MANCHESTER 1824

The University of Manchester

Cathie Marsh Centre for Census and Survey Research

Quantifying stability and change in ethnic group

CCSR Working Paper 2006-05

Ludi Simpson and Bola Akinwale
ludi.simpson@manchester.ac.uk

www.ccsr.ac.uk

Quantifying stability and change in ethnic group

Ludi Simpson[1] and Bola Akinwale[2]

December 2004

Submitted to *Journal of Official Statistics*

(NB: this paper has been accepted for publication with revisions, probably in 2007. The revisions exclude most of sections 5 and 6, which justify the 8-category classification recommended for use in the UK and are retained here in full)

Abstract

We investigate the stability of individuals' affiliation to ethnic group categories using the Longitudinal Study of England and Wales linked between the Censuses of 1991 and 2001. While membership of the White category is stable, between seven and nine per cent of those recorded in an Asian group in 1991 have changed to a different group by 2001, while 23% of the Caribbean and African groups have changed. We quantify the separate influences of question unreliability, changes in categories, and conscious change of affiliation, finding that the latter contributes little instability over the period 1991-2001. The unreliability of the question is significant, due partly to the ambiguity of the categories for some people, and partly to imprecise imputation of missing values. We find the best correspondence between the different classifications used in 1991 and 2001 using empirical measures of fit. We investigate compatible classifications using larger and smaller numbers of categories.

Key words: race; reliability; longitudinal; census

Acknowledgements

UK Census data is Crown Copyright. Ludi Simpson was supported by ESRC Fellowship R000271214. Thanks to Joanne Lindley for analysis of the Labour Force Survey yielding Table 2.

[1] Cathie Marsh Centre for Census and Survey Research, University of Manchester, Manchester, M13 9PL, UK.

[2] Longitudinal Study Analysis Unit, Office For National Statistics, Drummond Gate, London SW1V 2QQ, UK.

1. Introduction

Classification of the human population using the concepts of “race”, ethnicity, skin colour, cultural origin, or country of descent is a common but contested practice on all continents (Kertzer and Arel, 2002; Coleman and Salt, 1996). Although these concepts are not equivalent, they are used in similar ways, often as demographic variables in national censuses and surveys. In many countries, including the United Kingdom, a single classification measures several of the above concepts (Aspinall, 2002). To refer to these classifications this paper uses the term “ethnic group”, as in the censuses and surveys of the UK.

Ethnic group classifications are used to identify relatively distinct populations, and to monitor their social conditions. Official data show that differences in access to resources and power are typically related to ethnic group membership, such that the collection of ethnic group data is often justified as necessary for implementation of legislation aimed at reducing social disparities stemming from discrimination on racial or cultural bases. In many countries, ethnic group categories also identify groups of recent migrant origin and ethnic group data are used in debates on international migration policy. The size of group populations and their characteristics are monitored over time to assess the success or otherwise of anti-discriminatory and immigration policies. Salt and Coleman's survey

> ‘reveals considerable international differences of practice in census questions on ethnicity, race and related topics. These reflect the historical origins of nations and perceptions as to their ancient unity or ethnic diversity, the volume and origins of their recent immigration streams, and the policy responses to ethnic diversity arising from different national constitutional traditions and political pressures. In choosing to ask a census question on ethnic origin directed to minorities of recent immigrant origin, Britain has made itself unique in Europe’ (Salt and Coleman, 1996: 26).

They also note the similarity of policy context in Britain, the US and Old Commonwealth countries with regard to ethnic monitoring, targets and legal recognition of group rights. In different political and historical contexts the same groups are measured not through labels of cultural or family origin, as is current practice in the UK, but through measurement of parental and grandparental birthplace as was past practice in the UK and is current practice in France (Tribalat, 2004).

Used in these ways, ethnic group classifications imply a stable characteristic that is carried through an individual's lifetime. The growth, geographical settlement and conditions of populations defined by ethnic group questions, are interpreted as the movement of distinct groups across time and space, and their changing social conditions. However, the *in*stability of ethnic group has been found to be neither insignificant nor random in contexts outside the UK. In addition to the unreliability associated with recording answers, people respond to questions about ethnicity within the constraints of the categories offered to them, aware not only of their personal self-identification but also of the social acceptability of each category. The UK statistical office accepts that 'any ethnic group label is only valid for the period and context in which it is used' (ONS, 2003a: 11).

Implications of change in ethnic group

The recording of an ethnic or racial category may change in three ways fundamental to demography: between cohorts, at different ages, and at different time periods. For example, far more new people identified themselves as American Indian in the USA 1990 Census than could be consistent with the 1980 Census records (Passell 1993; Nagel 1995); similarly the increase of 46% in Aboriginal and Torres Strait Islanders between the Australian Censuses of 1981 and 1986 reflected 'an increase in the propensity to identify themselves as such in the Census' (Evans et al, 1993). In Trinidad, the count of young adult Africans grew rapidly after the political successes of the Black Power movement in the 1960s. These changes can be seen as reflecting *cohort* experiences of socio-political movements. *Age* also has an impact on stability of ethnic group, especially

since different systems record births, childhood and adulthood. Thus many babies registered as Mixed at birth were recorded as African in childhood by their parents in British Guiana in the 1940s (Kuczynski 1953: 180). Finally, changes over *time* are particularly associated with question changes, as we shall see in the case of the UK between the 1991 and 2001 censuses.

As multiple ethnic identities become more common and more reported, ethnic group will become increasingly difficult to measure, as has been noticed already for people of White ancestry in the USA (Waters 2000).This changeability poses problems for the use of an ethnic group classification to monitor social conditions and to assess the success of measures to combat discrimination. Once we accept that classifications and individuals' allegiance to them are unstable, can we compare statistics of ethnic group population size and conditions across time? What ambiguities and errors may occur and how can we best measure and minimise them? This paper addresses those questions specifically in relation to England and Wales over the time period 1991-2001.

The practical question for many researchers is the compatibility of groups from the 1991 and 2001 censuses, which in England and Wales used different questions; standard output from the 2001 census has sixteen categories compared to the ten from the 1991 census. Section 2 specifies and discusses and presents evidence on three sources of instability in ethnic group. In the following section, we then propose various measures of stability, and apply them using data for England and Wales to quantify stability and change in ethnic group. Next we address the possible methods of collapsing the 1991 and 2001 ethnic group classifications in the search for a stable set of common categories. We conclude with a discussion of the practical implications of our findings for social research, and the means of further understanding the use of ethnicity variables over time.

2. Ethnic group classifications: three sources of instability

There are three, conceptually distinct, sources of instability when ethnic group is measured for the same individuals over time: unreliability in measurement, change due to question changes and conscious changes in identity.

2.1. Unreliability

All survey measurement entails some unreliability: if an item is measured twice in the same way and under the same conditions, the outcome may be different because of unintentional respondent error, transcription error, or coding error. In the 2001 Census, there were 3.9 million answers coded from write-in answers to the ethnic group question (ONS, 2004a), more than expected and 8% of the enumerated population. Coding of write-in answers was most difficult for people who identified as mixed; there were errors associated with Asian groups in particular which resulted in significant numbers of people of Indian, Pakistani or Bangladeshi origin incorrectly being classified as 'Other Asian' (ONS, 2004a: #ethnic) Errors may also arise when an item is estimated or 'imputed' for a respondent who has not completed a question. Such imputation is a common practice in censuses in order to achieve multiple cross-tabulations based on a consistent denominator of the whole population (ONS 2003b). Imputation rates in the 2001 Census varied by ethnic group, with higher proportions of imputed values in the Mixed, Black and Asian groups (ONS, 2003b: 7). Overall, 2.9 per cent of responses to the ethnic group question were missing and therefore imputed into the 2001 Census database (ONS, 2003b: 5).

The 1991 Census Validation Survey directly measured the reliability of the ethnic group question, reporting the percentage of people changing their broad ethnic group between the 1991 Census and an interview conducted four to six weeks after the census. The results showed much less reliability for the Black and Other groups than for Asians and most reliability for the White group (Table 1). The residual group 'Other' was particularly unreliable with 20 per cent changing from it when asked a second time. Over

10 per cent of those who chose one of the Black groups in the 1991 Census changed to a group other than Black in the Census Validation Survey. The equivalent cross-tabulation from the matched records between census and Census Coverage Survey in 2001 is not available from the Office for National Statistics at the time of writing.

2.2. Question changes

Types of question change

Question changes between two points of data collection present respondents with the opportunity to provide an alternative response. In the measurement of ethnic group using census classifications, three main types of question change are apparent: first there is a different choice of categories between 1991 and 2001; second the question is laid out differently, including differences in the ordering of categories; and third different instructions are provided. Question change, in this context could be considered to include modifications to the way that answers written on a form are allocated between pre-coded or residual categories, as seemingly identical responses may yield different results in the context of a new question. In addition, differences in the mode of questioning (face-to-face interview, telephone interview, or self-completion) are likely to affect responses.

Changes to the England and Wales Census ethnic group question

Although the ethnic group question was asked directly in the 1991 and 2001 censuses, a variety of changes are evident from the reproduction of the two questions in Figures 1 and 2. In 2001 respondents were asked to tick or write in their 'cultural background', while in 1991 the note uses the terms 'descended' and 'ancestry', giving more emphasis to family rather than cultural origins. In 2001 tick boxes were grouped into five sections with space for a write-in answer within each one, while in 1991 there were just two write-in spaces. These changes can be considered developments of the original 1991 question, but they also introduced entirely new categories (Irish, and four Mixed options) and reordered the other options.

The new categories can be considered as intentional developments of policy. The inclusion of ‘Irish’ responded to lobbying for the recognition of poor social conditions on average not only for Irish-born but their families born in Britain (Walter 1998; Walls 2001); the inclusion of Mixed categories responds both to lobbying by groups concerned at the special issues faced by children of parents from different ethnic groups and the growing number of mixed-origin residents estimated before the census at more than 10% of all residents not of White origin (Aspinall 2001).

The expansion of ‘White’ in order to include ‘Irish’ had further consequences. While in Scotland the option of ‘Scottish’ was added (Walls 2001), the omission of ‘Welsh’ in the form used in Wales caused political debate; the title British attached only to a White tick-box was criticised when proposed because it related to nationality (Aspinall 2000) and was unfortunate according to some fieldworkers, because it attracted residents of British nationality who were not in fact White (Simpson, 2001: 10).

