Identity and Change 1991-2001

Bola Akinwale, Office for National Statistics [1]
Ludi Simpson, Cathie Marsh Centre for Survey Research, University of Manchester

To be presented at
IUSSP XXV International Population Conference, Tours, France, 18-23 July 2005

ABSTRACT

For two censuses, residents of Britain have been asked to identify their ethnicity. Researchers have generally taken the results as a demographic attribute that is stable through life but its instability has been found to be neither insignificant nor random. This paper examines the inconsistency in individuals' ethnic group identification in the 1991 and 2001 Censuses of England and Wales using data from the ONS Longitudinal Study. People respond to questions about ethnicity within the constraints of the categories offered to them, their own self-identification, and the social acceptability of each category. For these reasons, the recording of ethnic categories may change in three ways fundamental to demography: between cohorts, at different ages, and at different periods. This paper discusses three sources of inconsistency in measured ethnic group, then proposes and applies measures to quantify each one. Inconsistency in self-identification varies greatly between groups, which has practical implications for social research that uses these classifications and for our understanding of the use of ethnicity variables over time. In particular, this paper suggests that the compatibility of groups constructed from each census classification can be best exploited by an eight-category classification.

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 or the impact of targeted policies. 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. For example, in many countries, ethnic group categories also identify groups of recent migrant origin and ethnic group data are used

[1] bola.akinwale@ons.gsi.gov.uk
Office for National Statistics, 1 Drummond Gate, London SW1V 2QQ, United Kingdom

in debates on international migration policy. With the introduction of an ethnic group question directed at recent migrants in the 1991 Census, Britain stood out from most of its European neighbours, who were more concerned with recording ancient indigenous minorities (Coleman and Salt, 1996).

While the collection of ethnic group data is officially justified to aid the monitoring of group conditions, used in these ways the classifications imply a characteristic that is stable throughout the life course. However, the recording of an ethnic or racial category may change in three ways fundamental to demography: between cohorts, at different ages, and at different periods. For example, *cohort* experiences of change in ethnic identification can be seen in the rapid growth in the count of young adult Africans in Trinidad after the political successes of the Black Power movement in the 1960s. Similarly, 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). *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 England and Wales 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 propensity for change poses problems for the use of an ethnic group classification for the policy purposes described above. Once we accept that classifications and individuals' allegiance to them are changeable, we then face the challenge of whether it is possible to compare statistics of ethnic group size and conditions across time. In order to address this issue, the goal of this study is to quantify inconsistency in individuals' responses to the 1991 and 2001 Census ethnic group questions and propose a grouping of categories that maximises the compatibility of responses to the two questions. First, we consider the sources of inconsistency in responses to the England and Wales ethnic group questions.

SOURCES OF INCONSISTENCY IN ETHNIC GROUP IDENTIFICATON

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

Question changes

Question changes between two points of data collection present respondents with the opportunity to provide an alternative response to the same issue. 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 Box 1.

Box 1:Ethnic group questions in the Census of England and Wales

The 1991 Census ethnic group question

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.

The 2001 Census ethnic group question

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*

In fact, the Census measurement of ethnicity in England and Wales is premised on the understanding that ethnicity is not necessarily a stable attribute, and that it can be affected by the nature and wording of questions (Office for National Statistics, 2003a). Box 1 shows that there is a different choice of categories between 1991 and 2001. In 2001, the ethnic group question was revised to reflect evolution in the concepts and terminology used to describe ethnicity in Britain and in response to political developments. Respondents were newly presented with the option to identify themselves as being Irish or as having a mixed ethnic origin, using pre-coded tick boxes. The inclusion of the Irish category can be seen as a response to lobbying for the recognition of the poor social conditions, on average, of the Irish-born and their families born in Britain (Walls, 2001). The inclusion of Mixed categories was a response both to lobbying by groups concerned with the special issues faced by children of parents from different ethnic groups, and the growing number of mixed-origin residents, at more than ten per cent of all residents not of White origin before the Census (Aspinall, 2001). Other changes were intended to achieve a more efficient collection of data 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 heading 'Black or Black 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.

