MANCHESTER 1824

The University of Manchester

Cathie Marsh Centre for Census and Survey Research

Enhancing the population census: a time series for sub-national areas with age, sex, and ethnic group dimensions in England and Wales, 1991-2001

CCSR Working Paper 2007-11

Albert Sabater and Ludi Simpson
Albert.sabater@manchester.ac.uk

Ethnicity data from successive censuses are used to compare population change. This paper shows that such comparisons are often impossible, wrong or misleading. Distortions become more severe as the scale of areal units become smaller. The paper outlines the four main sources of confusion and applies solutions for England and Wales for 1991-2001. The paper presents methods that can be used to resolve these difficulties and produce more accurate results, and produces a consistent time series for single years of age, ethnic group and sex that can be aggregated from the smallest census output areas.

www.ccsr.ac.uk

Enhancing the population census: a time series for sub-national areas with age, sex, and ethnic group dimensions in England and Wales, 1991-2001

Albert Sabater and Ludi Simpson

albert.sabater@manchester.ac.uk

Cathie Marsh Centre for Census and Survey Research (CCSR), University of Manchester, Manchester, M13 9PL, UK

Albert Sabater: Post-Doctoral Research Fellow

Ludi Simpson: Professor of Population Studies

Enhancing the population census: a time series for sub-national areas with age, sex, and ethnic group dimensions in England and Wales, 1991-2001

Abstract

Ethnicity data from successive censuses are used to compare population change. This paper shows that such comparisons are often impossible, wrong or misleading. Distortions become more severe as the scale of areal units become smaller. The paper outlines the four main sources of confusion and applies solutions for England and Wales for 1991-2001. (1) Classifications including ethnic group and age changed between censuses; (2) non-response varies between ethnic groups, areas and ages and its treatment differs in each census; (3) students were counted at their home address in 1991 and at their educational address in 2001; (4) geographical boundaries used for standard census outputs changed. Each of these factors operates differentially on the outcome. Finally there is an additional problem of projecting census data taken on different dates of census years to comparable mid-year estimates each year.

The paper presents methods that can be used to resolve these difficulties and produce more accurate results, and produces a consistent time series for single years of age, ethnic group and sex that can be aggregated from the smallest census output areas. The enhanced dataset will be used for projections, assessment of employment and health trends, and estimation of migration in the intercensal decade. As an indication, after adjustment, Birmingham district's total population showed a 2% loss rather than the 1.7 percent gain derived from the direct 1991-2001 comparison; the Black Caribbean population shows a decrease rather than an increase.

Keywords: sub-national areas; ethnic groups; population change; England and Wales; non-response

Enhancing the population census: a time series for sub-national areas with age, sex, and ethnic group dimensions in England and Wales, 1991-2001

Introduction

Ethnic origin, age, sex and locality are fundamental dimensions for social analysis. Not only demographers are interested in how population composition is changing along these dimensions. Workforce participation, care of the elderly, political engagement, incidence of disease, the demand for housing, and the impact of increasingly diverse societies cannot be understood without relation to the composition of the population and its development over time. To the extent that counts of people and their age, sex and ethnic characteristics are unavailable, inaccurate or incompatible over time, all these analyses suffer. While ethnic origin (or race, ethnicity, ethnic group or heritage as it may be termed in different contexts) is more debateable as a demographic category than age or sex, it has become commonplace to prepare population tabulations differentiated by ethnic origin because of its centrality to policy debates and public services and the social analyses that inform them.

In the UK as in other countries where ethnic origin is a category used in population and social analysis the population census is the main source of comprehensive published statistics concerning the changing composition of the population both nationally and for smaller areas. The 1991 and 2001 Censuses of Population in England and Wales have provided comprehensive data of ethnic groups from national to local areas, thus stimulating analytical new research about the characteristics and distribution of the population (Dorling and Rees, 2003; Lupton and Power, 2004; Parkinson *et al*, 2006; Simpson, 2007).

However, census statistics are neither wholly accurate nor comparable over time. In order to judge social polarisation over time, Dorling and Rees use "1991 Census data which have both bee re-aggregated to 2001 local authority boundaries and meticulously adjusted for over a million people who were not recorded by that census" in order to make the two censuses "broadly comparable" (2003: 1289). Lupton and Power introduce their briefing on minority ethnic groups in Britain from

the censuses of 1991 and 2001 with warnings of several problems that beset an attempt to make use of the opportunity that a census time series appears to offer:

> "One is the problem of the use of different ethnic categories in 1991 and 2001, principally the introduction of 'mixed race' options in 2001. … Other problems arise … because comparisons of 1991 and 2001 Census data probably show greater increases in population than actually occurred, especially in urban areas where undercounting was worst. They also show artificially high increases in urban areas because the 2001 Census counted students at their term addresses, while the 1991 Census counted them at their vacation addresses. … For example, Liverpool's population declined by 3% according to the Census figures, and 7% according to the MYEs" (2004: 2-3).

Lupton and Power note that government Mid-Year Estimates (MYEs) provide consistent population definition over time but are not produced separately for 1991 for characteristics of population such as ethnic group, although these are more likely to be undercounted in the census. Their solution for neighbourhood analysis to adopt a set of population estimates funded by the Economic and Social Research Council which for 1991 included an element for non-response (Simpson S, 2002), but these are not consistent with the latest thinking on the total 1991 population from statistical agencies.

Some comparisons between censuses are misleading if inconsistencies between censuses are not allowed for. Later in this paper we shall show that some minority ethnic population increases according to the census are in fact decreases when full populations are compared over time. As well as changes in population definition, non-response and variable categories, Britain's predilection for changing boundaries for census output has misled analysts. A major government report on the state of English cities singled out Blackburn as the area with a significant increase in ethnic segregation, but the apparent increase was entirely an artefact of different boundaries used in the 1991 and 2001 Censuses (Parkinson *et al*, 2006; Simpson, 2007).

The contributions of this paper are to specify the problems of census tabulations as indicators of population change, and to overcome them. By providing complete and

consistent sub-national mid-1991 and mid-2001 population estimates for very small areas and single years of age, with sex and ethnic group disaggregation, we allow social researchers to undertake more analyses and to avoid misleading analyses.

Other countries face similar problems in constructing accurate time series of full population estimates with an ethnic group dimension, here reviewed through the experience of the USA, Canada, New Zealand and Australia. However, different approaches are taken. For example in the U.S., the mid-2000 population estimates are derived from the census usually resident population using a cohort component method, thus accounting demographic change (births, deaths and net migration) between Census day and mid-year in sub-national areas for each age, sex and race, and Hispanic origin group. The approach also takes into consideration the net movement of U.S Armed Forces overseas. One of the main challenges appears to be the recoding of each of the persons who identified themselves in the 'Some other race' category in the 2000 Census categories to one or more of the five Office of Management and Budget (OMB) race categories, which is used for the presentation of population estimates. Underenumeartion and duplication of persons were thought to balance each other, so that no adjustment was made (Siegel, 2002; USCB, 2006).

In Canada, quarterly population estimates are produced but without ethnic group (SC, 2006). However, population projections for visible minority populations for provinces and regions by age and sex for census years are generated by microsimulation. The base population used consists of a 20% census sample of permanent residents, which is adjusted for underenumeration (Bélanger and Caron Malenfant, 2005).

In New Zealand, population estimates are produced from the census usually resident population. The method is based on a cohort component method which takes into account demographic change between Census night and mid-year by ethnicity, age and sex for sub-national areas. Adjustments for residents temporarily overseas and for non-response are also made to the population estimates following a post-enumeration survey, taking into account national differentials by age, sex and ethnic group, but without area differentials (SNZ, 2007).

In Australia, population estimates of the resident population by age, sex and indigenous status are similarly derived from the census and from a census post-enumeration survey for sub-national areas. The latter is used to include an adjustment of census non-response to the population estimates by area, sex, age, country of birth and indigenous status, before an additional allowance for change between census day and mid-year is included (ABS, 2006).

In the UK, mid-year population estimates with an ethnic group dimension have rarely been produced. This was attributable to the lack of data classified by ethnic group prior to the 1991 Census (Haskey, 1988). The availability of 1991 Census data by ethnic groups led a number of researchers to the question of devising methods of estimating the ethnic composition of sub-national areas in 1981 by calculating the ethnic breakdown of the population by country of birth to provide disaggregated estimates of population change by ethnic group over the decade 1981-1991 (Owen, 1996, Rees and Phillips, 1996, Peloe and Rees, 1999). Estimation of 1991 census undercount for each ethnic group in sub-national areas was estimated in various ways (Simpson S, 2002; Mitchell *et al*, 2002).

The release of 2001 Census data with an ethnic group dimension represents the second time for which data for local areas with an ethnic group dimension is available to analyse the changing composition of ethnic groups in the UK. The Office for National Statistics (ONS) has published estimates of the mid-2001 population for ethnic groups for each local authority area of England (Large and Ghosh, 2006), and have revised their estimates of overall 1991 Census undercount (ONS, 2002) making previous work on this with an ethnic origin dimension less useful.

The next section of this paper specifies four challenges in creating consistent population estimates for 1991 and 2001, and our methods to overcome them. The resulting dataset is not only consistent with ONS population estimates for 1991 and 2001 but contains much more age and geographical detail. We then quality assure the results through their internal consistency and the external plausibility of the adjustments that have been made.

Examples of the impact on census analysis are given, using the example of Birmingham the largest local authority area of England and Wales, followed by a discussion of the general applicability of the methods and of the results.

