



ONC(ONS(SC))97/15

ONE NUMBER CENSUS STEERING COMMITTEE

Demographic analyses in support of a One Number Census

1. The attached paper describes the demographic research undertaken in support of a One Number Census. Within the One Number Census project the demographic analyses aim to produce:
 - a) A strategy to develop the best possible rolled-forward population estimates, to provide a benchmark comparison for the 2001 census;
 - b) A check on the accuracy of the adjusted census figures for 1971, 1981, and 1991 which formed the basis for subsequent rolled-forward population estimates;
 - c) Estimates of the "margin of error" for national population estimates, taking into account sampling and non-sampling errors in the source data. These will provide the "plausible range" for assessing the adjusted 2001 census figures.
2. The findings support the 1981 adjusted census results as the best basis for rolled forward estimates for a benchmark for the 2001 adjusted census results.
3. Sample based confidence intervals have been calculated for population estimates rolled forward from 1981 to 2001, based on the IPS survey results for 1975 to 1994. During this period there was a small net inward migration of 136,000 people, plus or minus 108,000. In addition there are likely to be non-sampling errors, the estimation of which is the subject of further investigation.
4. Priorities for further work are:
 - a) refinement and comparison of different migration estimates to come up with an estimate of total error (sampling plus non-sampling), and
 - b) research on the potential for demographic analysis at the sub-national level.
5. **The steering committee are asked to:**
 - a) **note the paper**
 - b) **provide any comments (at the forthcoming meeting or in writing by 10 December 1997) on the research.**

John Charlton
MQ Division
Office for National Statistics

D2/07
1 Drummond Gate
London SW1V 2QQ

November 1997

Demographic analyses in support of a One Number Census

John Charlton, Roma Chappell, Ian Diamond

Introduction

The aim of the one number census project is to provide a methodology that enables the ONS to produce census results that have been adjusted for coverage errors and are consistent, ideally down to small area level. The methodology makes use of a large Census Coverage Survey (CCS) and, possibly, administrative records. The adjusted 2001 census figures will be compared with demographic estimates, and if, at a national level, these come within the range of plausible values the adjusted 2001 census data will form the basis of the population estimates for 2001.

Within the One Number Census project the demographic analyses aim to produce:

- A strategy to develop the best possible rolled-forward population estimates, to provide a benchmark comparison for the 2001 census;
- A check on the accuracy of the adjusted census figures for 1971, 1981, and 1991 which formed the basis for subsequent rolled-forward population estimates;
- Estimates of the “margin of error” for national population estimates, taking into account sampling and non-sampling errors in the source data. This will provide the “plausible range” for assessing the adjusted 2001 census figures.

The demographic research will examine sub-national as well as national estimates. However since the availability of data at the two geographic levels is different, different approaches are required. The demographic research has begun with analyses at the national level, which is the main subject of the paper. Although the current research is concerned with England and Wales, we are in contact with Scotland and Northern Ireland and the methodology being developed can be applied in all countries in the UK.

1. Improving the methodology for the rolled-forward population estimates

This Section describes the ongoing research programme that aims to:

- Examine sources of error and ways of improving coverage (e.g. asylum seekers/visitor-switchers, and the armed forces)
- Examine the potential of administrative data for improving estimates for particular population age groups
- Investigate the method of extinct generations for adjusting the age distribution for the elderly

1.1. Each year the Office for National Statistics (ONS) produces estimates of the population by age and sex, at national level for England and Wales, and at sub-national level with England and Wales. The standard ONS methodology for producing annual mid-year population estimates is described in detail elsewhere (Population Statistics Division 1978, 1984, 1991; Charlton, Chappell, Diamond *et al.* 1997). These estimates are produced annually by rolling forward from the most recent census after allowing for underenumeration, taking into account the births, deaths, and net migration which have occurred in the intervening period. Thus in year t the population P_t is given by:

$$P_t = P_0 + \sum_i (B_i - D_i + I_i - E_i),$$

where P_0 is the base population and B, D, I and E are respectively the Births, Deaths, Immigrants and Emigrants in each subsequent year. For sub-national estimates there are a few other adjustments for definitional differences, e.g. location of students. When there is a new census, it becomes possible to assess the accuracy of the previous ten years of intercensal population estimates. This assessment of accuracy and any subsequent revisions to the population estimates is called “re-basing the estimates”, described for 1991 by Armitage and Bowman (1995). However, rebasing population estimates on the census can only occur if the census figures are accepted as definitive. For reasons explained elsewhere (Diamond *et al.* 1997) this was not the case in 1991. There, underenumeration was not identified by the census validation survey, but instead had to be estimated by demographic analyses which were based on the 1981 census. Thus the base for the current series of annual mid-year population estimates is the 1981 census.

1.2. Potential errors in the annual mid-year population estimates could arise from inaccuracies in: i) the base census figures; ii) birth/death counts; iii) estimates of international and iv) internal migration, including data on the armed forces and their dependents. The natural change component of the estimates (births minus deaths) is widely accepted to be reliable, and compulsory registration systems have been in force in the country since 1839. The main source of the international migration data is the International Passenger Survey (IPS) which has as its main aims the collection of data on movements to and from the UK by all travellers. Since migrants are a relatively small proportion of all travellers their sample size is small, about 1,300 immigrants and 600 emigrants in 1994. Thus the sampling error is high - around $\pm 10\%$ for each of the in- and out- migration flows for a single year. These errors, and non-sampling errors, are dealt with in detail in Section 3 of this paper. Further migration data, on asylum seekers and visitor switchers are supplied by the Home Office, since these are not estimated from the IPS. A visitor switcher is a person entering the UK who is admitted as a short-term visitor and then stays for a year or longer, including students, and some asylum seekers. Discussions are in progress with a view to improving the quality of these data. Contact has also been made with researchers at the London Borough of Hackney who have been investigating the number of refugees in their Borough. Ways to improve data on armed forces personnel for the annual population estimates are also being explored, for example by obtaining the data more frequently, or accessing additional sources.

1.3. Each decennial Census is liable to general underenumeration and miscounting of particular population sub-groups such as armed forces personnel, elderly people and babies. The sources of error in the Census base are largely concerned with underenumeration. The Census Validation Survey (CVS) following the 1991 census failed to detect the full extent of the underenumeration, and in the final analysis the rolled forward population estimates were deemed to be more reliable at the national level than the Census (Heady *et al.* 1994). Since, potentially, the 1981 CVS which used a similar methodology could have suffered from a similar problem there is a need to confirm that the 1981 Census provides the best base for rolled-forward national population estimates (see later). It is known that infants under one tend to be under-represented in censuses, and birth registration data have traditionally been used to produce population estimates in census year for this age group rather than relying on the Census. The elderly also tend to be under-counted in the census and the potential use of administrative data such as DSS pension data is being researched. Data from child benefit statistics may be used as a check on counts of children.

1.4. Internal migration is estimated using administrative registers - the National Health Service Central Register (NHSCR) and electoral roll data. The NHSCR is a system administered by ONS which contains records of all persons registered with a General Practitioner and includes data for National Health Service number, name, date of birth, sex and health authority area. When a person moves and changes their GP there is a re-registration which will be recorded in the NHSCR, but only when a person crosses a health service authority boundary. Changes in the electoral roll within local authority

areas are currently used to give indications of movements at a lower geographical level. We are evaluating Family Health Service Authorities (FHSA) registers at a post-code level as a potential source of migration data. If the quality of this information is better than current data it could be used to improve migration statistics.

1.5. The Government Actuary's Department (GAD) was consulted about using the method of extinct generations (Thatcher 1992; Das Gupta 1990, 1991), and especially estimation techniques for "almost extinct" generations. These techniques are ways of improving the accuracy of age distributions for the elderly. Analyses have shown that discrepancies with ONS population estimates only arise above age 90, where figures are not published at present. However if interest were to prompt publication at higher ages, and GAD intend to publish population projections with a top age group of 100 and over, then these methods may be appropriate for ONS.

Adjustments made to produce a new base for rolled forward population estimates

1.6 When population estimates for 1971- 1977 were revised in 1978 to take into account new methods and data, the 1971 census, adjusted for underenumeration, was taken as the base from which the estimates were carried forward (Population Statistics Division 1978).

1.7 The coverage checks following the Census in 1981 were the most thorough that had been made up to that time. These were:

- a post enumeration survey from which it was estimated that the Census had missed 214,000 persons.
- demographic analyses of "aged on" 1971 Census population less deaths and allowing for net migration, together with an estimate of the number of children under ten from birth records, taking mortality and net migration into account. Other administrative records (school roll statistics, child benefit statistics, and statistics on pensioners) were also used. These analyses indicated that 36,000 children had been missed by the Census, in contrast to the post enumeration survey estimate of 10,000. The conclusion was that the adjustment to the Census for underenumeration to form a base for the mid-year population estimates from 1981 onwards was 240,000 persons missed (214,000 - 10,000 +36,000). This is equivalent to an undercount of about 0.4%.

1.8 Comparisons of 1991 census results with the population estimates rolled forward from 1981 and other demographic analyses suggested a net undercount of about 1.2 million residents (Heady 1994). In calculating the base for rolled forward estimates from 1991 onwards an allowance was made for the underenumeration, including special groups for whom other reliable records of data existed (children under one, armed forces and their dependents, those aged over 85).

2. Corroborating the national population estimates using alternative methodologies

These analyses were undertaken primarily to establish the best base year for rolled-forward population estimates that will be used as a benchmark for the adjusted 2001 census results.

- Cohort analyses as an independent check, tracing births from 1911 onwards, making allowances for mortality and migration.
- Obtaining "best" estimates of international migration for cohort analyses.
- Comparison of cohort estimates with 1971, 1981 and 1991 censuses and population estimates.
- Analysis of people missing in censuses, using the ONS Longitudinal Study data.
- The most appropriate base year for rolled forward estimates?

Cohort Analyses

2.1 Cohort analysis is a way of producing population estimates which are independent of all censuses. Starting with births in each year following an initial base year, these birth cohorts are then “aged on”, subtracting deaths and allowing for net migration, to give an estimate of the numbers who remain. Thus if in year j the number of births is B_j , then the population from this cohort that remains at the beginning of year i is given by:

$$P_{ij} = B_j + \sum_{k=j..i-1} (I_{kj} - E_{kj} - D_{kj}),$$

Where D_{ij} deaths occur in this birth cohort up to year i , and I_{ij} and E_{ij} are respectively immigrants and emigrants born in year i . This approach is similar to that used in the current rolled forward estimates, except that the most recent census is not used as the starting point or base. A much longer series of births, deaths and migration estimates is of course required. We have used this approach as a method of corroborating the national rolled-forward population estimates.