In addition to the introduction of new categories, changes in self-identification could be caused by differences in question layout, differences in the ordering of categories, and differences in the instructions provided. 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 to the question mean that people could select different categories on the two occasions simply because of the changes to the options available. As such the composition of groups included in the classification at both Censuses may not be directly comparable.

Question change could also be considered to include modifications to the way that answers written on a form are allocated between pre-coded or residual categories. For example, in the 2001 Census, the 'Other Asian' category was composed mainly of those who wrote in a response under 'Asian or Asian British', whereas in 1991 'Other Asian' was created, from those who indicated any 'unmixed' Asian origin in write-in space, headed 'Any other ethnic group'. In 1991, responses to the write-in space 'Any other ethnic group' were divided between Asian groups and 'Others' creating ten standard published categories of ethnic group, from nine response spaces in the questions. In 2001 there were 16 standard published categories corresponding to the response spaces in the question.

Unreliability in measurement

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, more than expected, and amounting to 6.51 per cent of the population

(ONS, 2003b). Coding of write-in answers was most difficult for people who identified themselves as Mixed. There were also errors associated with the coding of South Asian groups, which resulted in significant numbers of people of Indian, Pakistani, or Bangladeshi origin incorrectly being classified as 'Other Asian' (ONS, 2003b).

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 2003c). Imputation rates in the 2001 Census varied by ethnic group, with higher proportions of imputations in the Mixed, Black, and Asian groups (ONS, 2003c: 7). Overall, 2.9 per cent of responses to the ethnic group question were missing and therefore imputed into the 2001 Census database (ONS, 2003c: 5).

Conscious change in self-identification

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. Specific events that trigger acceptance of new labels are not easily identified, but it appears that a reconsideration of identity may be prompted by changed personal circumstances, such as migration to a country with racialised discourses (Howard 2003; Samers 2003) or moves to an environment outside the household (Harris and Sim, 2002). 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 1991 and 2001 Censuses, there were no great shifts in political or social forces that could be expected to affect allegiance to ethnic groups. Perhaps the biggest observable shift of allegiance is motivated by the census itself, by its provision of Mixed options, which, as we shall see, were 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 inconsistency in reporting caused by question changes and consciously held allegiances.

MEASUREMENT OF STABILITY

The consistency of individuals' responses to the ethnic group question can be measured using the ONS Longitudinal Study (LS). The LS links Census data from 1971 onward with vital event data, such as births and deaths, and records of migration to and from England and Wales. Individuals are selected for the LS on the basis of their birth dates, using four dates in a year. The sample approximates to one per cent of the population of England and Wales or about 500,000 individuals at each Census. However, the quality of the dataset and reliability of results are dependent on the extent to which complete linkage is achieved. Failure to link records between the 1991 and 2001 Censuses was higher among minority ethnic groups, at 27 per cent, compared with 11 per cent for White people (Blackwell et al., 2003). In this study, use of the LS is restricted to those who were recorded in both the 1991 and 2001 censuses.

We are not aware of a methodical approach to measuring the consistency of ethnic group categories over time, and have created appropriate measures for this study. Table 1 shows the number n_{ij} of LS linked records that 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. Table 1 uses the standard classifications in which I_{91} is ten and I_{01} is 16.

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. We cannot measure the overall marginal fit using the LS because the categories are different at the two Censuses.

  $m = \sum_i (n_{.i} - n_{.i})^2 /(n_{.i} + n_{.i})$.

  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 degree of freedom.