Method

The four challenges for comparing 1991 and 2001 Census output

Although the 1991 and 2001 Censuses in Great Britain have measured the principal variables to compare populations over time and space, four standard but difficult problems of data harmonisation over time remain. These four problems are general to any country when comparing population estimates over time. Their impact for England and Wales for the period 1991-2001 is highlighted in the text and in Table 1, and described below with an indication of the solutions adopted to overcome them.

(1) Population definition. Who is included in the definition of population affects the population estimate published, even where several different 'population bases' have been used in fieldwork (UN, 1998). In England and Wales, two differences between practice in the censuses of 1991 and 2001 are significant, the enumeration of students and population date. Whilst the 2001 Census enumerated the whole population at the address of 'usual residence' including students at their term-time address, the 1991 Census enumerated students at their vacation address. The transfer of students from their vacation address to their term-time address in 1991 has a significant impact on assessment of population change, by increasing the 1991 population in areas with student campuses (often but not always within urban areas), and decreasing other areas from which students leave to study elsewhere. Because population estimates are usually made for mid-year (30th June) rather than Census day (a different day of April in 1991 and 2001), an additional allowance for timing is necessary to bring them both to the same population date. Although the net effect of timing is small nationally, its impact locally can be significant.

Table 1 Enhancements to comparisons between successive censuses

Enhancement, 1991 and 2001 censuses	Global impact, England and Wales	Examples of extreme impact
1. Population definition	53,975 net addition	
a. Students, transferred from vacation address to term-time address (1991 only)	213,628 net gain for 103 districts 159,653 net loss for 273 districts	14,500 net gain to Oxford 2,600 net loss from Wirrall
b. Population date, change from census day to mid-year		
1991: April 21 to June 30	43,094 net addition	974 net gain to Lambeth, 442 net loss to Brent
2001: April 29 to June 30	41,006 net addition	1,081 net gain to East Riding of Yorkshire, 1,746 net loss to Birmingham
2. Non-response not estimated within census output	In 1991, 1.6% addition In 2001, 0.5% addition	Pakistani addition of 6.7% in 1991, 2.1% in 2001. Manchester addition of 4.0% in 1991, 7.4% in 2001.
3. Demographic classifications		
a. Age, distribute broad age groups to individual ages	No net impact on population	Largest approximations in smallest areas where 5 age groups published for each ethnic group in 1991, 7 in 2001.
b. Ethnic groups 10 in 1991; 6 extra in 2001.	Of those in both censuses, 3.2% changed categories	77% of those recorded as Black Caribbean in 1991 were recorded as Black Caribbean in 2001, while a similar number moved from other groups to Black Caribbean.
4. Harmonisation of geographical units. Smallest 1991 areas converted to 2001 Census units	139 of 403 local authority boundaries and 4,398 of 9,527 electoral ward boundaries changed involving more than 1% of their population	The 2001 boundary of the district of York was created from the 1991 district boundary and parts of Harrogate, Ryedale and Selby in 1991.

(2) Treatment of non-response. Since it is widely accepted that no census will count the whole population, adjustments are usually made for undercount and in some countries for compensating overcount. In England and Wales in 1991 and 2001 the treatment of non-response in 1991 and 2001 was substantially different. In 1991 extra records for people in missed households were included in the census database and published output but a further 2% were estimated as missed from the census output (OPCS, 1993). In 2001 the One Number Census (ONC) integrated a more complete estimate of non-response in the published census counts for all areas, with further non-response limited to about 0.5% (Simpson, 2007). In both years, the non-response missed from census output was skewed towards young men, urban areas and minority ethnic groups. Plausible estimates based on evidence from post-enumeration surveys can been used to make allowances for this non-response.

(3) Demographic classifications. While not resulting any change to the total count of population, changes in recording and coding practices can render censuses incompatible, as happened in England and Wales with ethnic identification and age group categories. Whilst the 2001 Census recorded 16 ethnic group categories, including four mixed categories, the 1991 Census output included 10 ethnic group categories, with no mixed categories (Aspinall, 2008). Analyses of ethnic group stability over time using the ONS Longitudinal Study (LS) data showed that reliable comparisons over time can be made for five groups: White, Indian, Pakistani, Bangladeshi and Chinese and less reliable comparisons for the Black Caribbean and Black African groups (Bosveld *et al*, 2006; Simpson and Akinwale, 2007). The residual ('Other') ethnic groups of both 1991 and 2001 exhibit very low stability and, therefore, are not appropriate for comparisons. Classifications in which more groups are combined (such as 'Black') offer greater stability but less meaningful interpretation as they combine groups with very different demographic trajectories. Although date of birth is captured during census fieldwork, published output uses age bands which are not compatible between censuses. For example age 85 and over in 1991 and 90 and over in 2001 for electoral ward and further discrepancies for smaller areas.

(4) Harmonisation of geographical units. The geographical boundaries of most countries' administrative units change over time, in ways that prevent calculation of

population change directly from output of successive censuses In England and Wales, small geographical units have been most affected by geographical boundary changes. To achieve harmonisation of these geographical areas 1991 population estimates for the smallest census areas are proportionally converted, using the 2001 Census boundaries for districts, Standard Table (ST) wards and Output Areas (OAs) as target geographies.

In this paper, we describe a framework to solve these problems when comparing ethnic group populations across time and space for districts and smaller areas (wards and OAs) in England and Wales.

Solutions to create a consistent time series

When making population estimates consistent, there are choices regarding the target for consistency. Each estimate could refer to population on census day, to larger geographical areas which are common in both censuses, and to the broadest age groups which are consistent between censuses, to name three options which were not chosen but might have been appropriate had the aim been to create a small set of population estimates with greatest accuracy without regard to official population estimates.

Instead the population estimates created have been planned to be consistent with mid-year official population estimates and to be disaggregated to fine classifications of age and geography to allow re-aggregation for a variety of general uses. These decisions have resulted in the following constraints:

a) The population estimates are consistent with (i.e they add up to) the population estimates published by ONS without an ethnic group dimension, in 1991 for local authority districts and in 2001 for districts and electoral wards (ONS, 2002). This implies for example that census output must be adjusted in both censuses to include the impact of moving from the April Census day to mid-year, and that the 1991 Census must include an adjustment to transfer students from vacation to term-time address.

b) The population estimates are also consistent with the 2001 population estimates published by ONS with an ethnic group dimension for districts in England (Large and Ghosh, 2006). These assume for the same non-response rate for each ethnic group to distribute the extra 0.5% or 276,000 population estimated by ONS after the 2001 Census results were released, including in particular young men in urban local authorities. In order to examine a differential allocation of this extra non-response between ethnic groups, a second set of estimates for districts and wards have been derived too, which apply differential non-response rates based on those estimated within in the One Number Census.

c) 1991 population estimates are converted to the boundaries used in the 2001 Census output, including the smallest geographical unit, the OA. These include all electoral boundary reviews agreed by the end of 2003; although they are referred to here for example as '2001 areas' they in fact use boundaries existing in 2003 that may have changed since 2001. The population figures for each OA are not as accurate as for larger areas and are not intended to be used directly for estimates of population change but as the building bricks for larger scale analyses.

d) Single year of age to 90 and over is estimated for both 1991 and 2001, to allow subsequent aggregation to suitable age bands. Again, the estimates of population for single years of age in small areas for each ethnic group are not considered accurate in themselves, but enable the construction of age bands relevant to particular services and policies.

e) Ethnic group categories for 1991 and 2001 are maintained in their respective full detail for each set of estimates. The matching of categories to compare 1991 and 2001 is made subsequently by users of the data. In this paper an eight-category classification is used as advised by ONS (Bosveld *et al*, 2006) and Simpson and Akinwale (2007).

Meeting these constraints involves a set of technical solutions which are detailed in Sabater (2008). The appendix Tables 1-4 summarise the methods, which involve two

basic principles: (a) incorporating relevant evidence, and (b) scaling incomplete evidence to more reliable information, usually for larger areas. Prime relevant evidence is the census output itself, which is not rejected but is the rock which is built upon to fill its imperfections. Tabulations from the census and its post-enumeration surveys also provide information about the ethnic composition of students, the transfer of students from vacation address to term-time address, migration rates to estimate population change between census day and mid-year and differential levels of non-response by age, sex and ethnic group. Extensive work during the 1990s census by the Estimating with Confidence research programme, which created an accepted set of small area population estimates for mid-1991 (Simpson *et al*, 1997; Simpson L, 2002) has also been incorporated, improving on its internal consistency, extending its age detail and using revised estimates for larger areas of the level of non-response in the 1991 census.

Much of the evidence indicates the nature of adjustments required to the census, often in terms of rates and distributions, and for broad population categories, rather than absolute figures to apply directly to the detailed local age-sex-ethnic group categories which are the target of the exercise. The technique of scaling initial estimates based on such incomplete evidence to other more reliable information known for larger population categories is frequently used, and is termed 'fitting'. Where the more reliable information is known for more than one set of marginal sub-totals of a more complex cross-classification, the technique of Iterative Proportional Fitting is used.

To appreciate the complexity of combining data through use of fitting a variety of relevant evidence, the single example of ONS' downward revisions to non-response in 1991 will suffice. We shared these between local areas by adjusting only the previous estimate of non-response within each age-sex-ethnic group. Had the population as a whole been adjusted directly then the geographical pattern of non-response would have been implausibly changed.