Other changes were intended to achieve a more efficient collection in 2001 than the first attempt in 1991. Following feedback that many young people of Caribbean descent wished to be acknowledged as British, the 2001 labelling included headings ‘Black or Black British’ and ‘Asian or Asian British’. The Mixed categories can also be seen as an attempt to reduce the number of write-in answers from those who would have found the 1991 categories limiting.

Finally, the write-in questions were coded differently in 2001. In particular, in 2001 the ‘Other Asian’ category is composed mainly of those who wrote in a response under ‘Asian or Asian British’, whereas in 1991 ‘Other Asian’ was created differently, from those who indicated any 'unmixed' Asian origin in the second write-in box. In 1991, the division of responses to the second write-in box between ‘Asian’ and ‘Others’ created ten standard published categories of ethnic group from the nine response spaces in the question. In 2001 there were sixteen standard published categories corresponding to the response spaces in the question.

Further development of the question has already been recommended within the Office for National Statistics which runs the census, to include parallel questioning of 'national identity' with options of English, Welsh, Scottish, Irish, British or Other (ONS 2003a). Although various question formats were tested in focus groups and the test censuses before 2001 (Dixie 1998), the final choice of question format is also based on judgements of the current cultural and political context. Regardless of the origin, question change produces instability in the classification over time.

The impact of question change in the Labour Force Survey

The Labour Force Survey (LFS) asked both the 1991 and 2001 *Census* questions of the same people, permitting assessment of the impact of wording changes and unreliability. The survey was administered to one cohort in 2000-2001, with an interval of one year between questions (Smith, 2002). Table 2 gives the broad cross-classification of answers, omitting non-respondents.

The correspondence between the 1991 and 2001 questions, shown in Table 2, is similar to when the same question is repeated, tested in the Census Validation Survey (Table 1). The stability of the White category is high, unaffected by the introduction of three categories. The Other category, however, is exceptional, as many fewer people in the LFS gave the 'Other' response using the 2001 question. Unlike the census question, the LFS was administered by interview; for the 2001 question, respondents had opt for one of the five new broad categories before knowing if the finer sub-categories would suit them. The new Mixed category is taken mainly from those who previously had chosen White or Other labels.

2.3. Conscious choice of a different label

Change in ethnic group between two points in time may be due to shifts in consciously held identity, independent of changes to the question asked. In an epoch when the politics of identity are important both in national developments and on the world stage, conscious

change of identity is a strong focus of sociological study. The considerable literature on the social construction of racial and ethnic identity emphasises its dependence on personal and wider context over and above fixed demographic origins, which is also accepted by statistical agencies that create the classifications used in censuses. Jenkins distinguishes between 'two interacting but independent entailments' (1994: 218) in identity: nominal identity, (the name) and virtual identity (the experience). Conscious change in identity reveals the interaction between the 'nominal' and 'lived' ethnicity.

Specific events that trigger acceptance of new labels are not easily identified, but it appears that changed personal circumstances allow a reconsideration of identity, such as migration to a country with racialised discourses (Howard 2003; Samers 2003), an environment outside the household (Harris and Sim, 2002), and one would infer this might also apply to leaving the family home. “Greater anonymity leads to racial classifications that are more consistent with contemporary understandings of race”, was the broad conclusion from a study showing greater adoption of multiple racial origins in school than in teenagers’ homes at a time of public acceptance of diversity in the USA (Harris and Sim, 2002: 624). One expects that the degree of anonymity or social support offered by the household and community would also make a difference to the choice of group identification. Additionally, members of an immigrant community tend to ‘live locally but think globally’ (Anthias 1998; Clifford 1994), such that acceptable labels of identity are influenced by overseas and international events as well as by the local framework of statistical agencies.

Shifts of cultural acceptance of the American Indian label in the USA and the Black label in the Caribbean have already been referred to, but in Britain between April 1991 and April 2001, there were no great shifts in political or social forces that could be expected to affect allegiance to labels of ethnic group. Perhaps the biggest observable shift of allegiance is motivated by the census itself, by its provision of ‘Mixed’ as an option, which as we shall see was taken up by many who had chosen a single origin in the 1991 Census. If official labelling can itself encourage a shift of perception and identity, we cannot entirely disentangle the instability caused by question changes and consciously held allegiances. These changes in personal allegiance are of great interest but are not the

focus of the current paper; insomuch as British census data can identify them, they will be reported in a separate study of the social correlates of instability.

3. Measures of stability in ethnic group classification, and their application in England and Wales

The ONS Longitudinal Study of England and Wales (LS) measures together the three sources of instability described above: unreliability, question changes, and shifts in consciously held identity. We can assess the impact of conscious changes of identity as a residual component of change, after accounting for unreliability and question change. This section presents evidence quantifying stability in ethnic group using the LS, which is interpreted in the light of evidence from the Census Validation Survey (Table 1) and the LFS (Table 2) to quantify the amount of change in ethnic group that is attributable to each of the three sources of instability.

Data

The LS is a continuous, prospective record linkage study comprised of census records, vital events, and international migration data recorded by the National Health Service (NHS) Central Register, for approximately 1% of individuals since the 1971 Census (Hattersley and Creeser, 1995). The LS provides a powerful research data base of social and demographic change across the life course of individuals. The sample is based on four birth dates in each year and provides records for some 0.5 million individuals at any one time, together with details of their household members at the time of each census. However, the quality of the dataset and reliability of results is limited by the extent to which complete linkage is achieved (Blackwell et al., 2003).

Our use of the LS to study the stability of ethnic group is limited to those who were recorded in both censuses. Of individuals recorded in 1991 and not known to have died or emigrated, 88% were linked to records in 2001. The percentage linked between two censuses varied between ethnic groups, dropping below 80% for many of the groups other than White. However, the NHS Central Register is known to be a significantly incomplete record of migration (Hattersley, 1999), and therefore the majority of these unlinked records may be migrants not identified as such by the health service, and therefore not actually present at both censuses. Unsurprisingly there is evidence,

presented in the next section, that those who were not linked are more likely to be subject to unstable ethnic group, and therefore we must accept that our estimates of stability are optimistic.

Measurement of stability in ethnic group

The LS is the only data source which shows how individuals have changed their ethnic group over a period of ten years. Tables 3 and 4 show the number n_{ij} of LS linked records who were of ethnic group *i* in 1991 and ethnic group *j* in 2001. In what follows, n_{ii} is the number of records whose label in 2001 agrees with that of 1991; $n_{i.}$ and $n_{.i}$ are the total number of records of label *i* in 1991 and 2001 respectively, while $n_{..}$ is the total number of linked records. I_{91} and I_{01} are the number of categories in 1991 and 2001 respectively. In Table 3 I_{91} and I_{01} are both equal to 2, while Table 4 uses the standard classifications in which I_{91} is ten and I_{01} is sixteen.

The analysis of change in a categorical variable over time has precedents in various fields of study. Social mobility and the permeability of social barriers within a changing social composition has some similarity with the stability of an ethnic group classification that could be exploited with the technical approaches of log-linear modelling (Gilbert, 1993). There are less obvious but perhaps exploitable similarities with population geography, which may be seen as the classification of people into categories of residential address. Redrawn boundaries of administrative and governmental areas (Simpson, 2002a) have an impact akin to question change for ethnic group, while migration equates to the conscious movement across the boundaries between categories. We are not aware of a developed methodical approach to measuring the stability of ethnic group categories over time, and have created appropriate measures for this study.

For a specific group label *i* which appeared in the output for 1991 and for 2001:

The *stability* or *degree of fit* is the percentage of those with the label at the first time, who keep the same label. $s_i = n_{ii} / n_{i.}$

The *marginal fit* is the agreement between the populations at the two time points, expressed as a ratio of the second to the first. $m_i = n_{.i} / n_{i.}$

The *two-directional fit* is the percentage of those ever having the label, who keep the same label. It is symmetrical with respect to 1991 or 2001, unlike the previous two measures which rely on an ordering of the two time points. $t_i = n_{ii} /(n_{i.} + n_{.i} - n_{ii})$

For the entire classification, the measures of stability are derived from those for each group:

The overall *stability* is the total percentage of the population who have not changed labels. Note that the stability is the mean of the degrees of fit weighted by the original population. $s = n_{ii} / n_{..} = \sum_i (n_{i.} / n_{..}) s_i$

The *mean degree of fit* is the unweighted mean of the degrees of fit. $\bar{s} = \sum_i s_i / I_{91}$

The overall *marginal fit* is measured by the chi-squared statistic comparing the distribution between ethnic group labels on the two occasions. This is the only measure which requires the same category labels in each classification. $m = \sum_i (n_{.i} - n_{.i})^2 /(n_{.i} + n_{.i})$. The marginal fit measures the extent of divergence between the 1991 and 2001 distributions. If one assumes that they are each an independent manifestation of an underlying distribution estimated by their mean then the divergence can be tested statistically using i-1 degrees of freedom.

Table 3 shows that, even when restricted to two categories, ethnic group is not a completely stable classification, nor was the instability equal for the two categories. The degree of fit, for the 1991 White group (those who were recorded as White in both 1991 and 2001) was relatively high at 99.5 per cent. The remainder who moved from White to a different category in 2001 are small as a percentage of the whole. The degree of fit was rather lower for those other than White in 1991, at 94 per cent but because this category is smaller, the number of people who changed to White in 2001 was less than those moving from White to another category.

Taking the LS as an approximate one per cent sample, Table 3 suggests that over 350,000 people in England and Wales crossed the boundary between White and other ethnic groups between the 1991 and 2001 Censuses.The asymmetry of movement shows in the marginal fit. The White population decreased and the rest of the population increased during the period 1991-2001, purely due to shifts of individuals' recorded ethnic group. This is separate from the impact of births, deaths and migration on each population.

The two-directional fit emphasises the changes between categories. Nearly 1% of those who were White in 1991 *or* 2001, and over 10% of those who were in other categories in 1991 *or* 2001, moved across the White/other ethnic group boundary between the two censuses. The two-directional fit is always less than the degree of fit, as it expresses those who stayed in a category as a percentage of all those who were ever in that category, rather than just those who were in the category in 1991.

