Population Projections for Scottish areas (2014-based): Methodology Guide

Published on 27 October 2016

A National Statistics publication for Scotland

National Statistics are produced to high professional standards set out in the National Statistics Code of Practice. They undergo regular quality assurance reviews to ensure that they meet customer needs.

Contents

Introduction

This paper describes the methodology used to produce the sub-national population projections for Scotland.

The sub-national population projections (SNPP) provide figures by single year of age up to 89 and aggregated 90 and over by sex for the future population of Scottish areas, namely:

- Council areas;
- NHS Board areas;
- National Park areas; and
- Strategic Development Planning Authority areas.

They are produced every two years by the National Records of Scotland (NRS) to coincide with the production and release of the National Population Projections (NPP) for the UK and constituent countries by the Office for National Statistics (ONS).

The SNPP are fully consistent with the NPP at Scotland level.

Both the NPP and SNPP are trend based, making assumptions about future levels of fertility, mortality, and migration based on levels observed in the recent past. Therefore, they give an indication of what the future population, by age and sex structure, might be if recent trends continue and take no account of policy or development aims in local authorities or other factors such as the outcome of the EU referendum. The NPP also incorporates expert advice to form assumptions about future demographic behaviour.

The SNPP form the basis as input for other projections, such as household projections. They are used by central and local governments, and the health sector for informing policy, planning, and in resource allocation. For example, the projections are used for planning the provision of education and health services; by local authorities, together with other inputs, to produce their own projections for local planning; and by the Scottish Government for funding allocation to local authorities and NHS Board areas.

Overview of the Method

The population projections use the cohort-component method to project the population using a single year model – that is, a projection made by sex and single year of age for each future year. This is a well-established demographic technique, and is also used in the production of population estimates. In population estimates we take the components of population change (this is birth, deaths, and migration) in the year prior to the estimated year and apply them to the current estimated population to obtain the new estimated population.

However, with projections, we do not know the components of population change and therefore have to project them before we can project the population.

To project births and deaths we use age-specific fertility rates and age-specific mortality rates at Scotland level obtained from the NPP and apply them to the population at risk. We also apply a local scaling factor to account for the variation in Scottish areas.

To project international migration we take the estimates and apply time series analysis to project future flows. For migration around the UK we use past population estimates and

migration data to create rates. The rates are applied to the population at risk of migrating which generates projected migration.

1. The Geographies of Scotland

The National Records of Scotland (NRS) produce population projections for the 32 council areas, 14 NHS Health board areas, four strategic development plan (SDP) areas, and two national park areas.

As of April 2014, councils completely nest within health boards. However, SDP and national park areas are non-standard geographies and do not nest neatly within either of the other two geographies.

Processing Units

In order to produce consistent population projections for all areas we project the populations for 42 processing units. These processing units consist of councils, and (where SDP or national park areas intersect a councils) part councils. In total nine councils have been split in this manner, these are:

- Aberdeenshire;
- Angus;
- Argyll and Bute;
- Fife;
- Highland;
- Moray;
- Perth and Kinross;
- Stirling; and
- West Dunbartonshire.

Each area has been split into two parts with the exception of Perth and Kinross which has been split into three parts. Figure 1 and Figure 2 show how the council areas have been split.

Figure 1: Map showing the council areas and council area parts within strategic development plan area boundaries