Quality assurance

How can one give assurance about the quality of the results? Many of the assumptions made are plausible rather than certain, and subject to error. This error adds to the errors of recording, processing and imputation involved in the Census itself. The

potential for error in any one estimate will be greater for smaller populations, at least in percentage terms. This part of the paper argues that although there is no measured truth against which the accuracy of the results can be measured, internal checks and post-hoc validation can provide reassurance for users of the results.

First, we discuss the importance of plausible construction of the estimates to their validity, and the variety of checks which have been made to ensure that the results are internally consistent, and thus have a degree of *internal validity*. Second, we present results to allow users to judge whether the datasets and adjustments made to the census 'make sense' of what they expect. This provides a degree of face validity or *external validity.* These approaches to quality assurance are taken from notions of statistical responsibility (expressed well in Radical Statistics Education Group, 1982, which in turn draws on Cook and Campbell, 1979 and Bross, 1960).

Internal validity

Each element of the methods explained earlier on was subject to constraints to ensure internal consistency. In addition to full consistency with ONS estimates as discussed earlier, there are no negative populations although adjustments to census output may be negative, for example the net adjustments for transferring students from vacation to term-time address in 1991, and the impact of population change between each Census day and mid-year. All procedures have been constrained so that there are no negative populations in the final data sets using methods adapted from the treatment of marginal totals with positive and negative entries (Bryan, 2004).

In addition, wherever possible individual adjustments are estimated separately and in small areas, retaining their own coherence and maintaining known patterns of age, locality, sex and ethnic group. The treatment of non-response in 1991 described above is one example. Another is the conversion from 1991 to 2001 Census geographical boundaries, which has used the lowest geographical source unit possible (the 1991 Census Enumeration District) so that spatial differences in age-sex-ethnic group patterns are respected when constructing the 1991 population for 2001 boundaries.

Internal validity is also shown by using assumptions that are equally or more plausible than other possible assumptions, and by showing that where other assumptions are

equally or more plausible (but perhaps impossible to implement) the outcome would not have a misleading effect on users of the data. As an illustration, we were concerned that the extra 0.5% or non-response estimated after 2001 census output was finalised was distributed to ethnic groups in proportion their existing populations within census output, in the ONS estimates which we have used as a constraint. However, most of this extra non-response was due to incomplete enumeration which might be expected to show the same patterns of non-response as for the much larger volume of non-response already estimated by ONS within the census output. In particular a disproportionate omission of minority ethnic groups has been noted by ONS (ONS, 2003) such that nationally Indian, Pakistani, Bangladeshi, Caribbean, African and Chinese were each omitted from census enumeration at a rate more than twice the average, and that this differential was repeated in most local authority areas. Table 2 shows the population of England and Wales according to the ONS estimates and an alternative set which assumes the extra non-response was distributed to ethnic groups in the same way as estimated by ONS (2003). It shows that the difference in population estimates is small and does not make a significant difference to the assessment of population change over time. Nonetheless the assessment of population change for sub-national areas and in particular for age groups most affected by the extra non-response (men aged 20-39) may be significantly affected by this assumption.

External validity

There is no external truth by which to judge the time series, but its analysis may provide some 'face validity' of the results if it agrees with expectations. In this section we provide three analyses of the results, one highlighting the impact on population growth with and without enhancing 1991 and 2001 census totals for four minority ethnic populations in Britain, one focusing on national totals for each ethnic group and with detail of age and sex structure, and the other illustrating the impact of adjustments to ethnic groups locally. For the latter, population estimates for each ethnic group in Birmingham are shown.

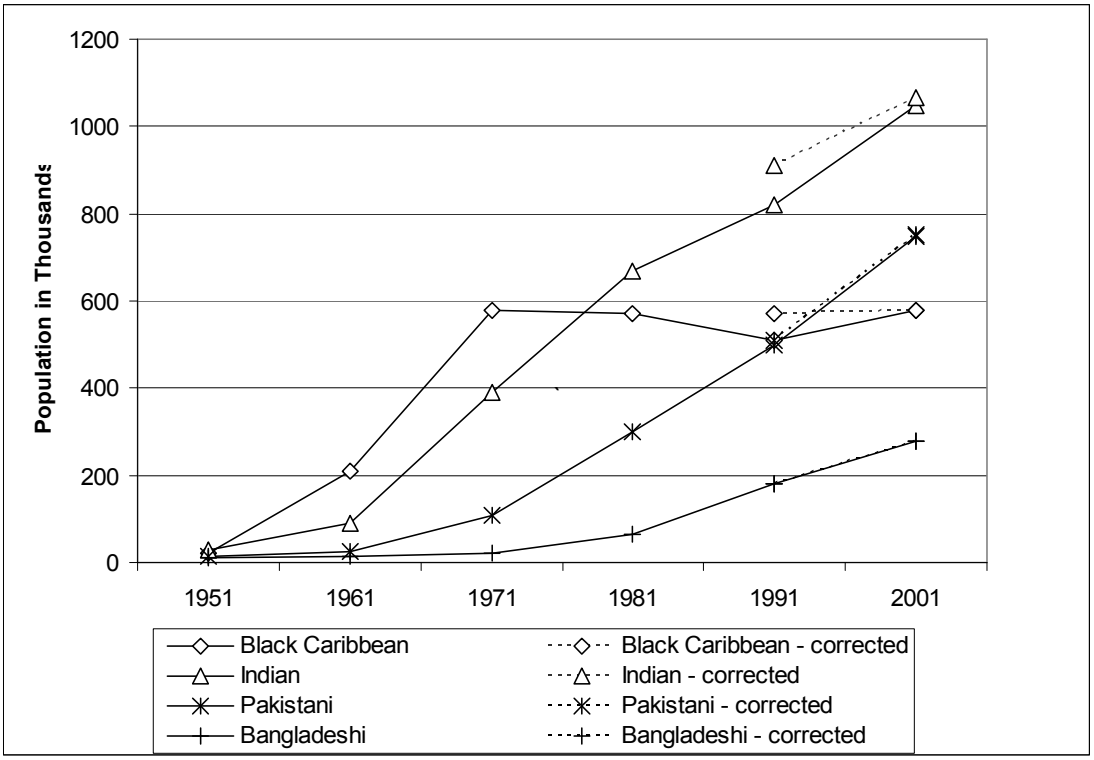
Table 2: Impact of alternative 2001 population estimates, England and Wales, total population and men aged 20-39

	Population in 2001		% change in population 1991-2001	
(a) Total population	ONS estimate	Alternative estimate	ONS estimate	Alternative estimate
Total	52,359,979	52,359,979	3.18%	3.18%
White	47,747,355	47,716,647	0.67%	0.61%
Black Caribbean	572,212	576,850	0.45%	1.27%
Black African	494,669	502,667	93.73%	96.87%
Indian	1,053,302	1,059,351	18.11%	18.78%
Pakistani	727,727	734,585	47.02%	48.41%
Bangladeshi	286,693	289,488	62.05%	63.63%
Chinese	233,346	236,090	34.74%	36.32%
Other	1,244,677	1,244,301	64.39%	64.34%
	Population in 2001		% change in population 1991-2001	
(b) Men aged 20-39	ONS estimate	Alternative estimate	ONS estimate	Alternative estimate
Total	7,386,875	7,386,875	-2.23%	-2.23%
White	6,565,396	6,547,140	-4.59%	-4.86%
Black Caribbean	92,583	95,300	-26.09%	-23.92%
Black African	101,225	106,775	41.11%	48.85%
Indian	189,420	193,071	12.24%	14.40%
Pakistani	130,173	134,045	54.02%	58.60%
Bangladeshi	51,683	53,288	99.50%	105.70%
Chinese	47,453	48,996	6.99%	10.47%
Other	208,942	208,260	36.05%	35.60%

The alternative estimate assumes different rates of non-response for each ethnic group (see text).

Figure 1 uses data reproduced from Peach (1996) and 2001 Census data as published to represent population growth between 1951 and 2001 for four separate ethnic minority populations. Additionally, the figure highlights the differences using corrected data for 1991 and 2001, with a significant decrease particularly among the Black Caribbean, slowing down rather than accelerating mainly as a result of the adjustment due to non-response not included in the 1991 Census. This adjustment also contributes to the more expected figure of growth of the Indian group in 1991.