The overall measures of fit for the dichotomous ethnic group classification suggest that the changes are not ignorable. The stability of 99.1% shows that one percent of the entire population changed their recorded ethnic group. This measure is weighted towards the relatively large and relatively stable White group. The mean degree of fit, without weighting for their population size, shows that over 3% of people changed their group. As we shall see, when considering the more useful published classification of finer ethnic groups, still higher proportions moved across ethnic group boundaries. Finally, although movement was in both directions, to and from White, the distribution of ethnic group also changed. Although the change was slight it was statistically significant and the -marginal fit will be a useful measure when comparing the consistency of different classifications later in this paper.

From Table 4 it is clear that some categories have much greater stability than others. The overall stability is less than when the population is divided into only two categories, because we now record as instability those who moved between two categories other than White. While 98.0% of the population kept the same group label in 2001 and 1991, the mean degree of fit across the ten categories of 1991 averages only 67%. For the people in

residual groups ‘Other Black’, ‘Other Asian’ and ‘Other’, many fewer than 50% remained in the same group. Even the group labels ‘Caribbean’ and ‘African’, less ambiguous as they are not residual to other categories, have stability of only 77%. One in each four who had recorded as Black African or Black Caribbean in 1991 was recorded in a different group in 2001. In Table 4 we cannot measure the overall marginal stability as the categories are different on the two occasions. The move away from White noted above is more than accounted for by those who moved to Mixed.

Components of change in ethnic group

We can measure the overall impact of conscious changes of identity, as a residual aspect of ethnic group change, after accounting for the impact of unreliability and question change. The evidence we have presented on the overall change observed in the LS, along with evidence on unreliability and question change from the 1991 Census Validation Survey (CVS) and LFS is summarised in Table 5. It suggests that the residual, change attributable to conscious changes in identity is relatively small.

It is clear that the question works well for the White population, identifying a group the vast majority of whom do not change from one occasion to another while the question has developed and an interval of ten years has elapsed. The CVS shows that 0.4% out of the 0.5% of change that does occur is due to unreliability of the question, its coding and imputation of omitted responses. There were significant changes of ethnicity associated with choice of the new Mixed labels but the numbers involved are small in relation to the size of the White group as a whole.

For the Black groups, the degree of fit of 2001 to 1991 responses is much lower, at 76%. Evidence from the CVS shows that unreliable responses alone caused about half of the overall instability among Black groups, or around 12 per cent.. Consideration of the LFS data shows that question change, which aimed to reduce the Other Black label and offer mixed African-White and Caribbean-White labels, added further instability of around 5%. If we take changes in consciously held identity as the remaining instability measured

by the LS, not attributable to question change and unreliable responses, these amount to around 6% of all individuals who were recorded as Black in 1991.

The Asian groups are much more stable in their identification with the labels offered by the Census with an overall degree of fit of 97.5% measured by the LS; again about half of this due to unreliable responses and their recording. The degree of fit based on the LFS, is lower than that measured by the LS, suggesting that the mode of administering the question (face-to-face interview in the LFS, versus self-completion of the Census form) may also have some impact on the stability of responses. There will be some conscious shifts in self-identified label over the ten years, but their number must be small according to these statistics.

The LS identifies those values for ethnic group which were imputed by Census Offices when not recorded by the respondent. The proportion imputed among all census records was reported above of 2.9%, but falls to 2.1% among Longitudinal Study members linked between the 1991 and 2001 censuses, confirming that census records that can be linked are among those with more complete responses. We can judge the success of imputation in 2001 by its agreement with the 1991 category; a complete match should not be expected because some of the 1991 values will themselves be imputed, but if imputation were completely successful the match should be no less than the degree of fit for all records given in Table 4. The final column of Table 6 shows this to be far from the case. The methodology works best for those with the White label in 1991, because they tend to live in all-White neighbourhoods. For all other groups, who tend to live in diverse neighbourhoods, the method is unsuccessful at least 50% of the time. Overall however, imputed values account for only 2.1% of linked members, and 9.6 % of the observed instability. Since a similar proportion of values of ethnic group were imputed in the 1991 census, erroneous imputation might contribute up to one fifth of instability.

4. Stability in context

Ethnic group measured by the census is less stable than life-time demographic variables sex and country of birth. Lesser stability is shown at younger ages, but the greatest stability is shown by those born in countries associated with the ethnic group labels.

4.1. Stability of ethnic group, sex and country of birth

Sex and country of birth are variables that one would expect to change over a life time only under exceptional circumstances. However both are subject to response and coding error; country of birth is additionally ambiguous in the context of geopolitical change. We have reduced each variable to a simple dichotomous classification and measure their stability from the LS, as in Table 7, which confirms that each variable is unstable.

Sex, an unambiguous biological descriptor and one of the variables whose completion was verified most closely in the field, has a very high degree of fit between measurement in 1991 and 2001. Only 0.2% of those recorded as male in 1991 were recorded as female in 2001 and the same small percentage of 1991 women were recorded as men in 2001. Perhaps one would expect no instability at all. Three quarters of the changes are attributable to missing values which were erroneously imputed in either 1991 or 2001 (Simpson and Akinwale, 2004). The remainder may be mostly due to other errors including respondent errors and errors introduced during processing.

Whether born in the UK or elsewhere was a less stable classification than sex. In particular,97.5% who in 1991 claimed to have been born outside the UK, claimed so in 2001. Only a quarter of this instability can be blamed on inaccurate imputation of missing Country of Birth in 1991 or 2001. For individual countries of birth, the degree of fit is lower, with many fits of 2001 response to 1991 response being under 90%. One's declared country of birth has considerable ambiguity; it is likely that the attention paid to its accurate self-recording depends on parochial circumstances at the time of completing the census, and also on the value laid on the question after political changes including devolution of powers to Wales, Scotland and Northern Ireland during the 1990s.

Instability is probably most likely for those who are not living in their country of birth, and those whose country of birth has changed its name or boundaries during their lifetime. It may also be the case that those whose country of birth is different from other relatives are most likely to report it unreliably.

The stability of an individual's ethnic group label is more akin to country of birth than to sex. The large White group is much more stable than the small Other group. We have seen above that imputation (in 2001 alone) accounts for around one tenth of this instability of ethnic group, although this proportion is greater for some groups than for others.

4.2. Age

Figure 3 shows three measures of stability for cohorts defined by their age in 1991. All ten ethnic groups recorded in 1991 are shown. As observed earlier, the two-directional fit is always lower than the degree of fit as it includes in its denominator all those who adopted the label at either time point. In general, the distinctions between groups, and the severe lack of stability for the residual groups, are evident at each age.

In general the degree of fit of ethnic group rises with age. The younger groups contain higher proportions of those born in the UK and those with parents of different origins, for whom the ethnic group question may be ambiguous. These factors may explain the strongest relation of stability with age for the Caribbean group.

In contrast to the general rise in stability with age, most groups show a decline for the very oldest group who were aged at least 60 in 1991 and surviving to 2001. While this cohort is small and for some groups is represented by less than 100 members of the Longitudinal Study, the decline is noticeable also for the larger Indian and Caribbean groups. It is not clear whether older people find the question more difficult to answer, or less acceptable. Figure 3 excludes imputed records, which therefore can be ruled out as a reason for the drop in stability for the oldest group.

The marginal fit (the third chart of Figure 3) highlights groups for whom movements between 1991 and 2001 did not balance. This is a particular issue for the Other Black and the Other groups which lost affiliation between the two censuses, arguably because the categories in 2001 gave people more appropriate tick-box choices, reducing the number in these residual groups. The larger instability for the oldest group is again noticeable. For example among the elderly nineteen were African on both occasions, a further nineteen only in 2001 and nine only in 1991.

4.3. Stability of ethnic group and whether born in Britain

As most of the groups identified by the ethnic group question have immigrant origins in the past fifty years, one might expect those who identify with those origins to be most comfortable and stable with the label. The census allows us to examine the country of birth of each person in 2001, along with the stability of their ethnic group between 1991 and 2001. Among the White group, for example, only 0.1% of those born in the UK changed their ethnic group to another, while 1.7% of those born outside the UK chose a group other than White. White is a very stable category but there are sub-groups of White people for whom the question is not so straightforward; these may include Arabs and Eastern Europeans born outside the UK.

Table 8 shows the percentage of those who changed from their 1991 ethnic group to another in 2001. It excludes those who moved to the new Mixed labels, whose move is associated with the changes in the census question. We have distinguished those born in countries or regions corresponding with an ethnic group label, thus Pakistan for ‘Pakistani’, China for ‘Chinese’, anywhere in Africa for ‘African’ and so on. It is clear that those born in a country presented in an ethnic group label have least difficulty identifying with it at each census. Those born the UK are generally more uncertain of their ethnic group label, while those born elsewhere in the world are least certain of all. Thus, the stability of identification with ethnic group labels depends on the complexity of the individual’s own history.

5. Amalgamation of 2001 groups to provide a classification compatible with 1991

Previous sections have established that when comparing data from 1991 and 2001 censuses, the two sets of ethnic groups only fit approximately, and have measured the extent to which this is the case. This section deals with the possible methods for collapsing 1991 and 2001 classifications to create a stable set of categories, which give the closest approximation to populations identified by both censuses. Assessment of the best amalgamation of categories in the ethnic group classifications will be made using data from the Longitudinal Study. The aim is to increase measures of stability of the classification described earlier (overall stability, mean degree of fit and marginal fit) and to provide a plausible amalgamation of categories that is conceptually meaningful to users.

The LS data have already been seen in Table 3, which is now presented as row and column percentages in Tables 9a and 9b respectively. Records with imputed values of 2001 ethnic group are included, since they are present in all Census output which would be compared across time. We have also examined the cross-tabulation without imputed values of 2001 ethnic group and come to the same conclusions as presented in this section.

5.1. Allocation of the new categories in 2001 to categories existing in 1991

Reducing the 16 ethnic group categories of 2001 to 10 for direct comparison with the 1991 classification can best be achieved as demonstrated in Table 10. The amalgamated categories provide a square ten-by-ten transition matrix from 1991 to 2001 categories carrying the same label.

This is a different allocation of Mixed groups from other ten-category schemes used by the Greater London Authority (Howes, 2003), Dorling and Rees (2003), and Dorling and Thomas (2004), which allocated in different ways but in particular added the 2001 category Mixed Asian/White to Asian Other. The suggestion here gives better arithmetical fits of the 1991 and 2001 categories (Table 11). It gives more distinction

between categories according to the real choices individuals made in answering the 1991 and 2001 censuses revealed by the LS and is consistent with the documented coding schemes.

