

Mid-2001 to Mid-2010 Small Area Population Estimates for Scotland: Methodology Guide



Published on 22 August 2019

Contents

| | |
|--|-----------|
| 1. An introduction to small area population estimates | 3 |
| Coverage and availability of population estimates | 3 |
| Uses of small area population estimates | 4 |
| Definition of the population | 4 |
| 2. Estimates rolled forward from 2001 | 6 |
| 2001 Census population | 6 |
| Components of change data | 6 |
| Migration data | 7 |
| Components of change in 2001 | 8 |
| Constraining and adjustments..... | 8 |
| 3. Rollback from 2011 | 8 |
| 90+ Populations | 9 |
| 2002-2010 MYE adjustments..... | 10 |
| Change-based adjustment..... | 10 |
| Adjustments to under 10s | 11 |
| Limitations of adjustment procedure | 11 |
| 4. Quality assurance | 12 |
| Enquiries and suggestions | 13 |

1. An introduction to small area population estimates

This paper describes in detail the methodology used by NRS to produce the rebased small area population estimates (SAPE) for 2011 Data Zones for mid-year 2001-2010, based on the 2001 and 2011 Censuses.

These are consistent with the mid-year population estimates for Scotland and its council areas already published for these years, and are comparable with SAPEs by 2011 Data Zones for other years, including the previously published 2011-2017 SAPEs and the simultaneously published 2018 SAPE.

Data zones are the Scottish Government's small area statistical geography. Data zones are based on Census populations, and so are updated every ten years with each Census. They are designed to cover the whole of Scotland and nest within council area boundaries. Data zones can be combined to create population estimates of larger areas, and are used to create estimates for larger geographies such as parliamentary constituencies and electoral wards.

After the 2011 Census, 6,976 data zones were created, designed to have populations of between 500 and 1,000 people in 2011 – though there are many data zones that have populations below or over these limits in years before or after due to changes in population.

The publication of 2001-2010 population estimates for 2011 Data Zones allows for comparisons to be made over wider time scales for small areas in Scotland. Previously, small area estimates were available for 2001 Data Zones from 1996 to 2014, however these data zones have different boundaries and do not allow for the creation of a time series with more recent data. It is now possible to compare populations of small areas across Scotland each year from 2001 to the most recent SAPE publication.

The most authoritative population estimates for Scotland come from the Census. This takes place every 10 years, with the most recent being held in March 2011. Population estimates from the census are updated each year with elements of population change in the previous 12 months to produce the annual mid-year estimates. These are considered the official estimates of the Scottish population.

The 2001-2010 estimates are based on combining data from the 2001 Census with indications of population change over each of the ten years from a variety of sources, including the International Passenger Survey (IPS), the National Health Service Central Register (NHSCR) and the Community Health Index (CHI). Adjustments have been made to ensure that totals by council area match the Mid-Year Population Estimates (MYE) for each year and that the time series remains consistent with the 2011 SAPE.

Coverage and availability of population estimates

Availability

Published estimates split by sex for all ages are rounded to the nearest 10 at council area level to avoid implying spurious accuracy and for ease of aggregation.

Data sets may be downloaded free of charge in Portable Document Format (PDF), Excel or Comma Separated Value (CSV) format.

Uses of small area population estimates

Mid-year small area population estimates currently have a wide variety of uses within central government, as well as being used by local authorities and health bodies, other public bodies, commercial companies and individuals in the private and academic sector.

These uses can be categorised into two broad groups:

- where the absolute numbers are of key importance. This may be in terms of allocating financial resources from central government, planning services or grossing up survey results. Some of the main central government uses are concerned with resource allocation, and
- where the population figures are compared with other figures such as the numbers of births or deaths in the calculation of rates and ratios.

Definition of the population

Mid-year small area population estimates relate to the usually resident population on 30 June of the reference year.

In simple terms, estimates of the usually resident population are estimates of people where they usually live. This does not always coincide with the number of persons to be found in an area at a particular time of the day or year. The daytime populations of cities and the summertime populations of holiday resorts will normally be larger than their usually resident populations.

The resident population is based on the population bases of the 2001 and 2011 Censuses, which measured where people usually live, as opposed to where they were on census night. Students and schoolchildren studying away from home were counted as a resident at their term-time address. If a member of the armed forces did not have a permanent or family address at which they were usually resident, they were recorded as usually resident at their base address. Residents absent from home on census night were required to be included on the census form at their usual/resident address. Wholly absent households were legally required to complete a census form on their return. No information is provided on people present but not usually resident.

