Revised methodology for setting the migration assumptions for the 2012-based national population projections

Table of Contents

1. Summary
In 2012 the Economic and Social Research Council (ESRC) Centre for Population Change (CPC) carried out a review of the migration assumptions setting methodology for the national population projections. The review suggested a number of methodological improvements which have been incorporated into the 2012-based national population projections. The main improvements include a general streamlining of the methodology and a move to modelling gross flows rather than net migration.

2. Introduction
This report documents the methodology used in producing the migration assumptions for the 2012-based national population projections. A new method for setting the assumptions has been incorporated into the projections based on an ESRC CPC review (the review is available here: http://www.ons.gov.uk/ons/guide-method/method-quality/specific/population-and-migration/population-projections/npp-migration-assumptions-methodology-review/index.html).

The previous assumptions setting methodology had been in place since the 1991-based projections and the review suggested that while these methods were in line with the current practice of many statistical agencies, they were not necessarily in accordance with the recommendations from academic literature, in particular surrounding the use of net migration levels. The methodology was also found to contain a number of 'patches' which had to be included to respond to unpredicted changes in trends (e.g. for migration from Central and Eastern Europe following the EU enlargement), or specific data situations. These patches were found to reduce the cohesion of the whole system of assumptions and sometimes necessitated making further 'tweaks'.
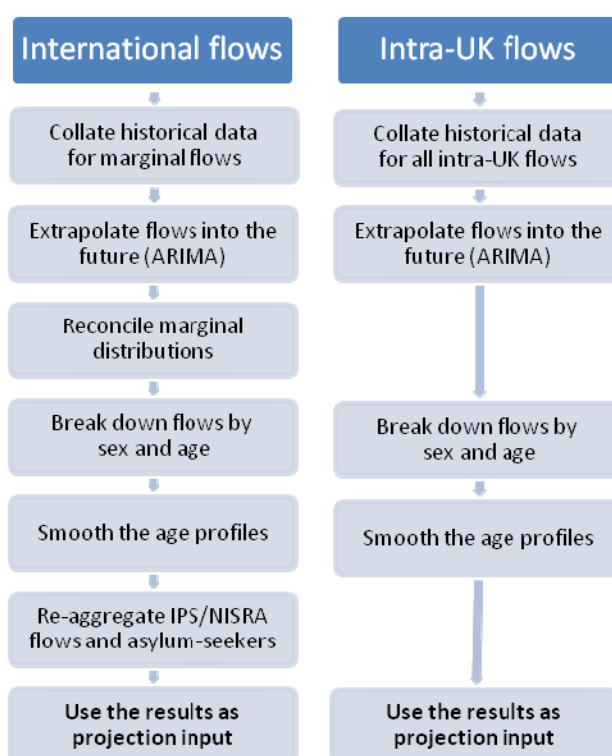
A user forum was also run in parallel to the review to ensure that users' views fed into the process. The users agreed with the principle of moving to gross flows as long as net migration levels were also available and some users expressed a preference for using rates for cross border flows. There was no strong desire to move to stochastic projections.

The review therefore ultimately recommended the use of gross flows (either volumes or rates) rather than net migration and a possible gradual shift from deterministic, argument-based variants to probabilistic projections with derived scenarios.

The first set of improvements, which have been incorporated into the 2012-based projections, were to streamline the existing argument-based scenarios into a new overall workflow, as follows:

1. Collate information on flows for the categories of interest (e.g. inflows by groups of countries, and by UK constituent country of destination; outflows by UK constituent country of origin), referred to as marginal flows, as well as for the matrix of intra-UK flows between the constituent countries;
2. For each flow, extrapolate the series with a chosen ARIMA time series model for the length of the short term assumptions (usually five to 10 years), and fix the obtained final values as constant for the remaining period of the projection;

3. Then, for all international inflow estimates and projections of flows, reconcile the extrapolated marginal information on different breakdowns by using log-linear models for contingency tables to estimate joint distributions by the characteristics of interest (e.g. migrants from EU15 countries to England);
4. Create smoothed age/sex distributions by averaging the sex ratios from the most recent five years of data and smoothing the age patterns for each flow by applying an appropriate variant of the Rogers-Castro model to averaged data from the same period;
5. Apply the age and sex distributions derived in stage 4 to the series extrapolated in stage 2, in order to obtain assumptions on specific migration flows by flow, sex and single-year age groups;
6. Re-aggregate the IPS-based flows and those of the asylum seekers. Derived quantities of interest (e.g. net migration) are also obtained at this stage.

