



Preserving the past | Recording the present | Informing the future

---

# **Mid-year Population Estimates for Settlements and Localities in Scotland: Methodology Guide**

**Published 12 March 2018**

---

## **Contents**

1. Introduction and Background	<b>3</b>
2. Postcodes, Boundaries and Densities	<b>5</b>
2.1 Previous Density Calculations	5
2.2 Updated Density Calculations	6
3. Current Methodology	<b>7</b>
3.1 Postcode Populations	7
3.2 Settlement boundaries	8
3.3 Locality boundaries	9
3.4 Estimates by sex and age	9
3.5 Quality Assurance	10
4. Publication	<b>11</b>
4.1 Comparability	12

## 1. Introduction and Background

Following the re-organisation of local government in May 1975, smaller local authorities, known as large and small burghs, ceased to exist. There was however, an on-going need expressed by users to know the population (and the characteristics of the population) of such areas.

By examining maps at the time of the 1981 Census, the National Records of Scotland (NRS), then the General Register Office for Scotland (GROS), defined localities as 'continuously built-up areas that had approximately 500 or more population at the time of the 1971 Census'. In many cases, localities within these built-up areas were formed based on the boundaries of former burghs.

By the time of the Census in 1991, NRS had digitised the boundaries of postcodes, which made calculation of population densities possible within postcodes. A postcode was defined to be urban if:

- it had been assigned to a locality in 1981, or
- it had five or more persons per hectare.

Otherwise it was classed as rural. Localities were then built-up from urban postcodes, also including non-residential areas which had come into existence since 1981 and which had been identified with the assistance of local authorities. Adjacent areas of urban postcodes were split into multiple localities based on the 1981 locality boundaries and from local authority suggestions.

For the 2001 Census, NRS developed a new process to identify settlements in Scotland. This definition was: 'a collection of contiguous high density postcodes bounded by low density postcodes (or water)' and whose population was 500 or more. A high density postcode, previously referred to as urban, was one with one of the following:

- more than 2.1 residential addresses per hectare; or
- more than 0.1 non-residential addresses per hectare.

The second condition was included so that non-residential parts of build-up areas such as industrial estates could be identified. Once settlement boundaries had been identified, they were then sub-divided into localities based on the 1991 Census locality boundaries.

From 2003 onwards a third condition was added to identify high density postcodes, expanding the definition to require one or more of:

- more than 2.1 residential addresses per hectare;
- more than 0.1 non-residential addresses per hectare; or
- an estimated population of more than five people per hectare.

Since 2001 NRS has continued to produce population estimates for settlements and localities at regular intervals to meet the need for population statistics for built-up areas in Scotland. These areas are generally more identifiable as the traditional towns and cities of Scotland than administrative areas such as Council areas, much of whose boundaries include land that is not built-up.

However, as a result of on-going housing and industrial development, there are groups of distinct towns and cities that have become close enough to be joined into one settlement. The Localities are published along with the Settlements to provide separate population estimates for distinct towns and cities within Settlements, based on previous settlement and locality boundaries and consultation with councils.

A list of common definitions and acronyms used in this document is available on the NRS [Glossary](#) of Terms page.

## 2. Postcodes, Boundaries and Densities

NRS produces and maintains digital boundaries and 'index files' of postcodes based on information received from Royal Mail on the addresses which have been assigned to each postcode. Although this information relates to both small and large user postcodes (a large user postcode is typically a single address which receives more than a set number of mail items per day), boundaries are only drawn for small user postcodes. The boundary is drawn to enclose all addresses which Royal Mail has assigned to the postcode, including any large user postcode located within its boundary, and has associated information such as residential addresses, non-residential addresses and area in hectares (the area is measured within the Mean High Water line). This is used in density calculations to determine whether a postcode is classed as 'high density' or 'low density'.

At the time of the 1991 Census, population estimates for settlements and localities were based on a calculation of persons per hectare to help distinguish between urban and rural areas. Updating the methodology to use the number of addresses within a postcode has meant that estimates can now be produced in years where there is no Census. This has also meant that the process is largely automated, which was the intention when the estimates were first produced, but doing so has also introduced uncertainty related to the input data which may need to be corrected by some manual intervention. This can be caused by the following:

- The way in which Royal Mail groups addresses into postcodes.
- The distinction between large user and small user postcodes, as only small user postcodes are given boundaries by NRS with large user postcodes linked to small user postcodes according to their geographical location.
- The way in which NRS draws small user postcode boundaries to include unpopulated land, which is because of the Census requirement that all of Scotland's land mass should be covered by postcodes.
- Royal Mail has not provided information on addresses and postcodes which is fully up-to-date.
- Information used by NRS to distinguish between residential and non-residential addresses may not be accurate.