2.2 A database was compiled of the available data, comprising:

- birth data, each year from 1911
- single year of age cohort life tables, each year from 1911 .
- deaths by single year of age, from 1950
- individual birth and death records, from 1968
- cohort life tables compiled by the Government Actuary’s Department from 1911.

The Government Actuary’s Department constructed cohort life tables from 1841 onwards, which we have used for the years prior to 1950. A comparison of cohort results using data where both life tables and actual numbers of deaths (or actual death records) were available indicated that little accuracy was lost through using life tables, and so for consistency these were used for the entire period 1911-1991. Those born in 1911 will be 90 by 2001, so there is little to be gained in using data prior to 1911. The methodology used for constructing migration series back to 1911, by single year of age and sex, is given in Annex A. Also described are the assumptions that had to be made.

Comparison of cohort estimates with 1971, 1981 and 1991 censuses and national population estimates

2.3 Using these data, comparisons have been made between:

- cohort analysis results, without taking into account the effect of migration;
- cohort analysis results adjusted for migration effects;
- unadjusted census data for 1971, 1981, and 1991;
- official population estimates.

2.4 Figures 1 to 6 show, for 1991, 1981, and 1971 the comparison of the cohort approach with census data and population estimates, with and without allowance for the effects of migration. Figures 1 and 2 compare the estimates for 1991, and also show 95% confidence intervals based on IPS sampling errors (see Section 3). Table 1 gives the figures for 1991. Points to note are the difference between the Census counts (data not adjusted for underenumeration) and the Population estimates (the *revised final* estimates for 1991). This equals (by definition) the underenumeration that was assessed for 1991 when the population estimates were made. Also, in Figure 1 a Census undercount of males in their early twenties is clearly visible when compared with the cohort estimates. The cohort estimates are higher for

this age group than the official estimates, but this was not the case in 1971 or 1981. Figures 3 and 4 compare the estimates for 1981, and Figures 5 and 6 for 1971. Once again it is clear that population estimates produced by both methods are similar. The difference between census results and both types of population estimates is much smaller than in 1991. This gives an indication of the likely greater accuracy of the 1981 census. There were more people found in the census than the cohort analyses predicted for ages 30-49 in 1971, 40-59 in 1981, and 50-69 in 1991. This could be because the migration figures failed to include a large number of immigrants in the 1960s, when there was a large net inflow (Coleman and Salt 1992). This will be investigated further using the Labour Force Survey.

Figure 1.

**Comparison of rolled-forward population estimates with cohort population estimates (Males - 1991):
Migration - IPS**

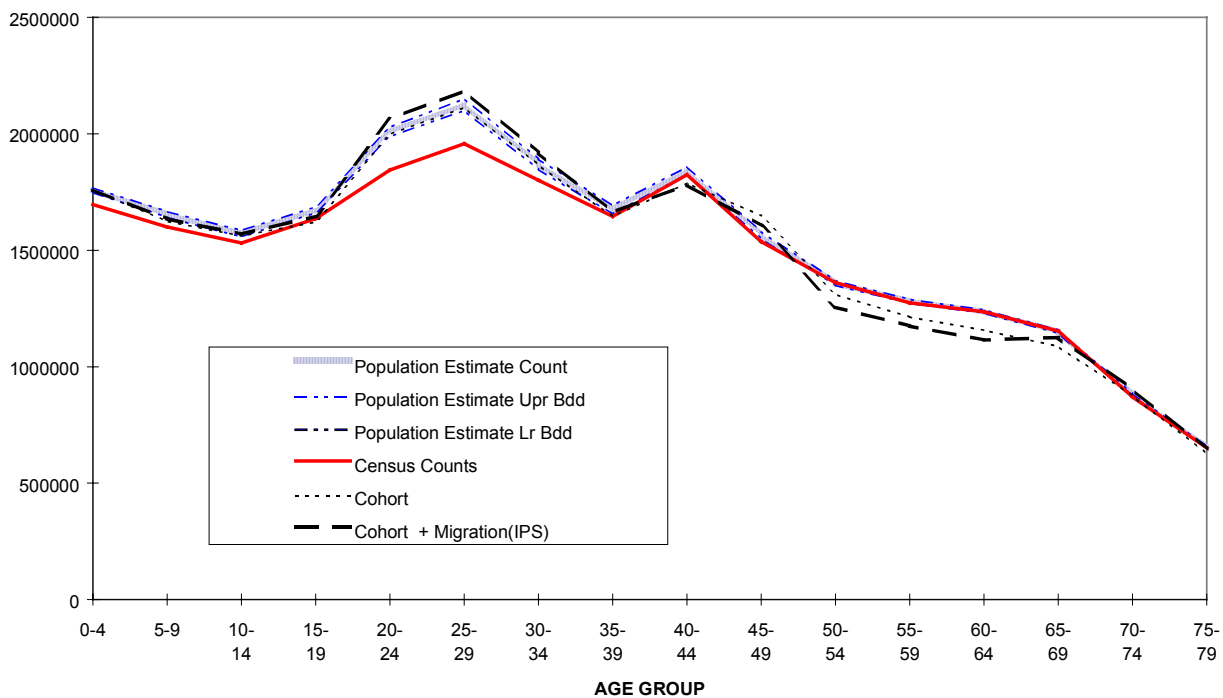


Figure 2.

Comparison of rolled-forward population estimates with cohort population estimates (Females - 1991): Migration - IPS

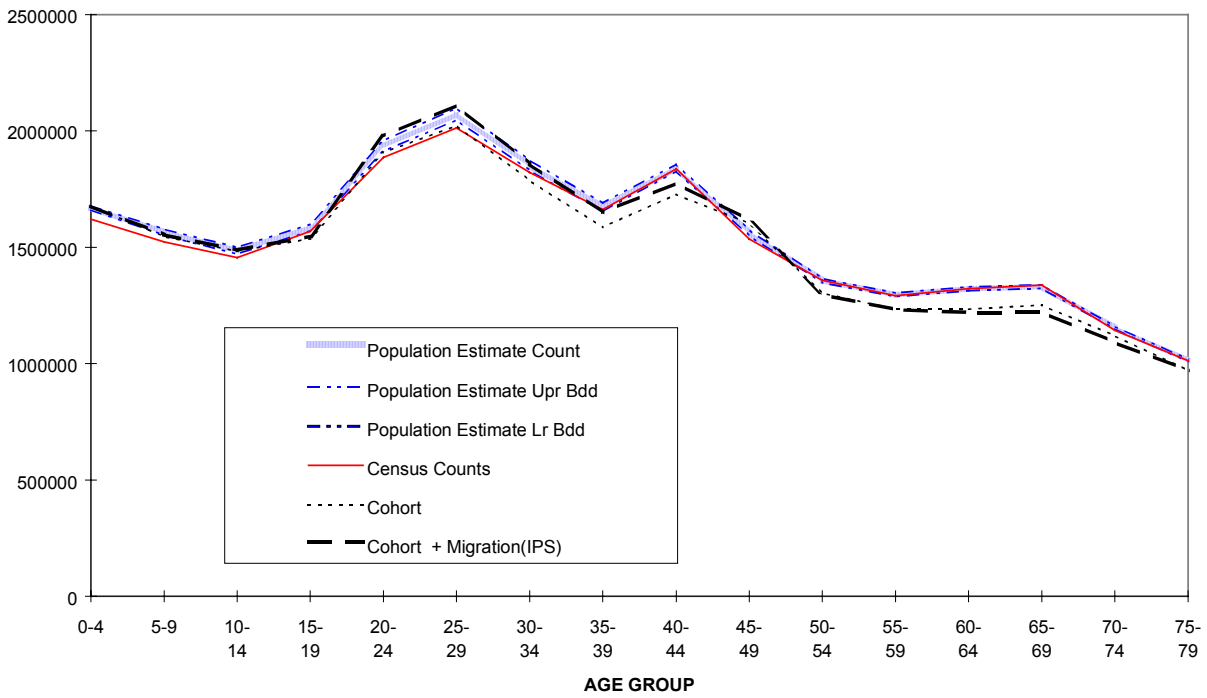


Figure 3.

Comparison of rolled-forward population estimates with cohort population estimates (Males - 1981): Migration - IPS

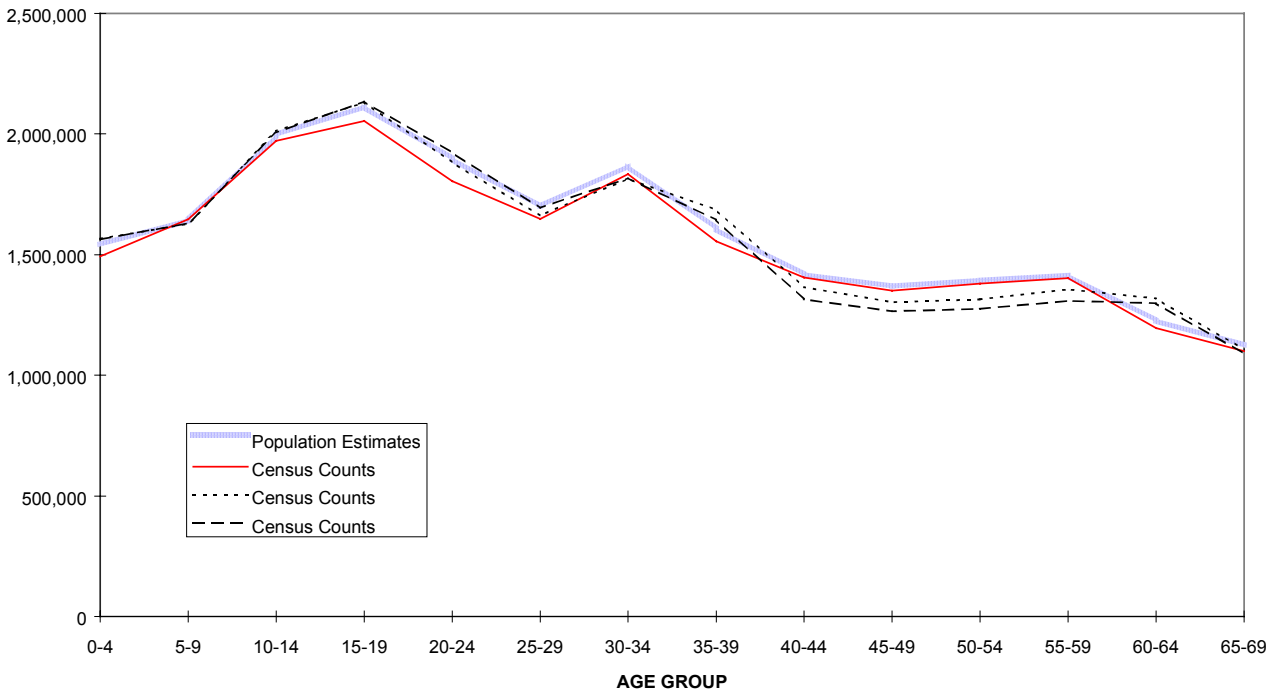


Figure 4.