Table 1: Transition in ethnic group between 1991 and 2001

Frequencies and percentages	Ethnic group in 2001															
	White or White British			Black or Black British			Asian or Asian British				Chinese or Other		Mixed				Total
Ethnic group in 1991	British	Irish	Other	Caribb-ean	African	Other	Indian	Pakist-ani	Bangla deshi	Other	Chinese	Other Ethnic Group	White & Black Caribbean	White & Black African	White and Asian	Other Mixed	
Frequencies																	
White	**380096**	**4704**	**5660**	165	68	51	126	99	29	113	67	127	367	98	487	325	392582
Black Caribbean	205	10	22	**2617**	22	264	8	5		12			194	7		26	3392
Black African	84	4	11	30	**956**	42	24	4	5	7		5	9	41		13	1235
Black Other	148	4	38	255	38	**103**	24	20	9	21		14	396	79	20	66	1235
Indian	156	10	34	16	22	4	**7825**	48	5	381		21		5	51	18	8596
Pakistani	96		10		7		49	**3960**	11	134		3			35	5	4310
Bangladeshi	40						17	14	**1543**	25					9	4	1652
Other groups - Asian	44		8	9	18	16	197	66	12	**451**	38	390	3		31	58	1341
Chinese	24						3				**1017**	48			6	20	1118
Other groups - Other	313	5	201	68	21	35	81	68	13	141	26	**144**	300	79	354	204	2053
Total	381206	4737	5984	3160	1152	515	8354	4284	1627	1285	1148	752	1269	309	993	739	417514
Per cent of 1991 group																	
White	**96.8**	**1.2**	**1.4**	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.1	100.0
Black Caribbean	6.0	0.3	0.6	**77.2**	0.6	7.8	0.2	0.1	0.0	0.4	0.0	0.0	5.7	0.2	0.0	0.8	100.0
Black African	6.8	0.3	0.9	2.4	**77.4**	3.4	1.9	0.3	0.4	0.6	0.0	0.4	0.7	3.3	0.0	1.1	100.0
Black Other	12.0	0.3	3.1	20.6	3.1	**8.3**	1.9	1.6	0.7	1.7	0.0	1.1	32.1	6.4	1.6	5.3	100.0
Indian	1.8	0.1	0.4	0.2	0.3	0.0	**91.0**	0.6	0.1	4.4	0.0	0.2	0.0	0.1	0.6	0.2	100.0
Pakistani	2.2	0.0	0.2	0.0	0.2	0.0	1.1	**91.9**	0.3	3.1	0.0	0.1	0.0	0.0	0.8	0.1	100.0
Bangladeshi	2.4	0.0	0.0	0.0	0.0	0.0	1.0	0.8	**93.4**	1.5	0.0	0.0	0.0	0.0	0.5	0.2	100.0
Other groups - Asian	3.3	0.0	0.6	0.7	1.3	1.2	14.7	4.9	0.9	**33.6**	2.8	29.1	0.2	0.0	2.3	4.3	100.0
Chinese	2.1	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0	**91.0**	4.3	0.0	0.0	0.5	1.8	100.0
Other groups - Other	15.2	0.2	9.8	3.3	1.0	1.7	3.9	3.3	0.6	6.9	1.3	**7.0**	14.6	3.8	17.2	9.9	100.0
Total	91.3	1.1	1.4	0.8	0.3	0.1	2.0	1.0	0.4	0.3	0.3	0.2	0.3	0.1	0.2	0.2	100.0
Per cent of 2001 group																	
White	**99.7**	**99.3**	**94.6**	5.2	5.9	9.9	1.5	2.3	1.8	8.8	5.8	16.9	28.9	31.7	49.0	44.0	94.0
Black Caribbean	0.1	0.2	0.4	**82.8**	1.9	51.3	0.1	0.1	0.0	0.9	0.0	0.0	15.3	2.3	0.0	3.5	0.8
Black African	0.0	0.1	0.2	0.9	**83.0**	8.2	0.3	0.1	0.3	0.5	0.0	0.7	0.7	13.3	0.0	1.8	0.3
Black Other	0.0	0.1	0.6	8.1	3.3	**20.0**	0.3	0.5	0.6	1.6	0.0	1.9	31.2	25.6	2.0	8.9	0.3
Indian	0.0	0.2	0.6	0.5	1.9	0.8	**93.7**	1.1	0.3	29.6	0.0	2.8	0.0	1.6	5.1	2.4	2.1
Pakistani	0.0	0.0	0.2	0.0	0.6	0.0	0.6	**92.4**	0.7	10.4	0.0	0.4	0.0	0.0	3.5	0.7	1.0
Bangladeshi	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.3	**94.8**	1.9	0.0	0.0	0.0	0.0	0.9	0.5	0.4
Other groups - Asian	0.0	0.0	0.1	0.3	1.6	3.1	2.4	1.5	0.7	**35.1**	3.3	51.9	0.2	0.0	3.1	7.8	0.3
Chinese	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**88.6**	6.4	0.0	0.0	0.6	2.7	0.3
Other groups - Other	0.1	0.1	3.4	2.2	1.8	6.8	1.0	1.6	0.8	11.0	2.3	**19.1**	23.6	25.6	35.6	27.6	0.5
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Source: ONS Longitudinal Study, author's analysis