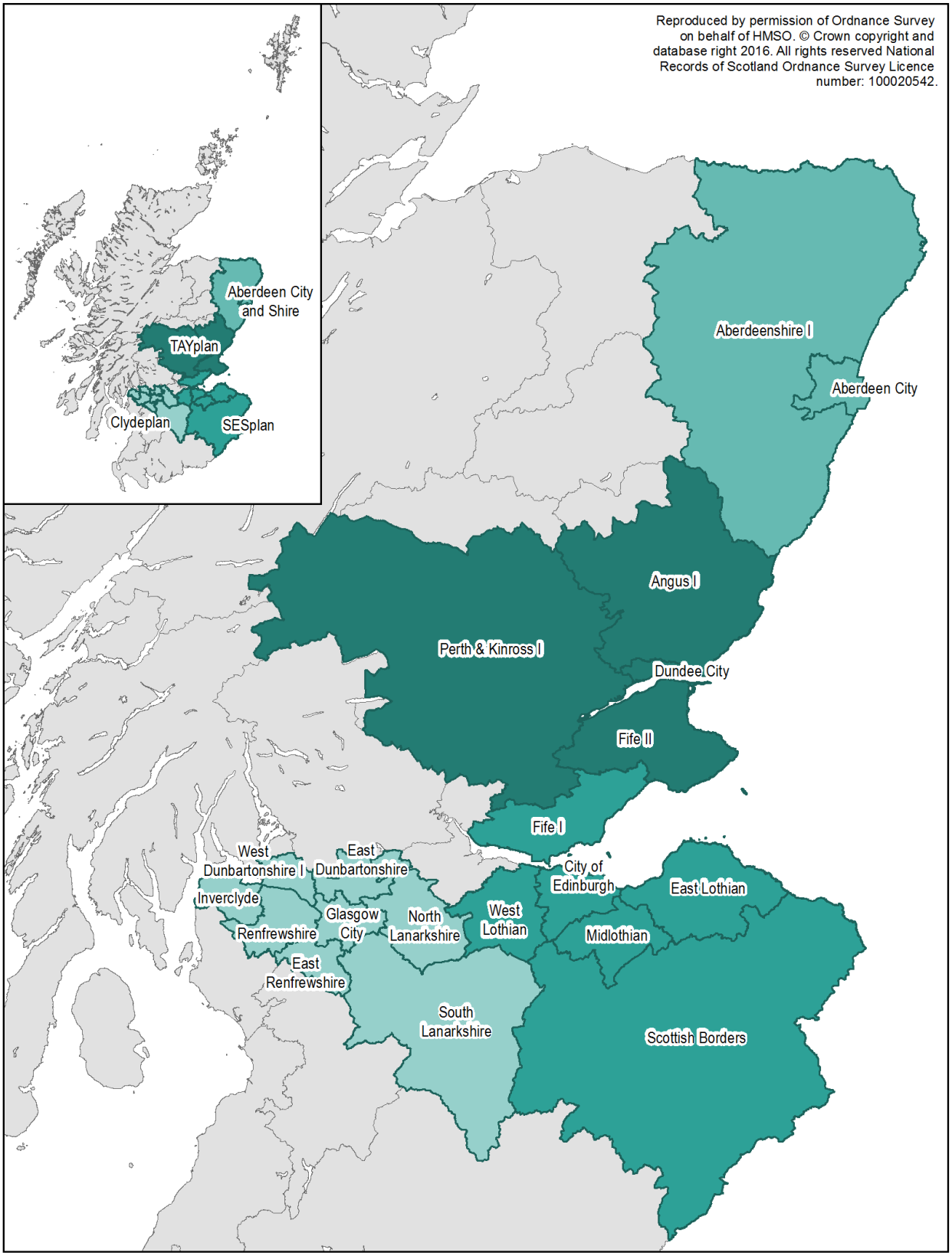
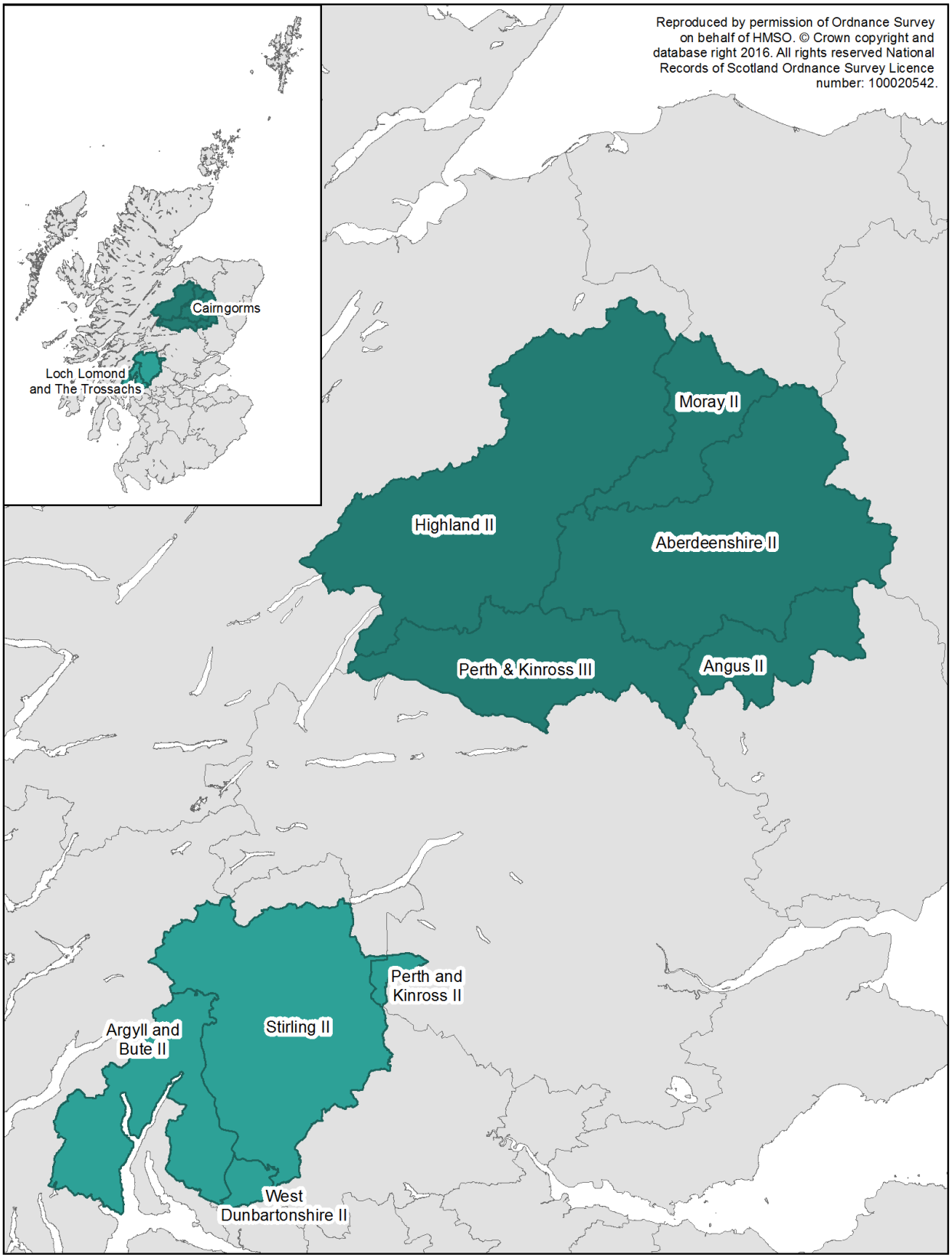