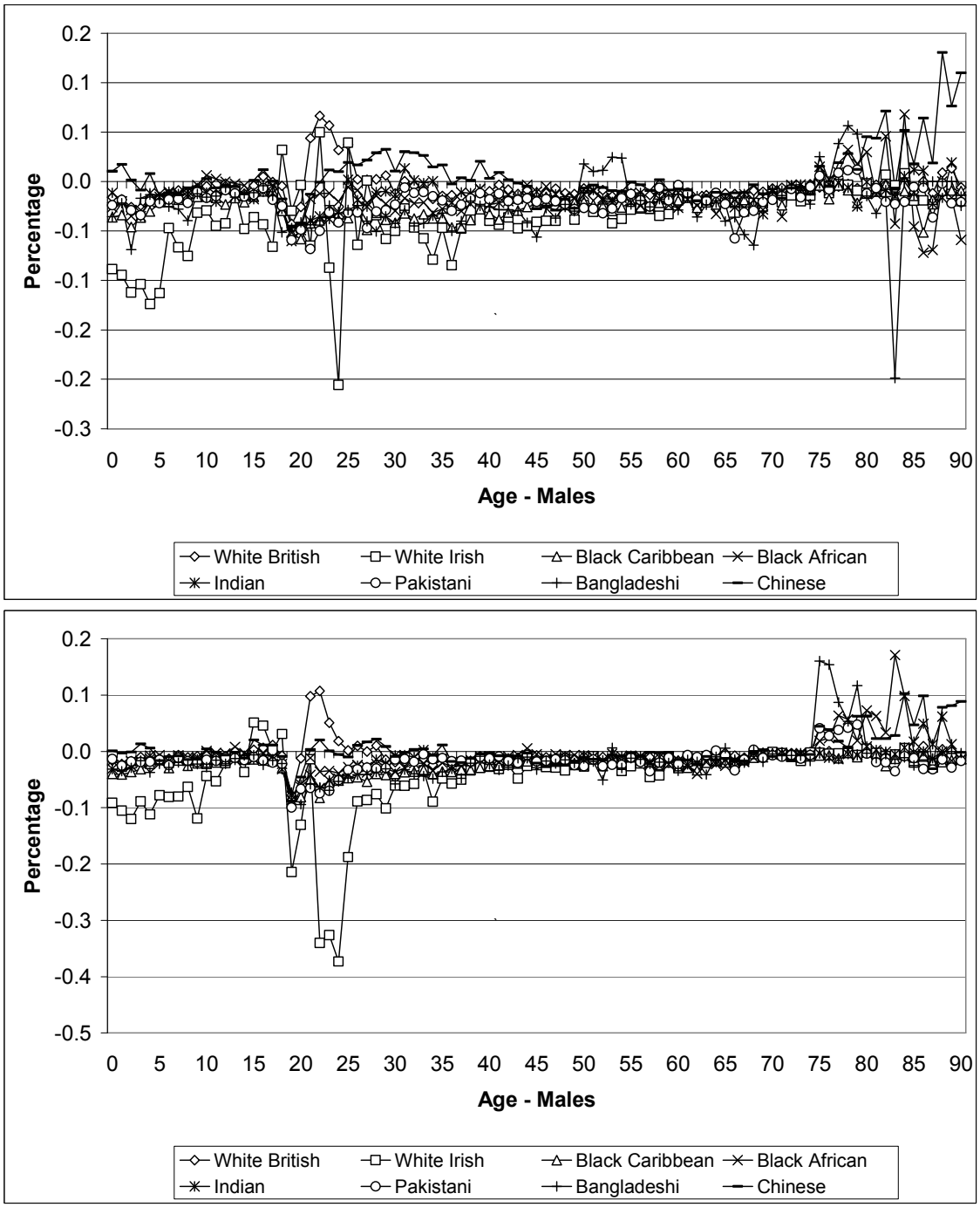
Figure 1: Growth of minority ethnic populations in Britain, 1951-2001



Source: Adapted from Lupton and Power (2004) using 1951-1991 data reproduced from Peach (1996: 9). 2001 data (without correction) from 2001 Census Key Statistics Table 6.

The following figures illustrate how the 2001 population estimates were derived from Census output, after aggregating the results for all districts to country totals. Figures 2 and 3 show the adjustments due to migration between England and the rest of the United Kingdom (UK) and international migration respectively during the 9 weeks between census day in April and mid-year, which is the element of change that has greatest impact on the population. These reveal the importance of emigration from England of the Irish group to other areas of the UK, most likely to Northern Ireland, as well to outside the UK ('international migration'), probably mainly to the Republic of Ireland. These together deduct between 1 and 5% of the Irish population aged in their mid-twenties, and will include the return of graduates after study in England.

Figure 2: 2001 percentage adjustment due to migration between England and the rest of the UK from Census day to mid-year by age, sex and ethnic group



The rises of population during this nine weeks period are principally due to net international immigration of Chinese men and women, Bangladeshi women and Black African men, also focused on young adult ages. The results in Figures 2 and 3 illustrate the assumptions used by ONS for estimates of mid-2001 population for districts of England, which we have replicated.

Figure 3: 2001 percentage adjustment due to international migration between Census day and mid-year by age, sex and ethnic group in England

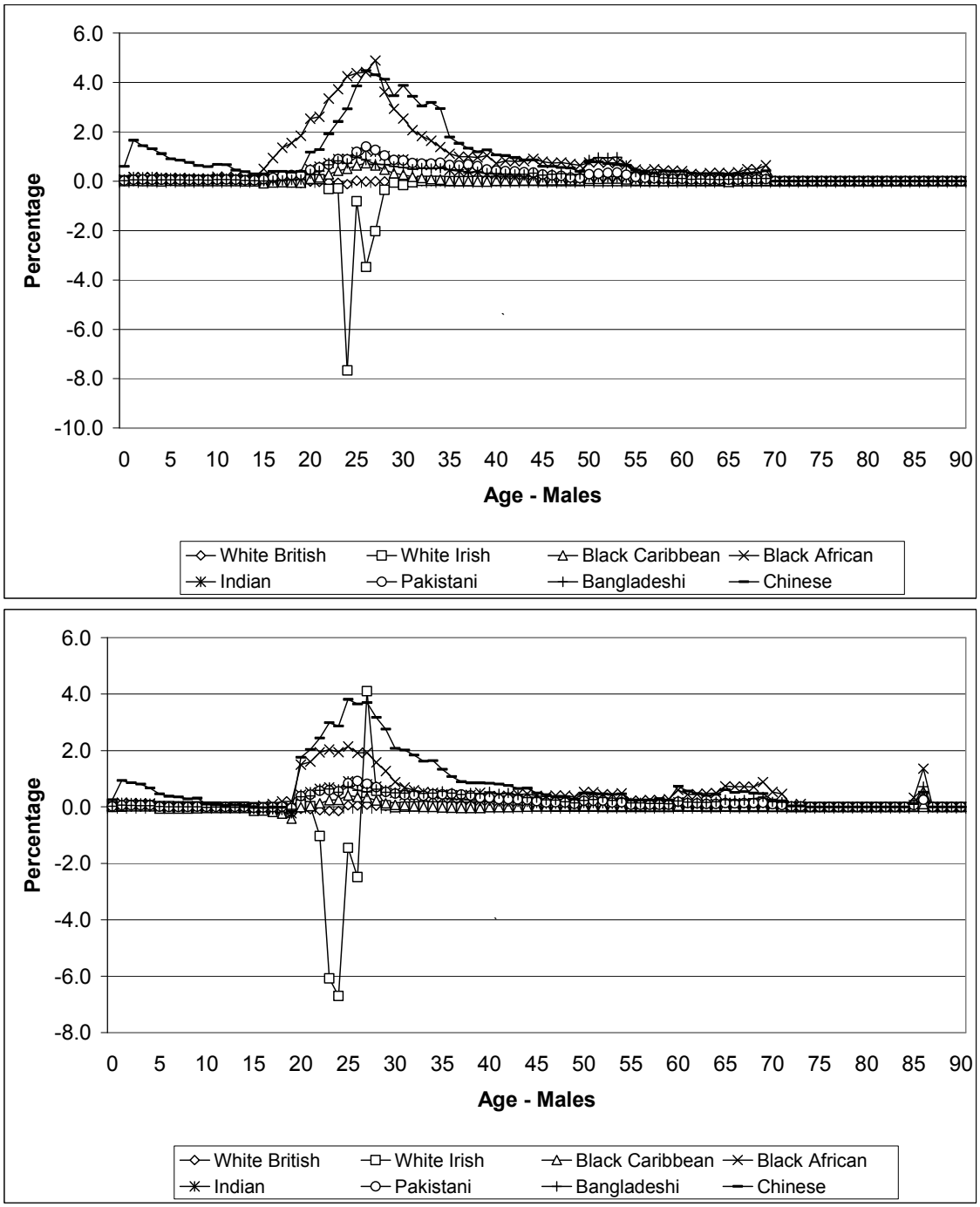


Figure 4 shows the adjustment of extra non-response identified by ONS after release of census results. The differences between ethnic groups are due to the concentration of each group in particular types of districts with different levels of extra non-response (the White British population is found more often in districts of lower non-response). As expected, this extra non-response is mainly concentrated among young male adults. The largest adjustment is for extra non-response among the Black

African group, with an increase over the published census population of more than 10% for those in their twenties and early thirties, solely due to their location mainly in London districts with high allocation of extra non-response.