5.1.1. Three White categories

White Britons, White Irish, Other White were introduced as tick box labels within a 'White' section in 2001. Table 9b shows that over 94% of those who chose each one of the 2001 White groups were also coded as White in 1991. There is no doubt on these grounds alone that the three categories should be amalgamated when comparing 1991 and 2001 census tabulations.

In passing however it should be noticed that the White Irish and Other White groups are small relative to the population of White Britons, together making up less than 3% of the amalgamated White group. The 1991 Other label contributes one quarter of its members to the 2001 White groups, 15% of them moving to White Britons and 10% to Other White. The addition of Other White to Other would arithmetically improve its degree of fit much more so than the degree of fit of White is improved, by 10% rather than by 1%. However, the addition of Other White to Other would overwhelm the latter category, and the suggestion is outweighed by the higher overall stability of the group, and the face validity of putting the three White labels together.

5.1.2. Four Mixed groups

In 1991 there was no Mixed tick box. Those who stated mixed origins by writing in a response were allocated to one of the ten standard categories according to complicated rules that are demonstrated in Tables A, B and C of OPCS (1994b). In 1991:

- Those with mixed descent were asked to identify with a single group in the note to the question (Figure 1).

- Those who could not identify with a single group were asked to write in their origins under the heading 'Any other ethnic group', but could and sometimes did write in under the heading 'Black – Other'.
- Those who wrote mixed origins of Asian and White were included in output under the 'Other' label, not the 'Other Asian' label, regardless of the write-in space they used. The 'Other Asian' label is made up of non-mixed write-in responses mainly with origins in the Far East and South East Asia.
- Those writing in mixed origins of African or Caribbean and White were included in output under the 'Other' label, if they wrote in the space under the 'Any other ethnic group' heading.
- Of those writing in mixed origins of African or Caribbean and White, only those using the space under the 'Black – Other' heading were included in the Other Black category.

Thus, on the one hand, those with mixed descent were asked first to identify with a single group, and, on the other hand, those who did write in mixed origins were allocated to different groups depending on where they wrote these origins. For these reasons the correspondence between 2001 Mixed categories and the existing 1991 categories is quite low.

Table 9 makes clear that despite the instruction in the question, many who in 2001 declared Caribbean/White or African/White origins had been recorded as a write-in answer under the 'Black - Other' heading in 1991. The largest absolute and relative contribution from a 1991 category to Caribbean/White is the 32% of 1991 Other Black (Table 9a). This accounts for 31% of the 2001 Caribbean/White group, which also has more than 10% of its members from each of White, Caribbean and Other (Table 9b). The largest relative contribution from a non-residual group is the 6% of 1991 Caribbean (Table 9a). In spite of the mix of 1991 antecedents of the 2001 Caribbean/White group, arithmetically the 2001 Caribbean/White should clearly be amalgamated with the 2001 Other Black label in order to compare with the 1991 Other Black category. The marginal fit is also improved.

Similarly, African/White contains significant numbers from the 1991 White, Black Other, Other and African groups. In fact more Whites of 1991 become African/White in 2001 than did any other group, and it could be argued that these two should be merged (Table 9b. However, in relative terms the Other Black group gives more of its 1991 members to African/White (6.4%) than the White group does (0.02%, Table 9a).

The 2001 Asian/White label has no 1991 home category except the Other group where write-in answers were allocated by coders of the 1991 Census returns. It should be allocated with the 2001 Other, in order to compare with the 1991 Other label. 85% of the Asian/White group were recorded in 1991 as White or Other, in roughly equal numbers. The Asian/White label is not used by many of those who in 1991 were labelled Indian, Pakistani or Bangladeshi.

The composition of the 2001 Other Mixed label has a pattern similar to Asian/White, but it has a greater variety of origins. 1991 Other contributes the most people relative to its own total (10%) although 1991 White contributes more in absolute numbers. Chinese and African also contribute significant numbers, although less than 3% of the Other Mixed total. The African origins may be northern African Arab origins. Thus Other Mixed is very mixed itself. The nature of the Mixed group is further discussed in section 6.1.

5.2. The eight-category 1991-2001 ethnic group transition tables

While Table 10, based on a ten ethnic group classification, generally achieves a high degree of fit, there are also significant changes between 1991 and 2001. In particular, for the three residual categories Other Black, Other Asian, and Other White, only one half or less of those who had these labels in 1991 retained them in 2001, even after the addition of the Mixed categories. Other Black, Other Asian, and Other White groups taken from 1991 and 2001 census output are describing different sets of people each year. As such we propose to amalgamate the categories Other Black, Other Asian and Other White into one diverse group, creating a new eight group classification shown in Table 12.

Before the addition of the mixed groups to Other and Other Black proposed earlier in Table 10, these categories were much smaller in 2001 than in 1991; whatever difficulties the new categories cause for comparison over time, they succeeded in reducing the number of people who did not find any of the tick boxes suitable to them. Indeed it is by design that residual categories are a mixture of populations which have not identified with the provided categories. However the Other and Other Black categories are now dominated by the mixed groups that have been added to them; their low degree of fit and poor marginal fit between 1991 and 2001 mean that neither can be considered as comparable categories.

In contrast, the Other Asian group has not been added to by any Mixed group, and its total number in 2001 is close to the 1991 total (96% of it). However, the group has changed very considerably. Only 34% of the 1991 Other Asian were also Other Asian in 2001. As noted earlier, the question was phrased differently in 2001, and Other Asian was associated with the Indian, Pakistani and Bangladeshi labels.

The 1991 category 'Other Asian' was constructed differently. It was the summation of three types of write-in answer whether they appeared below the 'Black – Other' or the 'Any other ethnic group' boxes: 'East African, Asian or Indo-Caribbean', 'Indian sub-continent', and 'Other Asian'. The third of these headings includes three quarters of the Other Asian category's population (OPCS, 1994b: Tables A, B, C). It is not surprising therefore that in 1991 only 20% of 'Other Asian' were born in the Indian subcontinent, but 40% were born in the Far East (Census table L51). The distribution of countries of birth for the Other Asian label in 2001 reveals a different composition. In 2001 more were born in the UK (31% compared with 22% in 1991), more born in the Indian subcontinent (37% compared with 20%), and far fewer born in the Far East (2% compared with 40%).

In summary, both in 1991 and in 2001, the category Other Asian is diverse, but while in 2001 it can be characterised as dominated by South Asian origins (the Indian subcontinent), in 1991 it can be better characterised as dominated by South East Asian and Far East origins. While its size in total has not changed, only one third of its 1991 LS

members remained in the group in 2001. As with Other Black and Other, apparent changes in aggregate Other Asian characteristics between 1991 and 2001 cannot be taken to reflect average changes in individuals' conditions.

The retention of three diverse residual categories with low degrees of fit has little sensible interpretation, and they would be better amalgamated. Table 12 shows this eight-category solution, originally proposed in Simpson (2002b). The degree of fit of the larger residual category is now better than was any of the three individual categories that make it up but it is still low at 63%. This eight-category classification is preferred to the ten-category classification for comparisons between 1991 and 2001 because of its greater simplicity and the avoidance of three residual groups that are all poorly defined and unstable over time. The larger residual group is now 1.4% of the Longitudinal Study members present in 1991 and 2001, bigger than any group except White and Indian. It is diverse, has changed its character between 1991 and 2001 and should not be interpreted as a cohesive group.

This classification is labelled Option (a) in Table 13 and compared with two alternative eight-category classifications, to address concerns that it may not be optimal. Option (b) keeps the African/White and the Caribbean/White with the African and Caribbean labels respectively, with which they have much in common as observed above. However, the overall levels of stability are less than with solution (a). The only advantage is an increase in the percentage of 1991 African and Caribbean groups that keep that label in 2001 (the degree of fit); however, this is undermined by an increase in the two-directional fit since the great majority of those two mixed groups were *not* identified as African or Caribbean in 1991. Option (c) puts instead the Other Black group in with the Caribbean group in both 1991 and 2001, since many young people with Caribbean-born parents opted for Other Black in 1991 (Owen, 1997). The overall levels of stability are not as good as in Option (a) and the degree of fit for the Caribbean label is also not as good, since the Caribbean group is mixed with the heterogeneous Other Black group.

5.3. Further amalgamation of 1991 and 2001 categories

Having reduced the 2001 ethnic group categories to eight that match 1991 categories as above, can improvements on compatibility be made by further amalgamations? In general, researchers agree that there is so much heterogeneity within broad categories that they should be avoided. Nonetheless, broad classifications are necessary in some situations. First, there are many geographical or social sub-populations in which individual groups are so small as to require amalgamation with others in order to yield reliable comparison of rates or modelling. Second, some analyses require matching census data to a non-census classification with fewer categories. Here we focus on a four-category classification to create ethnic group labels White, Black, Asian and Other. As before, we seek to reduce the number of people who moved categories between 1991 and 2001, and therefore be surer of the distinctions between categories.

There are many different ways of collapsing the 1991 and 2001 classifications to these four categories. A major source of instability for all of them is the permeability of their boundaries with the residual Other group. Table 14 gives the degrees of fit for alternative four-category solutions, and finds several with similar fit. None are clearly better than all others, but some general results emerge. Although we have noted that the 'Other Asian' category is constructed differently in 1991 and 2001, it has some overlap and is sufficiently big to unbalance the totals if allocated to different groups in each year. For example, if Other Asian is allocated to Other in 1991 but to Asian in 2001 (Option b), the marginal fit is very poor. The same is true if all Mixed categories are allocated to White in 2001 (Option d). Good fits are provided if the 2001 Mixed groups are allocated to their non-White label (option (a) in Table 14). One of the best fits is achieved by then allocating Other Asian and Chinese within the broad Asian category, leaving a smaller residual whose fit is poor, but improving the numbers fitting within other categories (Option c).

When relating census data to another source with broad categories, the objective should be to match those other data. When combining categories in a small sub-population, the objective should be to amalgamate only to give sufficiently large categories, which will be dependent on the sub-population concerned. For these reasons it is not sensible to

recommend an unconditional optimum four-category comparison of 1991 and 2001 census ethnic group.