Figure 2: Map showing the council areas and council area parts within national park area boundaries



2. Data Sources

The sub-national population projections (SNPP) use a variety of demographic statistics as inputs to calculate a projection. These data sources include:

- Estimates of population;
- Estimates of births;
- Estimates of deaths;
- Estimates of migration;
- Estimates of special populations; and
- Scotland level population projections.

What follows is a detailed description of the data sources and how they are produced/used by the population projections methodology.

All data sources have been collected for the period 2001 through to the base year (2014). With the exception of international migration totals, the data used in determining trends only includes data five years prior to the base year.

Estimates of Resident Population

The general population of an area includes all those usually resident there whatever their nationality. Usual residents temporarily away from home (for six months or less) are included, but visitors are excluded.

Members of Her Majesty's Armed Forces stationed outside Scotland are excluded but members of Her Majesty's and non-UK Armed Forces stationed in Scotland are included.

Supported asylum seekers and visitor swithers (people who enter the country intending to visit, but stay to become usual residents) now residing in Scotland are included as part of the general population. Migrant switchers (people who enter the country intending to become usual residents, but who stay for less than 12 moths) are not included in the general population.

Students are taken to be resident at their term-time address.

Population figures relate to 30 June of the year shown and ages relate to age at last birthday.

Estimates of Population

The National Records of Scotland (NRS) produce mid-year small area population estimates at data zone level on an annual basis. In order to estimate populations at processing units, the data zone estimates are taken and used to produce postcode estimates. The postcode estimates are then aggregated to best-fit processing units. This process follows the same methodology used to produce estimates of settlements and localities and more information can be found on the NRS website.

