

## Methodology for Producing Revised Population Estimates for 1992 to 2000

### **1. Introduction**

On 7 October 2004, ONS published a revised back series of population estimates for 1992 to 2000. These supersede both the interim revised national estimates for 1992-2000 published on 23 October 2003 and the revised national and subnational estimates for 1992-2000 published on 13 February 2003.

These new estimates:

- have been revised in the light of the Local Authority Population Studies (see [http://www.statistics.gov.uk/about/Methodology\\_by\\_theme/LAStudies.asp](http://www.statistics.gov.uk/about/Methodology_by_theme/LAStudies.asp)) and are fully consistent with the latest mid-2001 population estimates published on 9 September 2004 (<http://www.statistics.gov.uk/statbase/Product.asp?vlnk=601&More=N>).
- have been produced following in-depth research by ONS into the best methodology to use.

### **2. Starting Points for Producing a Back Series**

#### **2.1 Population estimates rolled-forward from the 1991 Census**

The starting point for all methodologies is the set of population estimates for 1992 to 2000 that were rolled-forward from the 1991 Census.

These rolled-forward estimates have been adjusted to take into account:

1. the change in the 1991 base arising from the 1991 Census adjustment (see [www.statistics.gov.uk/downloads/theme\\_population/Meth\\_popES\\_82\\_2000.pdf](http://www.statistics.gov.uk/downloads/theme_population/Meth_popES_82_2000.pdf).) and
2. for the first time, the rolled-forward estimates for 1992 to 2000 have also been updated to include the revisions to international migration estimates for the 1990s that were published in June 2003 (see [www.statistics.gov.uk/downloads/theme\\_population/Methodology%20for Revised International Migration Estimates.doc](http://www.statistics.gov.uk/downloads/theme_population/Methodology%20for%20Revised%20International%20Migration%20Estimates.doc).) This provides the most accurate base to which the remaining intercensal differences can be added.

#### **2.2 Intercensal difference**

The results from the 2001 Census published in September 2002 identified a difference of 1.1 million between the 2001 Census estimates for England and Wales and the 2001 mid-year estimates rolled-forward from the 1991 Census. By September 2003, a large part of the difference had been explained by the 1991 adjustment (351,000), the adjustments to international migration estimates for the 1990s (305,000) and the Longitudinal Study adjustment (193,000 – see link below). The Local Authority Population Studies and Longitudinal Study Consequential Adjustment published in September 2004 explained a further 63,000 of the difference. For further information on these adjustments, see [http://www.statistics.gov.uk/about/Methodology\\_by\\_theme/LAStudies.asp](http://www.statistics.gov.uk/about/Methodology_by_theme/LAStudies.asp) and [http://www.statistics.gov.uk/about/methodology\\_by\\_theme/revisions\\_to\\_population\\_estimates/downloads/methodology\\_for\\_revision\\_to\\_mid-2001.pdf](http://www.statistics.gov.uk/about/methodology_by_theme/revisions_to_population_estimates/downloads/methodology_for_revision_to_mid-2001.pdf)

This still leaves an unexplained intercensal difference of 209,000 in 2001. The divergence is believed to be due to varying definitions of ‘usual residence’ and remaining difficulties in estimating international migration. This difference of 209,000 needs to be allocated back over the preceding decade in order to produce a back series for 1992 to 2000 that is consistent with the population estimates for both 1991 and 2001. The following section describes possible ways in which this difference can be allocated.

### **3. Background to possible methods for producing the back series**

#### **3.1 Simple period methodology**

A period effect is where the difference is related to a particular age group, whatever the year in question. The effect is not aged back over time. The simple period methodology also assumes that the divergence has built up gradually and linearly over the decade and so one-tenth of the difference is allocated to that age group in each year. Males and females are considered separately.

Therefore using the simple period methodology, the difference for each age is split evenly across the rolled-forward estimates for that age in each of the preceding 10 years. For example, to allocate the 2001 difference of  $-9,500$  for males aged 30 back over the previous decade, 950 males would be taken off the number of males aged 30 in 1992, 1993 and each year up to 2000.

This methodology has the advantage of being relatively straightforward to understand and apply. The period approach is likely to hold in particular for younger ages, particularly where there is a clear change in behaviour at a particular age, for example, those taking 'gap years' at age 18 or 19. However for other ages it may represent an over-simplification of the situation since it ignores genuine variations by age that may accumulate over time.

#### **3.2 Simple cohort methodology**

Whereas the period effect assumes that differences have been built up in particular age-sex groups over time, cohort methodology assumes that the divergence has accumulated over time in age-sex cohorts. Under the cohort approach, the difference therefore needs to be aged back over time. The simple cohort methodology also assumes that the divergence has built up gradually and linearly over the decade and so one-tenth of the difference is allocated to that age-sex cohort in each year. Males and females are considered separately.

