

Kent Academic Repository

Full text document (pdf)

Citation for published version

Aspinall, Peter J. (2011) The utility and validity for public health of ethnicity categorisation in the 1991, 2001 and 2011 British censuses. *Public Health*, 125 (10). pp. 680-687. ISSN 0033-3506.

DOI

<https://doi.org/10.1016/j.puhe.2011.05.001>

Link to record in KAR

<http://kar.kent.ac.uk/28639/>

Document Version

Author's Accepted Manuscript

Copyright & reuse

Content in the Kent Academic Repository is made available for research purposes. Unless otherwise stated all content is protected by copyright and in the absence of an open licence (eg Creative Commons), permissions for further reuse of content should be sought from the publisher, author or other copyright holder.

Versions of research

The version in the Kent Academic Repository may differ from the final published version.

Users are advised to check <http://kar.kent.ac.uk> for the status of the paper. **Users should always cite the published version of record.**

Enquiries

For any further enquiries regarding the licence status of this document, please contact:

researchsupport@kent.ac.uk

If you believe this document infringes copyright then please contact the KAR admin team with the take-down information provided at <http://kar.kent.ac.uk/contact.html>

**The utility and validity for public health of ethnicity categorisation
in the 1991, 2001 and 2011 British censuses**

Peter J Aspinall BA MA

Reader in Population Health & Honorary Special Advisor to the London Health Observatory

Centre for Health Services Studies, University of Kent, George Allen Wing, CANTERBURY,
Kent CT2 7NF, UK.

Tel. 01227 827576

E-mail: peter.petjasp@gmail.com or P.J.Aspinall@kent.ac.uk

**The utility and validity for public health of ethnicity categorisation
in the 1991, 2001 and 2011 British censuses**

Abstract

Objectives: An attempt is made to evaluate the utility and validity of the ethnicity categorisations across the 1991, 2001, and 2011 British censuses for public health purposes.

Study design: Narrative review.

Methods: A review of journal literature and census and other policy reports was undertaken to assess specified criteria for the utility and validity of the censuses for public health.

Results: The census ethnicity categorisation satisfactorily captures the ethnic diversity of the population and adheres to the principle of self-identification in the labels used and underlying conceptual base. The stability of some of the categories (especially black groups and ‘mixed’) continues to be problematic for public health. Concealed heterogeneity in the white group has been partially addressed but remains in the black African group. Colour categories (‘white’ and ‘black’) have been retained in the 2011 census, with only limited objection amongst the communities they describe.

Conclusions: The complexity of the classifications and range of data on the dimensions of ethnicity have increased over the three decades. The breakdown of the ‘black African’ group, the shortcomings of ‘mixed’ categorisation, and how the white category is subdivided require further investigation.

Keywords: census, classification, categorisation, ethnic diversity, self-identification

Introduction

A question on ethnic group was asked for the first time in the 1991 Great Britain census. Revised ethnic group questions were asked in the 2001 censuses for England and Wales and Scotland that were similar in structure, as are those used in the 2011 census. Data from the 1991 and 2001 censuses have transformed our knowledge of minority ethnic health and healthcare by providing, for example, population denominator data for rates/ratios, information on the determinants of health and health inequalities, a template for collecting ethnic data in administrative systems and surveys, and, in Scotland, a means of populating ethnic group in administrative data that lacked it¹. Additionally, in 2001 a question on general health was added to that on limiting long-term illness, and a third more detailed health question in the 2011 Scotland census. These gains for population health have been widely acknowledged.

However, there has been less focus on the utility and validity of census ethnicity classifications and categories for public health and on census agencies' and respondents' understandings of the concept of ethnicity. Census ethnicity categories are now widely used in public health for the stratification of data such as that on health-related behaviour and the use of health services and in population profiling for public health risk assessment. Most debate has focused on the meaning of ethnicity and race and the reliability of these social constructions as proxies for other biological, social, and biosocial variables. There is general agreement that these concepts are crucial for assessing the risk of discrimination and disadvantage along the lines of race and ethnicity; moreover, the view that they should only be used as proxies for other variables that cannot be measured and when they are the most reliable proxies available is judicious².