Figure 4: 2001 percentage adjustment due to extra non-response as published by ONS experimental statistics by age, sex and ethnic group in England

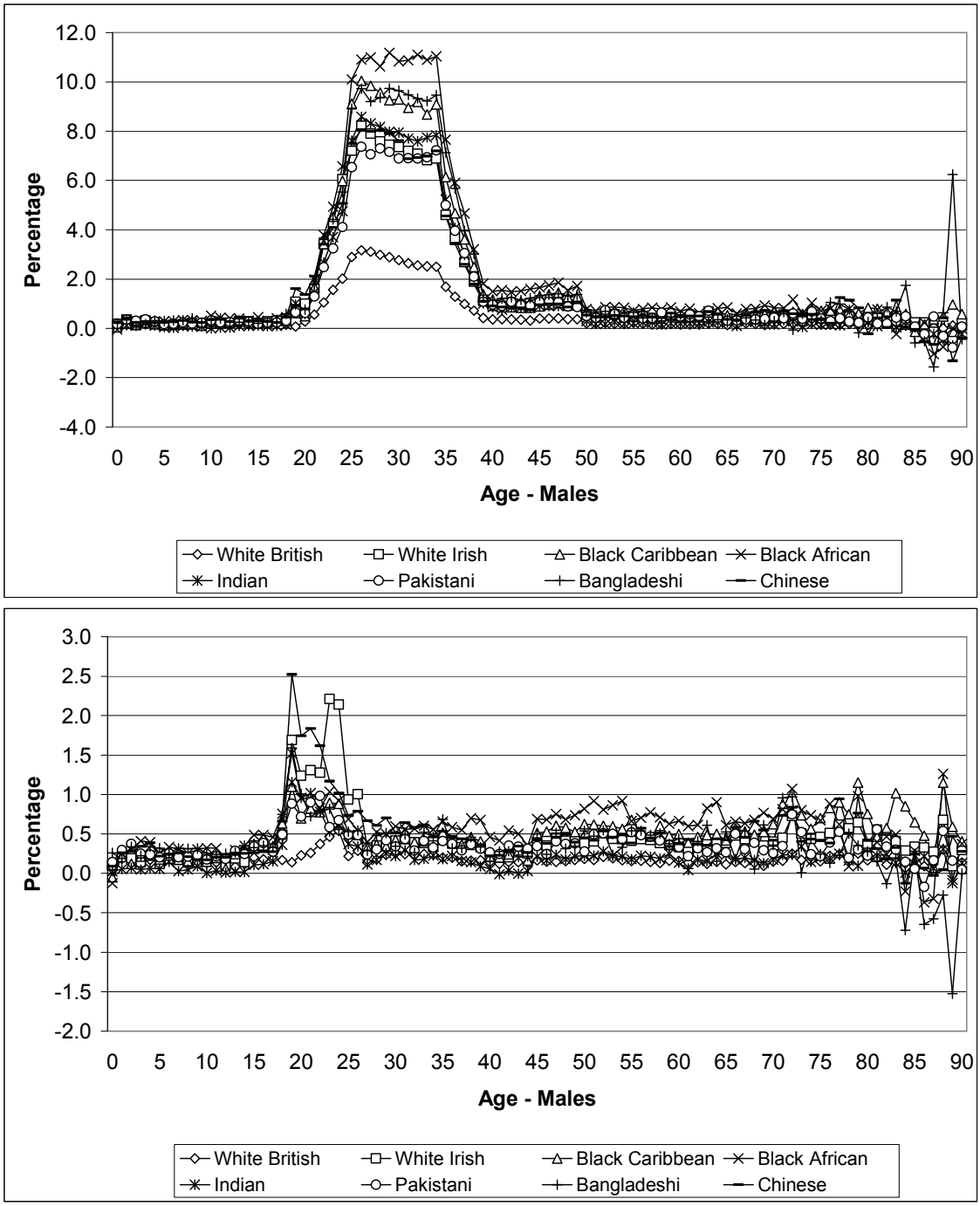
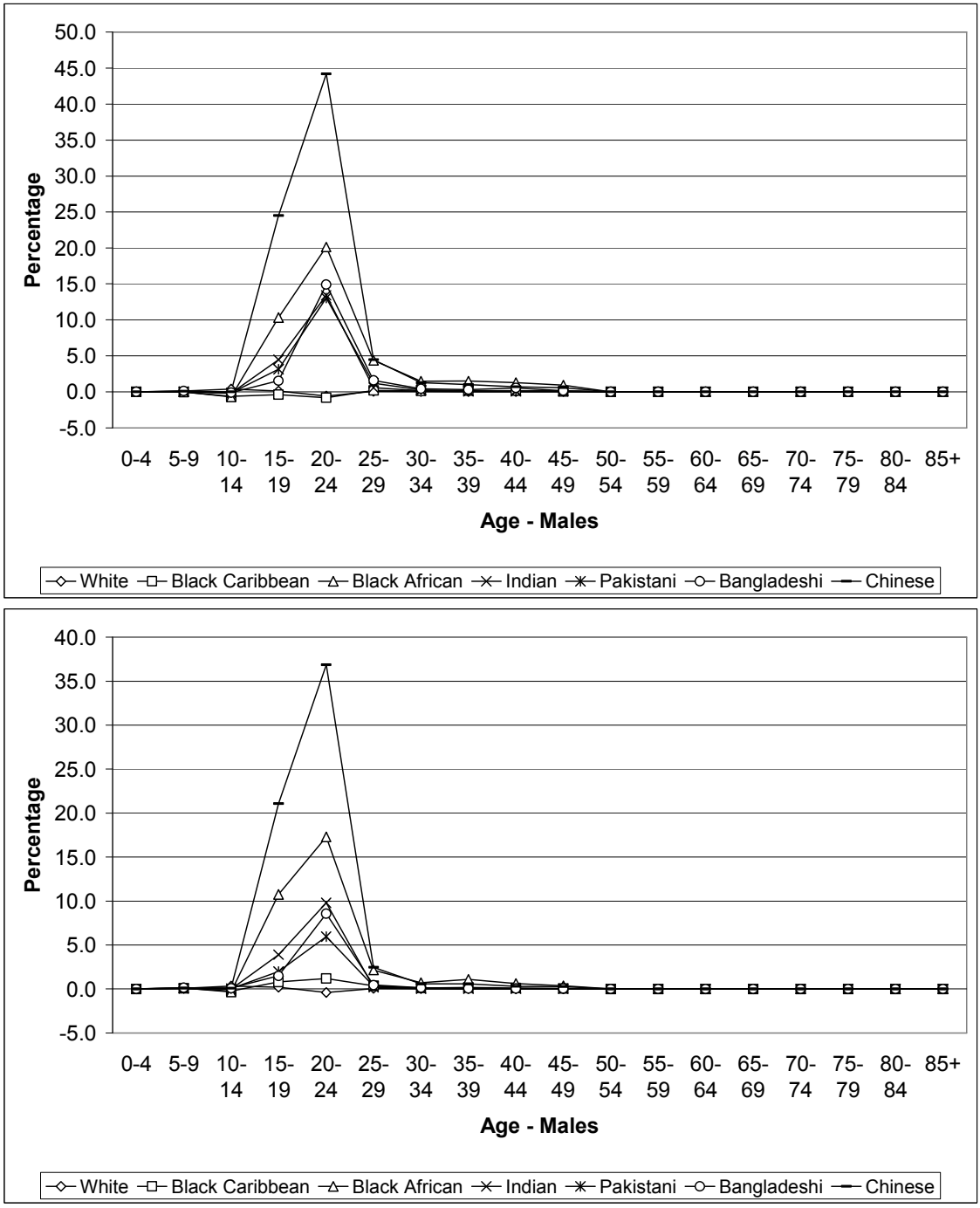


Figure 5 displays the percentage adjustment to the 1991 Census due to the student transfer from vacation to term-time address. Aggregated over all districts in England and Wales, this net impact of students otherwise resident outside England and Wales

presumably comprises mainly overseas students. The Chinese group experiences the largest addition of students with home address outside England and Wales, with an addition to the initial census population aged 20-24 of about 40% for both males and females.

Figure 5: 1991 percentage adjustment due to net student transfer by age, sex and ethnic group in England and Wales



In order to interpret sub-national population change between 1991 and 2001, Figure 6 shows the impact of adjusting each census for a consistent treatment of students, non-

response and population definition. The map showing census output is adjusted only so that 1991 figures refer to 2001 district boundaries. For this purpose we use a Universal Data Map (UDM) as this allows a more clear representation of those areas with large populations such as cities than conventional maps, which tend to highlight patterns in sparsely populated areas where few people live (Dorling and Durham, 2006).

Both the census and the full population estimates displayed in Figure 6 depict a widespread population growth of the non-White groups in districts in England and Wales. Many districts experience a growth in the total non-White populations of over 60%. The only areas showing a decrease are three districts where the USA armed forces have withdraw in significant numbers during the 1990s (Suffolk Coastal, Cherwell, and Forest Heath). The spreading out of cultural diversity is clear from the greater growth experienced outside the urban centres of London, the West Midlands and Yorkshire. Both maps show these two trends of minority population growth and spreading diversity, but a comparison of the two maps shows that the census output would be misleading on both trends. First, there are many more areas of slower population change indicated on the map of full population estimates because the better capture of non-response in the 2001 census wrongly appears to be population growth. The unadjusted census over-estimates increases in the non-White population. Second, the over-estimation of non-White population growth is mainly in the urban areas where census undercount is greatest. Thus the spreading of diversity is under-stated by the census. The full population estimates show more clearly that minority population growth is faster outside the urban areas.

Generally, when full estimates are used, the tendency is to slow population growth, with a shift towards less population increase, in particular because the 1991 Census is boosted by students living in the UK only during term time, and by a full allowance for non-response. The use of full estimates makes some districts not only reduce their population growth but suggest a population decrease rather than an increase. One such district is Birmingham, used in Tables 3 to 5 illustrate the local impact on population comparisons between 1991 and 2001 for each ethnic group. For this purpose ethnic groups are aggregated to eight categories as advised by ONS (Bosveld *et al*, 2006)

and Simpson and Akinwale (2007).[1] Birmingham is an ethnically diverse district, whose boundary changed during the decade, with significant non-response in the census, and which gains several thousand students overall in the transfer between vacation and term-time address in 1991.