For most people, defining where they 'usually' live for the purposes of the census is quite straightforward. However for a minority of people the concept of 'usual residence' is more difficult and it may be difficult to apply a general rule to assign people to where they are 'usually' living. Groups included in this category are:

- students;
- armed forces;
- prisoners;
- seasonal workers;
- contract workers and others who frequently move with their job;
- some people living in communal establishments;
- people sleeping rough;

- foreign students and au pairs;
- people with frequently used second homes in the UK or abroad;
- people who live and work away from a family home for part of the week;
- children who regularly move between a mother and father's home;
- adults who live with a partner for part of the time but maintain a separate residence; and
- any other groups of people with more than one residence.

The usual residence for students and certain members of the armed forces is specifically referred to in the definition of resident population for the 2011 Census given above. For other groups, guidance was provided either on the census form or in the enumerators' instructions.

In general, the definitions used in the Census are carried through into the population estimates, mainly because the census is used as a base for the population estimates. However, although efforts are made to ensure comparability of definitions in intercensal data and sources used in the population estimates, sometimes it is not possible to obtain data using the same definition as used in the census.

For example, in the International Passenger Survey (IPS), used to estimate international migration, a person is defined as an in-migrant and therefore a resident if they are intending to stay in Scotland for at least 12 months. However in the census, NRS made no specific adjustment for the presence of 6-12 months migrants among the persons counted in the census.

In practice, when compiling a population estimate, a number of data sources have to be used, each with its own definition of usual residence. These differences in definition are becoming increasingly important, and are the subject of constant research within the National Records of Scotland (NRS) and the Office for National Statistics (ONS).

Other outputs

National Records of Scotland also produces a range of other outputs and more information can be found in the [Statistics section](#) of the NRS website.

2. Estimates rolled forward from 2001

Small area population estimates for Scotland are made using the cohort component method. This is a standard demographic method and is used by several other national statistics institutions that also have access to high quality data sources for the components of population change.

The starting point for the estimates each year is the resident population on 30 June in the previous year. The previous year's armed forces and prisoner populations by sex, single year of age and data zone are subtracted from this figure. The population is then aged on by one year (for example all four-year-olds become five-year-olds one year later). Changes of population by sex, single year of age and data zone are then applied – primarily births, deaths, and migration – and then the armed forces and prisoner populations for the reference year are then added. For more details on the cohort component method as it applies to SAPE generally, refer to the [SAPE methodology guide](#).

2001 Census population

The 2001 Census is used as a starting point. First, population estimates are created by data zone for Census day (29 April 2001) by taking Census data at postcode level and summing these up using lookups from 2001 postcodes to 2011 Data Zones. These estimates are broken down by sex and single year of age – with the age calculated as at 30 June 2001 to avoid having to age the population on for the 2001 mid-year estimates.

2001 postcodes do not fit exactly into the boundaries of 2011 Data Zones, so there are some populations that may not be assigned to the correct data zone in this process. However, postcodes are usually small compared to data zones so the resulting errors will be minimal and postcodes were the smallest geographies that was practical for this purpose.

The estimates for Census day are then rolled forward nine weeks using the cohort component method to obtain estimates for mid-year 2001 (30 June).

An additional manual adjustment was made to three data zones in St Andrews, where an incorrectly coded postcode was detected for student accommodation. As a result, approximately 650 people are moved from data zone S01009734 to data zones S01009735 or S01009736.

Components of change data

The data for components of change used in previous SAPEs for this period was based on 2001 Data Zones, and so cannot be used directly.

Births and deaths data for 2011 Data Zones is obtained from Vital Events Statistics in NRS. Adjustments are made to correct for the possibility of misattributed births or deaths due to some postcodes that were split between data zone boundaries. The births and deaths data is also controlled to match council-level birth and death totals.

The number of prisoners each year are obtained from the Scottish Government by single year of age, sex, and prison. These prisons are then linked to data zones. If there has been a change to the number of prisoners in a council area compared to last

year, then an adjustment is made to other randomly selected data zones in the council area to balance out the change. This additional adjustment is generally very small.