A wide range of opinion exists on whether ethnicity can be reliably measured, including whether it can accurately reflect the most salient categories of group identity³, but with some

consensus on the fact that it should be self-assessed⁴. Some investigators point to the limitations of single, mutually exclusive categories for measuring a multidimensional concept and favour breaking down ethnicity into elements which can be separately measured and jointly or separately analysed. Such elements might include parentage or ancestry, national identity, language, religion, country of birth, and patterns of behaviour, friendship and association⁵. Others eschew this multidimensional approach that unpacks ethnicity as too simplistic, arguing that a global measure is needed as ethnicity is a context driven social and psychological concept^{6,7}.

At the interface between those asking ethnicity questions (census agencies) and those answering them, the diversity of ways in which ethnicity is understood by these actors is poorly documented. UK Census research into respondents' understandings of the terms 'ethnic', 'culture', 'nationality', and 'race' has shown that these concepts are generally quite separately defined but sometimes used interchangeably⁸. Most respondents also distinguished between 'ethnic origin' and 'ethnic group', recognising the importance of the former as referring to an individual's parental background while not necessarily associating themselves with a particular group. Parentage was much more important in determining ethnic group than where a person was born or language and religion in surveys of patient populations⁹. Again, amongst 'mixed race' respondents, the contribution of parental race/ethnicity to ethnic identity was far more salient than 'social factors' (such as the perceptions of the wider society, feelings of group allegiance, and identification by friends and peers)¹⁰.

The importance of the different dimensions of ethnicity to self-identity may vary across groups. For example, national identity (being 'British') was shown to be especially significant for black groups in the 1991 and 2001 Census free-text responses¹¹, while other research indicates that religion is prioritised amongst South Asian groups¹². Though non-response rates for the ethnic group question in censuses and surveys are now very low, utilisation of 'other' write-in categories is high and it is clear that many respondents need the cues of the

question's context, including list of categories, to answer it successfully: in one interview survey of unprompted ethnic group, almost a quarter of respondents replied that they did not know what an ethnic group was or were unsure which group they belonged to, while almost all selected a category when prompted with the census classification⁹.

Census ethnicity questions therefore need to be robustly designed and tested to ensure acceptability and understanding by those answering them. The Office for National Statistics (ONS) acknowledges that ethnic identification is a multifaceted and changing phenomenon which may include aspects of ancestry, country of birth, nationality, language spoken at home, religion, culture, skin colour, and national/geographic origin. However, in placing emphasis on the subjective nature of ethnicity, it argues that respondents will draw on these dimensions in ways that are relevant to them, while acknowledging that the ethnic group options presented to the respondent are not completely ones of self-identity¹³. Processes of group identification and social categorisation are not, however, separate but are mutually implicated in and feed back upon each other¹⁴.

Moreover, when the data collection instrument is a national, decennial census, there are additional issues to consider. The categorisation cannot always incorporate groups which may be numerically important at a local as opposed to a national level. Moreover, taking part in the census is compulsory so priority is accorded to such matters as acceptability, parsimony imposed by the limitations of questionnaire length, respondent burden, sensitivity with respect to personal information (the question on religion is voluntary), and optimal capture (the 2011 England Census questionnaire being printed in 56 other languages).

Methods

A narrative review of published literature accessed via Medline, Embase, King's Fund Database, DH Data, and Web of Knowledge, reports of the ONS and General Register Office

(Scotland) Census Development Programmes (CDPs), and other policy material was undertaken to address the research question: What is the utility and validity of ethnic categorisation in the 1991, 2001, and 2011 censuses for public health purposes? The assessment of a comprehensive body of knowledge on contested concepts, approaches to data capture, and reflective practice required the flexibility of a narrative review with broad coverage.

Results

The results of the literature synthesis are presented with respect to specified criteria of utility and validity for public health. Amongst utility criteria, capture of the complex ethnic diversity of the country is needed to facilitate studies on the health of populations. However, respondent understandings and burden and the need for data that does not incur the ‘small numbers’ and related confidentiality problems set limits to increased granularity. The categories should be measurable in an accurate and straightforward way and be reasonably stable/reproducible beyond the fluidity that is intrinsic to the concept of ethnicity. Poorly designed categories and classifications may create instability through inherent ambiguity or sheer complexity. Finally, the utility of the categories is ‘...ultimately a matter of the extent to which they contribute to the construction of cross-culturally testable hypotheses and theories’¹⁵.

