

Population estimates for local authorities in England and Wales, new methods: February 2018

Population estimates for local authorities in England and Wales: changes to methods planned for 2018.

Contact:
Pete Large
pop.info@ons.gsi.gov.uk
+44 (0)1329 444661

Release date:
1 February 2018

Next release:
To be announced

Table of contents

1. [Introduction](#)
2. [Internal migration](#)
3. [International emigration](#)
4. [Dependants of foreign armed forces](#)
5. [Other changes](#)
6. [Annex A: Emigration model](#)

1 . Introduction

Population estimates for local authorities in England and Wales (also called “mid-year estimates” or “MYEs”) are produced using the cohort-component method. This method takes the starting population from the most recent census and then applies the estimated changes due to the various “components of change” – births, deaths and migration – since that point.

The methods used to produce the estimates for 2016 (published in 2017) are described in the [population estimates methodology guide](#). Some changes to these methods are planned for the 2017 estimates due to be published in 2018. Where possible, we will also apply the new methods to produce a revised backseries of local authority estimates from 2012 to 2016, which will also be published in 2018. This revised backseries will not change the estimates for the countries of England (as a whole) and of Wales (as a whole). A revised series of small area population estimates for areas in England and Wales will also be published to maintain consistency with the local authority estimates.

Proposed changes to the methods were first described in Appendix 2 of the [population estimates methodology guide](#) published alongside the 2016 population estimates in June 2017. The methods outlined in this article are developed from those described in that guide in the light of further research and user feedback.

We always welcome comments on our methods and plans. If you want to tell us your thoughts on the methods described in this article, please email us at pop.info@ons.gsi.gov.uk.

2 . Internal migration

Internal migration relates to moves of people between areas in the UK. Estimates of internal migration are produced using a combination of administrative data on where people are registered with GPs and where students in higher education are recorded as living. We plan to make three changes in the way we estimate internal migration for the 2017 mid-year estimates (MYEs).

“Graduate” destinations

The fundamental approach to estimating internal migration within England and Wales is to compare people's area of residence on their health registration with that in the previous year. We know that one weakness of that approach was that people moving to (or leaving) higher education might be slow to update their registration so we would not identify all the moves into student areas, or into areas where graduates tended to move to after completing their studies. We have used several methods to try to account for these moves.

We improved our methods in 2013 by linking the health registration data with data from the Higher Education Statistics Agency (HESA). The HESA data showed where students were registered by their university as living and this allowed us to make more accurate estimates of the numbers of people moving to study in each area. However, it did not tell us where people (in particular, those people who were slow in updating their health registration) moved after completing their studies. Rather than simply assume that those people stayed in the area in which they studied (which would result in over-estimating the population of that area), we used a simple model in which people completing their studies and not updating their health registration record would be assumed to move back to their health registration address over time.

We plan to improve this model for the 2017 (and subsequent) MYEs by introducing a new end-of-studies methodology – the Higher Education Leavers Methodology (HELM). This method will distribute those higher education leavers who have not updated their Patient Register address after leaving higher education, using the movement patterns of students who have previously left higher education.

The method can be summarised as follows:

- identify those people who need to have their area of residence imputed; these will be health registration records previously linked to the HESA data where there is no longer a HESA record (so the person has left higher education) and where the health registration record has not been updated during the year
- identify similar people (that is, leaving higher education in that area and not updating their health registration in the first year) from three years previously and use the health registration record to estimate the distribution of destinations (we use data from three years previously so we can collect information from a largely completed cohort of movers; using three years is judged to be the best balance between using recent data to reflect current patterns and using older data to maximise the proportion of people who have updated their health registration)
- apply the estimated distribution to those people to be imputed; the random imputation is done in such a way that there will be no systematic biases in destinations chosen but the final distribution will be very close to the initially-estimated distribution

Though we do not yet have quantitative evidence for this, we can reasonably expect the estimates produced using the HELM to be more accurate than those produced using the existing method. Recognising that higher education leavers might disperse to any of the 348 local authorities in England and Wales will mean the internal migration estimates should better reflect reality. Furthermore, by not simply keeping the higher education leavers at their HESA address or returning them to their health registration address, we can also expect the methodology to improve the number of post-student-aged individuals remaining in “student” local authorities.