NRS believes, however, that, in adopting this methodology, the resulting uncertainty has been outweighed by the advantages of producing settlement and locality population estimates at more regular intervals. In doing so, this allows some of the uncertainty and subjectivity relating to the estimates to be addressed more frequently rather than only at the time of a Census.

### 2.1 Previous Density Calculations

The identification of a postcode as either urban or rural, based on its density, was introduced at the time of the 2001 Census. For a postcode to be classed as 'urban', it needed to meet either (or both) of the following conditions:

- The number of residential addresses per hectare exceeded 2.1.
- The number of non-residential addresses per hectare exceeded 0.1.

Two densities were used so that unpopulated areas, such as industrial estates, could be identified as well as residential areas. The minimum density of 2.1 residential addresses per hectare was comparable to the five persons per hectare threshold used to create localities in 1991, with the minimum density of 0.1 non-residential addresses per hectare being derived by using different densities and comparing the initial results with maps for Clackmannanshire Council area, which was chosen because of its small area.

In order to achieve a certain level of continuity with the 1991 localities, the density thresholds were revised downwards for Na h-Eileanan Siar and Shetland Islands Council areas to create several additional localities. In these areas, due to crofting and other factors, settlement patterns varied considerably from the Scottish norm with many dwellings in villages having several hectares of land attached. As a result, all or most of their constituent postcodes were not 'urban' as defined above. Instead, the density thresholds were revised downwards to ensure that at least 95% of postcodes in the 1991 localities remained as 'urban'. Additional localities with low population density were then created as they were considered to be villages acting as centres of local activity similar to higher density localities elsewhere in Scotland.

## 2.2 Updated Density Calculations

The first population estimates for settlements and localities to be produced in a non-Census year were for 2003. In producing these estimates, the different thresholds which were used to produce localities in the Na h-Eileanan Siar and Shetland Island Council areas for the 2001 Census no longer applied. Instead, the same thresholds were applied to all parts of Scotland.

To produce the settlement and locality estimates, the 2003 Small Area Population Estimates (SAPEs) and the Royal Mail's Postcode Address File (PAF) were used to generate populations for each small user postcode. These were then used as the basis for determining whether a postcode should be classed as 'high density' or 'low density', replacing the previous distinction between urban and rural postcodes. A postcode was defined as 'high density' if at least one of the following three conditions were met:

- The number of residential addresses per hectare exceeded 2.1.
- The estimated population per hectare exceeded five.
- The number of non-residential addresses per hectare exceeded 0.1.

These thresholds were the same as those used for the 2001 Census with the addition of the minimum population density which had been used at the time of the 1991 Census. Settlements were then identified as groups of contiguous high density postcodes with a total population of 500 or more and divided into localities where appropriate, based on the boundaries used for the 2001 Census.

### 3. Current Methodology

There have been six subsequent publications containing updated population estimates for settlements and localities. These were based on the [Small Area Population Estimates \(SAPes\)](#) for 2004, 2006, 2008, 2010, 2012 and 2016 using the same methodology by which the 2003 estimates were produced – postcode populations were estimated, high density postcodes identified and settlements created from these postcodes then divided into localities where appropriate.

#### 3.1 Postcode Populations

A population must first be estimated for each of the small user postcodes in Scotland. Although the density of a postcode is not entirely dependent on this calculation, as addresses from the Postcode Address File can be used for this purpose, the settlement and locality populations are estimated from the total of the individual postcode populations within their boundaries. The process goes through the following steps

- A population total for each of Scotland's 6,976 data zones (created from 2011 Census output areas) is taken from the SAPes.
- Data held by NRS on communal establishment populations is matched to data zones and subtracted from each data zone's population.
- The total number of households in each data zone is obtained from data held in the Postcode Address File.
- An average household size for each data zone is calculated, dividing the net data zone population by the total number of households in that data zone.
- The number of households in a postcode is multiplied by the average household size applicable to the data zone in which it is located.
- Communal establishment data which was subtracted earlier in the process is then added back to the postcode with which it is associated.

Every small user postcode will then have an estimated population which will feed into the identification of high density postcodes.

This same methodology is also used to estimate postcode populations broken down by sex and broad age groups (under 16, 16 to 64, and 65 and over).

Note that these postcode estimates depend on the assumption that household sizes, and the age and sex balance in each, will be broadly uniform across postcodes within a data zone.

The separate consideration of communal establishments helps eliminate some of the exceptions to this assumption, for example student halls which will have a very different age balance to the general population. However, our data on communal establishments is not always complete and so postcode estimates can sometimes still be skewed by their surroundings.