With respect to validity, a measure is considered valid if it is thought to measure the concept or property (in this case ethnic categorisation) which it claims to measure. Validity may depend on many factors, including the appropriate wording of classifications and categories. It thus requires that the categories be understandable and acceptable to those they describe, the ONS arguing that categories should be used that ‘...match people’s own preferred ethnic descriptions of themselves’¹³. The US Census Bureau, too, requires familiarity with and acceptance of the categories ‘if the principle of self-identification is to be honored’¹⁶.

Arguments of validity in this context rule out proxy reporting. However, entirely open response ethnicity data would result in statistics stratified by a myriad small categories of questionable utility¹⁷. Equally, the use of broad heterogeneous ethnic categories challenges validity, especially when the constituent populations are very different. Resolving such trade-offs is key to the work of CDPs and ultimately determines how useful the data will be for public health purposes.

Examining validity or measurement accuracy in the context of ethnicity is problematic in censuses. There is no (gold) standard - such as the form-fillers' unprompted self-descriptions - against which to assess validity of responses. Nor are other variables in the dataset, such as country of birth, religion, and small geography indicators necessarily informative about self-ascribed ethnic group, as evaluations of item imputed ethnicity through record linkage across censuses have shown¹¹. However, censuses are exceptional in having post-enumeration validation surveys after each census to check data quality and validity, although these do not always include the ethnicity question. Reliability - the reproducibility or consistency of the measure - is relevant in that a highly unreliable measure cannot be valid. The measure's reproducibility may, in turn, reflect its sensitivity (its ability to correctly identify most people that the category is intended to capture) and specificity (its ability to exclude most others).

The capture of the ethnic diversity of the population

Over the last two or three decades, Britain has become an increasingly diverse country by virtue of substantial flows of new migrants from a wider range of countries¹⁸, higher rates of inter-ethnic marriage and partnership formation¹⁹, and changes in ethnic identification, including use of mixed/multiple affiliations²⁰. While accurate data is needed on this ethno-cultural diversity to inform population-based health research, census ethnic categories cannot, of themselves, fully capture this diversity. The 2001 ethnic question was the longest on the census form and perhaps twenty categories represents an upper limit for self-completion. Census categorisation focuses on those collectivities that substantially reflect societal

perceptions of ethnic differences and provide a point of access to discrimination, disadvantage, and the longer-term historical processes of colonialism²¹. It does not encompass all the polyethnic minorities, such as Filipinos/Filipina now estimated by government to number 250,000. However, cross-tabulations of ethnic group by country of birth, religion (Christian, broken down in Scotland; Muslim; Buddhist; Sikh; Jewish; and Hindu), language, and national identity (English; Welsh; Scottish; Northern Irish; British; and a free text Other) – the set available in the 2011 censuses – will significantly facilitate our understanding of Britain’s current ‘superdiversity’²².

In the 1991 and 2001 CDPs, there were no explicit criteria for inclusion or exclusion, such as floor levels for the size of the group. In the England and Wales 2011 CDP a metrics- or rule-based ‘Prioritisation Tool’ has been used to rank the 22 candidates for the two available new categories, ‘Arab’ and ‘Gypsy or Irish Traveller’ winning out²³. The incentive to maintain comparability exerts a powerful influence in category selection for the next census but is always secondary to needed changes. Across all three censuses there has been significant continuity in the way the ‘African’, ‘Caribbean’, ‘Chinese’, and Indian sub-continent groups have been captured (table 1), their presence reflecting structural and historical processes that have shaped ethnic relations.

[table 1 here]

Developments in categorisation since 1991 have come about through either community pressure or user demand. The addition of ‘mixed’ categorisation in 2001 (impacting on some of the 1991 groups) was largely driven by need amongst users of census data to capture this diversity following the finding that 230,000 persons wrote in a ‘mixed’ description in the 1991 Census²⁴, rather than pressure from mixed race organisations as in the USA. Similarly, the frequency of ‘black British’ in free-text answers - and, to a lesser extent, ‘Asian British’ - found recognition in the overarching ‘Black’ and ‘Asian’ banners in 2001. Late in the 1991 CDP the endeavours of the Irish community to obtain an ‘Irish’ category came too late for

inclusion, the census agency providing a ‘born in Ireland’ count on many of the ethnicity tables²⁵. In the 2001 England and Wales CDP Irish organisations mounted a robust and eventually successful campaign based primarily on the group’s disadvantage (the born in Ireland group having the highest standardised mortality ratio amongst residents in England and Wales) to obtain official recognition for ‘Irish’, initially against resistance from ONS but with support from advisory group members. Degree of disadvantage in itself, however, has not always favoured candidate categories, ‘Gypsy or Irish Traveller’ having to wait till 2011 for inclusion.