Table 2 shows that, while most people chose with the same ethnic category at both the 1991 and 2001 Census, some groups were more inconsistent in their response. Unsurprisingly the residual 'Other' group was least consistent: just seven per cent of those in the 1991 'Other' category fell under the same label in 2001. The largest shift for the group was to the mixed White & Asian category (17.2 per cent), followed by the White British group (15.2 per cent). Of the 2001 'Other'group, 52 per cent were drawn from the 'Other Asian' group and just 19 per cent had had the same label in 1991.

The consistency of responses was highest for the 1991 White group, with over 99 per cent choosing one of the three White categories in 2001. While the 2001 White Irish group was comprised almost entirely of people who identified as White in 1991 (99.3 per cent), the 2001 'Other White' group included a notable percentage of people who identified as 'Other' in 1991 (3.4 per cent) in addition to those who were grouped as White (94.6 per cent).

Those who chose minority ethnic group labels in 1991 were more likely to change their response category in 2001 than those who identified as White, although the recording of responses among Asian groups was more consistent than for Black groups. Over 91 per cent of those who identified as Indian, Pakistani, Bangladeshi, or Chinese in 1991 chose the same group in 2001. By contrast, around 77 per cent of people who categorised themselves as Black Caribbean or Black African retained the same label in 2001. Almost 8 per cent of people categorised as Black Caribbean in 1991 shifted to the 'Black Other' label in 2001, while 5.7 per of 1991 Black Caribbeans opted for the new mixed White & Black Caribbean label. Black Africans were less likely than Caribbeans to opt for the 'Black Other' label (3.4 per cent) or their corresponding mixed label, White & Black African (3.3 per cent) in 2001.

The 1991 'Black Other' label, which has often been combined with Black Caribbean in analysis to increase sample sizes, is the most changeable group apart from the residual category 'Other'. Table 1 shows that a significant percentage of people in the 1991 'Black Other' group were distributed among a diverse range of categories in 2001. Just 8.3 per cent of the 1991 'Black Other' group chose this label in 2001, while most identified as either 'Black Caribbean' (20.6 per cent) or 'White and Black Caribbean' (32.1 per cent). In addition, more than 15 per cent of the 1991 'Black Other' group chose one of the 2001 White groups. This suggests a large amount of fluidity in the boundaries between the 'Black Other', Black Caribbean and White & Black Caribbean categories. A large number of those who had defined themselves as 'Black Other' in 1991 had written in a description of 'Black British' (Owen, 1996); the inconsistency of responses for the 'Black Other' group may be associated with the revised headings to the 2001 question, which explicitly labelled the Black groups, 'Black and Black British'. Additionally, significant numbers of people who were categorised as 'Black Other' in 1991 responded with the new White & Black categories in 2001. Over 31 per cent of White & Black Caribbeans, and 25 per cent of White & Black Africans originated from the 1991 'Black Other' group.

