Article

Population statistics research update: October 2017

A description of ONS research into improving standard population statistics outputs, including recent progress and plans for future research.

Table of contents

1 . Introduction

This article describes the research and development planned for our standard population statistics outputs as at October 2017. To help users quickly find the information they need, the article includes a description of each project similar to that provided in the previous research update in June 2017, followed by an update paragraph on progress since June.

A separate programme of work looking at alternative methods of producing statistics following the 2021 Census is available.

We welcome your comments on our proposed work and any suggestions for other research and development that would increase the value of our statistics to you.

The following sections will discuss and update on the progress of the various research and development planned for our standard population statistics outputs as at October 2017.

2 . Internal migration

The largest element of population change at the local authority level is internal migration – that is, people moving from one local authority in the UK to another. The methods for estimating internal migration were significantly improved in 2013 when the availability of new Higher Education Statistics Agency (HESA) data allowed a simpler and more reliable approach to be adopted for estimating migration of students. This further research has four main aims, which are detailed in this section.

Allowing the production of estimates for 2016 following the closure of the NHSCR system (extracts from which are used in calibrating the estimates)

The fundamental approach to producing estimates of internal migration moves is comparing a file (produced by linking the Patient Register to the HESA dataset of people in higher education) of where people were living at the start of the year with the corresponding file of where they were living at the end of the year. Any record with a change of address must have moved (that is, migrated) during the year.

However, this approach misses some types of moves – where people have moved several times during the year or where they were not present at the start, or the end, of the year (for example, babies born during the year). Extracts from the NHSCR data are used to adjust the initial internal migration estimates to reflect these moves.

As the NHSCR data source was discontinued in February 2016, the production of the 2016 estimates (published in June 2017) required an alternative method of estimating these additional moves. Our research indicated that using the previous year's adjustments was likely to be the best approach, both in terms of accuracy and simplicity and we adopted this method for the 2016 estimates. Estimates for later years should not require such adjustments.

Quality assuring the Personal Demographic Service (PDS) data source as an appropriate replacement for the Patient Register (PR) data source when the latter is discontinued

The PR data source currently used in producing the internal migration estimates is due to be closed in autumn 2017. However, the alternative PDS data source now available to us seems to have several advantages over the PR. We have started quality assuring the PDS data and checking that it is appropriate for use in producing the estimates, with a view to moving to using the PDS data in the mid-2017 population estimates.

Developing improved models for the destination of students after they graduate

The current methodology is thought to be an improvement on the previous method in that it is much more accurate in estimating migration of students to their place of study. However, there is still scope for improving methods for estimating the destination of students who move after leaving higher education.

At present, the assumption is that a student not updating their PR record upon leaving higher education will either stay in the local authority in which they lived while studying or return to their previous address as recorded on the PR, with an increasing probability of moving to their previous address as time goes on. We are developing an alternative approach of applying an Origin-Destination matrix to these students (so, for example, 10% of students in Southampton who don't update their PR record when leaving higher education move to Portsmouth).

Improving estimation of migration moves within the year not identified by comparing addresses at the start and the end of the year

As described previously in this section, the current methodology uses NHSCR data to adjust the initial estimates to allow for moves of people who were only present at either the start or the end of the year, but not both. We are investigating an alternative approach of directly identifying these moves, for example, by linking the start-year and end-year population stocks file with registrations of births and deaths.

Methods developed in the first of these four strands were used in the 2016 mid-year population estimates in June 2017. Methods developed in the remaining three strands are planned to be implemented for the 2017 mid-year population estimates in June 2018. If possible, we will also use the new methods to produce a revised back-series of population estimates from 2011 to 2016.

Update: October 2017

This project is on track for delivery of methods for 2017 and later years. Limitations in the available data sources, and thus in the derived internal migration estimates, mean we may not use the new methods to revise estimates for previous years. We published more information on the new methods in the methodology document accompanying the release of the 2016 mid-year population estimates in June 2017, and are looking at possible further developments of the methods to better meet user requirements for internal migration estimates.

3 . Local authority emigration

Estimates of emigration at the national level are derived from the International Passenger Survey (IPS), with adjustments made to reflect for asylum seekers. Disaggregating these estimates to the local authority level relies on complex regression models, which were first set up in 2011. This project aims to update these regression models to reflect as well as possible the current characteristics of migration at the local level. We have developed a proposed model to replace the existing model and are arranging for a further independent evaluation to ensure that this new model meets the quality standards required.

The proposed model is broadly similar to the existing model: both being based on a Poisson Regression approach. However, there are three differences between the models, which are detailed in this section.

Changing the explanatory variables

The proposed model uses a different set of explanatory variables to those used in the existing model. As with the existing model, the set of explanatory variables has been selected by a combination of manual selection and algorithmic stepwise selection, providing a compromise between strict statistical optimality (based on the estimation period) and intuitively plausible explanatory variables.