International flows	Intra-UK flows
Collate historical data for marginal flows	Collate historical data for all intra-UK flows
Extrapolate flows into the future (ARIMA)	Extrapolate flows into the future (ARIMA)
Reconcile marginal distributions	
Break down flows by sex and age	Break down flows by sex and age
Smooth the age profiles	Smooth the age profiles
Re-aggregate IPS/NISRA flows and asylum-seekers	
Use the results as projection input	Use the results as projection input

The specific improvements compared with the previous methodology are:
- The use of gross flows. The use of net migration levels in the previous methodology was found to be contrary to the recommendations from academic literature because net migration can be seen as an artificial measure that hides the difference between changes in migration propensities and population stocks, and blurs the regularities in the age profiles of migration intensity.
- A more streamlined process, meaning that the need for 'patches' and 'tweaks', which make the process more opaque, is reduced.
- The use of ARIMA modelling. The previous method used exponential smoothing to model the data, which was recognised in a 1992 Government Actuary's Department/Office of Population Census and Surveys (GAD/OPCS) review as being limited by having an arbitrary choice of smoothing parameter and over reliance on past data.

(The paper is available here: http://www.ons.gov.uk/ons/rel/npp/national-population-projections/occasional-paper-no--42/a-new-methodology-for-determining-migration-assumptions---part-1.pdf). The switch to formal time series modelling brings the methodology into line with current best practice. Also, a simple exponential smoothing model can be fitted using ARIMA so, effectively, the scope of the modelling is widened without losing anything compared with the old method.

- More flexibility and scope for consultation when choosing models for each flow due to the use of ARIMA, which can fit a variety of different models to the data.
- The smoothing of age profiles using a variation of the Rogers-Castro curve, which ensures that spurious peaks and troughs in the past data are not applied to the projections.

3. Detailed methodology used in the 2012-based national population projections

3.1 Collate historical data for marginal flows
The flows are modelled in five groups as set out below:
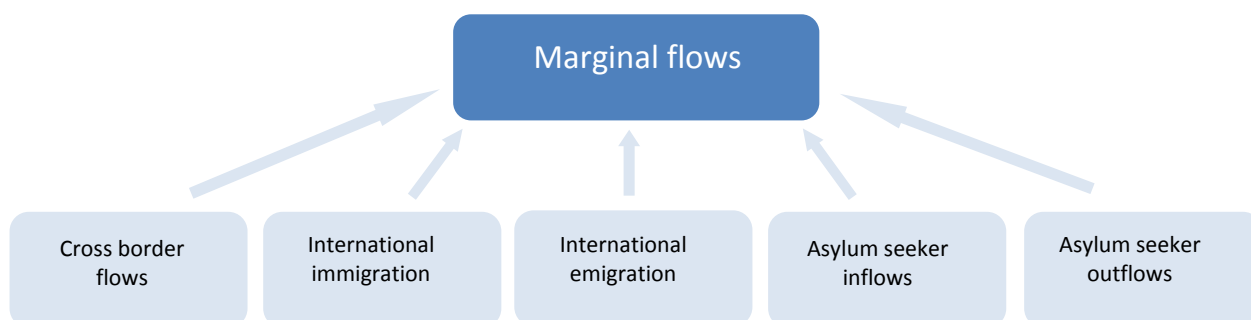
3.1.1 Cross border (intra-UK) flows
The data are obtained from the National Health Service Central Register (NHSCR) and cover the year ending June 1993 to year ending June 2012.

