

Methods used to revise the 1982-2000 annual mid-year population estimates for England and Wales.

Abstract

ONS produces mid-year population estimates annually, which are based on updating from the most recent Census. Therefore, whenever results become available from a Census, a new base is created for the population estimates. This has implications for the historic series, which needs to be revised to be consistent with both the past and the most recent Census. This note describes the methodology that has been used for revising the mid-year population estimates following the availability of results from the 2001 Census. The revised mid-1991 estimates were published on 13th February 2003, the 1992-2000 series on 27th February and the 1982-1990 series on 27th March 2003.

Introduction

ONS produces the annual mid-year estimates of the resident population of England and Wales using a cohort component method, which is a standard demographic approach. In simplified form, this involves taking a base population, ageing it on and then making an allowance for natural change and net migration. The base population is the most recent Census, so when results become available from a new Census, rebasing of the annual population estimates is required.

An important implication of the rebasing of mid-year estimates following availability of census data is that the previous intercensal series of population estimates need revision in order to bring them into line with the new Census. The scale of these revisions is likely to vary by area, age and sex. The revisions required following the availability of results from the 2001 Census were quite large in some cases, particularly where there was significant migration (both internally within the country and internationally) during the intercensal period, as this is the most difficult component of population change to estimate accurately. The revisions are needed to create a consistent series for those concerned with analysing change over time, either in population or in derived variables such as birth and death rates that require population.

The revised series replaces previously published population estimates.

Revising the 1982-2000 population estimates

The intercensal difference

The 2001 Census results showed that previous mid-year population estimates (MYEs) which were produced during the nineties were overestimates of the population. The difference between the 2001 rolled forward MYE and the 2001 Census based MYE was 1,140 thousand for England and Wales as a whole. This difference is due to:

- Overestimation of the mid-1991 base that was used to produce the rolled forward MYEs during the nineties. This was due to overestimation of the impact of 1991 Census underenumeration on the mid-1991 population estimates. With the information that is now available from the 2001 census it is estimated that the mid-1991 base was too high by 351 thousand people. Thus the revised 1991 base is lower by that amount.

- Overestimation of population change over the 10 year period since 1991, which is thought to be mainly due to underestimation of outward migration. This accounts for the majority of the remainder of the observed difference (789 thousand) which is to be explained.

As is usual after a census, ONS have revised the past intercensal MYEs to be consistent with the 2001 census. Since the 1991 base has been revisited, it was necessary to revise the population estimates for the 1980's and the 1990's. The revised MYEs for 1991 were published on 13 February 2003, the revised MYEs for 1992-2000 on 27th February and the revised MYEs for 1982 to 1990 were published on 27th March 2003.

Stages of the method

There are four stages in applying the method to produce revised estimates for E&W. These stages are:

- Revising the mid-1991 population estimates
- Applying the revision of the mid-1991 base to the 1992-2000 series
- Revising the 1992 to 2000 MYEs
- Revising the 1982 to 1990 MYEs.

These four stages are described.

Revising the 1991 MYEs

After the 1991 census it was apparent that there was a degree of underenumeration particularly for certain population groups e.g. young men, elderly women. To compensate for the underenumeration of young men, adjustments were made to bring the observed sex ratios in line with predicted sex ratios based on an extrapolation of the trend seen between 1971 and 1981. Local authority districts were grouped into six broad area types and the same adjustment factor was applied to each local authority within the group. Adjustment factors were calculated for different age groups.

Sex ratio evidence from the 2001 Census shows that this extrapolation was not appropriate for all areas.

The 1991 adjustments were therefore revisited using the following process at the level of Local/Unitary Authority district (LAD). The process used to revise the mid-1991 estimates was different for each of the three broad age-groups that were subject to revision, these were the under ones, ages 1-44, and ages 80 and over. No revision was made to the 45-79 year-olds as there were no additional adjustments for these ages in the 1991 MYEs for England and Wales.

The approach used for 1-44 year-olds assumes that the 1991 sex ratios can be interpolated from straight lines joining the 1981 and 2001 sex ratios. This procedure was followed by quinary age group for each local/unitary authority area, generating a series of derived 1991 sex ratios. Each of the derived sex ratios was then multiplied by the published 1991 female MYE to give a new male estimate. The new male estimates were constrained to ensure that they were no higher than the published 1991 male MYE and no lower than the 1991 census estimate. In cases where the census estimate was higher than the MYE, the over-riding constraint was that the new estimate was not higher than the MYE. Finally, the new quinary male estimates were apportioned to single year of age using distributions derived from the published 1991 MYEs.