There were other significant changes in recorded ethnicity associated with choice of the new mixed labels in 2001. Over two per cent of people who identified as Chinese in 1991 identified as White & Asian or 'Other Mixed' in 2001. Among the White group significant numbers of people changed to mixed groups and, although they are

small numbers in relation to the size of the 1991 White group as a whole, they represent 29 per cent of 2001 White & Black Caribbeans, 32 per cent of White & Black Africans, 49 per cent of the White & Asian group, and 44 per cent of the 'Other Mixed' group.

The measures of stability shown in Table 2 provide further evidence on variations in ethnic identification. While 98 per cent of the population were recorded with the same group label in 2001 and 1991, the degree of fit across the ten categories of 1991 averages only 67 per cent. From the degree of fit, it is clear that the residual groups are most changeable with just 8.3 per cent of 1991 'Black Others', 33.6 per cent of 'Other Asians' and 7 per cent of 'Others' retaining the same label in 2001. Even for Black Caribbeans and Africans, whose labels could be considered less ambiguous because they are not residual to other groups, around one in four was recorded in a different group in 2001.

Table 2: Stability of responses to the 1991 and 2001 ethnic group questions

	Measure of stability		
Ethnic group in 1991	Degree of fit	Two-directional fit	Marginal fit
White	99.5	99.1	99.8
Black Caribbean	77.2	66.5	93.2
Black African	77.4	66.8	93.3
Black Other	8.3	6.3	41.7
Indian	91.0	85.8	97.2
Pakistani	91.9	91.9	99.4
Bangladeshi	93.4	88.9	98.5
Other groups - Asian	33.6	20.7	95.8
Chinese	91.0	81.4	102.7
Other groups - Other	7.0	5.4	36.6
All groups	98.0		
Mean	67.0		

Source: ONS Longitudinal Study, author's analysis

The continuity of group composition shows in the two-directional fit, which not only measures movement between groups prospectively (from 1991 to 2001) but also retrospectively (from 2001 to 1991). Compared with the degree of fit (per cent of people who did not change from their 1991 group) the two-directional fit gives us some indication of the consistency of group composition, irrespective of whether we use the 1991 or 2001 classification to measure ethnicity. The two-directional fit is lower than the degree of fit for all groups, and is very low for the residual categories, demonstrating the inconsistency in the composition of these groups. In addition, the two-directional fit is much lower than the degree of fit for the Black Caribbean and Black African categories. This suggests that, aside from the consistency of responses in transitions from 1991 to 2001, these groups included large percentages of people in 2001, who were in a different 1991 category. For example, Table 1 shows that around 17 per cent of people categorised as Black Caribbean or Black African in 2001, were categorised to a different group in 1991. By comparison, among groups that have

much higher two-directional fits, around 6 per cent of Indians, 8 per cent of Pakistanis and 5 per cent of Bangladeshis were categorised to a different group in 1991.

We can assess the extent to which the introduction of new categories in the 2001 Census question impacted on the size of groups using the marginal fit, by expressing a 2001 group as a ratio to its 1991 size. Examination of the marginal frequencies in Table 1 shows that the larger number of categories in 2001 generally resulted in a reduction group size separate from the impact of births, deaths and migration on each population. The smallest proportionate reduction in group size was experienced by the White group, comparing the 1991 White category, with the three 2001 White categories (99.8 per cent). Table 2 shows that the marginal fit for the Black groups was lower than for the other main categories, with Black African and Black Caribbeans reduced by almost a quarter in 2001. South Asian groups experienced a small proportionate reduction in their size between 1991 and 2001 with marginal fits above 95 per cent, even for the residual 'Other Asian 'group. The Chinese category actually grew in size between the two censuses, by almost three per cent. By contrast, the size of the 'Black Other' and 'Other' groups were reduced significantly reflected in marginal fits of 41.7 and 36.6 per cent respectively. As Table 1 shows, many people who chose the Black Other and 'Other' groups in 1991, opted for one of the new mixed categories in 2001, this affected the size of these groups over time.