Table 1. Matrix of intra-UK flows

		Destination			
		England	**Northern Ireland**	**Scotland**	**Wales**
Origin					
	England	-	England to NI	England to Scotland	England to Wales
	Northern Ireland	NI to England	-	NI to Scotland	NI to Wales
	Scotland	Scotland to England	Scotland to NI	-	Scotland to Wales
	Wales	Wales to England	Wales to NI	Wales to Scotland	-

3.1.2 International immigration flows
The majority of the data are from the International Passenger Survey (IPS), including the adjustments made for 'switchers', and cover the year ending June 1976 to the year ending June 2012. The IPS is a voluntary sample survey of passengers travelling through airports, seaports and the Channel tunnel. It provides information on the number of people intending to stay in, or leave, the UK for 12 months or more. Estimates of migration of all citizenships between the UK and Irish Republic from 1991-2007 are provided by the Irish Central Statistics Office and are also incorporated into the immigration flows.

Data for inflow to Northern Ireland are obtained from the Northern Ireland Statistics and Research Agency (NISRA) and are based on health card data so implicitly include switchers, asylum seekers, Irish flows etc. They cover only the period from 1992 onwards.

Due to the survey-based origin of much of the data and the small cell sizes involved, only the marginal flows are modelled (**Table 2**). The joint flows which are of interest to users are calculated after modelling using iterative proportional fitting, with a moving average of the last five years of joint flow data used as the starting value.

Table 2. Matrix of international inflows to the UK

		Destination				
		England	**Northern Ireland**	**Scotland**	**Wales**	**Marginal flows**
Origin						
	EU15	EU15 to England	EU15 to NI	EU15 to Scotland	EU15 to Wales	EU15 to UK
	EU8/2	EU8/2 to England	EU8/2 to NI	EU8/2 to Scotland	EU8/2 to Wales	EU8/2 to UK
	New Commonwealth	New Commonwealth to England	New Commonwealth to NI	New Commonwealth to Scotland	New Commonwealth to Wales	New Commonwealth to UK
	Old Commonwealth	Old Commonwealth to England	Old Commonwealth to NI	Old Commonwealth to Scotland	Old Commonwealth to Wales	Old Commonwealth to UK
	Rest of World	Rest of World to England	Rest of World to NI	Rest of World to Scotland	Rest of World to Wales	Rest of World to UK
	Marginal flows	Total inflow to England *	Total inflow to NI *	Total inflow to Scotland *	Total inflow to Wales *	Total inflow to UK

Note: For modelling purposes, Malta and Cyprus are included in the EU15 grouping

It should be noted that only the country of destination marginal flows (denoted with an asterisk in **Table 2**) are used to create the assumptions that feed into the projections production system. The country of origin flows are modelled to meet user requirements and as a plausibility check. They are adjusted proportionately in order to sum to the same total UK migration figure as the country of destination flows.

3.1.3 International emigration flows
As with international immigration, the majority of the data are taken from the IPS, including the adjustments made for 'switchers', and cover the year ending June 1976 to the year ending June 2012. Estimates of migration of all citizenships between the UK and Irish Republic from 1991-2007 are provided by the Irish Central Statistics Office and are also incorporated into the emigration flows.

As with international inflow to Northern Ireland, data for the outflow from Northern Ireland from 1992 onwards are obtained from NISRA.

The following flows are modelled using these data:

- International outflow from England
- International outflow from Northern Ireland
- International outflow from Scotland
- International outflow from Wales

For previous projections, international immigration and emigration flows were modelled by citizenship (British v. Non-British) but the decision has been made to not

do this for the 2012-based projections because the numbers of British citizens immigrating are low by certain breakdowns (e.g. 1,000 or fewer migrate to Northern Ireland and the numbers become smaller still when broken down by country of origin). Given the fact that IPS data are survey-based, it was therefore decided that it would be inadvisable to model at this level. Furthermore, there appears to be little or no structure to model in a number of the British citizen flows, so modelling these would add little value.