It's important to note that some people do, in reality, remain in their local authority of study following higher education. The HELM does recognise and deal with this too, as the destination distributions still reflect a number of individuals staying in their local authority of study. Though we do not yet have quantitative evidence for this, we can reasonably expect the estimates produced using the HELM to be more accurate than those produced using the existing method. Recognising that higher education leavers might disperse to any of the 348 local authorities in England and Wales will mean the internal migration estimates should better reflect reality. Furthermore, by not simply keeping the higher education leavers at their HESA address or returning them to their health registration address, we can also expect the methodology to improve the number of post-student-aged individuals remaining in “student” local authorities.

In contrast to the current methodology, which continues to distribute students over time, the HELM distributes all higher education leavers to their imputed destination in the first year after they finished higher education. Although this will introduce some inaccuracy, as some of the moves informing the destination distributions took place in the second or third year after leaving higher education, this is offset to some extent by the fact that although those moves may have been recorded in the second or third year following leaving higher education, some of these will have been “lagged” moves that actually took place in the first year after leaving study. There is a further offsetting effect in that the destination distributions assume that any individuals who did not change address in any of the three years after leaving higher education remained in their local authority of study while some of these may have moved (but not updated their health registration).

As with the current methods, the approach of imputing place of residence for individual records has substantial advantages over making aggregate adjustments as any incorrect imputation would be automatically corrected when that person updates their health registration record. We plan to implement this method for the 2017 and subsequent MYEs.

Cross-border flows

The main internal migration methods are used to estimate internal migration between areas in England and Wales. Slightly different methods are used to estimate migration flows between areas in England and Wales (as a whole), and Scotland and Northern Ireland.

Prior to 2018, estimates for flows from Scotland and Northern Ireland into England and Wales were produced using health registration data obtained through the NHS Central Register (NHSCR) and Patient Register (PR) sources (unlike the estimates of moves within England and Wales, this data is not linked to higher education data to better estimate moves to and from study). We now obtain health registration data primarily through the Personal Demographic Service (PDS). Whilst the underlying data (relating to people registered with GPs) used in the estimates remains the same using the PDS, the level of detail available from that data source allows us to adopt a more straightforward approach.

For the 2017 MYE being published in 2018 and subsequent MYEs, we plan to estimate cross-border flows by using the counts of moves from Scotland and from Northern Ireland obtained from weekly extracts of PDS record changes. This will be the first use in the population estimates of data obtained through the PDS and we will publish a Quality Assurance of Administrative Data report on that data source in 2018.

Estimating within-year moves

The majority of internal migration moves reflect someone living in one area of England and Wales at the start of the year and another one at the end of the year. This type of move is called a “transition”. Not all moves are of this type, however. People may move multiple times within the year, babies might be born after the start of the year and move to a new address before the end of the year, and people present at the start of the year might move and then die or emigrate before the end of the year. These types of moves are collectively called “within-year moves”.

Within-year moves are calculated by estimating the ratio of within-year moves to transitions and then applying that ratio to the estimated number of transitions. Rather than assuming that this ratio is the same for all places and at all times, we use health registration data for each year and for areas within England and Wales to estimate these ratios. As noted previously, we now obtain health registration data via a different system to that previously used and this means we now have access to more detailed information on the area of the start- and end-place of within-year moves. However, we have not yet demonstrated that using this more detailed information results in more reliable population estimates. Therefore, we intend to follow the existing method of estimating within-year moves as closely as possible for the 2017 MYEs until we have completed further research ahead of the 2018 MYEs.

The inclusion of within-year moves in the internal migration estimates means that we will not be changing the target concept for these estimates as had been proposed in the [population estimates methodology guide](#) in June 2017.

3 . International emigration

Estimates of emigration at the national level are primarily based on the International Passenger Survey (IPS). Reliable direct estimates for individual local authorities are not possible from a simple combination of the available survey and administrative sources and since 2010 these estimates have been produced using a Poisson regression model. This produces estimates of emigration for each local authority by combining data on a number of factors that have been (statistically) identified as being correlated with (or, in a sense, “explaining”) emigration. These factors are called the “explanatory variables”.

In 2014, we launched a project to investigate whether improvements in the way we estimated local authority emigration were possible. There were two strands to this work:

- improvements to the current modelling approach – this looked at the specification of the existing regression model for total emigration from a local authority (LA) and whether changes in the structure or choice of explanatory variables could produce an improved model
- direct use of administrative data – this looked at whether an alternative approach based on allocating different streams of emigration (again, estimated using the IPS) directly to LAs using administrative sources (a similar approach to that used in the LA-level immigration estimates for the mid-year estimates (MYEs)) was possible

The second strand of work concluded that we did not yet have appropriate administrative data to develop such an approach. We therefore looked at improving the current model.