**Comparison of rolled-forward population estimates with cohort population estimates
(Females - 1981): Migration - IPS**

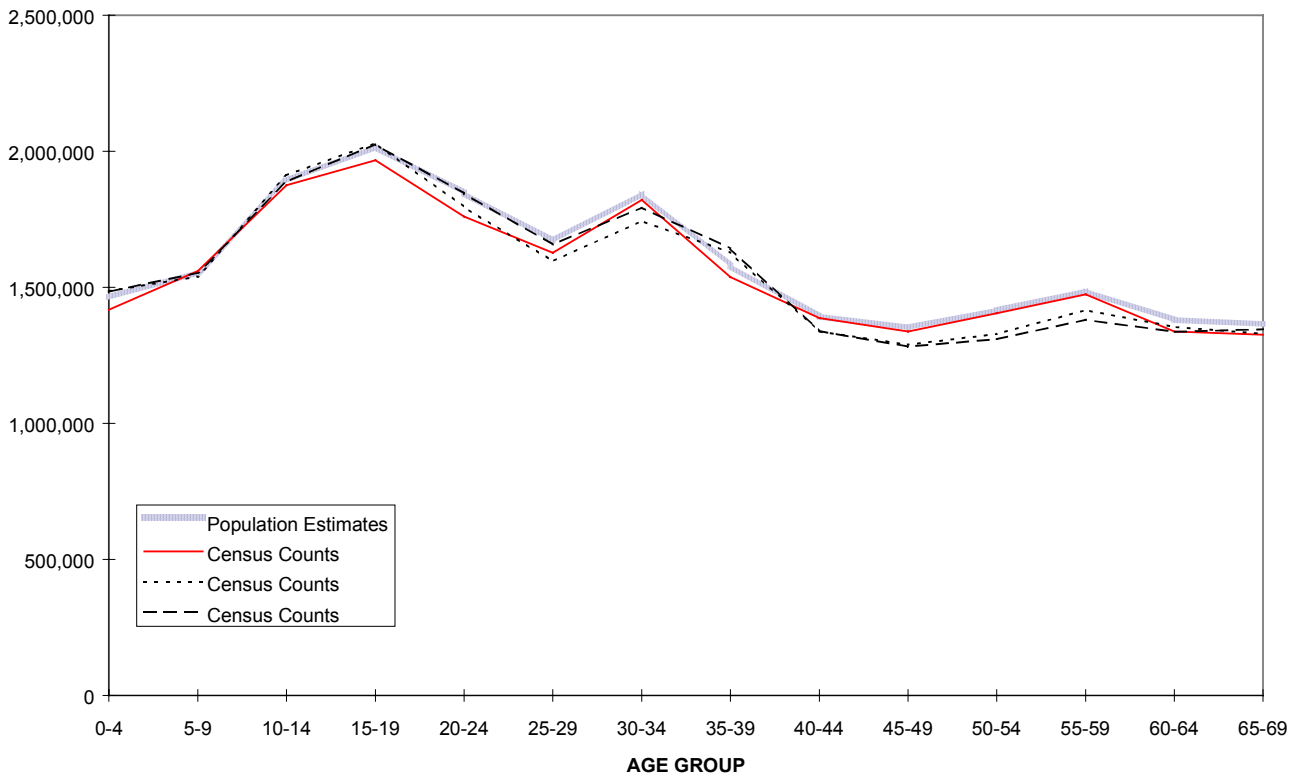


Figure 5.

Comparison of rolled-forward population estimates with cohort population estimates (Males - 1971)

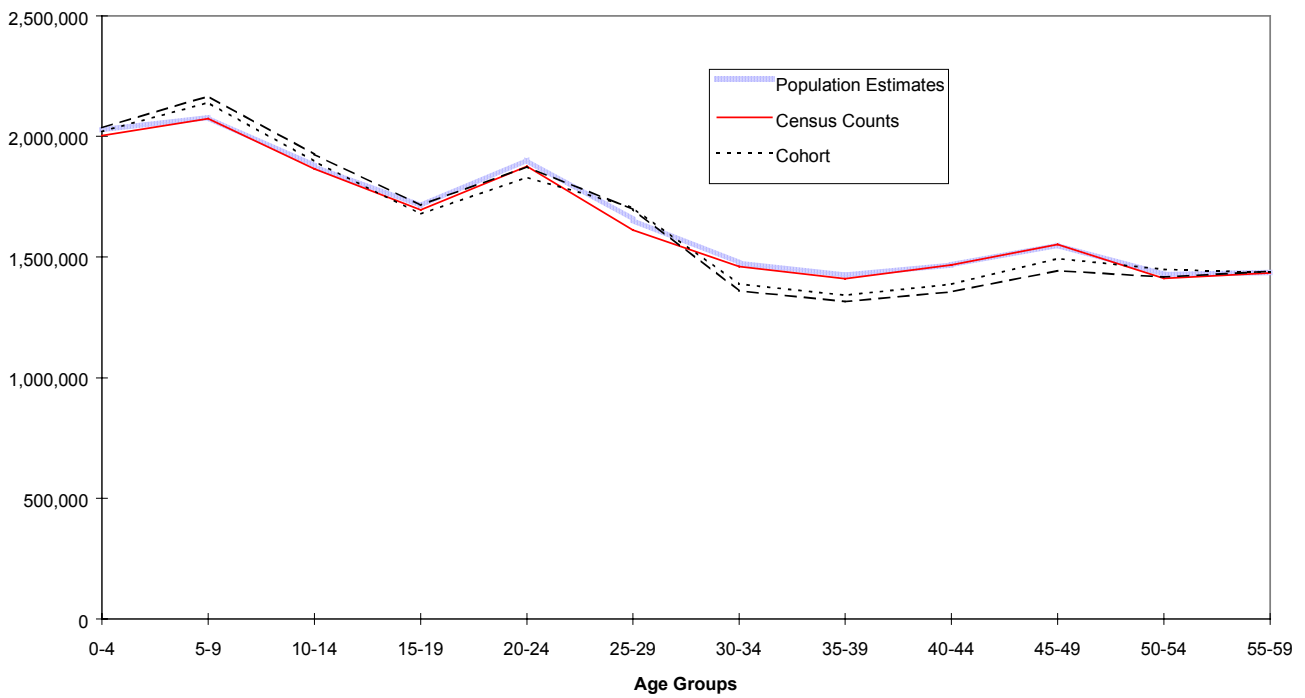
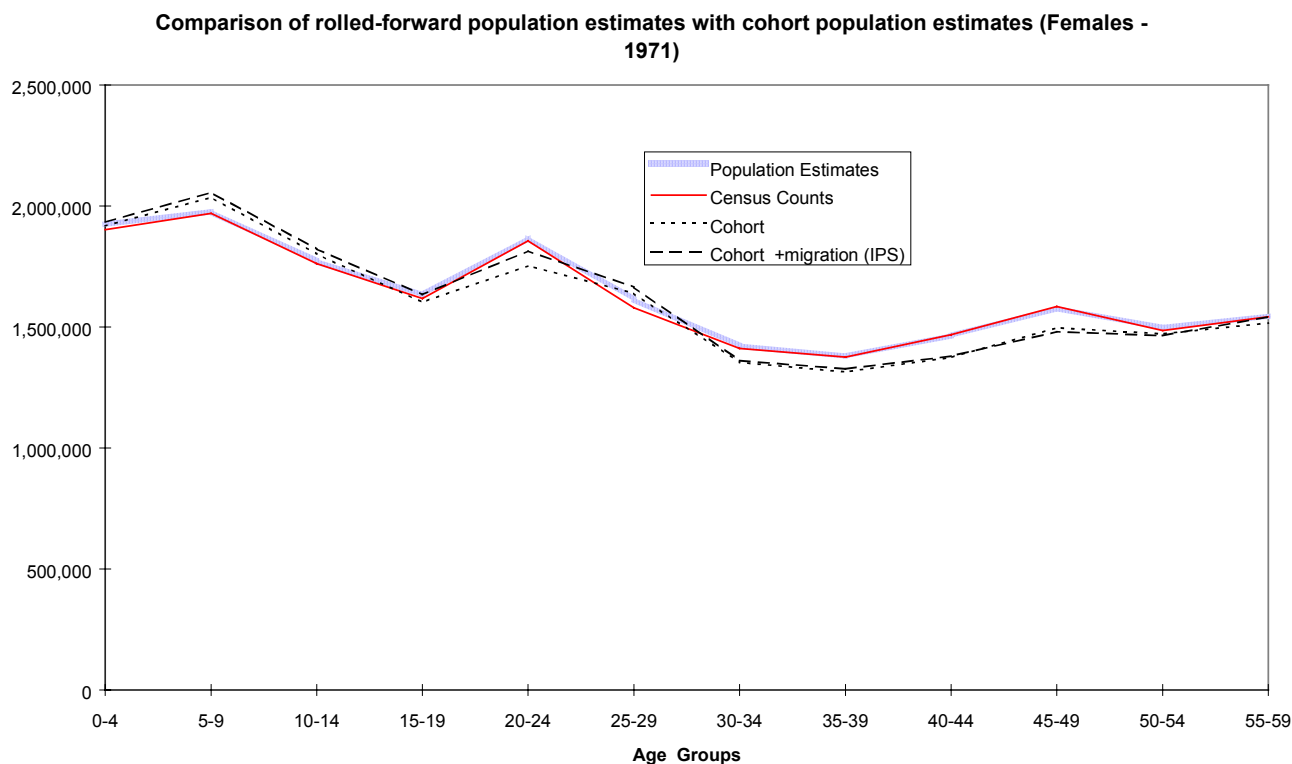


Figure 6.