The number of armed forces personnel were based on a combination of 2011 Census data, information from Defence Analytical Services and Advice (DASA) and data collected from armed forces bases. The armed forces population is then distributed to data zones using postcode-level armed forces populations from the 2001 Census. Some additional adjustments are also made to compensate for large changes between 2001 and 2011 in some data zones which were not accounted for by this method.

The number of asylum seekers at council area level by age and sex was calculated using Long Term International Migration (LTIM) estimate of asylum seeker flows from the Office for National Statistics (ONS) along with immigration information published by the Home Office. The number of asylum seekers in 2002 was calculated using the distribution of asylum seekers from 2002 on 2001 Data Zones. The 2001 Data Zones were matched to 2011 Data Zones based on the proportion of their total population that was in each overlapping 2011 Data Zone. Council area counts of asylum seekers for years 2003-10 were then distributed to 2011 Data Zones using the 2002 distribution.

Migration data

Migration estimates are derived from three key sources. The International Passenger Survey (IPS) provides information on overseas moves into and out of Scotland, and on asylum seekers. The National Health Service Central Register (NHSCR) is used to calculate moves between NHS Board areas within the UK. The Community Health Index (CHI) is used to estimate migration within health board areas, based on anonymised data from GP practice records linked to postcodes.

The migration data used differs from that used in the 2001 Data Zone SAPE in two ways. First, the migration data had already been rebased during the 2002-2010 Mid-Year Estimates rebasing to take into account corrections to external migration (refer to the [rebasings publication](#)), and so differs from that used in the 2001 Data Zone SAPE.

Second, the migration data had to be further modified for the 2011 Data Zone SAPE. During quality assurance, some errors in the processing of CHI data were found, affecting data zone level estimates of internal migration. These errors have been corrected for the 2001-2010 re-estimation, though previously published estimates for 2001 data zones are still affected and will not be updated. The issues encountered are listed below:

- Some large user postcodes (commonly used for addresses with large populations, such as student accommodation) were not correctly linked to their small user postcodes used to assign populations to data zones for records in 2001 to 2003. Several large user postcodes in the Edinburgh Riccarton campus of Heriot-Watt University were also missing a small user postcode link for other years.
- Issues were identified in the process to impute a postcode (and from that a 2011 Data Zone) to individuals without a recognised postcode. The random element of the imputation process did not work correctly which resulted in postcodes being imputed multiple times and high migration flows in/out of select

data zones. This mostly affected migration in/out of areas with a high number of student residents.

- A postcode for student accommodation which was registered incorrectly on a GP practice list resulted in inaccurate student age migration in data zones with high student populations in Edinburgh City.

Components of change in 2001

The components of change in the nine weeks between census day in 2001 and mid-year 2001 were calculated slightly differently from the other year-on-year components of change.

Migration, births and deaths are calculated in the same way with the period of interest reduced to the nine weeks in question. It was assumed that there was no significant change in armed forces, asylum seekers and prisoners between census day and 30 June 2001.

Constraining and adjustments

After all components of change have been applied, the data zone estimates by sex and single year of age are checked for any negative values that may have been introduced. For large negative values (-6 or less) these are set back to the previous years' population, while negative values between -5 and -1 are set to zero.

The data zone estimates are then constrained to the mid-year estimates for council areas by sex and single year of age.

Some manual adjustments are then made to specific data zones. Data zones containing hospitals are adjusted to account for births to people not resident in Scotland, and deaths of people not resident in Scotland. This particularly applies to the Borders General Hospital in Melrose, which is near the border with England.

Adjustments are also made to data zones with high levels of students (20% or more students in 2001). Migration in student areas is often difficult to capture, leading to situations where migration of students out of an area is often undercounted. If left unadjusted this affects the age distribution as individuals who should have been recorded as moving out stay in the data zone and are aged on each year. To correct for this, the age distribution of those aged 17-35 in these data zones is adjusted to match that in 2001, while not changing the total population. This means the total population will change with recorded increases and decreases in the student population, while keeping the relative age distribution the same.

The data zone estimates are then constrained again to the mid-year estimates for council areas by sex and single year of age, to ensure that the manual adjustments have not made the data zone estimates inconsistent with these.

3. Rollback from 2011

The roll-forward process is repeated for 2001-2011, with the estimates of each year based on the previous year's estimates and the components of change data. This set of rolled-forward estimates includes estimates for mid-year 2011.