3.1.4 Asylum seekers

The asylum seeker data are provided by the Home Office and the National Asylum Support Service, and cover the year ending June 1992 to the year ending June 2012. The flows are modelled at the Great Britain level due to the small numbers involved, and broken down to constituent country using the proportions from the last five years of data. Asylum seekers to and from Northern Ireland are not differentiated in the NISRA international flows so cannot be modelled separately. Northern Ireland asylum seeker inflows and outflows are therefore set to zero with their asylum seekers included in the international flows. The asylum seeker flows are therefore as follows:

- Asylum seeker inflow to England
- Asylum seeker inflow to Northern Ireland (set to zero)
- Asylum seeker inflow to Scotland
- Asylum seeker inflow to Wales

- Asylum seeker outflow from England
- Asylum seeker outflow from Northern Ireland (set to zero)
- Asylum seeker outflow from Scotland
- Asylum seeker outflow from Wales

3.2 Extrapolate flows into the future

A set of ARIMA models is fitted to each flow using the maximum likelihood method and the best models chosen on the basis of goodness of fit statistics and consultation regarding the plausibility of the resulting forecasts with National Records of Scotland (NRS), NISRA and the Welsh Government.

Orders of models between (0,0,0) and (1,1,1) only are used for extrapolation since demographic applications generally do not go beyond this parameter range and the goodness of fit statistics penalise models based on the number of parameters.

Where there is a noticeable and explainable discontinuity in the data (e.g. for the countries in the EU8/2 inflow grouping, which showed a large increase in inflow to the UK after A8 accession in 2004), an intervention is applied during modelling to take this into account.

The goodness of fit statistics examined are AICc and mean absolute percentage error (MAPE). In order to minimise the risk of unnecessarily switching models (e.g. from simpler to more complex) on the basis of a small improvement in these values, which effectively results in substituting one plausible model for another equally

plausible model, a rule of thumb suggested by Burnham and Anderson (2004)[1] was adopted. This splits the models into groups of those with substantial support, those with less support and those with little to no support. Where there is little difference in goodness of fit statistics (e.g. a number of models had substantial support), the simplest model is considered the best.

Forecasts are made for the duration of the short-term assumptions with the final short-term levels agreed after consultation with the four UK constituent countries. The long-term assumptions, which run to the end of the projection period, are set at a constant level equal to that of the last year of the short-term assumptions.

The forecasted totals for each flow are rounded to the nearest 500 to form the assumptions for each country and the assumptions for the UK are rounded to the nearest 5,000 in order to reflect the inherent uncertainty in setting the assumptions.

3.3 Reconcile marginal distributions

The international inflows are modelled separately by constituent country of destination and country group of origin (i.e. the marginal distributions as seen in **Table 2**) due to the small cell sizes present if the data are subdivided any further. The joint distributions (e.g. EU15 migrants to England) are however desired by users. Therefore, the assumed marginal totals for each year of the short-term projection are inputted into an iterative proportional fitting procedure where they are broken down proportionately to the required joint distributions using the moving average of the joint distributions from the last five years as the starting point.

3.4 Smooth and apply age-sex distributions

The flows of migrants are forecasted at a total level so need to be broken down by single year of age and sex before being fed into the projections production system.

For the intra-UK flows, age/sex distributions are created using NHSCR and NISRA single year of age data for each flow. For the international flows, age/sex data from various sources including the IPS, NRS, NISRA and the mid year population estimates international migration components of change are investigated in order to decide on suitable distributions. The international distributions are also applied to the asylum seeker flows because the asylum seeker data are available only by broad age groups. Comparison of the broad age distributions between the international and asylum seeker flows is carried out to confirm that the age distributions at this level are similar enough to justify the substitution.

For each flow, the assumed numbers of total migrants per year are broken down by sex using the average sex ratios from the last five years of data. Where necessary, the single year of age distributions for each sex for the last five years are then averaged and smoothed using an appropriate variation of the Rogers-Castro model incorporating a student peak where necessary to ensure that student migration patterns are retained in the age distributions (details of the model formulated by Wilson (2010) to incorporate student peaks can be found here: http://www.demographic-research.org/volumes/vol23/8/).

---

[1] http://www.sortie-nd.org/lme/Statistical%20Papers/Burnham_and_Anderson_2004_Multimodel_Inference.pdf

The smoothing outputs single years of age up to 89 years old only, so an assumption of zero migration is made for all ages 90+, as was the case for previous projections. The smoothed age distributions are applied to the assumed male and female migrants for each year of the projection to give assumed migrants by age and sex for each flow.

