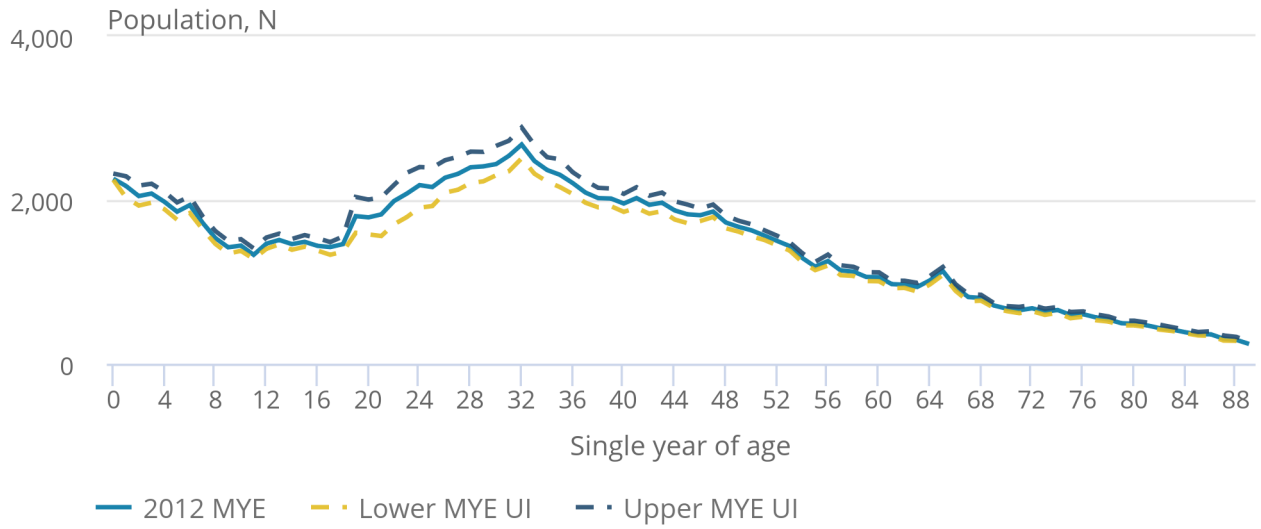## Figure 1: Mid-year estimates with uncertainty for males in Greenwich by single year of age

Source: Office for National Statistics

Figure 2: Mid-year estimates with uncertainty for females in Greenwich by single year of age

Source: Office for National Statistics