The rolled-forward estimates for 2011 are then compared with the already published 2011 SAPE, which was based on the 2011 Census and the recorded components of

change for the 95 days between Census day (27 March 2011) and the mid-year reference date of 30 June 2011.

For convenience, the 2011 estimates created by rolling forward from the 2010 estimates will be referred to as the 'rolled-forward' 2011 SAPE, while the already published version based on the census will be referred to as the 'Census-based' 2011 SAPE for the rest of this report.

Of the two versions, the Census-based 2011 SAPE is assumed to be correct, and so any differences are taken to be due to errors made somewhere in the roll-forward process between the 2001 and 2011 estimates.

The difference between the Census-based and rolled-forward SAPEs by sex, age and data zone are the total adjustments that need to be made to each cohort by sex, birth year and data zone over the decade. The aim is to spread these adjustments over the years judged to have the most chance of errors being introduced. This process will be based both on how close each year is to 2011 and on the changes in population to each cohort over the decade.

The reasons for choosing this method, and a summary of the size of the differences found between the two 2011 SAPEs, are detailed in Paper 7 presented to the Population and Migration Statistics Committee on 1 May 2019.¹

90+ Populations

First, the data set is expanded to create a time series for all birth years included in the data. Estimates are made and published by single year of age for ages 0-89, and a separate category for those 90+. In 2011, this means birth years from 2011 to 1922 and those born on or before 1921. In 2001, this means birth years from 2001 to 1912 and those born on or before 1911.²

In order to have a time series for all birth years, estimates are needed for 2001-2011 for all birth years from 1912 onwards, and an estimate for those with a birth year of 1911 or earlier. For example, to measure the adjustment to 89 year olds in 2006 (born 1917), we need the adjustment for 94 year olds in 2011, and the population for 90-94 year olds in 2007-2011.

These populations are calculated using the most recently published time series estimates from the Centenarians in Scotland publication³. Estimates are produced annually from deaths data and include yearly estimates for Scotland by sex and single year of age for those aged 90 to 104 at mid-year from 1981 onwards.

It is assumed that the proportion of the 90+ population of each age will be the same in each data zone as that for Scotland as a whole, and so the estimates for these older ages are made proportionately. For example, the number of 94 year olds in a data zone in a particular year is estimated as:

¹ <https://www.nrscotland.gov.uk/files//statistics/Events/pams-1-5-19/pams-1-may-19-paper-7.pdf>

² Due to the population being estimated at mid-year, birth years mentioned here are taken from the previous mid-year date to the mid-year date in the year of interest. So for example, we take a birth year of 2001 to mean all those born from 1 July 2000 to 30 June 2001.

³ <https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-estimates/centenarians-population-estimates/population-estimates-for-scottish-centenarians>

$$(94 \text{ year olds in data zone}) = (90+ \text{ year olds in data zone}) \times \frac{94 \text{ year olds in Scotland}}{90+ \text{ year olds in Scotland}}$$

2002-2010 MYE adjustments

The method is based on that used for rolling forward unattributable change in the 2002-2010 rebasing of the mid-year population estimates for Scotland and its council and health board areas⁴. This method worked in the opposite direction – the estimates from the 2011 Census were rolled back by applying the cohort component method in reverse to 2001, resulting in a rolled-back mid-year 2001 estimate that was compared to the already published Census-based 2001 estimate.

These differences (by birth year, sex and council area) were then distributed linearly across all years, reflecting the assumption that it was equally likely for the error to be made in each year. So, for example, 90 percent of the difference would be subtracted from the 2002 figure, 80 percent from the 2003 figure, and so on up to 10 percent being subtracted from the 2010 figure.

The adjustments to the 2001-2010 SAPEs will partially be based in a similar way on the number of years until 2011 (with more adjustments being made closer to 2011 in this case, instead of closer to 2001). However, the adjustments will also take into account the patterns of change that were estimated in the rolled-forward figures.

Change-based adjustment

For each year, the absolute value of the change on the last year is calculated then the sum of each of these absolute values for all years so far is calculated. This sums the change in either direction – for example, if in the second year the population has decreased in one year by 10 and increased in the next by 10, then the sum for the second year is 20, not 0. This sum is then taken as a proportion of the sum of absolute changes in all ten years.