2.5 A cohort approach is independent of censuses whereas the official population estimates are essentially based on a census. However the cohort approach requires migration estimates spanning a greater number of years. Results from the cohort analyses provide population estimates closely in line with 1971 and 1981 Census data, and most similar to the official population estimates for 1981. This corroborates the official estimates at national level, and suggests that 1981 official population estimates provide the best base from which to roll forward population estimates for the 2001 census benchmark.

An alternative check on which census provides the best basis for rolled-forward population estimates - Analysis of census non-respondents in the Longitudinal Study.

2.6 The ONS Longitudinal Study (LS) consists of linked census and vital registration data on a one per cent sample of the resident population of England and Wales. Selection into the LS is by birth date, and the study was designed as a continuous, multi-cohort study, with subsequent samples being drawn at each census, using the same selection criteria, and linked into the dataset (Hattersley and Creaser 1995). It includes the 1971, 1981 and 1991 Censuses, and there are no adjustments for underenumeration. Each year, new members are entered by virtue of being born on LS dates or by immigration (if born on LS dates) and exited by death or emigration. LS members are also traced to the National Health Service Central Register (NHSCR). The data on an LS member include everything collected in the censuses.

2.7 The LS can be used to identify the types of people who are apparently missing from censuses and can thus help to devise a strategy to minimise underenumeration. This is analogous to the Reverse Record Check used by Statistics Canada (Burgess 1988). There are some caveats, however:

- those missing may merely be not linked and this is most likely to be because they have the wrong date of birth on their census form (birth dates are more likely to be correct on the NHSCR than on

the Census) The census form is usually filled in for the household by one individual, who may record the date incorrectly;

- those missing may have been temporarily out of the country as part of an absent household on Census night;
- the LS does not cover Scotland, so those who move there will not be included in the linkage;
- the NHSCR records are only amended when people re-register with a different GP or de-register, for example when leaving England and Wales. Young healthy men may not always re-register promptly with their GP after moving, and those leaving the country will probably not de-register with the NHS.
- When an individual is absent from a census no data are available at that point in time. Information on age, sex, and country of birth will be available from other censuses, however.

2.8 Whilst bearing these caveats in mind, Figures 7 and 8 show estimates of those missing in 1971, 1981, or 1991 but present in the other censuses. These data exclude those who died or were known to have immigrated. Only individuals traced at NHSCR in 1981 and 1991 were included. LS “undercounts” are higher than the undercounts achieved in the censuses, partly because “missing” could also be due to non-linkage of records. Of the three censuses, 1981 had the smallest proportions of missing individuals. Also, young men (and to a lesser extent young women) aged 20-30 were most prone to be missing in the LS. This was the group suspected of being most prone to underenumeration in the 1991 census. The results also give a feel for the magnitude of the underenumeration problem in 1991. In the figures, the 70+ age groups have been excluded as the numbers in the sample are small and thus the estimates are not reliable. The graphs do not show the 15-19 age group for 1991 since they were born after the 1971 census and so could not fulfil the criterion of being in at least two of the three censuses. Figure 9 shows that those born outside the UK were more likely to be missed in a census. Some of these could however have temporarily returned to their place of birth.

Figure 7.

**Percentage of persons missing in 1971, 1981 or 1991 but present in the other two censuses
Males. Source: Longitudinal Study, England and Wales**

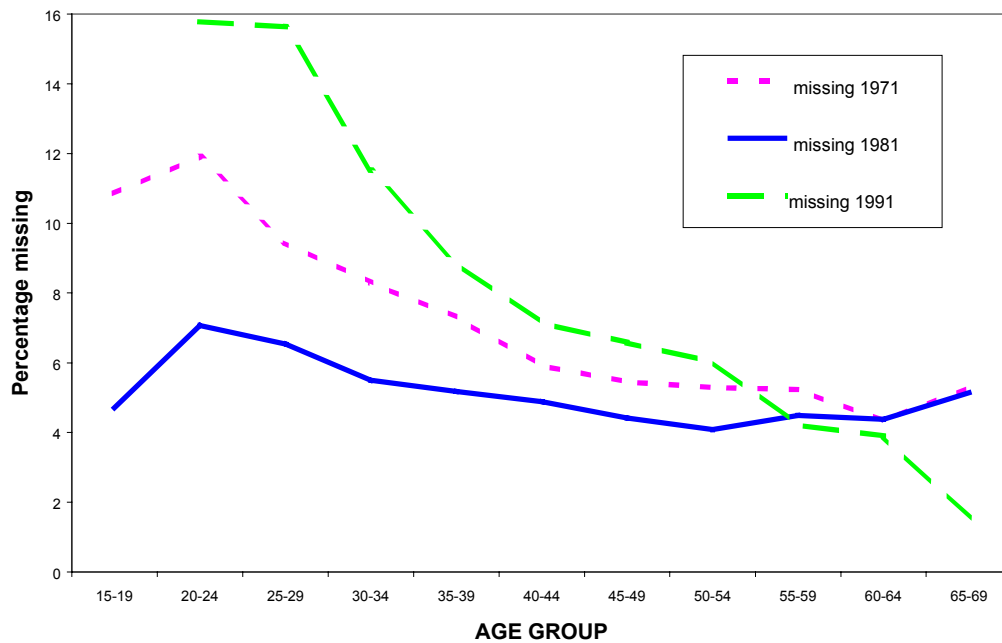


Figure 8.

**Percentage of women missing in 1971, 1981 or 1991 but present in the other two censuses
Source: Longitudinal Study, England and Wales**

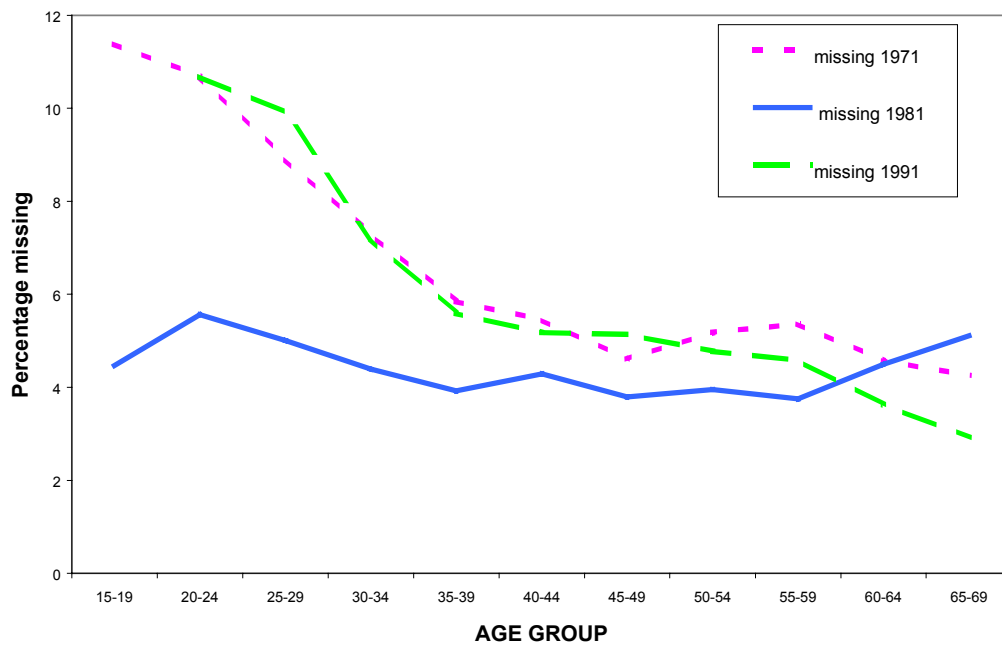
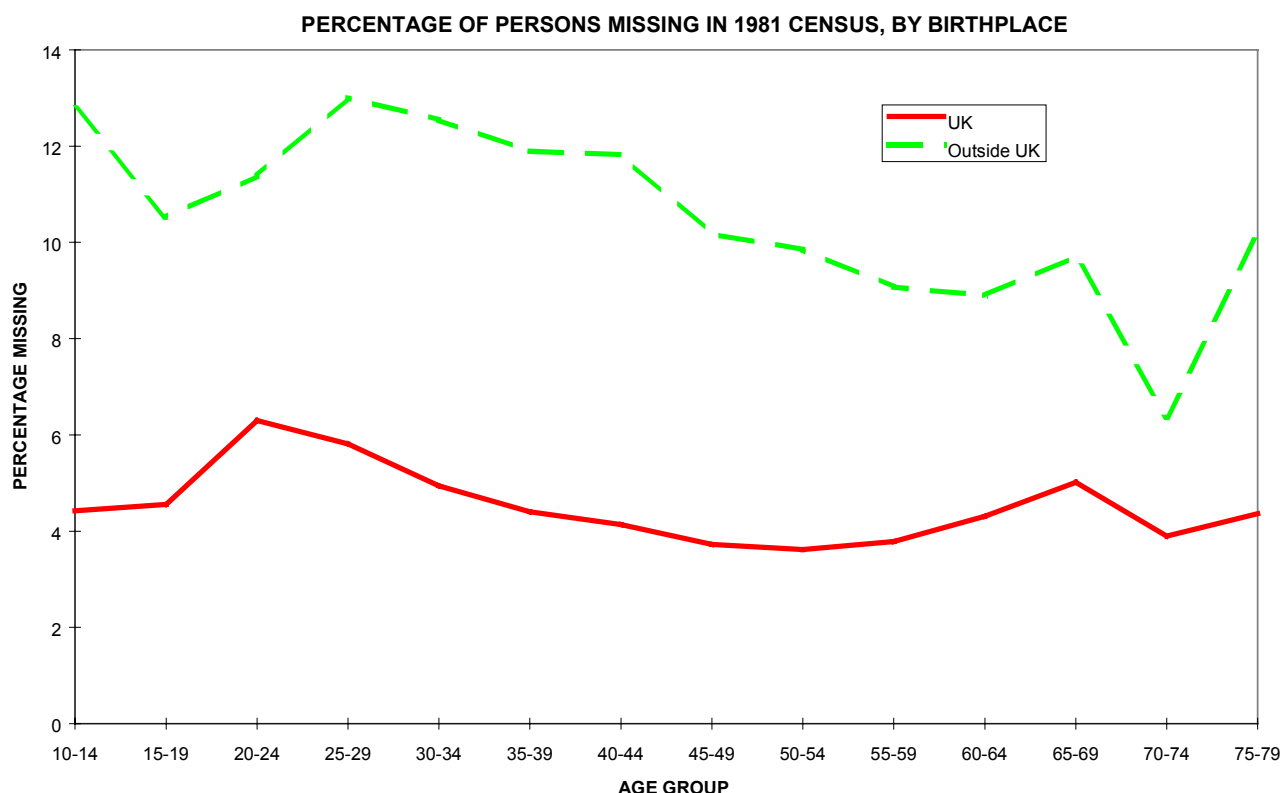


Figure 9.