6. Estimation of 2001 categories from 1991 data.

Up to this point we have discussed how to amalgamate the new 2001 groups into other groups to provide a direct comparison with 1991 labels. There is also interest in whether the new groups of 2001 can be reconstructed for 1991, much as the 1991 groups were estimated from 1981 census data using country of birth (Peloe and Rees, 1999). It might be hoped that proportional allocation to new groups will improve the predictive value of the remaining parts of other groups. For example, if parts of Other Black can be used to create 1991 equivalents of Caribbean/White and African/White, the remaining Other Black may be more closely equivalent to the 2001 Other Black.

A solution can be easily proposed from the transition matrix of table 9a, using the percentage of each 1991 category falling into a given 2001 category. Thus the 1991 Irish population would be computed as 1.2% of White plus 0.3% of Caribbean, and so on down the 'Irish' column. 1991 Mixed Caribbean/White population would be estimated as 0.1% of White plus 5.7% of Caribbean plus 0.7% of African plus 32.1% of Other Black, and so on. Any tabular output from the 1991 census could be converted to 2001 categories using these proportions.

These percentages from the LS represent the population of England and Wales who were resident in both 1991 and 2001. But they cannot be assumed to also represent those who have emigrated or died in the period, who are also in the 1991 population. It is still less likely that the same percentages hold for sub-populations. For example, the percentage of 1991 Whites who were labelled Irish in 2001 will be very different in each region, associated with twentieth century Irish labour migrants. Equally the proportion Mixed is likely to be much higher among younger people and in ethnically diverse areas. Further work with the Longitudinal Study and other sources may find strata within which the proportional allocation of 1991 groups to the new 2001 categories is more stable. One would expect that these strata are likely to be identified by age, country of birth and the ethnic composition of local areas.

6.1. Mixed

Earlier we showed the diverse 1991 allegiances of the 2001 Mixed groups. In 1991 there was only an opportunity to write in a mixed identity in 1991 rather than use tick-boxes; those who did write in a mixed identity were coded as Mixed before being allocated to one of the residual categories used in standard tabulations from the Census. These codes are available on the LS, showing that less than half of each group who in 2001 identified as Mixed had used that write-in opportunity in 1991, ranging from 26.1% of Other White to 49.7% of Caribbean/White. Conversely, of the 1,701 LS members who in 1991 were recorded with a write-in mixed description, 28% did not use the Mixed categories in 2001.

The Mixed group is young, half aged fifteen or younger compared to only one in five of the population as a whole. This is partly a product of mixed unions in recent years, the ultimate answer to concerns of segregation. However, it may also be the case that those identified as Mixed by their parents may not so identify later in life when completing their own forms.

It would be wrong to assume that all the Mixed group are products of multicultural Britain. More Mixed are born overseas than the rest of the population (20% compared to 9%), presumably often a result of UK emigrants returning with a family.

Overall, the introduction in 2001 of Mixed labels has given many people an opportunity to identify with a preferred label. It is difficult to compare backwards over time because of the lack of correspondence with 1991 categories, and it is likely to be difficult to compare to future censuses because people may grow out of the label as well as grow into it. How many generations will a Mixed allegiance last as it becomes genealogically more complex? The future of the Mixed label is dependent on how much it is used in public discourse in ways that are accepted by those who might adopt it. Mixed is one of the most ambiguous ethnic group labels.

6.2. Irish

The label Irish was introduced to the 2001 Census in order to better capture the population of Irish descent than did country of birth asked previously. Many tabulations of ethnic group from the 1991 Census showed the 'Born in Ireland' as an additional classification. However, the Ireland-born are relatively old, with more than two thirds aged 40 or older, compared to less than one half of the rest of the population. The second generation, children of immigrants from Ireland, are often in poor social conditions, which the inclusion of an Irish category was intended to measure (Hickman and Walter 1997). Nonetheless, although those declaring themselves as Irish include more children in 2001 than does 'Born in Ireland', they remain a relatively old population (Figure 4).

The conclusion must be that most of those born in the UK who are children of parents born in Ireland have chosen not to declare themselves as Irish in the 2001 Census. Figure 4 also shows that many of those who were born in Ireland themselves chose not to be counted as Irish in the UK Census. Irish is not the acceptable official label for very many of those with recent descent from Ireland. It would be possible to construct from the 1991 census data an equivalent to the Irish label of 2001 by using its relationship with country of birth and age and region, but this is a pointless exercise if the Irish label of 2001 has little relation to demographic descent and no clear cultural meaning. Statistics of residents of Britain with recent descent from Ireland remain elusive. A time series would continue to be limited to country of birth statistics.

In summary, the new categories introduced to the Census in 2001, Irish and Mixed, are both likely to be particularly susceptible to unstable affiliation. When an ethnic group category ceases to be of immediate importance to individuals or has an official existence with which the individual is not at ease, many will choose an alternative category. As already quoted: 'Any ethnic group label is only valid for the period and context in which it is used'.

7. Discussion

We have found that the ethnic group variable has more instability than other demographic variables which are considered constant over a life time, such as sex and country of birth. Instability varies between groups, exceeding 20% for the African and Caribbean categories. The instability in England and Wales is partly an inevitable result of the change in question between the 1991 and 2001 censuses. But an equally large role is played by the unreliability of the question, arising from the ambiguity that the categories offer to some respondents, leading to different responses when asked the same question on different occasions. While the question will continue to develop to reduce the numbers allocating themselves to residual 'Other' categories, there is evidence that the instability is also associated with characteristics that will grow in the population, including those not White but born in the UK.

Ethnic group is a classification imposed in Britain by the results of legislation and political debate. While it has a close relationship with aspects of cultural and demographic family background, which are also aspects of personal identity, the precise categories in use are a product of official statistics. In part, the official widespread use of particular labels trains the population to identify with them at least for those official purposes, but trails behind popular and political developments that give rise to new labels or de-legitimise existing labels. During the period between April 1991 and April 2001, there were no major political or cultural shifts such as those which have at one time increased the allegiance to Native American in the USA, to Black in the Caribbean, or to Aboriginal in Australia. Such political or cultural shifts may affect any group but are most likely to affect those who have an evident choice between labels, which includes Irish and Mixed in the UK.

By concentrating on the stability between 1991 and 2001, we have not investigated the properties of these new 2001 categories in depth, as others surely will using this same longitudinal dataset. To what extent can a demographic category of Irish descent be

constructed from Census data, and what relationship does it have to those who in 2001 declared their 'cultural background' to be Irish? To what extent do those with Mixed parentage choose that label, and how does this change when they have left the parental home? We have only touched the surface of the associates of instability, showing its relationship with age and most strongly with country of birth. Those not born in the countries named in Britain's ethnic group labels are less secure in adopting those labels, whether born in the UK or elsewhere overseas. The role of life events such as leaving home, marriage and parenthood, and of the composition of the household and community in which one lives, deserves a detailed and multivariate analysis.

There was greatest instability in the residual categories, labelled in 1991 'Other Black', 'Other Asian' and 'Other', of whom less than one third remained with the same label in 2001. This is a reflection of success and failure of the ethnic group question. On the one hand the 'Other Black' and 'Other' labels were less used in 2001 because the revised question's main categories were suitable for more people. On the other hand, those for whom the question is ambiguous are most likely to write in an answer on at least one occasion. These categories do not contain a stable set of people from one census to the next; a time series of these categories should not be interpreted as referring to a stable group and should only be displayed if accompanied by this caution.

We have found that seven categories (African, Bangladeshi, Caribbean, Chinese, Indian, Pakistani and White) have been relatively stable, when measured both by the proportion that stay in their category over 10 years (the degree of fit) and the lack of bias as a result of the changes that do take place (the marginal fit). The question can be said to work well for the White population. When the three sub-categories of White in 2001 are amalgamated, the degree of fit and marginal fit between 1991 and 2001 are 99.5% and 99.8% respectively. While other categories have high marginal fit (similar sizes at the two censuses among the same longitudinal population) none reaches a degree of fit over 94% (keeping the same label over the period). England and Wales residents showed more loyalty to the Asian labels – Bangladeshi, Indian, Pakistani and Chinese each had a degree of fit over 90% – than to the Caribbean and African labels. There are more mixed births and people born in the UK among the African and Caribbean categories and their

labels refer to diverse regions rather than individual countries; just 77% of 1991 Africans and Caribbeans kept their label in 2001. Others who chose the labels newly in 2001 made up the deficit of the Black categories.

None of the various attempts to amalgamate the Mixed and residual categories with these seven groups unambiguously improved their fit. This is because each of those Mixed and residual categories are composed of people from a variety of 1991 labels; for example while many Mixed African/White had indeed been African in 2001, more had been White, more had been 'Other', and more had been 'Black Other'; the addition of Mixed African-White enlarged the African group with those who had not been African in 1991 by far too much to consider it a compatible group over time. The same was the case for all other additions to the seven labels.

This conclusion, based on arithmetical criteria and discussion of these longitudinal data to advise construction of time series, may not be optimal for other situations. When analysing 2001 data alone, one does not need to collapse any of the categories. Depending on the hypotheses under consideration, and the use of other data, one may wish to amalgamate labels such as African-White and African, without attempting to compare conditions over time.

Our recommended comparison of 1991 and 2001 data using seven best-fitting ethnic group categories and one residual category will help to monitor changing social conditions. However this is not the whole story. Comparison of cross-sectional data has also to contend with fluid migratory patterns among some ethnic groups which raise 'doubts about the reliability of using cross-sectional data to predict future characteristics' (Blackwell, 2000: 7). Differential non-response in the two censuses (Simpson, 1996; ONS, 2004b), definitional differences arising from changes to questions other than ethnic group, and changes in administrative boundaries affecting sub-national studies, are other complexities by those constructing time series.

Research using a longitudinal study where ethnic group is measured on more than one occasion must make the additional decision of which occasion to prioritise. When ethnic

group is treated as a stratifying variable, to investigate variation in a life outcome such as employment or health, the nature of the mechanisms affecting the outcome should determine whether the most recent classification or an earlier classification is most suitable for use. Early-life influences on mortality might use an early record of ethnic group. Where the choice is not obvious and makes little difference to the resulting analysis, the most recent will be most understood and therefore suitable. When a study focuses on a single ethnic group, it will make sense to explore differences between those who have always been in the group and those who have incomplete recording in the group. We have found that ethnic group was usually imputed erroneously in the 2001 Census for those who were previously not in the White group. In the LS and other 2001 census microdata these imputed values are flagged; they should be excluded if ethnic group is an important variable in the analysis, and the impact of their exclusion can be investigated.