This can cause issues in smaller settlements made up of fewer postcodes where these issues can have a greater proportional effect. For example, in 2016 several

postcodes in the locality and settlement of Whitecraig in East Lothian have been estimated to have a large number of women, causing the settlement as a whole to have the highest estimated proportion of women in Scotland. This is most likely not a true reflection of its sex balance but rather due to these postcodes sharing a data zone with the campus of Queen Margaret University (in the nearby locality of Musselburgh), which has a large proportion of female students.

### 3.2 Settlement boundaries

The process of determining settlements requires the identification of contiguous groups of high density postcodes with a combined population which rounds to 500 people or more. A boundary for each settlement will then be produced, which is formed by the outer boundary of its constituent postcodes, whose individual boundaries have been digitally created by NRS according to data received from Royal Mail. When an outer boundary has been defined for a settlement, any 'holes' created by low density postcodes within the settlement, but surrounded by high density postcodes, will be filled in. The boundaries are also clipped to match the Ordnance Survey Mean High Water lines. The processing is carried out using [ArcGIS software](#).

Although the process of producing settlement boundaries is intended to be automated, there is a certain amount of subjectivity relating to unit postcode boundaries which may require manual adjustments to be made. For many settlements, particularly in rural areas, there will be postcodes on the outskirts of a settlement will include some buildings that are part of the settlement but also a large amount of rural land, and possibly some buildings that are far from the rest of the settlement but still share a postcode in common with buildings that are closer. The decision of whether to include or exclude such postcodes is made on a case by case basis by NRS, with input from local councils through the consultation.

As the postcode boundaries are drawn to include all land in Scotland, many coastal postcodes also contain islands, and for consistency these are included in the settlement boundaries. These islands can be far from the rest of the settlement, and may not be inhabited at all.

Because of these properties of postcode boundaries, the supplied boundaries of a settlement may contain more rural land than the boundaries of the built-up area it is attempting to approximate. As a result the 2016 publication has not included density figures for individual settlements or localities as these would not be reliable and may be too low in some cases.

The settlement boundaries also include some non-contiguous postcodes, where due to the arrangement of addresses a postcode has had to be split into two or more separate parts. In places where a non-contiguous postcode has one part within a settlement and one part elsewhere, the settlement boundary has been modified to exclude the part of the postcode out with the settlement. When only part of a postcode is in a settlement, the entire population of that postcode will be included or excluded as a whole, depending on which part of the postcode has been excluded.

This means that addresses considered in the population estimates will not always correspond exactly to the addresses within the settlement boundaries.

### 3.3 Locality boundaries

When settlement boundaries have been finalised, localities are then created by determining for each postcode in the settlement whether it was in the previous locality dataset and grouping them together based on the previous locality value. If a postcode has no locality value, then it is either assigned one based on the nearest locality to it or a new locality could be created. Information provided by local authorities as part of the process of quality assuring the population estimates is also taken into consideration.

When settlement boundaries have been finalised, the postcodes within the settlement are assigned localities. Locality boundaries were initially based on 1975 burgh boundaries and have been modified since then through local authority requests and NRS quality assurance. Some locality boundaries within settlements have also come into being as a result of two previously separate settlements becoming joined by one or more dense postcodes.

Aside from old burghs and settlements that have joined, the current locality boundaries are based on what portions of settlements local authorities have requested be considered as localities. This will cause some inconsistency in how settlements across Scotland are split into localities. This should not impact analysis or comparison of specific localities, but analysis of the set of localities in Scotland as a whole – for example, looking at the range of populations or age balances of all localities – is unlikely to be useful.

When locality boundaries are finalised, population weighted centroids are calculated for both settlements and localities.

### 3.4 Estimates by sex and age

Since the 2012 settlements and localities publication in July 2014, additional breakdowns of population estimates by sex and age have been introduced. These breakdowns have been created as a supplement to the population estimates described above and are created once the settlement and locality total population estimates and boundaries have been finalised.

Three age groups: those aged under 16, 16 to 64, and 65 and over are created for both sexes at postcode level using the method described in [section 3.1](#). Using a postcode look-up to localities, locality population estimates are created for these broad age groups.

The broad age group estimates are used as control variables to scale best-fit data zone population estimates to produce locality population estimates by sex and five year age groups.

Both the broad and five year age group estimates are then adjusted to be made consistent with the rounded total population estimates already created and finalised, and settlement populations are then created by aggregating the locality estimates.

To illustrate the additional methodology:

A locality is finalised, its population is 2,000, the locality contains 50 postcodes, and its boundary is contained within three data zones.

The population estimates for the 50 postcodes are aggregated by data zone to give the broad age groups by sex. These estimates are compared with the total population of the data zones.

For each data zone, the population estimated to be in the entire data zone is compared with the population within both the data zone and the locality, and a proportion is calculated.