2.9 We have considered whether it would be possible to use the LS to help to estimate the types of people missed by the Census in 2001. However a verified LS database it is only likely to be available one and a half years after 2001. Our provisional view is that any potential benefits of using the LS would be unlikely to warrant making a case for a speedier link.

3. Estimating the “margin of error” for the rolled-forward population estimates

The main source of error in population estimates is the data on migration. This Section covers:

- Estimation of sampling errors
- Estimation of non-sampling errors

3.1 As stated in 1.2 above, the key source of error in rolled-forward population estimates is the data on migration. International migration data are derived from the IPS, supplemented by administrative data on asylum seekers/ visitor switchers, and military personnel and their families. The other sources of error in national population estimates can be considered *de minimus*. The rest of this Section is concerned with migration data from the IPS for the period 1975 to 1994, which was readily available for analysis.

Below are some definitions of terms that will be used throughout this Section.

- Native - Someone born in England or Wales
- Resident Native - Someone born and living in England or Wales
- Immigrant - Someone arriving who has been living abroad for at least 1 year and intends to stay in England and Wales for at least 1 year.
- Emigrant - Someone leaving who has lived in England and Wales for at least one year and intends to live abroad for a least one year.

3.2 The IPS produces estimates of the number of migrants to and from England and Wales each year, based on stated intentions at the time of interview. It does not cover migration with the Irish Republic, and includes very few asylum seekers or visitor switchers. Asylum seekers usually go through different entry channels and are missed by the survey. A “visitor switcher” refers to a person entering the UK who is admitted as a short-term visitor and then stays for a year or longer, including students, and some asylum seekers. The sample is small (currently under 2,000 immigrants and emigrants are interviewed each year). Figures 10 and 11 show the estimates with 95 per cent confidence intervals, for all ages combined. The standard errors have been calculated based on the sampling fractions, and allowing for a survey design effect of 1.2. The sampling error increased from the early 1980s when the sampling fraction was reduced.

Figure 10.

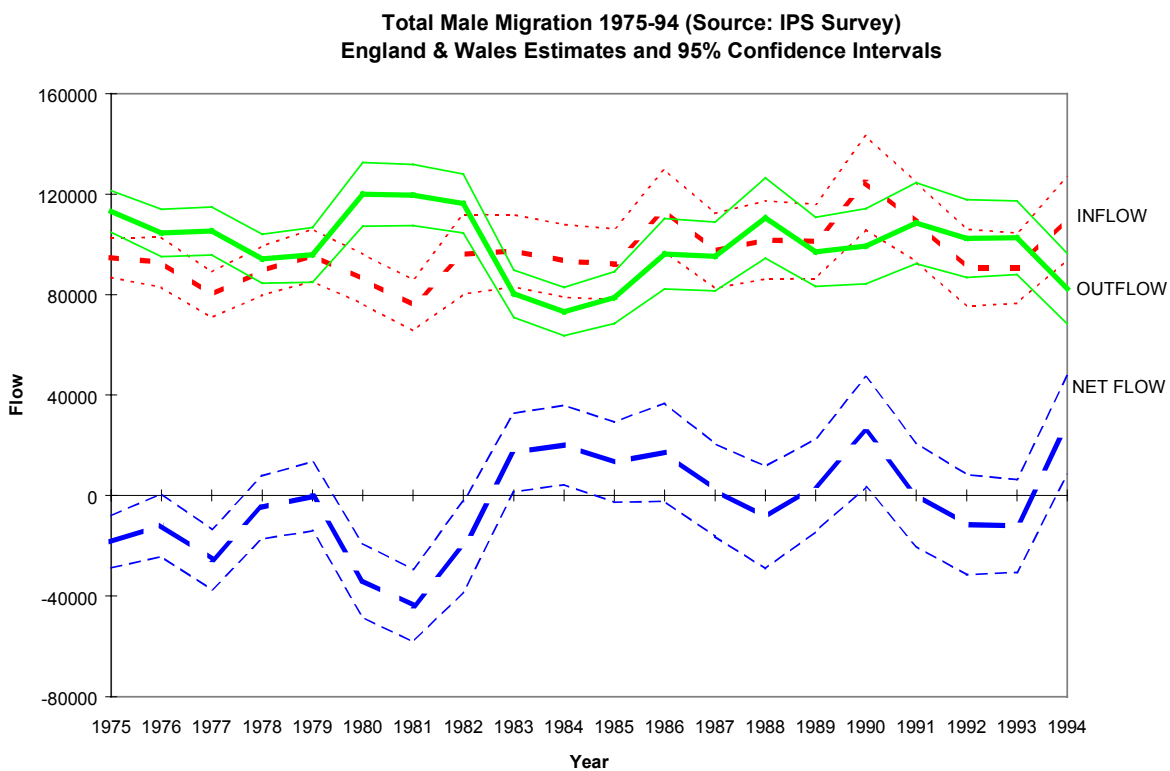
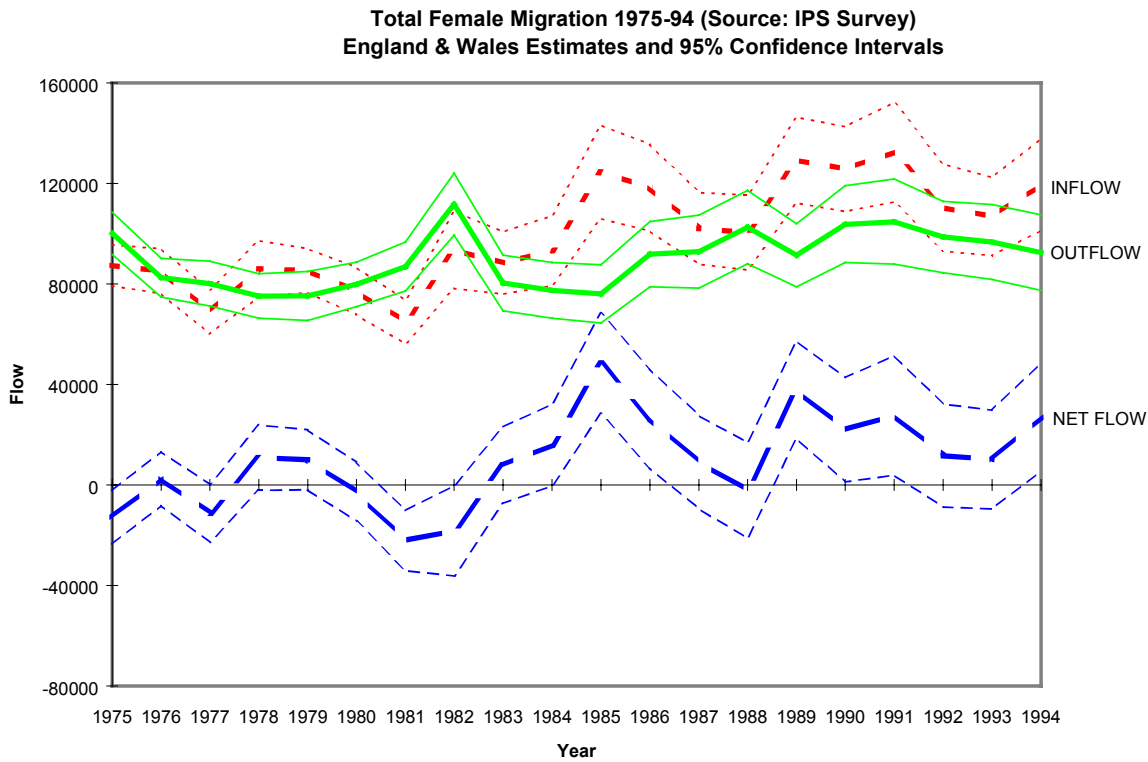


Figure 11.



3.3 Figure 12 shows the errors for migration by age group for males in a single year (1994). When the migration data are combined for several years (e.g. n years) to produce estimates of the net effect of migration, then, due to larger numbers in the sample, the relative sampling errors are reduced by a factor of $1/\sqrt{n}$ - see Figures 13 to 15. If the benchmark population estimates for the 2001 census are rolled forward from 1981, then 20 years worth of IPS data will have been used in the estimates for 2001, so the confidence intervals will be similar to those in Figure 15. From the point of view of making the estimates there is no difference between producing incremental estimates each year and using 20 years of migration data in one go to get from 1981 to 2001. However “final” as opposed to “provisional” versions of data could be used, which should increase accuracy. The confidence intervals may be reduced in relative terms by pooling the samples. The effect on the population estimates of applying the confidence intervals obtained from the IPS for 1981 to 1990 is shown in Figures 1 and 2.

Figure 12.

**Net Male Migration by age 1994 (Source: IPS Survey)
England & Wales Estimates and 95% Confidence Intervals**

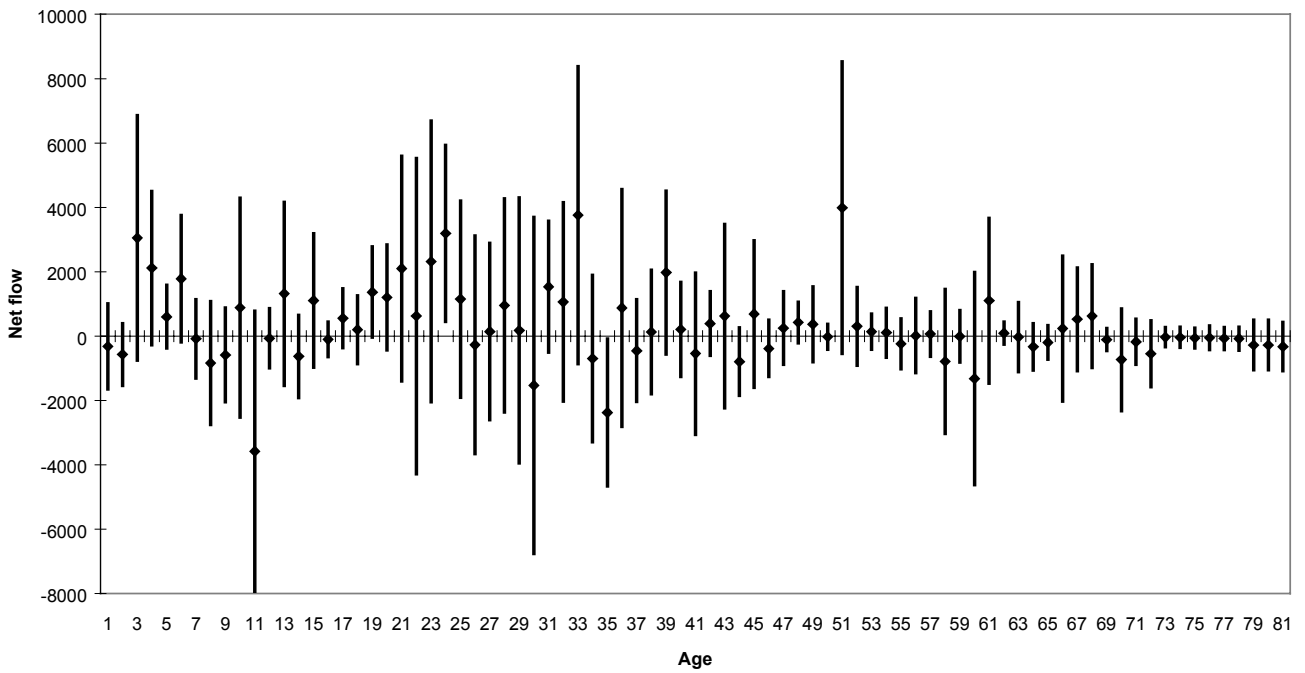


Figure 13.