To illustrate the simple cohort methodology, the difference for each age is split evenly across the preceding 10 years to the cohort that reached that age in 2001. For example, to allocate the 2001 difference of  $-9,500$  for males aged 30 back over the previous decade, the estimate for males aged 29 in 2000 would be reduced by 950, as would the estimate for males aged 28 in 1999 and so on back to males aged 21 in 1992.

The simple cohort method is more sophisticated than the simple period approach and is likely to be the most appropriate method for many age groups. However it may cause some unusual effects at single year of age level in younger age groups.

#### **3.3 Migration weightings**

One of the key difficulties with both the simple period and simple cohort methods is the assumption that the difference should be evenly split over the preceding ten years. In reality it is unlikely that the 2001 difference has accumulated linearly over the decade. If this is the case, it may not be reasonable to assume that one-tenth of the difference should be applied to each year.

Given that the remaining unexplained difference is likely to be attributable to remaining difficulties in estimating migration accurately and issues associated with the definition of usual residence, the level of migration is the key quantifiable factor that can be used to weight the allocation of the difference over the decade. The level of out-migration was chosen for investigation since a measure of gross rather than net flows was required and it is thought that the remaining difference is in part due to flows out of the country that are not captured by the current migration definition.

Since migration can vary by both age and time, ONS has investigated weighting the difference by age and by both age and time. Because of the nature of the variables, time weightings can be applied to either the period or cohort methods of allocating the difference, while age weightings can only be applied using the cohort methodology.

#### **4. Final method selected**

The final method selected by ONS for producing the back series is the cohort method, weighted for migration by both age and time, and with an additional period adjustment at ages 18 and 19.

The cohort method was chosen, as it is generally more reasonable to assume that divergences build up within cohorts over time. However various innovations have been included to address certain problems that have been identified with the simpler cohort model, in particular the assumption of linear divergence over the decade.

##### **4.1 Age weightings**

Since migration varies considerably by age, the allocation of the differences back over cohorts has been weighted according to the level of migration at each age. This means that, for example, the difference for males aged 20 in 2001 will be weighted more highly in 2000 and 1999 when the males were 18 and 19 than in earlier years when the males were aged 17 or below and had a lower propensity to migrate.

The national single year of age and sex distribution of out-migrants from the International Passenger Survey has been used to produce the weightings both at the national and the subnational level (since these data would not be sufficiently robust at single ages at the subnational level).

##### **4.2 Time weightings**

Variation in the level of migration over time is also a known issue; therefore it is reasonable to assume that the accumulation of the divergence over time may be correlated to the changing levels of migration over time. This means that, for example, the national difference at a particular age will be weighted more highly for the years ending mid-2000 and mid-2001 where migration outflows are relatively high compared to previous years.

At the national level, annual out-migration as measured by the International Passenger Survey was used to weight the allocation of the difference over calendar time. At the subnational level, the sum of internal outflows (to the rest of the UK) and international outflows from each Local or Unitary Authority was used to weight the allocation of that Authority's difference over calendar time.

##### **4.3 Period Adjustment.**

The period adjustment was introduced to address a specific issue for 18 and 19 year olds. Even with the time and age weighting, the resultant figures can still lead to estimates that would result in distorted rates calculated at single year of age for this age group. This is consistent with a significant period effect for the late teens i.e. the discrepancy is associated with the age rather than the cohort for this group of people. It is thought that this observation may well be related to 'gap years' taken abroad. Uncorrected, the method would lead to distorted rates calculations for younger ages.

In order to correct for this issue, a proportion of the difference observed at ages 18 and 19 has been allocated by a period calculation within the overall method. This proportion was determined by comparing the relative size of measured migration at 18 and 19 years old with migration levels at younger ages.

#### **4.4 Other Issues**

Due to these three adjustments, some constraining was necessary in order to ensure that total differences accumulated over each cohort added up to the difference observed in 2001.

Virtually all areas were calculated using the methodology set out above. However, substantial quality assurance of the back series indicated the need for a revised approach in three areas.

Low numbers in the City of London and the Isles of Scilly mean that smoothing by single year of age was required in some years in order to produce plausible results. The period adjustment was also removed in the Isles of Scilly since it led to distorted figures. In Kensington and Chelsea, the period effect leads to distorted figures at younger ages. This is thought to be due to the pattern and scale of the difference observed at younger ages. Therefore, for Kensington and Chelsea no period adjustment is applied in the calculations.

Population data for 1992 to 2000 and further information on population estimates, can be found at [www.statistics.gov.uk/popest](http://www.statistics.gov.uk/popest)

Michelle Butler, Julie Jefferies, and Jonathan Swan  
October 2004  
Population Estimates Unit  
Office for National Statistics