Using an "offset term"

The proposed model introduces an "offset term" in the regression equation to reflect the size of the population of the local authority. This term effectively changes the model from a direct model of counts to a model of rates (that is, modelling emigration as a proportion of the start population). Such a term is a standard feature of Poisson models where the "population at risk" is different for different observations.

Removing new migration geographies for outflows from constraining process

The current model uses a non-standard geography called new migration geographies for outflows (NMGos). An NMGo is a collection of local authorities that is treated as a single source of emigration when constraining to the IPS. The proposed model eliminates the use of NMGos. The removal of this constraint means we are making less use of the IPS data in allocating emigration, but also removes the practical problem that an over-estimate for one local authority could result in counter-balancing under-estimates in neighbouring local authorities. Removing the use of these non-standard geographies should also make the estimates more transparent and based on standard definitions.

Update: October 2017

This work is now complete and we published details of the proposed updated model in the methodology document accompanying the release of the 2016 mid-year population estimates in June 2017. We plan to implement this new model alongside the new internal migration methods for the 2017 mid-year population estimates in June 2018. We plan to use the new methods in a revised back-series of population estimates from 2011 to 2016.

4 . Sex and age distribution of international migrants

The mid-year population estimates currently use 2011 Census data in applying a sex and age distribution to international migrants at the local authority level. We investigated whether that is still the best approach or whether an alternative approach, for example, based on administrative data, should be adopted.

This work has concluded that the sex and age distribution of migrants has remained sufficiently stable since 2011 and that we should retain the current approach based on 2011 Census data rather than switch to other datasets, which are more up-to-date but less directly related to our target concept of the usual resident population.

Update: October 2017

This project is now closed.

5 . Population characteristics

We have acknowledged user interest in a new set of population estimates by ethnic group. We launched a project in September 2016 looking at whether such estimates (consistent with the mid-year population estimates) can be derived primarily from the Annual Population Survey (APS). If this is successful we would look to extend these methods to cover country of birth and nationality.

For many years we have published tables of the population by country of birth and by nationality as part of our suite of migration-related statistics. These estimates are derived from the APS. Since that survey does not cover some people not living in households, these estimates are not consistent with the mid-year population estimates (which cover all usual residents in an area). This project is looking at combining APS data with other data sources (in particular, the 2011 Census) to produce estimates of the population by country of birth, nationality and ethnic group, which are fully consistent with the standard mid-year population estimates.

Update: October 2017

In August 2017 we published a research report describing this work along with another report describing a method for using survey and administrative data to produce estimates by ethnic group as part of our research into developing an Administrative Data Census approach.

We are continuing to assess the quality and usefulness of these research outputs, and expect to make a decision in early 2018 on whether we should look to make the population characteristics estimates described previously a standard output.

6 . National population projections: uncertainty

We have considered whether to produce probabilistic measures with our biennial national population projections. The concept of a probabilistic approach would be to provide an indication of likelihood, allowing statements such as "there is an 80% chance of the UK population being between x and y million in 2030". It would also allow production of variants with similar likelihood measures.

We recognise the potential value of such an approach and thank the University of Southampton for supporting our research. However, for now we intend to focus our resource on developments such as sub-national variant projections, where there is a stronger user interest. Developing probabilistic projections is not a current priority.

We continue to welcome any views on probabilistic projections, or on projections more generally.

7 . National population projections: mortality assumptions

Assumptions on future mortality rates are one of the inputs into the National Population Projections (NPPs). We have been working with the University of Southampton (UoS) to investigate whether a purely model-based approach can produce appropriate projected rates.

At present, mortality assumptions in the NPPs are produced using a combination of a statistical model and expert advice from the ONS Demographic Analysis Unit and a representative of the Government Actuary's Department. This research has investigated modelling mortality improvement rates using a generalised additive model, with separate models for infant mortality and old-age mortality.

The proposed approach presents a number of advantages when compared with the current method: for example, it makes fuller use of available data and is more transparent and efficient. However, there are also possible disadvantages – there is a risk of inconsistency with previous projections and the model may be more difficult to explain to users.

Update: October 2017

Following testing and expert review, the NPP Committee supported the use of the new model for the 2016-based NPPs. The current method and the proposed method were run for the 2016-based projections to allow validation and comparison of results. Following the NPPs consultation and further testing of the UoS model, we decided to retain the current method to produce the mortality assumptions for the 2016-based projections.

Further testing and development of the UoS model will be carried out over the next two years with a view to transitioning to the model for the 2018-based projections. The projections will be published on 26 October 2017.

8 . Sub-national population projections for England

We are working on improvements to the methodology of the sub-national population projections (SNPPs), both to improve accuracy and to increase coherence with other population statistics. These will be incorporated in the methods used for the 2016-based sub-national population projections for England, scheduled for May or June 2018.

Update: October 2017

Since the June 2017 update we have completed further work on the distribution of armed forces and international migration. Around the start of 2018 we will share more detailed information on all methodological developments affecting the sub-national projections.