**Net Male Migration by age 1975-84 (Source: IPS Survey)
England & Wales Estimates and 95% Confidence Intervals**

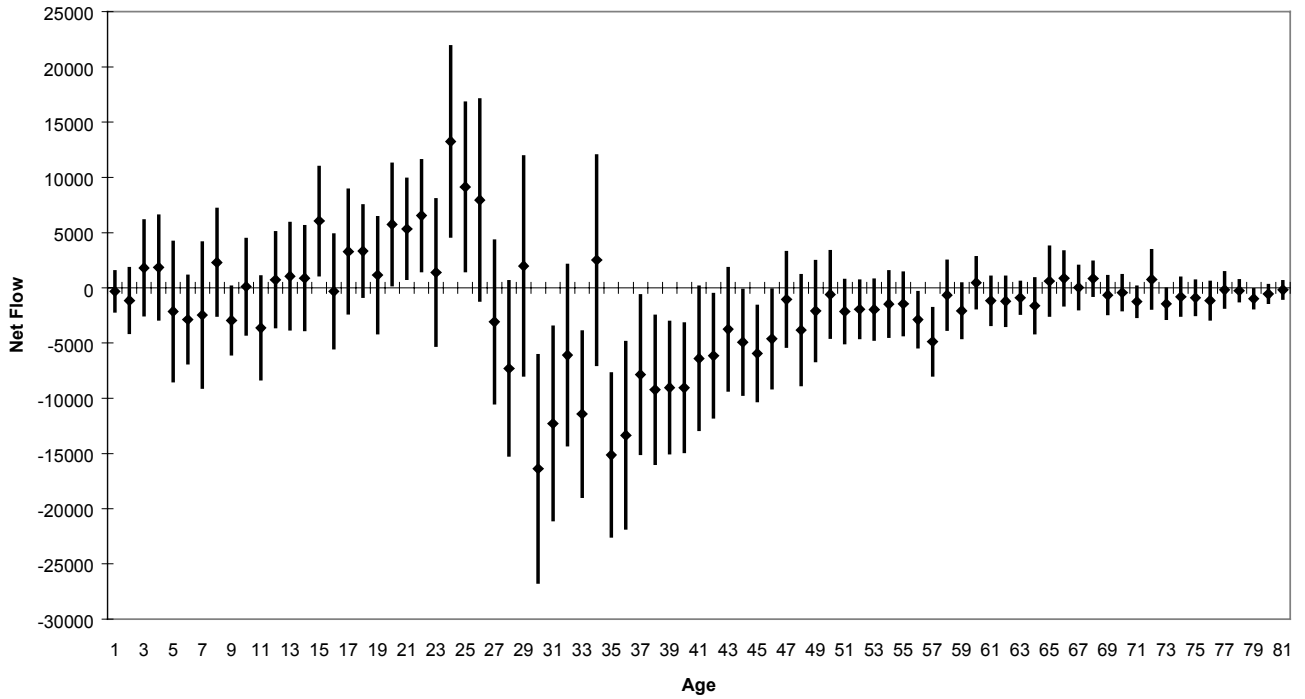


Figure 14.

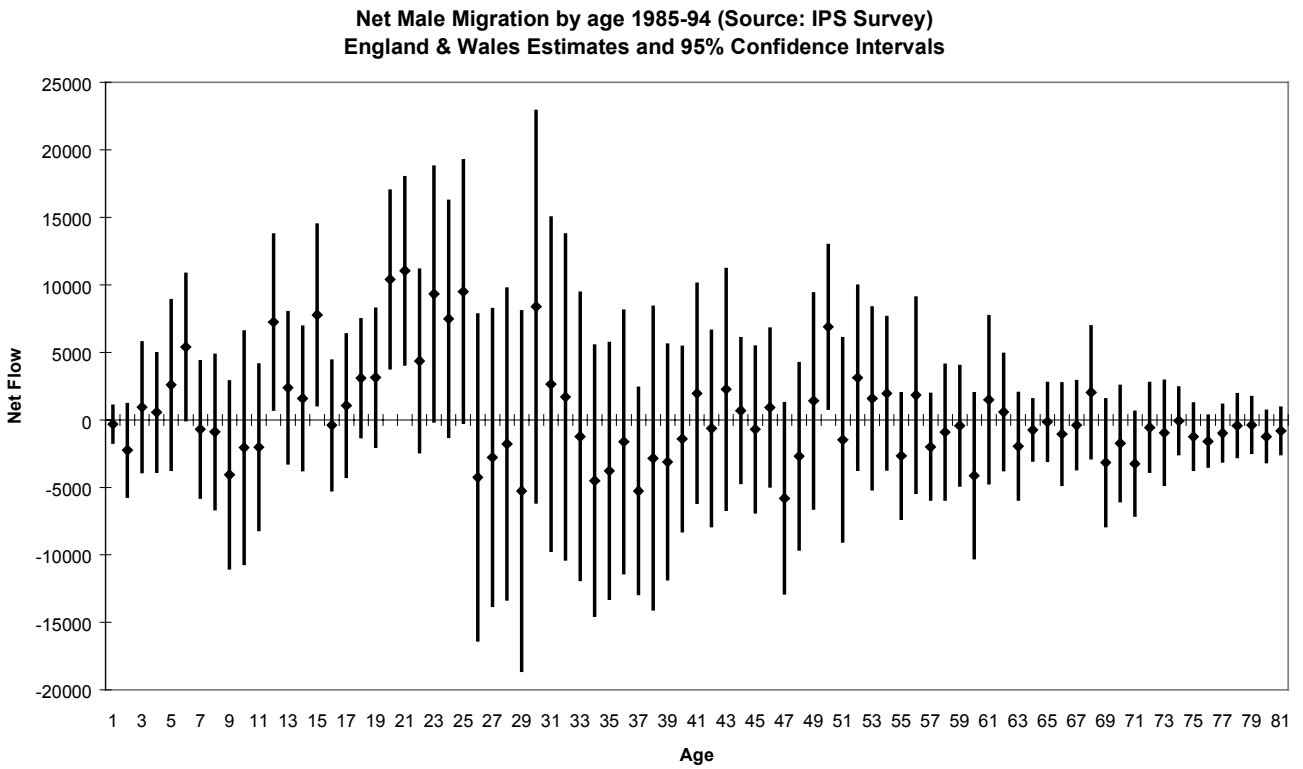
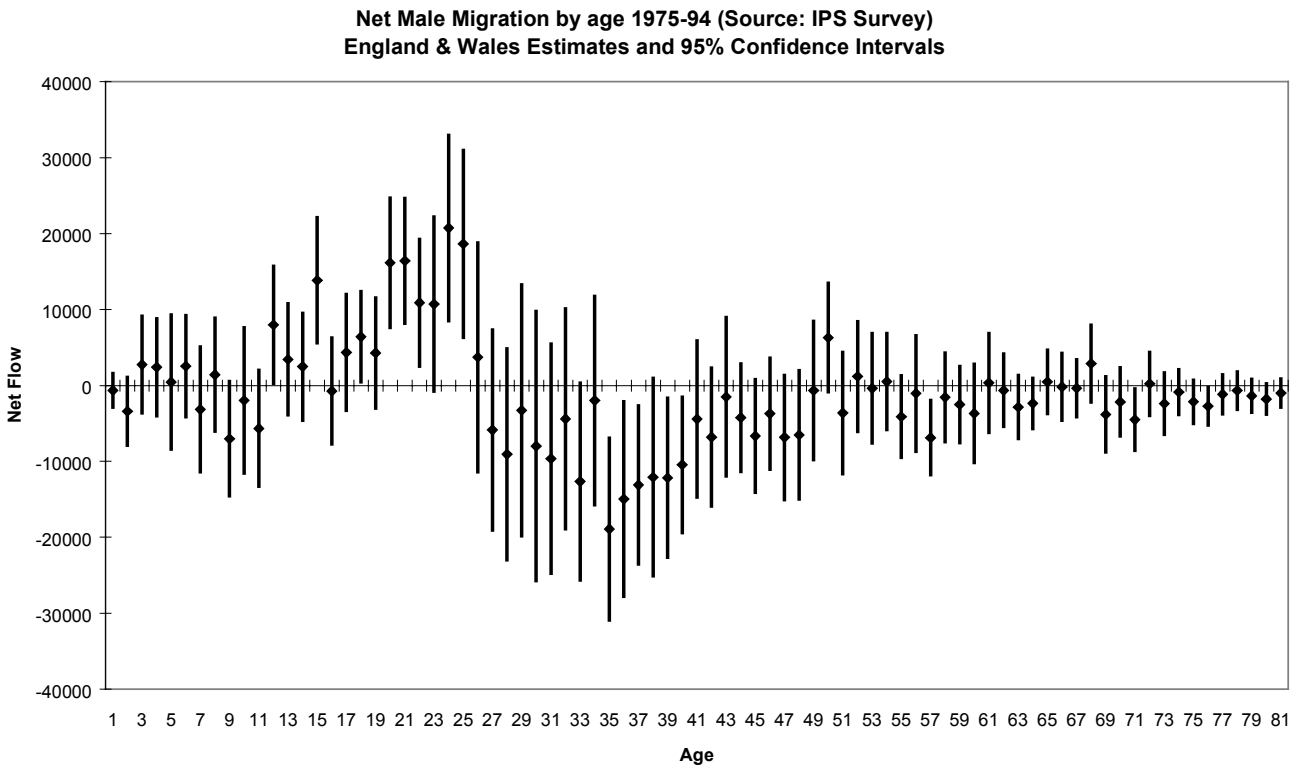


Figure 15.