There are certainly significant ambiguities in the ethnic group question in UK Censuses which have not been resolved. On the contrary, they are likely to grow in time as the dimensions of colour and country-of-origin that underlie the current question are not clear for increasing numbers of individuals of diversely mixed parentage or born in the UK. The main streams of posts 1950 immigration which the 1991 categories capture may be maintained by increasing appeal to family origin and descent in parents or grandparents to countries of birth outside the UK. But interest in the question will include new streams of immigrants and their descendants, from within Europe and the Middle East, which the question does not identify. Thus monitoring whether conditions have improved for particular groups may prove more rather than less difficult in the future.

Note

The full census titles of each category of ethnic group include both the section heading and the tick-box label in 2001, and are long to the extent of burdening discussion. Standard shortened labels are used in this paper are as on the right below:

2001:	
White: British	White Briton
White: Irish	Irish
White: Other White	Other White
Mixed: White and Black Caribbean	Caribbean-White
Mixed: White and Black African	African-White
Mixed: White and Asian	Asian-White
Mixed: Other Mixed	Other Mixed
Asian or Asian British: Indian	Indian
Asian or Asian British: Pakistani	Pakistani
Asian or Asian British: Bangladeshi	Bangladeshi
Asian or Other Asian British: Other Asian	Other Asian
Black or Black British: Black Caribbean	Caribbean
Black or Black British: Black African	African
Black or Black British: Other Black	Other Black
Chinese or Other Ethnic Group: Chinese	Chinese
Chinese or Other Ethnic Group: Other Ethnic Group	Other
1991:	
White	White
Black Caribbean	Caribbean
Black African	African
Black Other	Other Black
Indian	Indian
Pakistani	Pakistani
Bangladeshi	Bangladeshi
Chinese	Chinese
Other groups – Asian	Other Asian
Other groups – Other	Other

References

Anthias, F. (1998). Evaluating 'diaspora': beyond ethnicity? *Sociology* **32**(3): 557-580.

Aspinall, P, (2000) The challenges of measuring the ethno-cultural diversity of Britain in the new millenium" Policy and Politics, 1: 109-118.

Aspinall, P (2001) Operationalising the collection of ethnicity data in studies of the sociology of health and illness. Sociology of health and illness, 23(6): 829-862.

Aspinall, P (2002) 'Collective terminology to describe the minority ethnic population: the persistence of confusion and ambiguity in usage. *Sociology* 36 (4): 803-816.

Blackwell, L (2000) Fragmented life course: the changing profile of Britain's ethnic populations. *Population Trends* **101**: 6-10.

Blackwell, L., Lynch, K., Smith, J. and Goldblatt, P. (2003) *Longitudinal Study 1971-2001: completenes of census linkage; Series LS no. 10*, Office for National Statistics, London.

Clifford, J. (1994). Diasporas. *Cultural Anthropology* **9**: 302-338.

Coleman, D and Salt J (1996) The ethnic group question in the 1991 Census: a new landmark in Brisih social statistics, in D. Coleman and J. Salt (eds) Ethnicity in the 1991 Census: Demographic Characteristics of the Ethnic Minority Populations, Volume 1 of 4, London: Office of Population Censuses and Surveys

Dixie, John(1998) The ethnic and religious questions for 2001: research and responses. Patterns of Prejudice 32(2): 5-14.

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.

Dorling, D. and Thomas, B. (2004) *People and Places: A 2001 Census Atlas of the UK.* Bristol: The Policy Press.

Evans, J, Kahles, D and Bate C (1993) *1991 Census data quality: Aboriginal and Torres Strait Islander counts,* Australian Bureau of Statistics, Population Census Evaluation, Census working Paper 93/6.

Gilbert, Nigel (1993) *Analyzing tabular data : loglinear and logistic models for social Researchers*. London: UCL Press.

Harris, D. R. and J. J. Sim (2002). Who is multiracial? Assessing the complexity of lived race. *American Sociological Review* **67**: 614-627.

Hattersley, L. and R. Creeser (1995). *Longitudinal Study 1971-1991: history, organisation and quality of data; OPCS Series LS no.7*. London, HMSO.

Hattersley, L (1999) LS User Guide 18: International Migration data in the LS. London: Centre for Longitudinal Studies.

Hickman, M. and Walter, B. (1997) *Discrimination and the Irish community in Britain*, Commission for Racial Equality, London.

Howard, D. (2003). Reappraising race? Dominicans in New York City. *International Journal of Population Geography* **9**: 337-350.

Howes, Eileen (2003) 2001 Census Key Statistics: Ethnicity, religion and country of birth. DMAG briefing 2003/9. London: Greater London Authority.

Jenkins, R. (1994) Rethinking ethnicity: identity, categorization and power, *Ethnic and Racial Studies*, **17**(2): 197-223

Kertzer, D. and D. Arel (2002). *Census and identity: the politics of race, ethnicity and language in national censuses*. Cambridge, Cambridge University Press.

Kuczynski, R.R. (1953) *Demographic survey of the British colonial empire* Vol III, London: Oxford University Press

Nagel, J. (1995). American-Indian ethnic renewal - politics and the resurgence of identity. *American Sociological Review* **60**(6): 947-965.

ONS (2003a). *Ethnic group statistics, a guide for the collection and classification of ethnicity data*. London, Office for National Statistics.

ONS (2003b). Census 2001 Review and evaluation, edit and imputation: evaluation report Tichfield: Office for National Statistics.

ONS (2004a) Quality for Data Capture and Coding - Evaluation Summary. Web report, http://www.statistics.gov.uk/census2001/data_capture_and_coding_evr.asp

ONS (2004b) onc_imp_rates_all_areas.xls File available on request. London: Office for National Statistics.

OPCS (1994a) 1991 Census: Quality Check. Monitor SS 94/2. London: Office of Population Statistics and Censuses.

OPCS (1994b) Great Britain 1991 census Topic Report on Country of Birth and Ethnic Group. London: HMSO.

Owen, D (1997) Black Caribbean. Ethnicity in the 1991 Census, volume 2. London: The Stationery Office.

Passell, J. (1995) *Discussion of racial identity classification and its effect on the undercount*, in 1993 Research Conference on undercounted ethnic populations, Washington: US Bureau of the Census, 345-53.

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

Samers, M. E. (2003). Diaspora unbound: muslim identity and the erratic regulation of Islam in France. *International Journal of Population Geography* **9**: 351-364.

Simpson, L (2001) Census fieldwork – the bedrock for a decade of social analysis. Manchester: CCSR occasional paper 22, University of Manchester.

Simpson, L. (2002a). Geography conversion tables: a framework for conversion of data between geographical units. International Journal of Population Geography 8(1): 69-82.

Simpson L. (2002b) The starting population for population projections by ethnic group. Chapter 5 (pp 73-80) in John Haskey (ed) Population projections by ethnic group, a feasibility study, Studies in Medical and Population Subjects no. 67, 2002, The Stationery Office.

Simpson, L and B Akinwale (2004) Longitudinal Study imputation and instability: sex, country of birth and ethnic group. Beta test report. London: Office for National Statistics.

Simpson, S. (1996) Non-response to the 1991 census: the effect on ethnic group enumeration. In *Ethnicity in the 1991 Census Volume: Demographic characteristics of the ethnic minority populations*(Eds, Coleman, D. and Salt, J.) Office of Population Censuses and Surveys, London, pp. 63-79.

Tribalat, M. (2004). An estimation of the foreign-origin populations of France in 1999. Population-E, 59(1), 49-80.

Walls, Patricia (2001) Religion, ethnicity and nation in the census: some thoughts on the inclusion of Irish ethnicity and Catholic religion. *Radical Statistics* **78**: 48-62.

Walter, Bronwen (1998) Challenging the Black/White binary: the need for an Irish category in the 2001 Census. Patterns of Prejudice 32(2): 73-86.

Waters, M. (2000). Immigration, intermarriage, and the challenges of measuring racial/ethnic identities. *American Journal of Public Health* **90**(11): 1735-1737.

Table 1. Reliability of ethnic group in the 1991 Census, % choosing each category in validation survey

		Quality check 4-6 weeks after the 1991 Census				
		White	**Black**	**South Asian**	**Other**	**All =100%**
1991 Census	**White**	**99.6**	0.1	.	0.3	12,017
	Black	1.5	**88.0**	1.6	8.8	264
	South Asian	0.2	0.1	**98.7**	1.0	669
	Other	9.6	11.3	1.0	**78.1**	130

1991 Census Validation Survey, reproduced from OPCS (1994a)
Each % based on weighted sample numbers; cell counts not given in the report.

Table 2. 1991 and 2001 ethnic group questions asked of the same cohort in the Labour Force Survey

		2001 categories asked in 5th wave								
		White	**Mixed**	**Asian**	**Black**	**Other**	**Total 91**	1991%	Degree of fit	Marginal fit
1991 categories asked in 1st wave	**White**	**65621**	218	102	59	104	66104	94.1%	99.3%	99.9%
	Black	71	84	8	**887**	24	1074	1.5%	82.6%	96.1%
	South Asian	52	23	**1751**	2	12	1840	2.6%	95.2%	115.3%
	Other	277	297	261	84	**338**	1257	1.8%	26.9%	38.0%
	Total 2001	66021	622	2122	1032	478	70275	100.0%		
	2001%	93.9%	0.9%	3.0%	1.5%	0.7%	100.0%			
	Stability		97.6%							
	Mean degree of fit:		76.0%							

Labour Force Survey 2001

Table 3. Ethnic group in 1991 and 2001: White and All others

	White	All others	Total 1991	1991 distribution %	Degree of fit	Marginal fit	Two-directional fit
White	**390,460**	2,122	392,582	94.0%	99.5%	99.8%	99.1%
All others	1,467	**23,465**	24,932	6.0%	94.1%	102.6%	86.7%
Total 2001	391,927	25,587	417,514				
2001 distribution %	93.9%	6.1%					

Stability 99.1%
Mean degree of fit: 96.8%
Marginal fit (1 df) 9.0 p=.0026
ONS Longitudinal Study