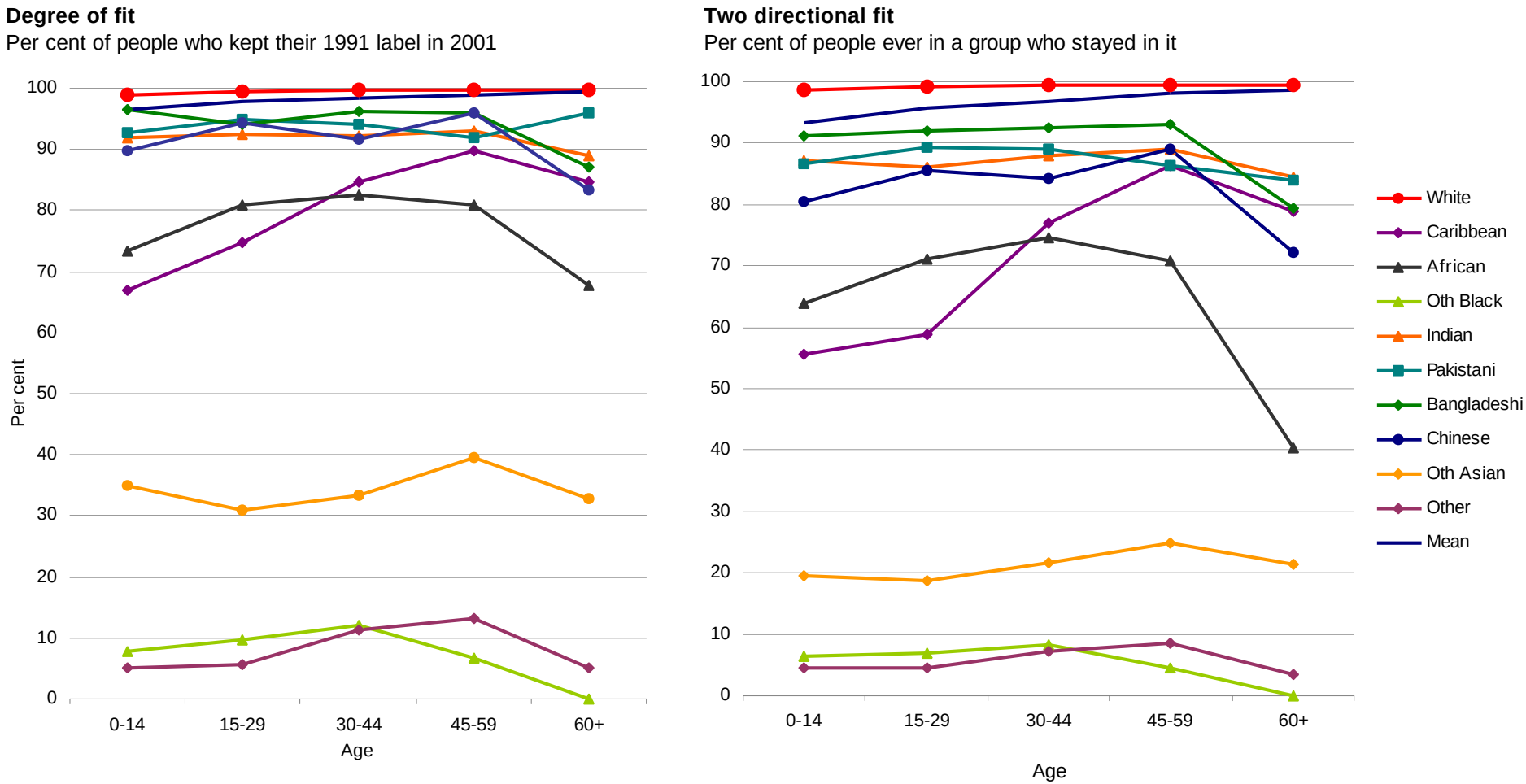
Instability by age

The figure below shows two measures of stability: the degree of fit and the two-directional fit, 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.