Population estimates for England, Wales, and Northern Ireland are also used in the processing. Estimates are aggregated to form rest of UK population estimates. The

population estimates for England and Wales are produced by the Office for National Statistics (ONS), more information can be found on their website. The population estimates for Northern Ireland are produced by the Northern Ireland Statistics and Research Agency (NISRA), more information can be found on their website.

Population estimates are used in the calculation of rates for rest of UK, and within Scotland migration. The most recent population estimates are also used as the base from which new populations are projected.

Estimates of Special Populations

Two types of special populations are used in the mid-year population estimates, these are:

- Estimates of the prison population; and
- Estimates of Her Majesty's Armed Forces resident in Scotland.

The prison population is estimated to be those people sentenced to incarceration for six months or more.

The data is that used in the mid-year estimates processing and is available at prison level then aggregated to processing unit.

Estimates of Her Majesty's Armed Forces include numbers of home and foreign armed forces living in barracks and in army quarters. The estimates do not include armed forces dependents; they are included in the general population.

The data is available at data zone level and is aggregated to processing unit. Please note that due to the relatively small size of the population out-with specific military bases, the effect of using data zone aggregation rather than postcode aggregation is negligible.

Estimates of Births and Deaths

Data on births are obtained from the compulsory civil registration system administered by NRS and the Local Registration Offices. Data is provided for live births by sex and because registration of a birth may legally take place up to 21 days after the birth, the data received refer to the date of birth rather than the date of registration. Births are recorded at age zero and allocated to the area of usual residence of the mother.

Similar to births, data for deaths are obtained from the compulsory civil registration system. Data refer to the date of death rather than the date that the death was registered. Deaths are allocated to the area of usual residence of the deceased. The age at death is calculated to be the age that the deceased person would have been on 30 June. No adjustments are made for non-resident deaths.

A postcode level dataset is provided and used to aggregate births by sex to processing areas, and to aggregate deaths by single year of age and sex to processing units.

Estimates of births in England, Wales, and Northern Ireland are aggregated to form rest of UK births. The vital events estimates for England and Wales are produced by the Office for National Statistics (ONS), more information can be found on their website. The vital statistics for Northern Ireland are produced by the Northern Ireland Statistics and Research Agency (NISRA), more information can be found on their website.

Births are used in the production of migration rates for zero year olds, and in the calculation of projected births. Deaths are used in the calculation of projected deaths.

Estimates of Migration

Migration estimates are produced by NRS as part of the mid-year population estimates process.

International migration figures consist of the following components:

- Moves between Scotland and overseas;
- Visitor switchers;
- Migrant switchers; and
- Asylum seekers.

Within Scotland migration consists of moves between areas within Scotland.

Rest of UK migration figures consist of moves between Scotland and the other constituent countries of the UK.

Further information on the migration methodology for the mid-year population estimates can be found on the NRS website.

Flow data at record level is aggregated to processing units. Migration estimates are used in the projection of international, rest of UK, and within Scotland migration.

Scotland Level Population Projections

At Scotland level the sub-national population projections are consistent with the National Population Projections that are produced by ONS on behalf of the National Statistician and the Registrars General of Scotland and Northern Ireland.

The underlying assumptions were agreed in liaison with the devolved administrations following consultation with key stakeholders in each country and expert advice. Local authority assumptions for each component sum to the national projections assumptions.

Data from the national population projections used in the sub-national population projections process, as well as where they are used is provided in Table 1.