Estimation of non-sampling errors for the IPS

3.4 As with any survey, in addition to sampling errors, the IPS also has non-sampling errors. People interviewed at ports of entry are asked about their intentions. They may not necessarily tell the truth, or may change their minds as to how long they will stay. It is suspected that the IPS picks up a higher proportion of emigrants than immigrants because of language problems, but it would be an expensive study that could quantify such a tendency. Also more foreign tourists appear to arrive than leave, perhaps due to the greater length of the exit questionnaire. More British tourists appear to arrive than leave. In the IPS a correction factor is applied to compensate for this for tourists but not for migrants. The IPS only monitors arrivals and departures during the daytime, and the assumption is made that those on night flights are similar. Some sources of bias in the IPS are reviewed by Bulusu (1991).

3.5 In order to obtain some estimate of the possible size of non-sampling errors in the IPS the survey data have been compared with alternative estimates of migration based mainly on:

3.5.1 The Labour Force Survey (LFS)

The annual labour force survey in 1991 was used to estimate the number of immigrants in England and Wales. The LFS collects country of birth data as well as the date of arrival of the individual in the UK.

3.5.2 The Sample of Anonymised Records (SARs)

The SARs are a 2% sample of the 1991 Census. The SARs are used to estimate the immigrant and resident native populations in England and Wales as at 1991.

3.5.3 Longitudinal Survey (LS)

The LS is a study containing linked census and vital events data on a 1 per cent sample of the population of England and Wales. It was started with a sample drawn from the resident population of England and Wales enumerated at the 1971 Census. The LS contains country of birth data which can be used to estimate the immigrant and resident native populations in 1981 and 1971, as with the SARs.

3.6 These sources of data are being used to produce estimates of migration to check the estimates produced from the IPS data.. Two types of comparison are required – net immigration up to a census date for comparison with the IPS plus historic time series data used with the cohort analyses, and short period immigration for checking the 10- or 20-year immigration data that are used with rolled-forward estimates. The research is in progress.

4. Sub-national analyses

Sub-national analyses include:

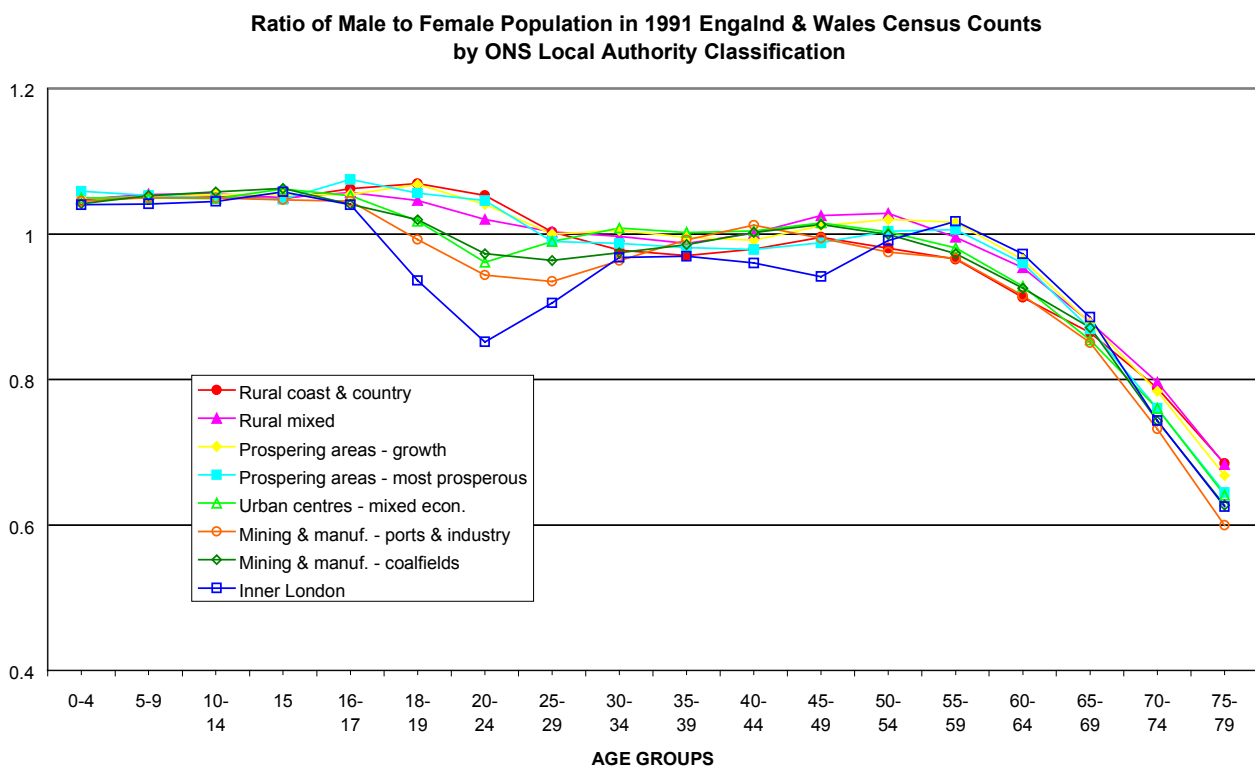
- Identifying potential improvements in sub-national migration data (e.g. through using FHSA data)
- Identifying locations of transient population groups, e.g. armed forces and students.
- Identifying hard-to enumerate areas, and methodologies for detecting undercounts in certain types of area (sex-ratio analyses etc.).

4.1 Methodologies for detecting undercounts in different types of areas

In 1991, sex ratios were examined for ten groups of local authorities (Inner London, Outer London, major metropolitan areas, other metropolitan areas, cities, remote areas, mixed urban rural, new towns, resort/retirement, and one other). The strategy used in 1991 was, for each group, to estimate target sex

ratios for 1991 based on their sex ratios in 1971 and 1981 together with national change between 1981 and 1991. Using these target sex ratios it was then possible to estimate age sex specific underenumeration for each group. However, as the national estimates in 1991 were based on target sex ratios it is somewhat implausible to use the same target sex ratios in 2001 as there is likely to have been real change over the twenty years from 1981. Current research is focusing on the potential to use sub-national sex ratios as a check on the accuracy of census counts for different types of area and to explore the extent to which such sex ratios change over time. Data which adjust for census undercounts for small areas in 1991 (*Estimating with Confidence*, Simpson *et al.*, 1995) could be used as one benchmark of what the sex ratios for particular types of areas should be (although based ultimately on 1971 and 1981 ratios). One approach that may be explored is to use cluster analysis to classify different types of areas according to their age/sex distributions, and possibly other characteristics. Then if they do not fit the expected pattern in 2001 this may signal a potential problem. This work is only at an initial stage. It may, in addition, be used to help inform the choice of stratifying variables for the census coverage survey. Figure 16 shows for the 1991 census how sex ratios varied by age, for different types of local authority. Here local authorities were grouped according to a cluster analysis based on 36 census socioeconomic characteristics. A similar analysis will be undertaken using data that has been grossed up to compensate for underenumeration.

Figure 16.



5. Conclusions

The cohort analyses confirm that the adjusted 1981 census forms an appropriate base population for estimates up until the 2001 Census. For the national level confidence intervals for the 2001 population estimates can be calculated based on the sampling error inherent in the International Passenger Survey, and for the total population these amount to about $\pm 50,000$. Further work is required to quantify the potential extent of non-sampling error, and this is being estimated by comparing IPS data with other sources for migration. Quantified sampling and non-sampling error will enable plausible national

population ranges to be constructed, against which the 2001 adjusted census results can be validated. The extent to which further demographic analyses are required at a sub-national level is currently under consideration.

References

Armitage B and Bowman J (1995) Accuracy of rolled forward population estimates in England and Wales 1981-1991. OPCS Occasional Paper 44

Bruce, S and Elliot, D Sampling Errors and Sample Optimisation on the IPS Vol. 1: Methods, Vol. 2: Results Internal ONS Reports 1993

Bulusu L (1991). A review of migration data. OPCS Occasional Paper 39, London, OPCS.

Burgess R (1988). Evaluation of reverse record check estimates of undercoverage in the Canadian census of population. Survey Methodology 14: 137-156.

Charlton, J and Murphy, M. The Health of Adult Britain 1841-1994 Vol 1, Ch.2. ONS/TSO, 1997.

Charlton J, Chappell R, Diamond I, Spencer C, Colman S (1997). Demographic analyses for a one number census. Paper ONS (ONC(SC))97/04 for One Number Census Steering Committee of June 1997.

Coleman, DA (1987), United Kingdom Statistics on immigration: development and limitations, *International Migration Review*, 21:1139-69.

Coleman D, Salt J (1992). *The British Population*. Oxford University Press, Oxford.

Das Gupta P (1990). Reconstruction of the age distribution of the extreme aged in the 1980 census by the method of extinct generation. Paper presented at the 1990 Joint Statistical Meetings, Anaheim, California, August 6-9 1990.

Das Gupta P (1991). Reconstruction of the age distribution of the extreme aged in the 1980 census by the method of extinct generation. American Statistical Association 1990 Proceedings of the Social Statistics Section, 154-159.

Diamond, I, Teague A, Thorogood D, Brown J, Buckner L, Codd W, Chappell R, Charlton J (1997). Developing a One Number Census in the United Kingdom. Statistics Canada Symposium, November 1997.

Hattersley L, Creeser R (1995). Longitudinal Study 1971-1991. History, organisation and quality of data. LS no 7. OPCS/ HMSO, London

Heady P, Smith S, Avery V (1994). 1991 Census Validation Survey: coverage report. OPCS.

International Migration 1994 Series MN No. 21 ONS

Mitchell BR (1992), International Historical Statistics, Europe 1750-1988. Third Edition, McMillan Press, London

OPCS 1981 Census General Report. HMSO

Population Statistics Division (1978). Population Estimates 1971-77. Population Trends 13: 10 – 12, 1978.

Population Statistics Division (1984). Population Estimates 1961-81. Population Trends 35: 30 – 33, 1984.