The distinctions between groups and the severe lack of consistency in reporting for the residual groups are evident at each age. In general, the degree of fit for each 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 response consistency with age for the Caribbean group.

In contrast to the general rise in consistency 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 fewer than 100 members of the LS, the decline is still noticeable for the larger Indian and Black Caribbean groups. It is not clear whether older people find the question more difficult to answer, or less acceptable. The figure excludes imputed 2001 records, which therefore can be ruled out as a reason for the drop in consistency of reporting for the older group.

Figure 1: Age and stability of ethnic group



IMPLICATIONS FOR SOCIAL RESEARCH

Allocation of groups for comparison between 1991 and 2001

There are several possibilities to ensure the maximum comparability between data based on the 1991 and 2001 ethnic group classifications. Tables 1 and 2 showed that there was much inconsistency in recorded ethnic group if we simply compare the groups based on the original Census output classifications. Reducing the 16 ethnic group categories of 2001 to 10 for direct comparison with the 1991 classification can best be achieved as demonstrated in Box 2. The amalgamated categories provide a square ten-by-ten transition matrix from 1991 to 2001 categories carrying the same label. This suggestion provides good arithmetical fit of the 1991 and 2001 categories, makes distinction between categories according to the real choices individuals made in answering the 1991 and 2001 censuses, and is consistent with documented coding schemes. However, it does mean that the mixed categories are not identified separately because they were not included in the 1991 classification.

Box 2: Ten-category ethnic group classification

Presentation group	1991 categories	2001 categories
White	White	White British White Irish White Other
Indian	Indian	Indian
Pakistani	Pakistani	Pakistani
Bangladeshi	Bangladeshi	Bangladeshi
Other Asian	Other Asian	Other Asian
Black Caribbean	Black Caribbean	Black Caribbean
Black African	Black African	Black African
Other Black	Other Black	Other Black White and Black Caribbean White and Black African
Chinese	Chinese	Chinese
Other ethnic group	Other	Other White and Asian Any Other Mixed

The three mixed categories are split between the 'Black Other' and 'Other' categories in the proposed grouping. Table 1 shows that there is relatively low correspondence between the 2001 Mixed categories and the existing 1991 categories; 'Black Other' who changed to White & Black Caribbean (32.1 per cent) are largest single group who changed to a mixed category. This is due to, on the one hand, differences in coding decisions taken in 1991, which meant that people who wrote in White & Black origins were more likely to be identified as a distinct group, and on the other hand, the instructions on the 1991 Census form which advised people to chose a single group. Those who did write in a mixed identity on the form were coded as 'Mixed' before

being allocated to one of the residual categories used in standard tabulations from the Census. However, this depended on where people wrote their response. Those who wrote in mixed White & Black African or White & Black Caribbean origins in the space under the 'Black Other' heading were included in the 'Black Other' category. Those who wrote in any other mixed origins, or who used the space under the 'Any other ethnic group' heading were coded as 'Other ethnic group'. Thus, people were allocated to different groups depending on where they wrote in their answer. For this reason, it seems appropriate to split the mixed groups between the 'Black Other' and 'Other' categories.

While there are strong diagonals produced from comparing two 10-group classifications, there are also significant off-diagonals. In particular for the three residual categories 'Black Other', 'Asian Other' 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. One solution is to amalgamate the least consistent 2001 groups into one diverse group, creating a new eight-group classification, shown in Box 3. It is simply based on isolating the groups where one half or less retained their 1991 label in 2001 ('Black Other', 'Asian Other', and 'Other Ethnic Group'), or where less than half of a 2001 group originated from a single 1991 group (White & Black Caribbean, White & Black African, White & Asian, and 'Other Mixed'). This classification maximises the fit between the two measurement points, retains meaningful groups for analysis, and ensures that, conceptually, the groups contain the same sets of people at both points. The trade-off is the inclusion of the 'Black Other' group and mixed White & Black groups in the diverse 'Other' group. However, Table 1 and the discussion above demonstrate that the 2001 categories included in the 'Other' group for the proposed eight-category classification were recorded in a diverse range of groups in 1991, and are each relatively small analysis groups.