Table 1: Use of the national population projections in the sub-national population projections

Component of National Projection	Use in Sub-National Methodology
Age-Specific Fertility Rate	To produce birth projections
Age-Specific Mortality Rate	To produce death projections
Asylum Seeker Distribution	To constrain asylum seeker projections
Asylum Seeker Totals	To constrain asylum seeker projections
Births by Age of Mother	To constrain birth projections
Births by Sex	To constrain birth projections
Deaths	To constrain death projections
International Migration	To constrain international migration projections
Population	To constrain population projections
Rest of UK Births	To project migration rates
Rest of UK Migration	To constrain rest of UK migration projections, and to project migration rates
Rest of UK Population	To project migration rates

3. Methodology for Producing Population Projections

The projections start with the population estimates for the base year, disaggregated by single year of age, sex, and area. The base population is then projected one year ahead. This is done in the following way:

1. International Migration is projected using time series analysis, and historical age and sex distributions.
2. Rest of UK and within Scotland migration is projected for non-zero year olds using rates produced using historic estimates of population and migration.
3. An estimate of the numbers of non-zero year olds surviving to be one year older is made by applying age-specific mortality rates and local area scaling factors to the population at risk to give the number of deaths.
4. The above components of population change, together with the starting population, are combined in the cohort-component method to form a projection of the population of non-zero year olds one year from the base date.
5. The number of births in the year is produced using age-specific fertility rates, and local area scaling factors to an average of the female population of child bearing age at the start and end of the year.
6. Steps 2 to 4 are repeated for 0 year olds and combined with the non-zero year olds to create a projection of the population one year from the base year.
7. The process can then be repeated.

The reason for separating the zero year old migration is to account for the presence of the end population of females in the calculation of births. As such the whole projections process must be completed for non-zero year olds and then repeated for zero year olds once the births have been projected.

The Cohort-Component Method

The sub-national population projections use the cohort-component method to project the population by one year. This is a standard demographic method and is used in the production of population estimates. It does this for each age and sex, and for each processing unit.

The process can be summarised as follows:

1. Start with the population at the beginning of the year.
2. Subtract the special populations present at the beginning of the year.
3. Age on[1] the remaining population.
4. Add the births which have occurred in the year
5. Subtract the deaths which have occurred during the year.
6. Add on the net migration that occurred during the year.
7. Add on the special populations present at the end of the year.
8. This produces the population at the end of the year.

Footnote

1) The civilian population in all areas is aged on one year to become the appropriate age in the following year of the projection. For example 10-year-olds in Aberdeenshire in 2014 will become the 11-year-olds in Aberdeenshire for 2015.

When projecting the population we start with the base year, this is the most recent year for which we have population estimates. We then apply the components of population change, which we have projected in order to generate the first year of the projection.

Members of Her Majesty's Armed Forces, and the population in prison are considered to be special populations. The age and sex structure of these populations remains fairly consistent over time and is a-typical of the underlying population of an area.

In order to project the special populations, estimates from five years prior to the base year are averaged. The resulting distributions are removed at the start, and added on at the end of each year in the projection.

Births

The process for projecting births involves applying Scotland level age-specific fertility rates (ASFR), and local fertility scaling factors (FSF) to the population at risk.

The process of creating local scaling factors for each processing unit is detailed in the Annex.

The population at risk is defined to be an average of the female population at childbearing ages (aged 15 to 46) taken from the start of the projection period and the end of the projection period. This can be represented as follows:

$$\frac{(Population\ at\ Start + Population\ at\ End)}{2} \times FSF \times ASFR$$

This projects the number of births for the year. The births are allocated to each gender on a ratio of 1.05 males to 1 female, the same ratio as used for the NPP. This method ensures that the total number of projected births for Scotland for each year of the projection period agrees with the figures from the NPP.

Once the births have been calculated for each processing unit they are constrained to the totals obtained from the NPP. More information on the technique used can be found in the Annex.

Deaths

The process for projecting deaths is very similar to that for projecting births. We apply the age-specific mortality rates (ASMR), and local mortality scaling factor (MFS) to the population at risk. The process for creating local scaling factors for each processing unit is detailed in the Annex.

In the case of projecting deaths the population at risk is half of the net migrants (rounded down to avoid part people) to the area added to the population at the start of the year. This can be represented as follows:

$$(Population\ at\ risk + \left\lfloor \frac{Net\ migrants}{2} \right\rfloor) \times MSF \times ASMR$$

Once the deaths have been calculated for each processing unit they are constrained to the totals obtained from the NPP. More information on the technique used can be found in the Annex.

Rest of UK and Within Scotland Migration

Rest of UK migration is migration between Scotland and the other countries that constitute the UK. Within Scotland migration is migration between areas of Scotland.