As with previous projections, a constant age distribution methodology is used, with the same age distributions being applied for all years of the projection.

3.5 Re-aggregate the flows and calculate net migration levels

The assumptions for each flow, expressed as numbers of migrants by single year of age and sex for each year of the projection, are then combined to produce net migration levels by age and sex for each UK constituent country. These are then rounded to the nearest even number before being fed into the projections production system.

Even number rounding is required due to the fact that the projections methodology specifies that the number of net migrants to be added to obtain the population aged x+1 at the end of the projection year consists of half of those migrating during the year at age x and half of those migrating during the year at age x+1.

These final rounded net figures are the principal migration assumptions.

4. Variant projections

High and low variants are set after consultation with the devolved administrations and key stakeholders. The variants are currently established as a set level above and below the principal. ONS explored using a more objective method to set the variants e.g. using alternative models or confidence intervals around the net figures but no one method gave a satisfactory solution for all countries and the UK as a whole.

5. Outputs

The net migration levels for each UK country by age and sex for each year of the projection period are fed into the projections production system.

6. Future research

ONS will continue to review the methodology in conjunction with the devolved administrations and plans are in place to research the use of rates for emigration and cross border flows and to consider setting the variants using a more objective method.

ONS also plans to redevelop the NPP production system to allow greater flexibility in terms of inputs into the system.

ONS welcomes feedback on the new methodology and future research. Email: projections@ons.gsi.gov.uk. Phone: +44 (0)1329 444652.

7. Glossary

ARIMA time series modelling
Autoregressive integrated moving average modelling. Time series data are used to predict future trends. ARIMA modelling can take into account trends, seasonality, cycles, errors and non-stationary aspects of a data set when making forecasts.

EU2
The EU2 (formerly known as the A2) are the two countries that joined the EU on 1 January 2007: Bulgaria and Romania. EU2 nationals currently have certain restrictions placed on them; in the first 12 months of stay, working Bulgarian and Romanian nationals are generally required to hold an accession worker card or apply for one of two lower-skilled quota schemes. Other Bulgarian and Romanian nationals can apply for a registration certificate, giving proof of a right to live in the UK.

EU8
The EU8 (formerly known as the A8) are the eight central and eastern European countries that joined the EU on 1 May 2004: Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovakia and Slovenia. The EU8 does not include the two other countries that joined on that date: Cyprus and Malta. EU8 nationals previously had restrictions on their rights to work and were required to register under the Worker Registration Scheme, but since 1 May 2011 EU8 nationals now have the same rights as other workers from the EU and EEA.

EU15
Austria, Belgium, Denmark, Finland, France, Germany, Greece, Republic of Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain and Sweden. The UK is excluded.

International Passenger Survey (IPS)
The International Passenger Survey (IPS) is a survey of a random sample of passengers entering and leaving the UK by air, sea or the Channel Tunnel. Over a quarter of a million face-to-face interviews are carried out each year. The IPS is carried out by ONS.

New Commonwealth
The New Commonwealth statistical grouping consists of African Commonwealth countries (Botswana, Cameroon, The Gambia, Ghana, Kenya, Lesotho, Malawi, Mauritius, Mozambique, Namibia, Nigeria, Seychelles, Sierra Leone, Swaziland, Tanzania, Uganda, Zambia and Zimbabwe), Indian subcontinent countries (Bangladesh, India, Pakistan and Sri Lanka), and other Commonwealth countries in the Asian, Caribbean, and Oceania regions.
It also includes British Dependent Territories and British Overseas citizens. Rwanda was admitted to the Commonwealth in November 2009, but the definition for this statistical grouping has remained unchanged. Zimbabwe withdrew from the Commonwealth in December 2003, but again the definition for this grouping also remained unchanged following this.

NISRA
Northern Ireland Statistics and Research Agency.

NRS
National Records of Scotland.

Old Commonwealth
The Old Commonwealth statistical grouping consists of four countries: Australia, Canada, New Zealand and South Africa.

Switchers
These are responders to the IPS who intended to remain in or out of the UK for 12 months, but actually spent less than a year (migrant switchers) and those who believed they would be staying or leaving for less than a year but actually spent longer (visitor switchers).