Figure 6: Percentage population change between 1991 and 2001 for non-White groups for 2001 districts in England and Wales

Table 3 shows the adjustments to the 1991 Census output by ethnic group. Overall, the impact of adjustments adds forty four thousand residents to the 1991 Census as published, pushing the total population over 1 million. A significant enlargement to the area of Birmingham added almost nine thousand residents from neighbourhoods that were almost entirely White. The largest contributor to the adjustments is non-response, particularly among ethnic groups other than White, which adds twenty-eight thousand residents missed by the census in 1991. This represents 65% of the total

[1] The tables use the following allocation of 1991 and 2001 categories. White: 1991 White, 2001 White British, White Irish and White Other. Caribbean, African, Indian, Pakistani, Bangladeshi, Chinese: in both 1991 and 2001 the single categories with these labels. Other: in both 1991 and 2001 the remaining categories which are residuals or Mixed.

addition to the 1991 Census as published, and is presented in the table together with the much smaller timing adjustment.[2]

The impact of transferring students from vacation to term-time address represents a gain too for all ethnic groups in Birmingham, adding in total six thousand residents. The differences between ethnic groups partly reflect their different age structures, and partly the procedure used to add students which recognises that students moving to Birmingham will not reflect the local ethnic composition.

Table 3: Adjustments to the 1991 Census output by ethnic group, Birmingham district

	Census 1991 as published	Full population estimate with 2001 boundaries
Total	960,686	1,004,502
White	753,937	772,094
Black Caribbean	44,769	53,717
Black African	2,797	3,627
Indian	51,057	55,512
Pakistani	66,081	71,055
Bangladeshi	12,733	13,693
Chinese	3,318	3,961
Other	25,994	30,843

Table 4: Adjustments to the 2001 Census output by ethnic group, Birmingham district

	Census 2001 as published	Full population estimate with 2001 boundaries	Alternative estimate
Total	977,105	984,642	984,642
White	687,406	691,952	689,703
Black Caribbean	47,832	48,075	48,442
Black African	6,205	6,430	6,542
Indian	55,749	56,245	56,517
Pakistani	104,018	105,137	106,188
Bangladeshi	20,836	21,062	21,257
Chinese	5,110	5,230	5,273
Other	49,949	50,511	50,720

The alternative estimate assumes different rates of non-response for each ethnic group (see text).

Table 4 shows the same information but for adjustments to the 2001 Census output. Here there is no impact of boundary change or student transfers as these are already included in the Census output. Overall, the impact of adjustments adds only seven thousand residents to the 2001 Census as published. This is mainly the result of extra non-response not included in the census output, which adds nine thousand residents to

[2] Non-response and the adjustment from census day to mid-year were not distinguished in the method as applied to ethnic groups in 1991. For the total population timing accounted for 384 residents and non-response for 28,212 in Birmingham.

the initial census output, and adds slightly greater proportions to the minority ethnic groups than to the White population.

Table 5 illustrates the impact of the work on population comparisons in Birmingham between 1991 and 2001. The changes when using a full population estimate are significant, giving a different – and we would argue more accurate – assessment of population change. In particular, the census output for Birmingham as a whole suggests a gain in population of 1.7%, but after taking into account the adjustments, this is seen to be a slight loss of 2.0%. The reduction in growth noted above is apparent for all minority groups. An apparent Black Caribbean increase of 7% is seen with full population estimates to be a loss of 10%, while a Black African increase of 122% is reduced to 80%. A White loss of 9% becomes a loss of 11% when using the full population estimates. A study of Tables 3 and 4 shows that the minority ethnic groups' growth is over-estimated by the published census largely because of the greater non-response in the 1991 Census, while the under-estimated loss of the White population is due largely to the enlargement in district boundary.

Table 5: Population change by ethnic group, Birmingham district

	Censuses 1991 and 2001 as published	Full population estimates with 2001 boundaries	Full population with alternative estimate and 2001 boundaries
Total	1.7%	-2.0%	-2.0%
White	-8.8%	-10.4%	-10.7%
Black Caribbean	6.8%	-10.5%	-9.8%
Black African	121.8%	77.3%	80.4%
Indian	9.2%	1.3%	1.8%
Pakistani	57.4%	48.0%	49.4%
Bangladeshi	63.6%	53.8%	55.2%
Chinese	54.0%	32.1%	33.1%
Other	92.2%	63.8%	64.4%

The alternative estimate assumes different rates of non-response for each ethnic group (see text).

The estimates as building bricks

The dataset provided as full population estimates contains far too more detail than can be reasonably expected to be validated in every respect. The detail of single years of age, and of very fine geographical detail is provided instead to allow aggregation to larger populations. Larger populations, because they incorporate more of the evidence directly available for broader age groups and larger areas, are more likely to be accurately estimated. The aggregation to appropriate age groups is straightforward. This short section addresses a means of aggregating the estimates to areas with ad hoc

geographical boundaries that may be of interest to particular policy programmes and urban studies but are not those used in the 2001 census output. In this way, the time series of population estimates for ethnic groups is available for *any* area and any requirement of age groups.

A common example of non-census geographical units required for many purposes is provided by boundary reviews agreed after 2003, the year reflected in 2001 Census output and these estimates. However, updating such estimates for current boundaries can be a straightforward procedure using an appropriate Geographical Conversion Table (GCT). Simpson (2002b) and Norman *et al* (2003) have devised methods to convert between different geographical systems which can be used in the face of boundary changes over time, and shown that the accuracy is good so long as the source units (in this case the 2001 census OAs of around 200 households) are small compared to the target geographical units. The following example is taken from Birmingham, a district whose ward boundaries changed in 2004. A GCT defined and constructed by Norman (2007), containing weights from postcodes points and a count of addresses at each valid postcode was used to transfer data from the source area (OAs) to the target areas (2004 wards).

Table 6 gives an example of this transfer for the full mid-1991 estimates. The table shows that the 2001 OA 00CNGM0106, which was fully included in Sutton New Hall ward, is re-allocated in two different 2004 wards, Sutton New Hall and Sutton Trinity. The table has an extract from the OA 2001 to ward 2004 GCT. The count of the address-weighted postcodes in the source-target overlaps indicate that 0.025 of data for 00CNGM0106 be allocated to 00CNJE and 0.975 of 00CNGM0106 be allocated to 00CNJF. Since the OA overlap is mostly with Sutton Trinity, the intersection estimate is almost entirely allocated in this ward (4 * 0.975).

Table 6: GCT extract: 2001 OAs to 2004 electoral wards, Birmingham district

a.) Source/target links and conversion weights

Source geography OA01	Ward 2001 code	Ward 2001 name	Target geography Ward 2004	Name Ward 2004	Source / Target intersection address	Source unit total addresses	Conversion weight
00CNGM0106	**00CNGM**	**Sutton New Hall**	**00CNJE**	**Sutton New Hall**	**3**	**120**	**0.025**
00CNGM0106	**00CNGM**	**Sutton New Hall**	**00CNJF**	**Sutton Trinity**	**117**	**120**	**0.975**
00CNGM0107	00CNGM	Sutton New Hall	00CNJF	Sutton Trinity	124	124	1
00CNGM0108	00CNGM	Sutton New Hall	00CNJF	Sutton Trinity	130	130	1
00CNGM0109	00CNGM	Sutton New Hall	00CNJF	Sutton Trinity	116	116	1
etc.	*etc.*	*etc.*	*etc.*	*etc.*	*etc.*	*etc.*	*etc.*

b.) Converting OA data to 2004 ward geography

Source OA01	Mid-1991 estimate White males aged 20	Conversion weight to 00CNJE and 00CNJF	Intersection estimate for White males aged 20 in 00CNJE and 00CNJF
00CNGM0106	4	0.025	0.096
00CNGM0106	4	0.975	3.744

Discussion

We have presented methodological issues and solutions to enable robust comparisons of population composition between the 1991 and 2001 censuses. Here we discuss the nature of the advances represented by this paper and its outputs, the potential that its outputs imply for other studies, and implications for future censuses and demographic studies.

This paper has shown that census data as published are not suited to analyse population change of ethnic groups in districts and smaller areas in England and Wales. The introduction of adjustments which take into account changes in the population definition, changes in the treatment of non-response, changes in ethnic identification and age group categories and boundary harmonisation, can significantly change the interpretation of population change of ethnic groups over time and space. In the work reported here, we have demonstrated that full population estimates can be used to improve analysis of population change over time and space.

The production of a quality-assured time series is challenging, especially when an ethnic group dimension is included. The work presented demonstrates that it is possible to address these issues and incorporate detailed age and ethnic group information into a set of consistent population estimates for very small areas. This detail allows aggregation to areas and age bands appropriate for any applied study. The population estimates together with data quantifying the adjustments made to the censuses have been deposited for general research use.

This paper provides 'face validity' for its mid-1991 and mid-2001 population estimates, giving evidence of how population change can be misleading without taking into account the adjustments for timing, non-response and students transferred at their term-time address. The results support the case that analysis of population change between 1991 and 2001 is clearly improved by including adjustments to census output and creating a consistent population time series. The demonstration that the data set makes a difference to sub-national comparison of population change of ethnic groups over time represents a wakeup call to those who have used raw census publications uncritically, and a positive one because a solution is provided by these new data sets.

General trends are altered, including the more rapid spread of population diversity. But specific conclusions are changed too; Birmingham's apparent population growth in the 1990s and Blackburn's increasing segregation in that decade are both false results based solely on census output. However, the census is by no means rejected. Its local detail is an irreplaceable ingredient of good social science. The message is rather one of caution, and of advice to construct improved analyses wherever that is possible.