Table 4. Ethnic group in 1991 and 2001: classifications as published

1991 . categories:	2001:White			Mixed				Asian or Asian British				Black or Black British			Other		Total 91	1991 distribution %	Degree of fit	Marginal fit	Two-directional fit
	White Briton	Irish	Oth White	Car-Wh	Afr-Wh	Asian-Wh	Oth Mixed	Indian	Pakistani	Bangladeshi	Oth Asian	Caribbean	African	Oth Blac	Chinese	Other					
White	**380,096**	**4,704**	**5,660**	367	98	487	325	126	99	29	113	165	68	51	67	127	392,582	94.0%	**99.5%**	99.8%	99.1%
Caribbean	205	10	22	194	7	-	26	8	5	-	12	**2,617**	22	264	-	-	3,392	0.8%	**77.2%**	93.2%	66.5%
African	84	4	11	9	41	-	13	24	4	5	7	30	**956**	42	-	5	1,235	0.3%	**77.4%**	93.3%	66.8%
Oth Black	148	4	38	396	79	20	66	24	20	9	21	255	38	**103**	-	14	1,235	0.3%	**8.3%**	41.7%	6.3%
Indian	156	10	34	-	5	51	18	**7,825**	48	5	381	16	22	4	-	21	8,596	2.1%	**91.0%**	97.2%	85.8%
Pakistani	96	-	10	-	-	35	5	49	**3,960**	11	134	-	7	-	-	3	4,310	1.0%	**91.9%**	99.4%	85.5%
Bangladshi	40	-	-	-	-	9	4	17	14	**1,543**	25	-	-	-	-	-	1,652	0.4%	**93.4%**	98.5%	88.9%
Chinese	24	-	-	-	-	6	20	3	-	-	-	-	-	-	**1,017**	48	1,118	0.3%	**91.0%**	95.8%	81.4%
Oth Asian	44	-	8	3	-	31	58	197	66	12	**451**	9	18	16	38	390	1,341	0.3%	**33.6%**	102.7	20.7%
Other	313	5	201	300	79	354	204	81	68	13	141	68	21	35	26	**144**	2,053	0.5%	**7.0%**	36.6%	5.4%
Total 2001	381,206	4,737	5,984	1,269	309	993	739	8,354	4,284	1,627	1,285	3,160	1,152	515	1,148	752	**417,514**	100%			
2001 distrib'n %	91.3%	1.1%	1.4%	0.3%	0.1 %	0.2 %	0.2%	2.0%	1.0%	0.4%	0.3%	0.8%	0.3%	0.1%	0.3%	0.2%	100%				

Stability 98.0%
Mean degree of fit: 67.0%
ONS Longitudinal Study
Bold indicates those whose 2001 ethnic group agrees with their 1991 label.

Ta ble 5. Degree of fit to the 1991 census ethnic group question

	1991 Census Validation Survey	LFS with 2001 question	LS 1991-2001
White	99.6%	99.3%	99.5%
Black	88.0%	82.6%	76.1%
Asian	98.7%	95.2%	97.5%
Other	78.1%	26.9%	37.9%

Derived from tables 1, 2, 4

Table 6: Imputation of ethnic group 2001, for each 1991 ethnic group

	All records		Ethnic group imputed in 2001		Imputed to a different category in 2001	
Ethnic group in 1991	N	%	N	% of group	N	% of imputed
White	392,582	94.0%	8,027	2.0%	201	2.5%
Black Caribbean	3,392	0.8%	167	4.9%	117	70.1%
Black African	1,235	0.3%	48	3.9%	45	93.8%
Black Other	1,235	0.3%	74	6.0%	73	98.6%
Indian	8,596	2.1%	238	2.8%	121	50.8%
Pakistani	4,310	1.0%	182	4.2%	91	50.0%
Bangladeshi	1,652	0.4%	67	4.1%	40	59.7%
Chinese	1,341	0.3%	43	3.2%	16	88.9%
Other groups - Asian	1,118	0.3%	18	1.6%	36	83.7%
Other groups - Other	2,053	0.5%	80	3.9%	80	100.0%
All groups	417,514	100.0%	8,944	2.1%	820	9.2%

ONS Longitudinal Study

Table 7. Stability of sex, country of birth and ethnic group

Classification		Number in 1991 Census	Number not in this category in 2001	Degree of fit	Stability	Mean degree of fit
Sex	Male	198,427	376	99.8%	99.8%	99.8%
	Female	218,339	356	99.8%		
Country of birth	UK	384,076	1,243	99.7%	99.5%	98.6%
	Overseas	27,937	702	97.5%		
Ethnic group	White	392,582	2,122	99.5%	99.1%	96.8%
	Other	24,932	1,467	94.1%		

ONS Longitudinal Study

Table 8. Percentage retaining 1991 ethnic group in 2001, and country of birth

1991 ethnic group	Born in country corresponding with ethnic group label	Born in the UK	Born elsewhere
Caribbean	5.3	20.4	45.9
African	5.2	26.8	65.2
Indian	4.5	10.9	40.0
Pakistani	3.9	5.9	18.3
Bangladeshi	3.0	5.1	25.0
Chinese	0.7	5.0	14.0

ONS Longitudinal Study, excluding imputed records

Table 9. Transition between ethnic group in 1991 and 2001: standard classification, row and column percentages

(a) % of 1991	2001: White			Mixed				Asian or Asian British				Black or Black British			Other		
1991:	White Briton	Irish	Other White	Caribb-White	African-White	Asian-White	Other Mixed	Indian	Pakistani	Banglade shi	Other Asian	Caribbean	African	Other Black	Chinese	Other	Total 91
White	**96.8%**	**1.2%**	**1.4%**	0.1%	0.0%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100%
Caribbean	6.0%	0.3%	0.6%	5.7%	0.2%	0.0%	0.8%	0.2%	0.1%	0.0%	0.4%	**77.2%**	0.6%	7.8%	0.0%	0.0%	100%
African	6.8%	0.3%	0.9%	0.7%	3.3%	0.0%	1.1%	1.9%	0.3%	0.4%	0.6%	2.4%	**77.4%**	3.4%	0.0%	0.4%	100%
Other Black	12.0%	0.3%	3.1%	32.1%	6.4%	1.6%	5.3%	1.9%	1.6%	0.7%	1.7%	20.6%	3.1%	**8.3%**	0.0%	1.1%	100%
Indian	1.8%	0.1%	0.4%	0.0%	0.1%	0.6%	0.2%	**91.0%**	0.6%	0.1%	4.4%	0.2%	0.3%	0.0%	0.0%	0.2%	100%
Pakistani	2.2%	0.0%	0.2%	0.0%	0.0%	0.8%	0.1%	1.1%	**91.9%**	0.3%	3.1%	0.0%	0.2%	0.0%	0.0%	0.1%	100%
Bangladeshi	2.4%	0.0%	0.0%	0.0%	0.0%	0.5%	0.2%	1.0%	0.8%	**93.4%**	1.5%	0.0%	0.0%	0.0%	0.0%	0.0%	100%
Chinese	2.1%	0.0%	0.0%	0.0%	0.0%	0.5%	1.8%	0.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	**91.0%**	4.3%	100%
Other Asian	3.3%	0.0%	0.6%	0.2%	0.0%	2.3%	4.3%	14.7%	4.9%	0.9%	**33.6%**	0.7%	1.3%	1.2%	2.8%	29.1%	100%
Other	15.2%	0.2%	9.8%	14.6%	3.8%	17.2%	9.9%	3.9%	3.3%	0.6%	6.9%	3.3%	1.0%	1.7%	1.3%	**7.0%**	100%
2001 total	91.3%	1.1%	1.4%	0.3%	0.1%	0.2%	0.2%	2.0%	1.0%	0.4%	0.3%	0.8%	0.3%	0.1%	0.3%	0.2%	100%

(b) % of 2001	2001: White			Mixed				Asian or Asian British				Black or Black British			Other		
1991:	White Briton	Irish	Other White	Caribb-White	African-White	Asian-White	Other Mixed	Indian	Pakistani	Banglade shi	Other Asian	Caribbean	African	Other Black	Chinese	Other	Total 91
White	**99.7%**	**99.3%**	**94.6%**	28.9%	31.7%	49.0%	44.0%	1.5%	2.3%	1.8%	8.8%	5.2%	5.9%	9.9%	5.8%	16.9%	94.0%
Caribbean	0.1%	0.2%	0.4%	15.3%	2.3%	0.0%	3.5%	0.1%	0.1%	0.0%	0.9%	**82.8%**	1.9%	51.3%	0.0%	0.0%	0.8%
African	0.0%	0.1%	0.2%	0.7%	13.3%	0.0%	1.8%	0.3%	0.1%	0.3%	0.5%	0.9%	**83.0%**	8.2%	0.0%	0.7%	0.3%
Other Black	0.0%	0.1%	0.6%	31.2%	25.6%	2.0%	8.9%	0.3%	0.5%	0.6%	1.6%	8.1%	3.3%	**20.0%**	0.0%	1.9%	0.3%
Indian	0.0%	0.2%	0.6%	0.0%	1.6%	5.1%	2.4%	**93.7%**	1.1%	0.3%	29.6%	0.5%	1.9%	0.8%	0.0%	2.8%	2.1%
Pakistani	0.0%	0.0%	0.2%	0.0%	0.0%	3.5%	0.7%	0.6%	**92.4%**	0.7%	10.4%	0.0%	0.6%	0.0%	0.0%	0.4%	1.0%
Bangladeshi	0.0%	0.0%	0.0%	0.0%	0.0%	0.9%	0.5%	0.2%	0.3%	**94.8%**	1.9%	0.0%	0.0%	0.0%	0.0%	0.0%	0.4%
Chinese	0.0%	0.0%	0.0%	0.0%	0.0%	0.6%	2.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	**88.6%**	6.4%	0.3%
Other Asian	0.0%	0.0%	0.1%	0.2%	0.0%	3.1%	7.8%	2.4%	1.5%	0.7%	**35.1%**	0.3%	1.6%	3.1%	3.3%	51.9%	0.3%
Other	0.1%	0.1%	3.4%	23.6%	25.6%	35.6%	27.6%	1.0%	1.6%	0.8%	11.0%	2.2%	1.8%	6.8%	2.3%	**19.1%**	0.5%
2001 total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