The assumption that sex ratios follow a straight-line trend between 1981 and 2001 is plausible for most local/unitary authorities. However this method may not work well in a few areas with large populations of armed forces. In particular, where there has been a change in the armed forces population after 1991 that was sufficient to impact on the sex ratio of 15-44 year olds. In these cases steps were taken to ensure that the revised mid-1991 estimates reflect the 1991 armed forces population. Thus less men were taken out of these areas than would be the case in a straight application of the sex ratio interpolation method.

The approach used for 0 year-olds was similar to that used for 1-44 year-olds. However sex ratios by single year of age at the level of local authority are prone to wide fluctuation and an irregular sex ratio in either 1981 or 2001 would have led to an implausible interpolated sex ratio in 1991. The interpolation of 1991 sex ratios for 0 year olds was, therefore, conducted at the level of Government Office Region. The resulting sex ratios were then applied to each local/unitary authority female estimate to give new male estimates. These estimates were lower and upper bounded in the same way as the estimates for 1-44 year-olds.

It follows that the revision procedure for ages 0-44 calculated new male estimates only. The female estimates were left unchanged from those published post 1991.

The approach for 80+ year-olds was carried out at a national (England & Wales) level. The number of people in the 2001 published MYEs aged 90+ by single year of age were added to the number of deaths at each age between mid-1991 and mid-2001 to give estimates by single year of age and sex for 1991. It was assumed that the net effect of international migration for these age groups was zero. The differences between these estimates and the published estimates were then apportioned sub-nationally using 1991 unadjusted Census data. New 1991 MYEs were then calculated by deducting the apportioned differences from the published MYEs.

Tables 1 and 2 show as expected the largest revisions to 1991 were in areas/age-sex groups that received the largest underenumeration adjustments in 1991. This was for England, for males aged up to 44 and females aged 85 and over.

Table 1: Total 1991 revision by sex

Country	Persons	% of total revision	Males	% of total revision	Females	% of total revision
England and Wales	-351,500	100%	-313,800	89%	-37,700	11%
England	-333,000	95%	-297,500	85%	-35,500	10%
Wales	-18,500	5%	-16,300	5%	-2,200	1%

Table 2: Total 1991 revisions by age (note: not all age groups cover the same number of years)

Age	Revision	% of total revision
0	-4,000	1%
1-14	-48,700	14%
15-19	-17,800	5%
20-24	-54,400	15%
25-29	-91,800	26%
30-34	-53,800	15%
35-39	-20,000	6%
40-44	-14,500	4%
45-79	0	0%
Males 80+	-8,600	2%
Females 80+	-37,700	11%
Total	-351,500	100%

In addition the rebasing revisions are mainly in the areas that received additional adjustments for underenumeration post 1991. Table 3 shows that these were the main metropolitan areas, non-metropolitan cities and Inner London in England as well as the three main local authorities Swansea, Cardiff and Newport in Wales.

Table 3: Total 1991 revisions by 1991 Area Type

1991 Area Type	1991 Adjust. Factor		1991 Rebasing Revision	% of total revision	Ave revis	No of Areas
	M20-24	M25-29				
Main Met Areas	1.21	1.25	-50,500	14%	-5,600	9
Inner London	1.19	1.22	-28,100	8%	-2,000	14
Non-Met Cities	1.24	1.16	-49,100	14%	-2,000	24
Cardiff, Newport, Swansea	1.24	1.16	-6,300	2%	-2,100	3
Outer London	1.09	1.11	-32,000	9%	-1,700	19
Other Met Areas	1.05	1.10	-41,400	12%	-1,500	27
Other Non-Met Areas	1.05	1.05	-132,000	38%	-500	261
Rest of Wales	1.05	1.05	-12,200	3%	-600	19

The individual local authority areas with the highest 1991 revision are shown in Table 4. The Local/Unitary Authorities receiving the largest revisions were all in area types that received a large adjustment in 1991: 6 are main metropolitan areas, 3 are Inner London Boroughs and 1 is a non-metropolitan city.