As a formula, this is

$$\frac{\sum_{y=2001}^Y |Change_y|}{\sum_{y=2001}^{2011} |Change_y|}$$

where $Change_y$ is the difference in population between years y and $(y - 1)$. This provides a measure of how much movement there has been in the population so far relative to how much movement there is over all ten years. It is assumed that cohorts that have experienced more changes so far will have a larger chance of having additional changes that are not recorded. In the few cases where there is no change recorded in any year, this fraction is set to 1.

This is combined with a measure of the number of years since 2001 to give the formula:

$$Adj_Y = (Adj_{2011}) \times \frac{Y - 2001}{10} \times \frac{\sum_{y=2001}^Y |Change_y|}{\sum_{y=2001}^{2011} |Change_y|}$$

⁴ <https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-estimates/mid-year-population-estimates/mid-2002-to-mid-2010-revision>

where Adj_Y is the adjustment that needs to be applied to the estimate in year Y (Adj_{2011} being the rolled-forward figure subtracted from the Census-based figure).

After the adjustments are made, all estimates of those aged 90 and over are summed for each year, data zone and sex so that the estimates can be consistently published each year by single year of age for 0-89 year olds and a separate total for those aged 90 and over.

The adjusted estimates were then constrained again to the mid-year estimates by year, sex, single year of age, and council area.

Adjustments to under 10s

For those who are under 10 years old in 2011, the calculation is modified slightly to take into account of the fact that they were born after 2001. Instead of distributing the changes over 10 years, changes are distributed among the years that the cohort was alive. This includes the first year, to account for potential errors in distributing the newborn population to data zones.

So, instead of $\frac{Y-2001}{10}$, we use

$$\frac{Y - (2011 - N)}{N},$$

where N is either 10 if there is data for all years, or otherwise the number of years data is available. For example, for the cohorts:

- born in 2010 (2 years of data) this will go from 1/2 in 2010 to 1 in 2011
- born in 2005 (7 years of data) this will go from 1/7 in 2005 to 1 in 2011
- born in 2002 (10 years of data) this will go from 1/10 in 2002 to 1 in 2011.
- born after 2002 (11 years of data) this will go from 0/10 in 2001 to 1 in 2011.

The change-based second fraction of the equation remains the same, except it takes the sums from the first year of data rather than 2001.

Limitations of adjustment procedure

The available data for making a judgement as to how to distribute corrections was limited, and so the roll-back adjustment methodology provides only a best guess as to when errors were introduced, that relies on several assumptions.

In particular, the adjustments can only take into account the errors noticed in the cohorts in 2011. For example, if a population has moved into a particular data zone in 2003 and then left that data zone in 2006, but has been mistakenly recorded as moving in and out of a different data zone, then the error will not be detectable in 2011.

It is feasible that this scenario may have happened in some student areas, where large numbers of people of similar ages generally move into an area and then move away within a few years. When large differences in student ages were found, it was generally not possible to tell whether it was a one-off error for that cohort or a pattern of students being mistakenly registered to a data zone when they actually lived elsewhere.

4. Quality assurance

For each year, the data zones with the largest overall change (both absolute change and percentage change) were inspected and attempts made to find explanations for these changes. This was done by inspecting the CHI extracts used to construct migration data, demolition records, registrations of housing with the Scottish Assessors Association and other sources.

Population counts were compared with counts of households by data zone for 2001, 2005, 2008, and 2011 to get an estimate for average household size within a data zone, and those with unusually large average household sizes were inspected to determine if there was a plausible explanation for this (for example a prison or student accommodation).

Data zones that had negative values for specific estimates after applying the various components of change were inspected, to determine the reason for these negative values being introduced. These were usually areas where large-scale demolitions or clearings of buildings had been carried out. These could be due to people who moved out before the census but registered with a doctor some years after, or due to moving out of people who had not registered there but did register with a doctor at their destination.

Enquiries and suggestions

Please contact our Statistics Customer Services if you need any further information.

Email: statisticscustomerservices@nrscotland.gov.uk

If you have comments or suggestions that would help us improve our standards of service, please contact:

Alan Ferrier
Senior Statistician
National Records of Scotland
Room 1/2/12
Ladywell House
Ladywell Road
Edinburgh
EH12 7TF

Phone: 0131 314 4530

Email: alan.ferrier@nrscotland.gov.uk

© Crown Copyright

You may use or re-use this information (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence. Further information is available within the [Copyright and Disclaimer section](#) of the National Records of Scotland website.