Annex A provides some detail on the model specification.

The proposed model has the following differences to the currently used model.

Offset term

The current model uses explanatory variables expressed as counts to model emigration as a count. The proposed model includes an “offset term” representing the (previous year’s) population of an area. This transforms the model from a model of counts to a model of rates – in effect modelling the “risk” of a resident emigrating over the year. This is a standard approach adopted for such models and ensures that the modelled emigration remains related to the population at risk.

Removal of NMGoS as a constraint

The current model constrains the initial local authority estimates to the three-year average IPS estimate for the New Migration Geographies for Outflows (NMGoS). The NMGo areas are groups of local authorities designed as an interim geography between local authority and region and intended to help disaggregate the IPS data to subregional levels. NMGoS are not a standard geography and it was not clear that their inclusion in the methods improved the quality and usefulness of the estimates. Therefore, the proposed model does not constrain the initial local authority estimates estimates to NMGo figures. The removal of this constraint means we are making less use of the IPS data in allocating emigration but also removes the practical problem of an over-estimate for one area being directly responsible for corresponding under-estimates in neighbouring areas.

Explanatory variables

The proposed model uses a different set of explanatory variables to those used in the existing model. The proposed set of variables, together with those used in the existing model, are shown in Annex A. The data sources used in the specification of the current model were the best available at the time but it was always planned that the variables used were to be reviewed as new data sources became available. The proposed variables have been selected by a combination of manual selection and statistical procedures. As with the current model, the set of explanatory variables chosen remains the same for each year, but the “weight given” to each variable can change to reflect current patterns.

We plan to implement this change in methods in the population estimates for 2017 and subsequent years and in the revised backseries of estimates for 2012 to 2016.

4 . Dependants of foreign armed forces

The standard methods of estimating internal and international migration are thought to work well for most local authorities but less well for the few local authorities with sizeable populations of foreign armed forces. While the forces personnel are correctly reflected using “special populations” (that is, populations estimated directly from administrative sources rather than updated through the usual cohort-component method), the families (dependants) of those personnel are not always accurately reflected as they have much less interaction with UK administrative sources and emigration flows are too small, and in too few areas, to be accurately estimated using the local authority emigration model.

Table 1 summarises the estimated size of the foreign armed forces and dependants population for the five most affected local authorities.

Table 1: Foreign armed forces and dependants as a proportion of total population, selected local authorities, 2015

Local Authority	Foreign Armed Forces population (including dependants)	Total population	Thousands	
			Foreign Armed Forces population (including dependants) as % of Total population	
Forest Heath	11.4	63.7	17.9	
East Cambridgeshire	1.4	87.3	1.6	
St Edmundsbury	1.3	112.5	1.2	
King's Lynn and West Norfolk	1.4	151	0.9	
Breckland	1	135.5	0.8	

Source: Office for National Statistics

The issues described lead to the following anomalies in the population estimates for affected areas:

- lower than expected estimates of females in the age 20s to 30s bracket from uncounted new dependants (not “captured” when moving in)
- higher than expected estimates of females in the age 30s to 40s bracket due to those dependants counted in the 2011 Census being aged on without being migrated out
- high counts of children born since the 2011 Census being aged on without being migrated out

We plan to address these issues by adopting the approach of special populations for the dependants of foreign armed forces as well as the personnel themselves.

We have obtained administrative data relating to foreign armed forces' dependants for years since 2011 (and expect to obtain similar information for future years). We will combine this data with those for the personnel themselves to derive the "foreign armed forces-related population". As the administrative data relates to forces at each base, we transform this to a residence basis by using a base-to-residence matrix derived from the 2011 Census (this is the same approach as currently used). Calculating the adjustment necessary to reflect changes in the special population is then a matter of subtracting last year's special population and adding the special population for this year.

The exception to this approach is the method for estimating zero year-olds. At the beginning of the process of calculating the MYEs, all zero-year-olds of the previous year's special population must be subtracted. This is to avoid ageing on any zero-year olds that will be accounted for at the end of the MYE calculation process through addition of the current year's special population one-year-olds.