Population Statistics Division (1991) Making a Population Estimate in England and Wales. OPCS Occasional Paper 37

Population Statistics Division (1993). How complete was the 1991 Census? Population Trends 71: 22 – 25.

Rocke, L and Goodwin,G. Developing the IPS Sample Design. Internal ONS Interim Report 1996

Rowntree JA. Population estimates and projections. Population Trends 60:33-34, 1990.

Simpson, S. (1993) Measuring and Coping with local underenumeration in the 1991 Census. Paper to the conference on 'Research in the 1991 Census', Newcastle. Paper available from the author.

Simpson, S., Tye, R. and Diamond, I. (1995) What was the true population of local areas in mid 1991? Working Paper10, Estimating with Confidence Project, University of Southampton.

Simpson, S. and Dorling, D. (1994) Those missing millions: implications for social statistics of non-response to the 1991 Census. Journal of Social Policy, 23(4), 543-67

Thatcher AR (1992). Trends in numbers and mortality at high ages in England and Wales, Population Studies, 46: 411-416.

Werner, B (1984) Infants aged under 1 in the Census, 1861-1981. Population Trends, 38, 18-24

Table 1: Comparison between various population estimates for 1991

1991		Male population estimates														
		0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74
Population	Estimate (official)	1757893	1654075	1571444	1674145	2008052	2126000	1870001	1674197	1842877	1565256	1359304	1280040	1237199	1150351	879534
Population	Est Upr Bound	1767889	1666677	1585417	1686995	2028397	2150983	1892466	1692007	1858984	1580026	1369221	1288463	1244183	1157481	884543.3
Population	Est Lr Bound	1747897	1641473	1557471	1661295	1987707	2101017	1847536	1656387	1826770	1550486	1349387	1271617	1230215	1143221	874524.7
Census	Counts	1696098	1599581	1532019	1636124	1844532	1957808	1800696	1645575	1824168	1537361	1360870	1274953	1235649	1152977	870015
Cohort		1758208	1623973	1561427	1621658	1998427	2113090	1867006	1644901	1788785	1647876	1310570	1213542	1156674	1086779	875247.4
Cohort	+Migration(SARs)	1750987	1611785	1547550	1600765	1961760	2060395	1820559	1596559	1728358	1582123	1251115	1157928	1098499	1038949	836686.3
Cohort	+Migration (IPS +SARs +Asylum/V.S)	1758600	1634474	1569035	1645388	2063143	2185264	1918520	1663716	1781564	1605315	1256529	1174804	1114631	1125426	903822.1

1991		Female population estimates														
		0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74
Population	Estimate (official)	1669278	1565104	1485077	1585457	1935303	2073767	1852790	1672712	1843012	1563198	1357326	1296160	1321912	1331324	1154410
Population	Est Upr Bdd	1678214	1577975	1499623	1600256	1957895	2099206	1874791	1690306	1857825	1575987	1366377	1303874	1330103	1338455	1160052
Population	Est Lr Bdd	1660342	1552233	1470531	1570658	1912711	2048328	1830789	1655118	1828199	1550409	1348275	1288446	1313721	1324193	1148768
Census	Counts	1620130	1523042	1455517	1569340	1886843	2013101	1822012	1662607	1837947	1536262	1358987	1291916	1323056	1337539	1144379
Cohort		1677817	1546436	1483286	1536952	1906585	2023729	1788138	1585844	1727894	1602532	1303496	1234150	1233627	1251493	1120079
Cohort	+Migration (SARs)	1592931	1502499	1476865	1596321	1797979	1874997	1702024	1586108	1781205	1486401	1345558	1285344	1317927	1317781	1121174
Cohort	+ Migration (IPS +SARs +Asylum/V.Sw.)	1676810	1554757	1488584	1547237	1979318	2111236	1858684	1650624	1775878	1615960	1296625	1231939	1220382	1221341	1092002

% Differences from population estimates

1991		Male population estimates														
		0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74
Population	Estimate (official)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Population	Est Upr Bdd	0.6	0.8	0.9	0.8	1.0	1.2	1.2	1.1	0.9	0.9	0.7	0.7	0.6	0.6	0.6
Population	Est Lr Bdd	-0.6	-0.8	-0.9	-0.8	-1.0	-1.2	-1.2	-1.1	-0.9	-0.9	-0.7	-0.7	-0.6	-0.6	-0.6
Census	Counts	-3.5	-3.3	-2.5	-2.3	-8.1	-7.9	-3.7	-1.7	-1.0	-1.8	0.1	-0.4	-0.1	0.2	-1.1
Cohort		0.0	-1.8	-0.6	-3.1	-0.5	-0.6	-0.2	-1.7	-2.9	5.3	-3.6	-5.2	-6.5	-5.5	-0.5
Cohort	+Migration (SARs)	-0.4	-2.6	-1.5	-4.4	-2.3	-3.1	-2.6	-4.6	-6.2	1.1	-8.0	-9.5	-11.2	-9.7	-4.9
Cohort	+ Migration (IPS +SARs +Asylum/V. Sw.)	0.0	-1.2	-0.2	-1.7	2.7	2.8	2.6	-0.6	-3.3	2.6	-7.6	-8.2	-9.9	-2.2	2.8

		Female population estimates														
		0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74
Population	Estimate (official)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Population	Est Upr Bdd	0.5	0.8	1.0	0.9	1.2	1.2	1.2	1.1	0.8	0.8	0.7	0.6	0.6	0.5	0.5
Population	Est Lr Bdd	-0.5	-0.8	-1.0	-0.9	-1.2	-1.2	-1.2	-1.1	-0.8	-0.8	-0.7	-0.6	-0.6	-0.5	-0.5
Census	Counts	-2.9	-2.7	-2.0	-1.0	-2.5	-2.9	-1.7	-0.6	-0.3	-1.7	0.1	-0.3	0.1	0.5	-0.9
Cohort		0.5	-1.2	-0.1	-3.1	-1.5	-2.4	-3.5	-5.2	-6.2	2.5	-4.0	-4.8	-6.7	-6.0	-3.0
Cohort	+ Migration(SARs)	-4.6	-4.0	-0.6	0.7	-7.1	-9.6	-8.1	-5.2	-3.4	-4.9	-0.9	-0.8	-0.3	-1.0	-2.9
Cohort	+ Migration (IPS +SARs+Asylum/VS)	0.5	-0.7	0.2	-2.4	2.3	1.8	0.3	-1.3	-3.6	3.4	-4.5	-5.0	-7.7	-8.3	-5.4

Annex A Estimating international migration for cohort analyses.

Annual migration data were available from the International Passenger Survey (IPS) on international migration between England and Wales and other countries, excluding Scotland and Ireland, by single year of age and sex, for the years 1975 onwards. Since the cohort approach requires migration data from the start date (1911), IPS data had to be used in combination with earlier less comprehensive data.

For 1911 to 1974 estimates of the net flow of persons into and out of England and Wales, by single year of age and sex, were produced as follows:

- an historic series of migrants between Britain and the rest of the world from 1911 to 1974 (Mitchell 1992), not broken down by age and sex, was taken as the basis. Mitchell's figures were based on administrative port statistics up until 1964 and omitted arrivals of UK citizens, citizens of the dominions and colonies, and the Irish (Coleman and Salt 1992). From 1964 onwards IPS data were available and used instead. There were no figures available for 1939-45.
- Since the distribution by 5-year age groups and sex of immigrants and emigrants was stable over the 20-year period 1975-94 for different 3-year periods, the data for 20 years were combined, and used to provide an age/sex distribution for immigrants and emigrants by single year of age. This distribution was used to apportion the earlier data.
- The ratios of immigrants and emigrants to and from England and Wales to immigrants and emigrants to and from Britain were calculated using 20 years of IPS data. The IPS covers the constituent countries of Britain - 94% of immigrants to Britain settled in England and Wales and 89% of emigrants from Britain left from England and Wales. These conversion factors were applied to obtain England and Wales figures for each year from 1911 to 1974.
- Immigrants and emigrants were "aged" and estimated deaths (based on cohort life tables) were removed for each year to obtain estimated net migration up to each census date 1971, 1981, and 1991.
- For 1991, final adjustments were made to take account of migration flows to and from Eire, Northern Ireland, and Scotland. For Eire the Irish census was used to estimate migration from the UK to Ireland, and data on migration in the opposite direction came from the GB SARs (adjusted for underenumeration). The Irish Census provided movements from the UK to Ireland, and on the basis of the populations of the constituent UK countries it was assumed that 88 per cent of migrants from the UK to Ireland came from England and Wales. For migration to and from Northern Ireland and Scotland the GB SARs were used. In addition, data supplied by the Home Office on asylum seekers/ visitor switchers who had entered the country between 1982 and 1991 were added to the figures. For 1981 and 1971 a similar approach was adopted, but using the ONS Longitudinal study (LS) data, a 1 per cent sample from the censuses, instead of the SARs (Hattersley and Creeser 1995). However the data currently available only distinguishes between 'born in the UK' and 'not born in the UK'. If justified, more detailed data on country of birth could be obtained from the LS

to enable more accurate adjustments to be made.

2.3 The method outlined above by necessity made four assumptions:

- The age by sex distributions for immigrants and emigrants was the same in the period 1911 to 1974 as it was for 1975-94.
- The proportions of emigrant and immigrants from and to Britain relating to England and Wales were constant over time
- The death rates of the native population were the same as those for immigrants and emigrants. This assumption has little impact on the final estimates because the number of deaths is small, and there would need to be a large difference in mortality rates to make a difference.
- There are no figures supplied for migration during World War II (1939-1945). These have been set to zero for this analysis - likely to be accurate, on the basis that few people moved during WWI.

Table A1. Sources of data for migration estimates - Immigrants to England and Wales

Data source	Country of origin	Census years where data could be used
UK IPS (small sample) *	All except Northern Ireland, EIRE and Scotland	(1971),1981, 1991
E&W Longitudinal Study (LS)	All countries of birth	1971,1981,1991
British SARs	All countries of birth	1991
UK Labour Force Survey	All countries	1984 onwards

Table A2. Sources of data for migration estimates - Emigrants from England and Wales

Data source	Destination country	Census years when data could be used
IPS*	All except Scotland, Northern Ireland and EIRE	(1971),1981, 1991
UK Longitudinal Study (difference from cohort figures)	All countries of birth	1971,1981,1991
British SARs	Scotland	1991
British SARs (difference from cohort figures)	All countries	1991
Northern Irish SARs	E&W to Northern Ireland	1991
Irish Census	GB to Ireland	1981,1991
Irish Labour Force Survey	GB to Ireland	1991