Migration is calculated for each processing unit, and for the rest of the UK as a whole. Due to the availability of data it is not possible to model migration between processing units and constituent countries of the UK separately. Migration is modelled by single year of age and sex for each flow between areas.

The migration is created using a multi-region rates based model. Rates are calculated for each movement between areas. The rate is created from the population and migration exhibited in the five years prior to the year that is being projected. Rates are calculated and then averaged. The average rate is then applied to the population at risk and this projects the out migration from that area.

The projected migration from area A to area B can be represented as follows:

$$\frac{Previous\ migration\ from\ area\ A\ to\ area\ B}{Previous\ population\ of\ area\ A} \times Current\ population\ of\ area\ B$$

Once all flows have been projected it is then possible to aggregate flows by origin, or destination to calculate the out, or in migration respectively.

For rest of UK migration the projected age and sex distributions obtained for each area are then constrained to the Scotland totals obtained from the NPP. More information on the constraining technique can be found in the Annex.

In addition, the within Scotland migration outflows are constrained to the remaining population once international and rest of UK outflows have been removed. This is to prevent the within Scotland migration from projecting more out migration that there are people in the area.

International Migration

International migration is migration between Scotland and countries other than the UK. It includes migration by supported asylum seekers who are modelled separately from other international migrants.

It is assumed in the National Population Projections (NPP) that international migration will move from levels seen in the base year to a long-term trend several years later. We call the time taken to reach the long-term trend the run-in period. In the 2014-based population projections, the run-in period was seven years with the long-term trend beginning in 2021.

We project international migration at processing units in three stages:

1. Project total in and out flows by sex.
2. Adjust averaged historical age distributions to projected totals.
3. Adjust projected age and sex distributions to NPP distribution.

In stage 1 we take historical estimates of total international migration to and from a processing unit, split by sex, and apply time series analysis to them.

The time series technique that is applied is Auto-Regressive Integrated Moving Averages (ARIMA). More information on ARIMA modelling can be found in the Annex. For international migration totals the simplified model involving just the Auto-Regressive (AR) component was chosen as it provided the most robust results when implementing models.

We then select the best AR model based on statistical tests of fit, and subjective examination of the projected output. If either the model fails the tests or the projected output looks a-typical in comparison with estimated data then the data is re-examined and a model that passes the quality tests is chosen instead.

Once a model is chosen, several years' worth of projected data is output. The number of years projected represents the run-in period with the last projected data point the long-term assumption.

The next stage in the process is to take an average of estimated international migration by single year of age and sex. These distributions can then be constrained to the totals created in the previous step.

The projected age and sex distributions obtained for each area in the previous stage are then constrained to the Scotland totals obtained from the NPP. More information on the constraining technique can be found in the Annex.

Supported asylum seeker migration is taken from the NPP and we assume that all the migrants coming to Scotland will migrate to Glasgow City Council. Similarly, we assume that all asylum seekers migrating out of Scotland, migrate out from Glasgow City Council.

4. Consistency and Quality Assurance

The National Records of Scotland (NRS) have processes in place to assess the quality and comparability of its data across the UK.

Comparisons with projections from other parts of the UK

A paper titled Sub-National Population Projections Across the UK has been published on the Office for National Statistics website. The paper describes the methods used by the different countries of the UK to produce sub-national population projections and it includes a note on comparability. Information on the national projections methodology is also available on the Office for National Statistics website. Please note that this paper will be updated to reflect the new Scottish methodology in due course.

Quality Assuring the Data

When the Population and Migration Statistics team within the NRS assemble the data for the various components, checks are carried out and comparisons made with council level data to gauge consistency and completeness of coverage.

The data are then processed electronically to produce the sub-national projections. Quality assurance takes place throughout this process. Quality assurance of outputs are also made with previous projections, taking account of changes in trends since the previous projections were published.

NRS also have processes in place to check the suitability of the administrative sources used in producing population estimates. More information can be found in the following paper Information about Quality Assurance Arrangements for Administrative Data on the NRS website

5. Annex

This section is intended to explain in more detail some of the statistical techniques that are used to produce the components in the population projections.