Table 4: Totals by area - LAs/UAs with most people removed by size

Rank	LA/UA	Revision
1	Birmingham	-10,600
2	Leeds	-10,500
3	Sheffield	-8,900
4	Bradford	-6,200
5	Liverpool	-5,100
6	Bristol	-4,800
7	Haringey	-4,400
8	Newham	-4,000
9	Wandsworth	-3,700
10	Stoke-on-Trent	-3,700

Some of the areas that received a large 1991 adjustment do not have many people removed in the current revision. For example, the Inner London boroughs of Southwark and Lambeth are two but not the only examples. This arises because the original adjustments, although not sufficiently fine tuned, were appropriate in some places. Some areas that did not receive a high extra adjustment post 1991 have received a moderate rebasing adjustment now. However, no area has been revised below the level of the 1991 Census base.

Finally, as shown in Figure 1, the method produces revised 1991 estimates with sex ratios for the key age-sex-groups (males 20-44) that are between those for mid-2001 and mid-1981, and that are less than in the original mid-1991 estimates.

Fig. 1: Comparison of sex ratios, England & Wales, 1981-2001

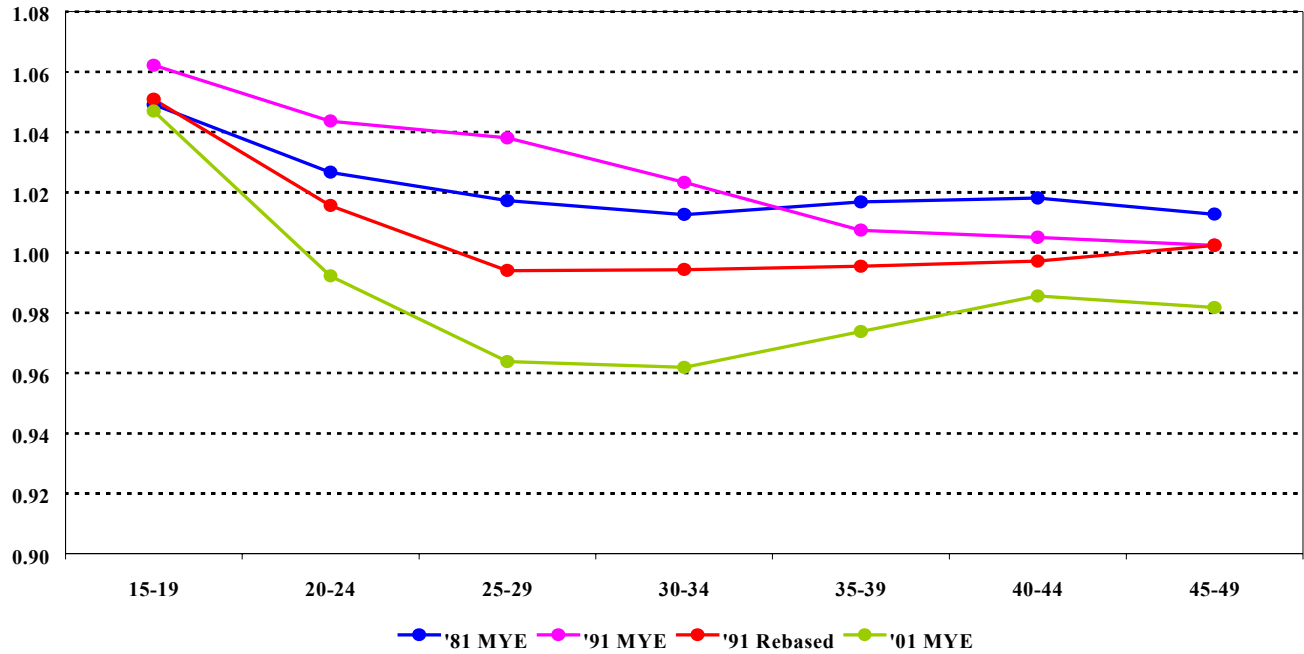
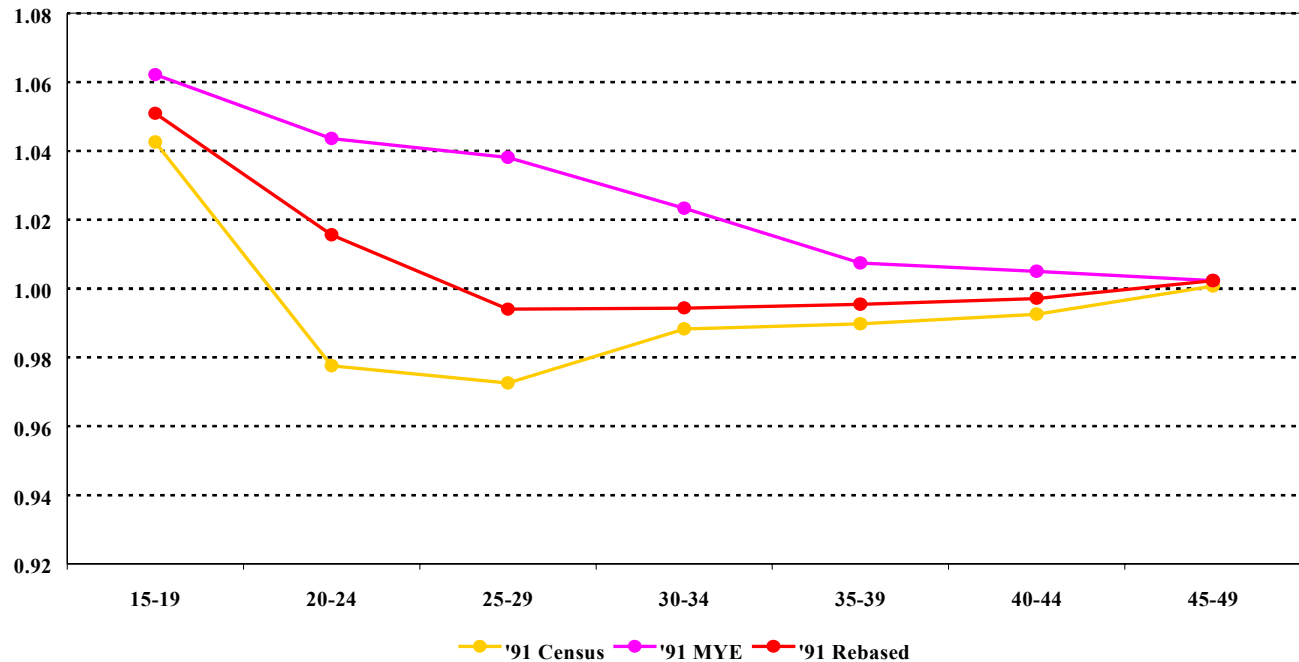