9 . Quality assurance of administrative data sources

Each of our regular statistical releases is accompanied by a Quality and Methodology Information report, which summarises information on the quality of the published statistics and the methods used to produce them. In addition, we are preparing reports on the quality assurance procedures adopted for each of the main administrative data sources used in producing our statistics.

This work is being conducted within the framework of the UK Statistics Authority's Quality Assurance of Administrative Data Sources Toolkit. All of the administrative data sources feeding in to our statistics have been assessed on the risk of quality issues having an impact on the statistics. This has allowed us to set out appropriate levels of documentation needed for each data source. The initial series of quality assurance of administrative data (QAAD) reports was published in early 2017 and links to the reports are available in the previous Population statistics research update for June 2017.

Update: October 2017

We are planning updates of the published reports, and additional reports including ones covering the Personal Demographic Service (PDS) and the Department for Work and Pensions' Customer Information Services data sources, to appear in the first half of 2018.

10 . Household projections for England

We announced in January 2017 that we were taking responsibility with immediate effect for the production and publication of the household projections in England, previously produced by the Department for Communities and Local Government (DCLG). We hope that this transfer of responsibility will further develop the consistency between the household projections and the national and sub-national population projections (SNPPs), and allow us to produce them with increased efficiency.

Following a discussion with DCLG on their experience of producing the household projections, and a consultation with users, we are evaluating some proposed changes to the methods with a view to including any changes in the 2016-based household projections due for publication in summer 2018.

Update: October 2017

We have been developing a new processing system through which potential methodological changes outlined in the consultation and consultation response can be tested and compared. Using the 2014-based SNPPs we will be able to test different variations in methodology, providing research and analysis for ONS and a collaborative group of experts to discuss.

11 . Adjustment for Foreign Armed Forces dependants

Whilst the methods for producing the mid-year population estimates work well for most local authorities in England, we recognise that they may not work so well for areas with a large number of Foreign Armed Forces, where their dependants are not reflected in available data sources in the same way as most international migrants. We are looking at whether we can use administrative data to reflect migration patterns of these dependants more accurately.

As with our research into internal migration and emigration described in this article, any methods developed through this work are planned to be implemented for the 2017 mid-year population estimates in June 2018. If possible, we will also use the new methods to produce a revised back-series of population estimates from 2011 to 2016.

Update: October 2017

We have developed a method for dealing with this issue by treating these dependants as a "special population" (as we do the Armed Forces themselves) in the mid-year population estimates and the sub-national population projections (SNPPs), and plan to implement this method for both the 2017 mid-year estimates and the revised back-series of estimates. We also provisionally expect to use a similar approach in the 2016-based SNPPs.

12 . Sexual identity estimates: local authorities

In October 2016 we published experimental official statistics on sexual identity of the population of the UK and its constituent countries and regions. We then investigated the feasibility of providing similar statistics, using information from the pooled APS dataset for 2013 to 2015, to local authority level.

Update: October 2017

In April 2017 we published experimental research estimates of sub-national sexual identity for the combined years of 2013 to 2015. Using the three-year APS pooled dataset to produce sub-national estimates of sexual identity does not seem to be a robust method, particularly for English counties and local authorities of Great Britain. The method means that, for a few counties and a substantial number of local authorities, we cannot publish a robust estimate of their sexual identity distribution.

Increasing the sample size, either through a larger surveyed sample or through increasing the number of pooled years with a dataset could improve sample sizes enough for us to produce and publish robust estimates for all counties and local authorities. However, rolling together more years will affect the timeliness of the estimates. We will monitor the sample sizes for sexual identity at the sub-national level within the next APS three-year pooled dataset to assess whether these estimates are robust enough to publish.

Sexual identity estimates for 2016 at national level were published on 4 October 2017.

13 . Gender identity

Work on gender identity is being conducted jointly by the Population Statistics and Census teams within ONS.

The 2021 Census topic consultation identified a need among a number of data users for information about gender identity for policy development and service planning; especially in relation to the provision of health services. These requirements are strengthened by the need for information on those with the protected characteristic of gender reassignment as set out in the Equality Act 2010. The Gender identity topic report on the consultation findings provides further information.

As we do not currently collect data on gender identity on any of our social surveys, research and testing work will inform our position on this topic. The Gender identity research and testing plan sets out the work we will do to help us determine how to meet user needs for information on this topic.

On 13 January 2017 we published the Gender identity update paper. This addresses our commitment to review the Trans Data Position Paper, which we published in 2009. The update outlines developments around the topic of gender identity. It covers:

- legislation

- the Women and Equalities Committee Transgender Equality inquiry

- data collection and question development worldwide

- our research, testing and findings so far

- the next steps and future work we will be undertaking

More information on our work on this topic is provided in our gender identity page.

Update: October 2017

In September 2017, we held a workshop with data users to confirm understanding of data requirements for gender identity. We explored the extent to which a range of possible questions would meet the data need for a count of the trans and non-binary population and discussed priorities for data collection. A report will be published in due course.

A further update on the gender identity work will be published later on in 2017.