ONS Longitudinal Study

Table 10. Ten-category classification

Presentation group	1991 categories	2001 categories	Degree of fit	Marginal fit	Two-directional fit
White	White	White Briton Irish Other White	99.5%	99.8%	99.1%
Indian	Indian	Indian	91.0%	97.2%	85.8%
Pakistani	Pakistani	Pakistani	91.9%	99.4%	85.5%
Bangladeshi	Bangladeshi	Bangladeshi	93.4%	98.5%	88.9%
Other Asian	Other Asian	Other Asian	33.6%	95.8%	20.7%
Caribbean	Caribbean	Caribbean	77.2%	93.2%	66.5%
African	African	African	77.4%	93.3%	66.8%
Other Black	Other Black	Other Black Caribbean / White African / White	46.8%	169.5%	21.0%
Chinese	Chinese	Chinese	91.0%	102.7%	81.4%
Other	Other	Asian / White Other Mixed Other	34.2%	121.0%	18.3%

Table 11. Measures of stability of ten 1991 ethnic groups when compared to 2001 classifications

					Overall Stability	Mean degree of fit	Chi-squared fit of margins
16 categories as in 2001 Census					4.5%	57.1%	
14 categories. 3 White categories together					98.0%	67.0%	
10 categories, allocating 2001 Mixed to other categories as follows	2001 African /White	2001 Caribbean /White	2001 Asian /White	2001 Other Mixed			
Greater London Authority 2003	Other Black	Other Black	Asian Other	Other	98.1%	72.1%	568.7
Dorling and Rees 2003, Polarisation	African	Caribbean	Asian Other	Other Black	98.1%	68.7%	1007.3
Dorling and Thomas 2004, Atlas	African	Other Black	Asian Other	Other Black	98.1%	71.3%	1319.3
Identity and Change Project 2004	Other Black	Other Black	Other	Other	98.2%	73.6%	279.1

ONS Longitudinal Study

Table 12. Eight-category classification

Presentation group	1991 categories	2001 categories	Degree of fit	Marginal fit	Two-directional fit
White	White	White Briton Irish Other White	99.5%	99.8%	99.1%
Indian	Indian	Indian	91.0%	97.2%	85.8%
Pakistani	Pakistani	Pakistani	91.9%	99.4%	85.5%
Bangladeshi	Bangladeshi	Bangladeshi	93.4%	98.5%	88.9%
Caribbean	Caribbean	Caribbean	77.2%	93.2%	66.5%
African	African	African	77.4%	93.3%	66.8%
Chinese	Chinese	Chinese	91.0%	102.7%	81.4%
Other	Other Black Other Asian Other	Caribbean / White African / White Asian / White Other Mixed Other Black Other Asian Other	62.8%	126.6%	38.3%

Table 13. Stability of alternative 8-category classifications

Overall measures of fit:		Overall stability	Mean degree of fit	Chi-squared marginal fit
	(a)	98.5%	85.5%	160.7
	(b)	98.4%	84.3%	174.5
	(c)	98.5%	84.7%	553.1
Fit for specific categories:		African	Caribbean	Other
Degree of fit	(a)	77.4%	77.2%	62.8%
	(b)	80.7%	82.9%	44.2%
	(c)	77.4%	70.0%	63.5%
Marginal fit	(a)	93.3%	93.2%	126.6%
	(b)	118.3%	130.6%	92.5%
	(c)	93.3%	79.4%	157.5%
Two-directional fit	(a)	66.8%	66.5%	38.3%
	(b)	58.7%	56.1%	29.8%
	(c)	66.8%	64.0%	32.7%

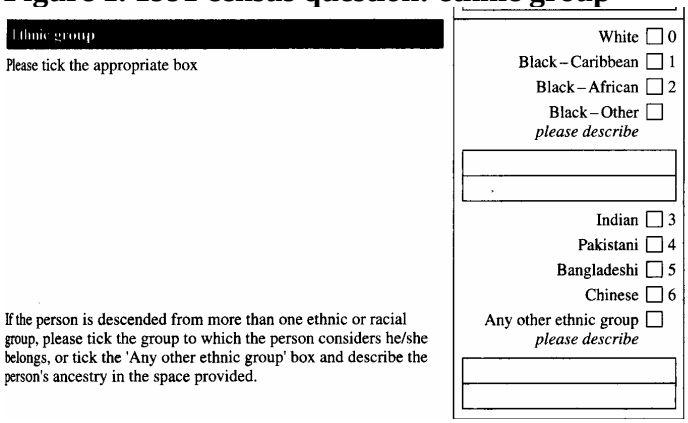
ÓNS Longitudinal Study
Classifications: (a) All Mixed and residual categories with Other
(b) As (a) except African/White and Caribbean/White with African and Caribbean respectively
(c) As (a) except Black Other with Caribbean

Table 14. Stability of alternative 4-category classifications

	Overall stability	Mean degree of fit	Chi-squared marginal fit
Option (a)	98.59	83.77	82
White: White Briton, Irish, Other White **Black**: Caribbean, African, Caribbean/White, African/White, Other Black **Asian**: Indian, Pakistani, Bangladeshi, Asian/White **Other**: Other, Other Mixed, Chinese, Other Asian			
Variations from option (a)			
(b) Other Asian in Asian for 2001 census only	98.58	81.42	641.85
(c) Other Asian and Chinese in Asian both censuses	98.63	74.05	126.79
(d) All Mixed in White	98.62	76.68	1205.88
(e) All Mixed in Other	98.56	82.44	183.56
(f) All Mixed and residuals in Other	98.59	85.25	139.67

ONS Longitudinal Study

Figure 1. 1991 census question: ethnic group



Ethnic group

Please tick the appropriate box

White 0
Black – Caribbean 1
Black – African 2
Black – Other
please describe

Indian 3
Pakistani 4
Bangladeshi 5
Chinese 6
Any other ethnic group
please describe

If the person is descended from more than one ethnic or racial group, please tick the group to which the person considers he/she belongs, or tick the 'Any other ethnic group' box and describe the person's ancestry in the space provided.

Figure 2. 2001 census question: ethnic group

8 What is your ethnic group?

Choose ONE section from A to E, then ✓ the appropriate box to indicate your cultural background.

A White
British Irish
Any other White background, *please write in*

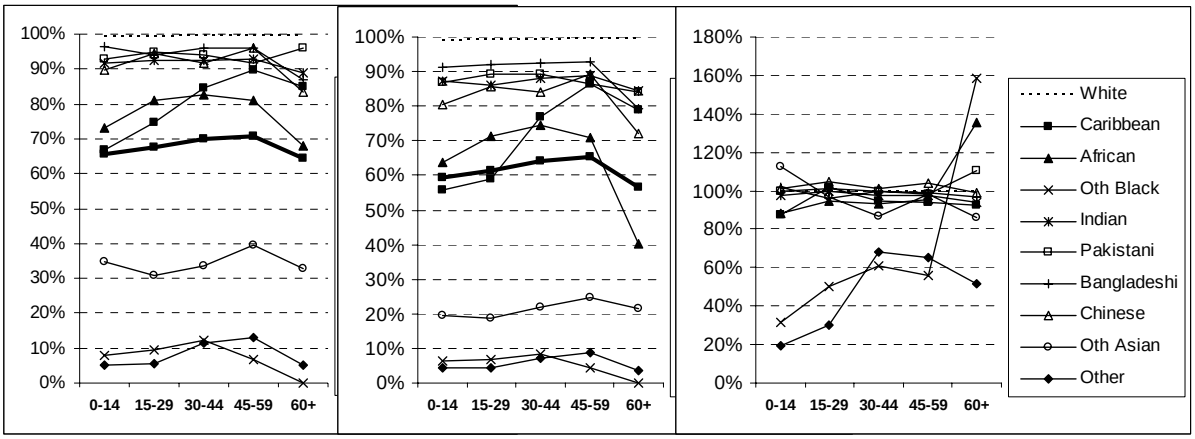
B Mixed
White and Black Caribbean
White and Black African
White and Asian
Any other Mixed background, *please write in*

C Asian or Asian British
Indian Pakistani
Bangladeshi
Any other Asian background, *please write in*

D Black or Black British
Caribbean African
Any other Black background, *please write in*

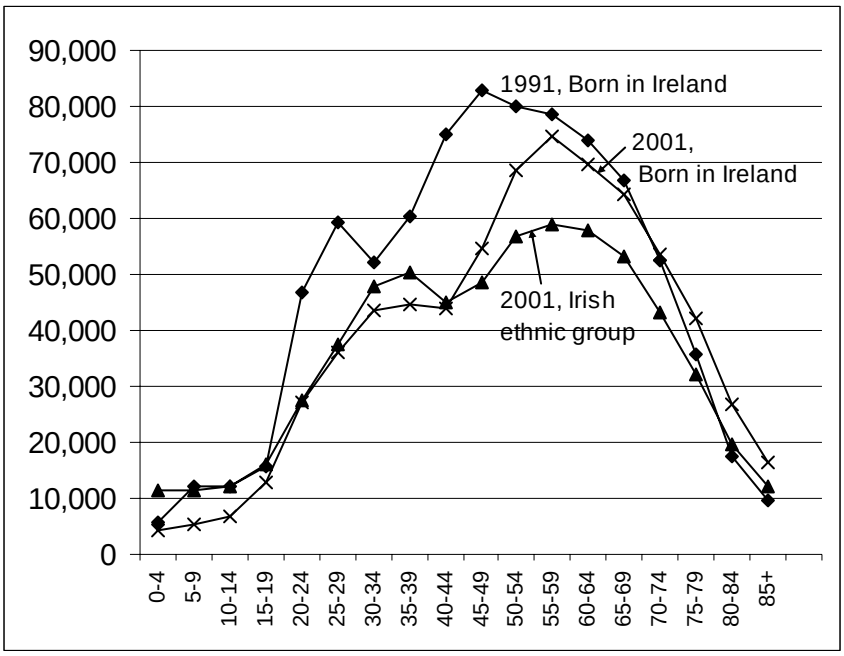
E Chinese or other ethnic group
Chinese
Any other, *please write in*

Figure 3. Age and stability of ethnic group.



Source: ONS Longitudinal Study. Age as in 1991.
(a) Degree of fit: the percentage of the 1991 group which kept the same ethnic group.
(b) Two-directional fit: the percentage who kept their ethnic group, of all those who were in 1991 or 2001 of that ethnic group.
(c) Marginal fit: the ratio of 2001 divided by 1991 population.
The bold line shows the median stability in parts (a) and (b).

Figure 4. Irish country of birth and ethnic group: England and Wales



Source: 1991 Census Topic Report Table EG2; 2001 Census Commissioned Table C0004.

[NB: chart should be drawn with horizontal lines at each point, joined by vertical lines, a 'step chart'.