Fertility and Mortality Scaling Factors

Age-specific fertility rates and age-specific mortality rates are obtained from the NPP. These rates are for Scotland and do not represent the range of fertility, and mortality rates in different processing units. In order to represent this diversity local scaling factors are produced. The scaling factors are produced using data from the base year and used throughout the projection.

Fertility scaling factors are calculated for each processing unit , to determine the fertility scaling factor we compare the expected number of births that would have been experienced by an area if the Scottish age-specific fertility rates were applied, with the actual number of births that were observed and define the scaling factor from the results.

Expected births are calculated by applying the Scotland level age-specific fertility rates to the female population at childbearing ages (aged 15 to 46) in each area. An average births figure is calculated for each area using the observed births from the five years preceding the projection period, and these are scaled to the Scotland births figure from the first year of the NPPs. The scaled averages are then divided by the number of expected births mentioned above and the result is the local fertility scaling factor.

The different scaling factors can be interpreted as follows:

- Scaling factor less than 1 – Fertility is less than that of Scotland.
- Scaling factor equals 1 – Fertility is the same as Scotland.
- Scaling factor greater than 1 – Fertility is greater than that of Scotland.

Mortality scaling factors are calculated in a similar way to fertility scaling factors, however, deaths are compared instead of births. The mortality scaling factors are calculated for each processing unit, both sexes, and for three age groups: 0 to 59, 60 to 79, and 80 and over.

Confidence intervals are calculated around the mortality scaling factors. In instances where the confidence intervals in a processing unit overlap then there is statistically no difference between the scaling factors. In these instances the ages are combined and a new scaling factor is calculated using the new age group. The test of confidence intervals is then repeated until only statistically different scaling factors remain.

Auto-Regressive Integrated Moving Average Time Series

There are three components of an ARIMA model: the auto-regressive, integrated, and moving average components.

The auto-regressive (AR) part of the model looks at available data and derives a model based on previous data points. In most situations an AR model will look at the previous data points in a series to model the current point and the number of points that the model looks at is the coefficient of the model. For example, an AR(1) model looks at the previous

data point to model the current data point, while an AR(2) model looks at the previous two data points to model the current point.

The moving average (MA) part of the model works similarly to the AR component however, it derives a model based on the previous error terms in the data points. As in the AR model, a coefficient is determined based on the number of errors that the model looks at. For example, an MA(3) model would look at the error terms from the previous three data points.

Finally, the integrated component of the model looks at whether the data is stationary. Stationarity is where the probability distribution of a data point is the same irrespective of the time at which the it is collected. To model a time series the data should be stationary. Integrating the time series can account for the presence of non-stationarity.

The three components of ARIMA models combine additively and as such it is possible to combine one, two, or all three of the components in the final model.

Constraint

Throughout the projections there are periods in the processing where components are constrained to the Scotland totals produced by the national population projections (NPP) in order to make them consistent with the NPP.

The same method is applied irrespective of the component being constrained.

As a worked example, consider Table 2. There are five data points 7, 5, 2, 6, and 9 and we want to constrain them to the total 35 (all in bold).

Table 2: Illustration of constraint

Stage	Data					Total
Original	**7**	**5**	**2**	**6**	**9**	29
Proportions	0.24	0.17	0.07	0.21	0.31	
Constrained	8.45	6.03	2.41	7.24	10.86	**35**
Rounded	8	6	2	7	11	34
Adjusted	9	6	2	7	11	35

The process begins by computing the proportion of each unconstrained value to the unconstrained total. The original data, in the example, add to 29, and the proportion of this that each data point represents is shown in row two (Proportions) of the table.

This proportion is then multiplied by the total the data is being constrained to, this generates a constrained total. In our example we multiply 35 by the proportions which gives the constrained values in row three (Constrained).

However, the constrained total may not consist of whole numbers and so the constrained data is rounded, as seen in row four (Rounded) of the table.

Rounding may produce a different total to the constrained total and so 1 is added or subtracted from the rounded values that were closest to being rounded in the opposite direction until the rounded totals equal the desired total. The rounded totals in our example only add to 34 and so 1 is added to the first data point as displayed in row five (Adjusted).

The addition of one is done to the first data point because the data was not rounded up and of the three other data points that were not rounded, the decimal part of this data was closest to 1.