Population studies and a broad range of social, employment and health studies gain from the improved data resource provided by consistent mid-year population estimates. Population studies of fertility, migration and mortality are enabled at a sub-national scale by estimates of age-structure and its change over time, of which Finney and Simpson (2008) is one example. Apart from these analyses, consistent estimates serve as the base for population projections and forecasts.

Many social science applications require population denominators in epidemiology, employment, crime and other substantive studies. Any study of service demand or of discrimination relies on an appropriate population for each group compared. Many analyses of change over time are simply not possible without a consistent and detailed set of population estimates using the same age and area classifications for more than on time period. Research measuring residential segregation and diversity at different points in time is sensitive to the changing boundaries of residential areas, as shown by Simpson (2007).

The UK 2001 Census was the first worldwide to attempt to measure non-response so completely that a full allowance could be incorporated in the census outputs themselves (ONS, 2003; Brown *et al*, 1999). It was largely successful and the estimates presented in this paper for 2001 gain significantly from those efforts which are being extended the next UK Census in 2011. Without the inclusion in the census of an allowance for non-response, published output would be biased in particular against men, young adults, and minority ethnic groups. This paper has shown that estimates of population change suffer when bias in the census enumeration is not fully corrected.

However, even if a consistent approach to non-response is included in census and equivalent outputs in the future, it may be too much to expect area boundaries and ethnic group classifications to remain static over time. Methods of harmonisation will remain a priority, highlighting the importance of longitudinal resources which allow the impact of boundary and classification changes to be quantified and at least approximately overcome. The availability of updated boundary and address files for statistical purposes, and of the linked census records over time will be essential to future harmonisation of population estimates just as they were to the current work.

The methods used in this paper focused on estimation of the greatest detail possible, for the smallest census areas of around one hundred households, for single years of age, and for the fullest ethnic group classification available in each census year. While the estimates of population are less reliable for smaller populations, and certainly not reliable for each cell of a cross-classification of age, sex, area and ethnic group, the advantage of this detailed approach to detail is evident when considering comparison to results from the next census. Its output cannot be predicted in advance, so the flexibility given by detailed estimates is the only means of coping with this uncertainty. A similar approach has begun to be taken by the ONS itself, whose small area population estimates since 2001 have extended to sub-district levels and are now held as a database of estimates attached to each unit postcode and single year of age, ready for aggregation to the unpredictable boundaries set by government neighbourhood social programmes (ONS, 2007). At present these do not extend to an ethnic group dimension between census years, although government reports have suggested that there is demand for such detail (Social Exclusion Unit's PAT18; Community Cohesion Panel, 2004).

The production of very detailed population estimates does however beg the question of their accuracy for differently sized populations. There is a need to develop statistical approaches which add a measure of reliability to each estimate. Only then will the relationship between population size, characteristics and accuracy of population estimates be available to guide users of population statistics towards robust results.

Acknowledgements

- The original Estimating with Confidence (EwC) data are copyright of the EwC project which was funded by the ESRC, award number H519255028.
- The research uses 1991 and 2001 Censuses, the National Statistics Postcode Directory (NSPD) and GIS boundary data obtained via MIMAS CASWEB and EDINA UKBORDERS which are academic services supported by ESRC and JISC.
- The Census, official Mid-Year Estimates, NSPD and vital statistics data for England and Wales have been provided by the Office for National Statistics and the digital boundary data by Ordnance Survey. These data are copyright of the Crown and are reproduced with permission of the Controller of HMSO.

Appendix: Technical summary

Precise details of datasets and methods adopted to create mid-1991 and mid-2001 population estimates are contained in Sabater (2008). This appendix contains the essential details to establish the strategies and prior research on which the estimates are based.

Table A1: 1991 population for 2001 OAs by age and sex without an ethnic group dimension.

The research makes use of the EwC population estimates for mid-1991 and its adjustments to the census, provided for census EDs (Enumeration Districts) and five-year age groups (Simpson *et al*, 1997). These are transferred to the 2001 census geography of OAs, ONS' revised estimate of non-response is included and single years of age added.

Table A2: 1991 population for 2001 OAs by age and sex: adding the ethnic group dimension.

Non-response in the census is assumed to be greater for minority ethnic groups, in line with the evidence contained in the 1991 SOCPOP dataset (Simpson, 2002a). The ethnic composition of students living away from home, and the single year of age structure within five-year age groups is based on relevant evidence.

Table A3: 2001 population for 2001 districts by age and sex with an ethnic group dimension.

The ONS estimates are used for districts in England (Large and Ghosh, 2006), which separately allocate to ethnic groups change between census day and mid-year, and the 257 thousand extra non-response not included in census outputs. For Wales, timing and non-response were allocated together. For all districts, the ethnic group composition of non-response is the same as the population ethnic group composition, for each age and sex group within each LAD. A separate set of estimates using disproportionate non-response among minority ethnic groups as evidenced by imputation within the Census has also been prepared (Sabater, 2008).

Table A4. 2001 population for 2001 electoral wards and OAs by age and sex with an ethnic group dimension.

Census tables and ONS population estimates with aggregated information on one dimension (age groups rather than single year of age, the total rather than separate ethnic groups) are combined using Iterative Proportional Fitting. The change of date to mid-year, and extra non-response, are in effect allocated proportionately to smaller sub-populations, from the district ethnic group estimates (Table 3) and ward estimates without and ethnic group dimension published by ONS (2007).

References

ABS (2006) *Indigenous Estimated Resident Population – Method of Calculation*. Canberra: Australian Bureau of Statistics publication.

Aspinall, P. (forthcoming - 2008) 'The future of ethnicity classifications', *Journal of Ethnic and Migration Studies*.

Bélanger, A. and Caron Malenfant, É. (2005) *Population projections of visible ethnic groups, Canada, provinces and regions, 2001-2017*. Ottowa: Statistics Canada publication.

Bosveld, K., Connolly, H. and Raha, C. (2006) 'Measurement and classification', in Dobbs, J., Green, H. and Zealey, L. (eds) *Focus on Ethnicity and Religion*. London: HMSO, 1-18.

Bross, I. (1960) 'Statistical criticism', *Cancer,* 13: 394-400.

Brown, J., Diamond, I., Chambers, R., Buckner, L. and Teague, A. (1999) 'A methodological strategy for a one number census in the UK', *Journal of the Royal Statistical Society Series A*, 162(2): 247-267.

Bryan, T. (2004) 'Population Estimates', in Siegel, J. and Swanson, D. (eds) *The methods and materials of demography*. San Diego: Elsevier Academic Press, 523-560.

Community Cohesion Panel (2004) *The End of Parallel Lives? The Report of the Community Cohesion Panel*, London: Home Office, online at http://www.communities.gov.uk/documents/communities/pdf/153866, accessed 14 January 2008.

Cook, T. D. and Campbell, D. T. (1979) *Experimental and quasi-experimental designs for research,* Chicago: Rand McNally.

Dorling, D. and Durham, H. (2006) *Universal Data Map cartogram of the UK*. Sheffield: Social and Spatial Inequalities Research Group, University of Sheffield.

Dorling, D. and Rees, P. (2003) 'A nation still dividing: the British census and social polarisation 1971 – 2001', *Environment and Planning A*, 35(7): 1287-1313.

Haskey, J. (1988) 'The ethnic minority populations of Great Britain: their size and characteristics', *Population Trends*, 54: 29-31.

Large, P. and Ghosh, K. (2006) *Population Estimates by Ethnic Group. Methodology Paper*. London: Office for National Statistics, online at http://www.statistics.gov.uk/downloads/theme_population/MethodologyforPEEG.pdf, accessed 14 January 2008.

Mitchell, R., Dorling, D., Martin, D. and Simpson, L. (2002) 'Bringing the missing million home: correcting the 1991 census small area statistics for undercount', *Environment and Planning A*, 34(6): 1021-1035.

Norman, P. (2007) *Geographical Conversion Table for 2001 to 2004 electoral wards in Birmingham*. Leeds: School of Geography, University of Leeds.

Norman, P., Rees, P. and Boyle, P. (2003) 'Achieving data compatibility over space and time: creating consistent geographical zones', *International Journal of Population Geography*, 9(5): 365-386.

Norman, P., Simpson, L. and Sabater, A. (forthcoming, 2008) 'Estimating with Confidence and hindsight: new UK small area population estimates for 1991', *Population, Space and Place*.

ONS (2002) *Methods used to revise the 1982-2000 annual mid-year population estimates for England and Wales*. London: Office for National Statistics, online at http://www.statistics.gov.uk/downloads/theme_population/Meth_PopES_82_2000.pdf, accessed 14 January 2008.

ONS (2003) *Imputation rates by variable*. London: Office for National Statistics, online at http://www.statistics.gov.uk/census2001/imputation_rates_by_variable.asp, accessed 14 January 2008.

ONS (2007) *Project on Small Area Population Estimates for England and Wales*. London: Office for National Statistics, online at http://www.statistics.gov.uk/about/methodology_by_theme/sape/default.asp, accessed 14 January 2008.

Owen, D. (1996) 'Estimating local change in the population of minority ethnic groups, 1981 to 1991', Warwick: Centre for Research in Ethnic Relations, University of Warwick, Working Paper No. 1.