Community involvement was formalised in the 2011 CDP through a ‘special populations’ (renamed ‘diversity’) advisory group, with representatives from African, Arab, Gypsy, Irish, Jewish, Muslim, and Sikh organisations as well as bodies like the Equality and Human Rights Commission²⁶. Although there was a community liaison programme ahead of the 2001 Census²⁷, the influence of community opinion has come late to CDPs in Britain, while special advisory groups have exerted a powerful influence on categorisation in the US race question since the 1990 Census.

Category labels and the principle of self-identification

Selecting appropriate ethnic labels is important as they can influence data quality if unfamiliar to or not locally grounded in the terms communities use. Similarly, terms are avoided that split the choices of specific groups, to maintain the principle of ‘mutual exclusiveness’. Cognitive testing in the 2001 CDP showed that those of Indian sub-continent origin did not understand the term ‘South Asian’, thinking it applied to SE Asian origins⁸. In the 2011 census, ONS changed the label for the mixed categories from ‘mixed’ to ‘mixed/multiple ethnic groups’ to meet the preferences of the ‘mixed’ population. These mainly reactive programmes of cognitive research, focus groups, small scale testing, and the large Census Test and Census Rehearsal are the modus operandi for developing categorisation. More

importance is accorded to the terminology community members favour by the US Census Bureau, including exploratory large-scale population-based surveys²⁸. In the British approach use is made of the census free-text responses to identify emerging identities and groups, but data on how people describe themselves unprompted in their own words is scarce^{9,29} and almost absent in surveys³⁰.

Conceptual bases of identity or origin

The 1991 Census question used the conceptual base of ‘ethnic group’ but referenced concepts of descent, ancestry, racial group, and belonging in the instruction, suggesting a ‘global’ measure of ethnicity. In 2001 the question retained the concept of ‘ethnic group’, but the 16 tick boxes embedded within the five main ‘pan-ethnicities’ (‘White’, ‘Mixed’, ‘Asian or Asian British’, etc.) were termed ‘cultural background’. This resulted in the NHS Data Dictionary changing the term used for the options from ‘groups’ to ‘categories’ to reflect their officially constructed nature³¹. Further, whether terms like ‘Indian’ are sustainable, in the absence of reference to religion, as ‘cultural background’ descriptors³² rather than ‘countries or regions of family origin’³³ is questionable. ‘Ethnic group or background’ was agreed for the 2011 census tick boxes, replacing ‘cultural background’. ‘Family origins’, suggested for the 2001 Census³⁴, was not tested as ONS wished to prioritise the subjective measure of ethnic identity.

Given the focus on identity, scrutiny has fallen on the degree of agency (that is, the capacity for willed action, in this case uncoerced and unmanipulated choice making) accorded to the form-filler by the question. Morning³⁵ has argued that the wording of the 2001 census question, ‘What is your ethnic group?’, embodies ‘an essential being ethnic, as opposed to a constructed belonging to an ethnicity’ in the wording ‘To what ethnic group do you belong?’, a distinction between ‘a more essentialist concept of ethnicity as objectively given, and a more constructionist understanding of ethnicity as socially and thus subjectively developed’.

The 2011 questions yield to the respondents' agency in making a choice by inviting the respondent to 'tick **one** box to best describe your ethnic group or background'. Some US evidence does indicate that it is the categories that are important and that the public frequently treat differing concepts as part of the same semantic domain³⁶. Indeed, in Britain the Fourth National Survey of Ethnic Minorities found substantial overlap between 'ethnic group' and 'family origin' responses³⁷, though, as noted, cognitive research for the 2001 Census did show that some participants distinguished between 'ethnic group' and 'ethnic origin'³⁸. A conceptual base of identity is therefore likely to have widespread utility across public health usages with only minor selective attribution effects.