Box 3 : Eight-category ethnic group classification

Presentation group	1991 categories	2001 categories
White	White	White British White Irish White Other
Indian	Indian	Indian
Pakistani	Pakistani	Pakistani
Bangladeshi	Bangladeshi	Bangladeshi
Black Caribbean	Black Caribbean	Black Caribbean
Black African	Black African	Black African
Chinese	Chinese	Chinese
Other ethnic group	Other Black Other Asian Other	White and Black Caribbean White and Black African White and Asian Any Other Mixed Other Black Other Asian Other

Implications for social research

The comparability of ethnic group data from 1991 and 2001 Censuses is an important concern for social researchers. For example, in a recent compendium examining the feasibility of projecting the England and Wales population with ethnic group as dimension, many contributors expressed concern with the comparability of data over time and the ability to model changes in identity that may be associated with life stages, and therefore age (age being a key concern of population projections) (Haskey, 2002).

The recommendation from results presented in this paper has been to use seven categories that are most comparable between 1991 and 2001, leaving a single residual category that is not meaningfully comparable over time. Research using a longitudinal study where ethnic group is measured on more than one occasion must make the 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. It is advisable to test the sensitivity of analyses to the choice of ethnic group classification, and, particularly 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 not.

More generally, data from the two censuses allow conditions to be compared at two time points. In the analysis of time series data, interpretation of ethnic group is affected by consideration of the fluidity between group boundaries, and the fact that people may have 'migrated' into or out of an ethnic group category during the decade. We have seen that this issue most affects Black Caribbeans and Africans. A related point is whether the group actually contains the same people at both time points. Comparisons need to acknowledge and if necessary adjust for the different composition of each category. Census data show that the 1991 ‘Other Asian’ group was dominated by people with Far East and South East Asian origins, with 40 per cent of people born in the Far East and 20 per cent born in the Indian Subcontinent (OPCS, 1993). In 2001, on the other hand, rather more of this group were born in the UK (32 compared with 22 per cent in 1991) (ONS, 2003d) and the Indian Subcontinent (37 per cent) compared with the Far East (20 per cent). This suggests that, in addition to differences between groups in individuals' propensity to change ethnicity, researchers must situate their understanding of group composition historically.

REFERENCES

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

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

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

Haskey, J. (ed) (2002) Population projections by ethnic group. A feasibility study. Studies on Medical and Population Subjects No 67. London: The Stationery Office.

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

Kertzer D. and Arel D.(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.

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

Office for National Statistics (2003b) *Census 2001 Review and Evaluation. Quality of Data Capture and Coding: Evaluation Report.* Titchfield: Office for National Statistics.
http://www.statistics.gov.uk/census2001/pdfs/data_capture_and_coding_evr.pdf

Office for National Statistics (2003c). *Census 2001 Review and Evaluation. Edit and imputation. Evaluation report*. Titchfield: Office for National Statistics.
http://www.statistics.gov.uk/census2001/pdfs/editimputevr.pdf

Office for National Statistics (2003d) Census 2001. *National Report for England and Wales* pp 122-125. London: HMSO.

OPCS (1993) *1991 Census Ethnic Group and Country of Birt*h. pp415-416. London: HMSO.

Owen, D (1996) 'Size, structure and growth of the ethnic minority populations'. In Coleman D. and Salt J. (eds) (1996) 'The ethnic group question in the 1991 Census: a new landmark in British social statistics' in Coleman D. and Salt

J.(eds) *Ethnicity in the 1991 Census. Volume One. Demographic Characteristics of the Ethnic Minority Populations*. London: Office of Population Censuses and Surveys

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.

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

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.

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