Parkinson, M., Champion, T., Evans, R., Simmie, J., Turok, I., Crookston, M., Katz, B., Park, A., Berube, A., Coombes, M., Dorling, D., Glass, N., Hutchins, M., Kearns, A., Martin,

R. and Wood, P. (2006) *The state of the English cities*. London: Office of the Deputy Prime Minister, Vol. 1.

Peach, C. (1996) 'Introduction', in Peach, C. (ed) *Ethnicity in the 1991 Census*. London: HMSO, 2: 1-24.

Peloe, A. and Rees, P. (1999) 'Estimating ethnic change in London, 1981-91, using a variety of census data', *International Journal of Population Geography*, 5(3): 179-194.

Radical Statistics Education Group (1982) *Reading between the numbers, a critical guide to educational research.* London: British Society for Social Responsibility in Science.

Rees, P. and Phillips, D. (1996) 'Geographical spread: the national picture', in Ratcliffe, P. (ed) *Ethnicity in the 1991 Census: Social geography and ethnicity in Britain*. London: HMSO, Vol. 3.

Sabater, A. (2008) *Estimation of ethnic groups in sub-national areas for analysis of population change, England and Wales, 1991-2001*. Manchester: Cathie Marsh Centre for Census and Survey Research, University of Manchester, PhD Thesis.

Siegel, J. (2002) *Applied Demography. Applications to Business, Government, Law, and Public Policy*. San Diego: Academic Press.

Simpson, L. (2002) 'Geography conversion tables: a framework for conversion of data between geographical units', *International Journal of Population Geography*, 8: 69–82.

Simpson, L. (2007) 'Ghettos of the mind: the empirical behaviour of indices of segregation and diversity', *Journal of the Royal Statistical Society A*, 170(2): 405-424.

Simpson, L. (forthcoming - 2008) 'Title to be confirmed', *Journal of Ethnic and Migration Studies*.

Simpson, L. and Akinwale, B. (2007) 'Quantifying stability and change in ethnic group', *Journal of Official Statistics*, 23(2): 185-208.

Simpson, L., Cossey, R. and Diamond, I. (1997) '1991 population estimates for areas smaller than districts', *Population Trends*, 90: 31-39.

Simpson, S. (2002) 'Dealing with census undercount', in Rees, P., Martin, D. and Williamson, P. (eds) *The Census Data System*. London: HMSO, 171-184.

SC (2007) *Difference between Statistics Canada's census counts and population estimates*. Statistics Canada, online at http://www12.statcan.ca/english/census06/reference/Census_briefs_census_vs_estimates.cfm , accessed 14 January 2008.

SNZ (2007) *Population Base at 30 June 1996, 2001 and 2006*. Statistics New Zealand, online at http://www.stats.govt.nz/tables/population-estimates.htm, accessed 14 January 2008.

UN (1998) 'Principles and Recommendations for Population and Housing Censuses', *Statistical Papers*, New York: Department of Economic and Social Affairs, Statistics Division, Series M, 67.

USCB (1998) *United States Population Estimates by Age, Sex, Race, and Hispanic Origin Method*. U.S. Census Bureau, online at http://www.census.gov/popest/topics/methodology/2006_nat_meth.html, accessed 14 January 2008.

TABLE A1: 1991 population for 2001 OAs by age and sex and without an ethnic group dimension

Issues	Strategies	Methods / assumptions
Population definition		
Student transfer	Move students from vacation to term-time address.	The EwC transfers for students are used based on student location and government estimates for districts, for each ED.
Timing	Move from census day (21 April) to mid-year (30th June).	The EwC adjustment for timing is used. This assumes the same proportional demographic change (ageing, births, deaths and migration) between Census day and mid-year as for the containing district.
Boundary harmonisation	Census counts and EwC adjustments for 1991 EDs are transferred to 2001 OAs, and summed to larger areas.	A GCT is used to convert EwC data from the 1991 EDs to 2001 OAs. The ED-OA GCT is constructed from a postcode list provided by ONS, linked to both 1991 EDs and 2001 OAs and a count of addresses at each valid postcode. Overlaid ED and OA digital boundaries improve the quality of the allocation (Norman *et al*, 2008).
Non-response	The difference between EwC mid-1991 population and ONS revised mid-1991 population (estimated by ONS only for 2001 districts) constitutes the ONS revision to non-response.	The ONS revision to 1991 non-response in each 5-year age-sex group (mostly a reduction among men aged under 40) is distributed to a 2001 district's OAs in proportion to the original non-response estimated by EwC.
Ethnic group categories	Not applicable: Table 2 shares the population to ethnic groups.	
Age group categories	Single year of age (91 categories) is created from 5-year age groups.	Each OA five-year age-sex group is spread to single years of age using Iterative Proportional Fitting (IPF) to maintain consistency with the district single year of age from ONS. Initial values include births 1990-91 to represent those aged under 1 at mid-1991, and the more detailed composition of those aged 15-19 and 85+ given in 1991 census output.

TABLE A2: 1991 population for 2001 OAs by age and sex: adding the ethnic group dimension

Issues	Strategies	Methods / assumptions
Non-response	The pattern of different non-response rates for each group is taken from the ward SOCPOP estimates (5-year ages, sex, ethnic group).	The SOCPOP dataset for wards includes a non-response adjustment based on differential non-response specific to each 5-year age group, sex, ethnic group and ward. This is distributed to EDs assuming the same rate of age-sex non-response in each of the ward's EDs. It is further adjusted in the next step after conversion to OAs by making consistency with the all-group non-response in the OA.
Boundary harmonisation	Estimates for 1991 EDs are transferred to 2001 OAs.	As Table 1. The population estimates from the previous step, now for OAs, are scaled to agree with the OA all-group population estimates estimated from Table 1 after deducting the student adjustment.
Population definition *Student transfer* *Timing*	Move students from vacation to term-time address. Move from census day to mid-year (30th June).	For districts, OPCS adjustments to add and to remove students from each 1991 district are converted to 2001 district boundaries using suitable GCTs. The ethnic group composition of students removed because studying elsewhere is assumed the same as the resident population. The ethnic group composition of students added because at vacation elsewhere is assumed to be that of the 2001 ethnic group distribution of students, separately for males and females and ages 18-24 and 25+. These adjustments are distributed to wards and OAs, rather crudely but maintaining consistency both with the all-group student adjustment in each OA, and the ethnic group student adjustment in each higher level. Lack of evidence prevents a more precise allowance for each group. The timing adjustment is subsumed into the adjustment for non-response; it is small relative to each of the other adjustments.
Ethnic group categories Age group categories	Target categories are those used in 1991 (10 categories). Single year of age (91 categories) is created from 5-year age groups.	Student information from 2001 uses allocation to the 10 1991 groups advised by Simpson and Akinwale (2007) Using IPF, the OA 5-year age-sex-ethnic groups are made consistent with the OA all-group single year of age-sex totals (Table 1), the interaction (single-year by ethnic group) coming from the demographically smoothed SOCPOP data for the 1991 electoral ward containing the OA.

TABLE A3: 2001 population for 2001 districts by age and sex with an ethnic group dimension

Issues	Strategies	Methods / assumptions
Ethnic group categories	Target categories are those used in 2001 (16 categories).	None needed.
Age group categories	Target categories are available for districts in England (91 categories), and estimated for districts in Wales.	For districts in England, ONS Commissioned Census table C0533 has single year of age. For Wales, the 22 age categories of Census table ST101 are spread to single year of age to 89 and 90+, by demographic smoothing and then constrained to the single year of age published for the all-groups total.
Population definition		
Student transfer	Students already at term-time address in 2001.	None needed.
Timing	Move from census day (29th April 2001) to mid-year (30th June).	For districts in England each component of change between Census day and mid-2001 (births, deaths migration) was estimated by ONS by single year of age, sex and with an ethnic group dimension. For Wales, 2001 Mid-year estimates by single year of age and sex, without an ethnic group dimension, consistent with those released by ONS in September 2004 are used. The timing transfer is subsumed to the allocation of extra non-response, assuming equal proportional impact on each group, within each age-sex-district category.
Non-response	Extra non-response as estimated by ONS (September 2004) assuming same non-response rates for each ethnic group.	For all districts, the ethnic group composition of non-response is the same as the population ethnic group composition, for each age and sex group within each LAD. A separate set of estimates using disproportionate non-response among minority ethnic groups as evidenced by imputation within the Census has also been prepared.
Boundary harmonisation	Target geographies are those used in 2001 (376 districts).	None needed.

TABLE A4: 2001 population for 2001 electoral wards and OAs by age and sex with an ethnic group dimension

Issues	Strategies	Methods / assumptions
Boundary harmonisation	The targets are 2001 standard output (ST) wards, and OAs	ONS estimates for CAS wards are aggregated to ST wards.
Ethnic group categories	Target categories are those used in 2001 (16 categories).	None needed.
Age group categories	Single year of age (91 categories) is created from 5-year age groups (wards) and 7 broad age groups (OAs).	IPF is used to estimate more detailed cross-tabulations consistent with less detailed information. For example, the ward all-group 5-year age-sex group population estimates from ONS and the all group district single year of age-sex totals from ONS are constraints to the interaction (single-year by wards) from the Census table ST001.
Population definition		
Student transfer	Students already at term-time address in 2001.	None needed.
Timing	Move from census day to mid-year (30th June). Timing already incorporated in ONS ward population estimates (September 2004)	The use of IPF as above, allocates the timing adjustment already in district estimates to smaller areas proportionately.
Non-response	Extra non-response already incorporated in ONS ward population estimates (September 2004)	The use of IPF as above, allocates the non-response adjustment already in district estimates to smaller areas proportionately.