For example, one of the three data zones is not entirely contained within the locality – it covers 20 postcodes, 10 of which are in the locality and 10 of which are out of it. This data zone is estimated to have a population of 100 males of age 65 or over. However, the postcodes within the locality are estimated to have only 50 males of age 65+. The proportion is equal to 0.5 and so the 10 postcodes within the locality are estimated to contain exactly half of the 65+ population of the data zone as a whole.

The computed proportion is used to scale the data zone estimates. Here, the data zone estimates for males in the 65 to 69, 70 to 74, 75 to 79, 80 to 84, 85 to 89, and 90+ age groups are all halved to produce the estimates for these groups for the part of the locality within this data zone.

These estimates for the localities split by data zones are then aggregated to produce the estimates for each locality, and the broad and five-year age groups are adjusted so that their totals are consistent with the population estimate of 2,000.

This methodology was quality assured for the 2012 publication against output area data from the 2011 Census and was found to be broadly comparable.

### 3.5 Quality Assurance

The population estimates for settlements and localities, along with other population and demographic statistics, have been designated as National Statistics by the United Kingdom Statistics Authority (UKSA). To ensure that they continue to comply with the Code of Practice for Official Statistics requirement of sound methods and assured quality, the estimates are subjected to a thorough process of quality assurance.

Any major increases or decreases in population for any settlement or locality are checked to ensure that they can be reasonably explained and are consistent with population trends in the surrounding areas. Settlement populations which either rise

above or fall below key population thresholds of 3,000, 10,000 and 125,000, and which inform changes to the Scottish Government's Urban Rural Classification, are also checked.

Local authorities have also been invited to participate in a consultation process to provide advice on settlement boundaries and how localities should be identified within larger settlements. Five consultations have taken place to date, at the time of the 1991 Census, following the publication of the 2003 estimates and during quality assurance of the 2010, 2012 and 2016 estimates. Any changes which have been made as a result of these consultations have been consistent with the methodology used to produce the population estimates - settlement and locality boundaries can only be defined by their constituent postcode boundaries and not by other geographical features.

#### **4. Publication**

The estimated settlements and locality populations have been provided on the NRS website in eight main tables and ten additional tables, which are available in Excel and csv format (in addition, the main tables are available in pdf format).

Tables 1a and 1b of the main tables contains settlement and locality look-ups. All settlements and localities have been listed with their relevant Council area. A specific naming convention is applied within these tables:

- Settlements with multiple localities, one of which has the same name as the settlement as a whole, are suffixed by the words 'Settlement of', for example 'Edinburgh, Settlement of'. This is to help distinguish these from settlements which are exactly equivalent to a single locality of the same name.
- Where a settlement crosses a Council area the settlement has been listed with each Council area separately. Settlements and localities falling into this category are suffixed by the word 'part' in brackets, for instance 'Liff (part)'. Small parts of settlements and localities – those of less than 100 people – have not been listed.

This naming convention is not used in any of the other tables provided.

The remaining main tables, Tables 2a, 2b, 2c, 3a, 3b, and 3c contain population estimates for settlements and localities by broad age groups. Tables 2a, 2b, and 2c provide population estimates for settlements by persons, males, and females respectively. Similarly, Tables 3a, 3b, and 3c provide population estimates for localities by persons, males, and females respectively.

The additional tables provide five year age breakdowns of populations and include additional analysis on age and sex breakdowns.

The tables containing population estimates are also supported by a series of geographic shape files which display the settlement and locality boundaries by using Geographic Information System (GIS) software.

The additional data provided in the tables by sex and five year age groups have been derived from data zone populations and are provided at unit level. However, it is not implied that the population estimates are accurate to this level of detail. The population figures are estimates that have gone through a number of stages of processing, each of which may impact on the quality of the estimates. Also, there are limitations with the administrative data sources used to produce the figures which may increase the uncertainty in the estimates. For more information refer to sections [3.1](#) and [3.2](#). In addition, data zone population estimates are constrained to the age/sex distribution at Council area level.

#### 4.1 Comparability

Users of the estimates are advised that they are not directly comparable with those of previous years and should exercise caution if using them to identify population trends. Although increases in population may be attributed to newly built housing, it may also be caused by the inclusion of existing housing which was previously separated by a low density postcode, or vice versa for population decreases.

Further issues with comparability have been introduced due to errors affecting the age distribution of the mid-year population estimates from 2002 to 2014. For more information, refer to the [Population](#) section of the NRS website.

We welcome feedback on this paper and the methodology used. If you have comments or suggestions, or if you need any further information, please contact our Statistics Customer Services.

Email: [statisticscustomerservices@nrscotland.gov.uk](mailto:statisticscustomerservices@nrscotland.gov.uk)