Figure 2 shows the original published mid-1991 estimates, the 'raw' 1991 census counts and the new revised mid-1991 estimates. It demonstrates that the rebased mid-1991 estimates still include adjustments for underenumeration in 1991 but that they are more modest than those applied previously.

Fig. 2: Comparison of sex ratios, England & Wales, 1991



Rebasing the 1992-2000 series

Before revising the estimates for 1992-2000, they have to be rebased in the light of the revised mid-1991 estimate. The total amount of the 1991 revision is an absolute change that is constant for all years from 1992 to 2000, but, of course, it had to be aged before it could be applied in subsequent years. In other words an adjustment to the 1991 estimate for 80 year old men of -500 would become -500 to 81 year old men in 1992, -500 to 82 year old men in 1993 and so on up to 2001 when it would be -500 to 90 year old men.

Revising the 1992 to 2000 MYE series

ONS researched various methodological options for revising the previous intercensal series. The research was based on a literature review that included web-based resources outlining the methods used in other National Statistics Institutes. The review identified possible approaches and the most appropriate method was the one that involves distributing observed differences as a function of time elapsed since the previous Census. This method is consistent with the notion that errors are likely to have accumulated over the intercensal period. The research and information about the chosen method were written up in a Population Trends article published in September 2002, by Duncan, Chappell, Smith, Clark and Ambrose, 2002

http://www.statistics.gov.uk/downloads/theme_population/PT109.pdf.

Although in theory, there are a number of methodological options for producing a revised series of population estimates, it is clear that not all the options are viable. In terms of those that are practical,

there is little benefit to be gained by choosing a more complicated method over a simpler one. In order to minimise for users the period in which there is a discontinuity in the historic series, particularly between mid-2001 and mid-2000, one key requirement is that a revised historic series should be available in the shortest possible time after publication of the mid-2001 estimates. In view of these considerations, the chosen method was the clearly practical alternative. It has all the advantages that follow from taking a straightforward approach, it is viable in a short time frame and it is transparent (see Appendix A for more details of the approach and a worked example). It should be noted, however, that the method has been applied alongside an analysis of available national migration data so as to ensure that the results obtained are broadly consistent with other ONS estimates of the components of change.