However, when the current year special population is added at the end of the MYE calculation process, none of the zero year-olds should be added. The zero year-olds in the current year special population will already have been counted into the population because they were born in the UK and are part of the births data that is added to the MYEs as a matter of course. Some additional special population zero-year olds will have been born outside the UK and migrated in within the last year and won't be counted, but these would be broadly balanced by those zero year-olds born to the special population in the last year who then migrate out of the country, assuming a broadly similar resident special population over the year. There may be larger variations in this fraction if bases increase or decrease their personnel significantly.

We expect the new methods to substantially address the previously-identified anomalies for affected local authorities, maintaining age profiles more similar to those seen in the 2011 Census and stabilising sex ratios.

We plan to implement this change in methods in the population estimates for 2017 and subsequent years and in the revised backseries of estimates for 2012 to 2016.

5 . Other changes

We are also using the release of a revised backseries as an opportunity to introduce some very minor improvements, for example, where new data has become available or to improve consistency of the methods. These improvements include:

- changing the method used to estimate the age and sex distribution of asylum seekers at the local authority level for 2012 and 2013 to be consistent with that used for later years
- revising the local authority distribution of immigration distribution for 2015 and 2016 in the light of administrative data not available at the time those estimates were originally produced and using a more refined method of linking data
- a very minor change to the Coventry/Warwick adjustment, taking on data not originally available to derive a more reliable distribution of affected students between the two local authorities in 2014 and 2015

6 . Annex A: Emigration model

This section describes the proposed Poisson regression model for estimating emigration for local authorities in England and Wales.

The proposed model consists of two stages: the use of the regression model to produce initial estimates for each local authority, then the constraining of those initial estimates to regional totals from the International Passenger Survey (IPS).

Model specification

Production of initial estimates

The proposed regression model is:

EQN 1

$$y_{est,i} = (w_{est,i})$$

EQN 2

$$w_{est,i} = a_i + B'X$$

where:

$y_{est,i}$ is the initial estimate of the number of emigrants leaving LA i.

$w_{est,i}$ is the log of the initial estimate of the number of emigrants leaving LA i.

a_i is the "offset term", defined as the log of the population of LA i at the start of the year.

B' is a row vector of estimated parameters with values constant for all LAs (but varying each year). The values of the parameters are estimated by initially fitting the model with the log of the average of the most recent three years of IPS data for that LA as the dependent variable.

X_i is a column vector of (logged) explanatory variables (or covariates) with values relating to LA i.

When $w_{est,i}$ has been calculated using EQN 2, this is exponentiated using EQN 1 to derive $y_{est,i}$, that is the initial estimate of emigration from LA i.

Constraining the initial estimates

Once the initial estimates have been produced these are constrained by simple scaling to the IPS regional total for the current year.

EQN 3

$$y_{fin,i} = y_{est,i} \times (m_R / \text{SUM}_{\text{All } i \text{ in } R} (y_{est,i}))$$

where:

$y_{fin,i}$ is the final estimate of the number of emigrants leaving LA i .

m_R is the IPS estimate of emigrants leaving Region R in the year.

Explanatory variables

The explanatory variables used in the current and in the proposed models are set out in Table 2.

Variables in the proposed model were selected using a combination of manual selection (where variables are forced into the model) and algorithmic stepwise selection where variables are iteratively added to and removed from the model according to statistical criteria designed to maximise the reliability of the model.

Table 2: Explanatory variables used in the current and in the proposed emigration models

Description of variable	Current	Proposed
2011 Census		
Usual residents with limiting or long-term illness		Y
Usual residents of Mid-or Far East-Asian country of birth		Y
Usual residents of Oceania country of birth	Y	Y
Usual residents of North American country of birth	Y	Y
Usual residents of South-Asian country of birth		Y
Usual residents living in hostels	Y	Y
Usual residents of African country of birth	Y	
Annual Population Survey		
Accommodation owned with mortgage		Y
Accommodation rented		Y
Retired people		Y
Employed aged 16 and over	Y	Y
Birth Registrations		
EU2 country of birth of mother		Y
EU8 country of birth of mother		Y
Higher Education Statistics Agency (HESA) Student Record		
Students of EU (EU2, EU8 & EU15, excl. UK) nationality in Higher Education in their final year of study, aged 20 to 25		Y
Migrant Worker Scan (MWS)		
Migrant Workers of EU8 nationality	Y	
Migrant Workers of EU8 nationality aged 26-59		Y
ONS Mid-Year Population Estimates		
International in-migration estimates for the year of interest		Y
Patient Register		
Registered patients aged 5 to 15		Y

Source: Office for National Statistics