The series for 1992 to 2000 has been revised in respect of the intercensal error in estimating population change. The intercensal error is estimated by comparison of the rolled forward 2001 MYE with the 2001 census based MYE. The rolled forward 2001 MYE was first adjusted for the 351 thousand people recognised as being overestimated due to the incorrect 1991 underenumeration adjustment. The remaining difference is mainly due to underestimating emigration over the ten year period, this is an error that has accumulated throughout the decade. The process used to distribute the intercensal discrepancy is to apportion it by age and sex assuming that it is a function of time and taking account of ageing. The method applies the intercensal error as a flat profile over time, this is broadly consistent with the pattern of out migration over the decade. However, the profile obtained will not be completely flat because it is not possible to apportion the error in estimating people under 10 years old over the whole decade. In practice the profile obtained is to apportion slightly more than 10% between the last three years of the decade and slightly less than 10% between the first three.

For further discussion of the method see the Population Trends article referred to above.

Revising the 1982 to 1990 MYE series

The 'intercensal discrepancy' for E&W in 1991 was recomputed by comparing the MYE for 1991 based on the 1991 Census to the final revised MYE for 1991 produced in stage 1. This gave a new discrepancy in 1991 of 351 thousand and it was necessary to revise the series back to 1982 to account for this. The discrepancy was applied uniformly to the years 1982 to 1990 using the same method as was applied to make the revisions for 1992 to 2000.

The series that has been revised is the series for 1982-1990 that was published after the availability of final revised mid-1991 population estimates rather than the original series that was published during the 80s. A Population Trends article published in 1994 explained the post-1991 revisions that were made to the 1982-1990 series: *Bob Armitage. ONS, Retrospective revisions to population estimates for 1981-90. PT77, Autumn 1994.* Further revisions are now being made to this series to allow for revisions to the mid-1991 population estimates that have been made in the light of the results of the 2001 Census. In the most recent revisions to 1982-1990, all areas down to local authority level have been revised downwards to some extent, although the changes in some areas are very small. In addition, the latest revisions provide a series for 1982-1990 that is on 2001 boundaries. Therefore, in comparing the latest revised estimates to the previous series for the 80s at subnational level, there are also differences apparent that are due to boundary changes between local areas that occurred during the 1991 to 2001 decade.

The 1982-1990 interim revised population estimates for England and Wales and for the UK that were published in November 2002 were based on the original series that were published during the 80s. As already explained, this did not take account of improvements made post 1991. The post 1991 revisions altered the pattern of change between some years during the decade, therefore the pattern of change in the final revised estimates is not the same as in the interim revised estimates. This is most noticeable for 1987. The latest revisions are the final revised estimates for 1982-1990 and they replace previous estimates for these years.

Appendix A: Applying the time-based method for producing revised estimates

Using the chosen method, the correction for each year is derived by interpolating linearly between zero at the start of the period (mid-1991) and the observed discrepancy (the difference between the two mid-2001 estimates) calculated at the end of the period. These interpolated corrections are then combined with the original mid-year estimates on a cohort basis (i.e. the correction for 40 year olds in 2001 is applied to 39 year olds in 2000, to 38 year olds in 1999 etc). As a hypothetical example, consider a particular single year of age by sex in which the ONC based mid-2001 estimate is 2,500 and the 1991 Census based mid-2001 estimate is 2,700. The difference between these estimates (-200) is then divided by the number of years in the intercensal period (10) to produce an adjustment factor. This factor is multiplied by the number of years elapsed since the starting point (mid-1991) to produce an adjustment for each particular mid-year. These adjustments are then added to the mid-1991 based estimates, aged-back, to produce the revised estimates. The process is summarised below.

Difference between ONC-based mid-2001 estimate and 1991 Census based mid-2001 estimate = $2,500 - 2,700 = -200$

Adjustment factor (AF) = $-200/10 = -20$

Date	1991-based Estimate	Intercensal Adjustment	Revised Estimate
Mid-1991	2200	0	2200
Mid-1992	2250	-20	2230
Mid-1993	2350	-40	2310
Mid-1994	2600	-60	2540
Mid-1995	2450	-80	2370
Mid-1996	2550	-100	2450
Mid-1997	2800	-120	2680
Mid-1998	2500	-140	2360
Mid-1999	2650	-160	2490
Mid-2000	2675	-180	2495
Mid-2001	2700	-200	2500

The revision method maintains the profile of change over the intercensal period but shifts the profile up or down according to the direction of the discrepancy.