Stability / reproducibility of the categories

Categories that have high test-retest consistency have greater utility and validity for public health, such stability being especially important in the estimation of health risks. The 1991 Census Validation Survey (CVS) provided some evidence of consistency: 99.6% of those who identified as 'White' in the census did so in the CVS and the proportion consistently identifying was also high in the Indian subcontinent groups (98.7%), but lower in the Black groups (88.0%) and Other groups (78.1%)³⁹. No CVS that included ethnicity was undertaken after the 2001 census. However, transitions between ethnic categories in the ONS Longitudinal Study (LS) showed strong stability between 1991 and 2001 amongst people who identified as White, Chinese, or with one of the South Asian groups in 1991⁴⁰. Consistency of response over the decade was lower for the Black African and Black Caribbean groups.

Concerns over issues of consistency should not limit the use of the census ethnic categories in public health research. Use, in itself, contributes significantly to the evidence base on validity and utility and catalyses improvements in data quality. However, it is premature to assume that the data are now satisfactory. Even with the more stable South Asian categories, evidence of instability in some subgroups is worrying, especially for those residents born in the UK, a

segment that is increasing in size⁴¹. Among UK-born persons recorded as Pakistani or Bangladeshi in 1991, more than 5% were recorded as a different group in 2001. More than 10% of UK-born Indians and more than 20% of UK-born Africans and Caribbeans were recorded as a different group in 2001. While the full matrix of transitions is not available by whether born in the UK or elsewhere, such data is available at ethnic group level. For example, the most common alternative category in 2001 by those identifying as Indian in 2001 was 'Other Asian' (4.4%), the first time this category had been offered on the census form; the next most common being 'White British' (1.8%)⁴⁰. Similar shifts were found amongst those who ticked Pakistani in 1991 (3.1% to 'Other Asian' and 2.2% to 'White British') and Bangladeshi in 1991 (1.5% to 'Other Asian' and 2.4% to 'White British'). In seeking to explain the shift to 'White British', ONS argues that these respondents may have interpreted the category as meaning culturally 'British', as 'White' appeared in the heading rather than tick box category¹¹.

With respect to 'mixed' ethnicity, which has a greater potential for inconsistency, an analysis of patients with more than one spell in hospital episode statistics for the period 2003/04-2005/06 showed high instability: 'White and Black African Mixed' patients had only a 70 per cent chance that a following admission would be coded in the same way, compared with 95 per cent with the 'White British' group⁴². This lack of consistency for some categories requires further investigation and underpinned arguments promoting ethnic family origins over identity³⁴. A post-2011 census validation survey would provide an important test of short-term category reproducibility, to set against longer-term change (over a decade), established through the Longitudinal Study.

Addressing concealed heterogeneity

The need to address concealed heterogeneity in the categories, the 'fallacy of homogeneity'⁴³, has been a persistent public health user demand, as systematic within-group heterogeneity

might correlate with different health outcomes, as in the Black African group⁴⁴. In 2001 concealed heterogeneity was addressed through category subdivision, the 'White' category being broken down to separately distinguish 'British' (but 'Scottish' and 'Other British' in Scotland), 'Irish', and a free text 'Other White' background. In the 2011 Census, 'Gypsy or Irish Traveller' has been added and the 'British' option changed to 'English/Welsh/Scottish/Northern Irish/British', a 'Polish' category being included in Scotland only. A pragmatic category breakdown for the 'Mixed' (now 'Mixed/multiple') category is provided in the England and Wales 2001 and 2011 Censuses (important differences in standardised ratios for not good health being reported across the four categories⁴⁵), Scotland using free-text. The 'Asian' categories have remained little changed, besides the addition in 2001 of a free text 'Any other Asian background', 'Other Asian' being a manufactured output category in the 1991 Census. However, Chinese was added to the 'Asian' set in Scotland in 2001⁴⁶ and in England and Wales in 2011.

The ability to cross-tabulate the ethnic categories by other census variables does provide a point of access to finer granularity. Religion, first asked in 2001, enabled Muslim, Sikh, and Hindu Indian 'cultural backgrounds' to be identified (the Fourth National Survey of Ethnic Minorities showing important variations in diagnosed disease rates and health-related behaviour across these ethno-religious categories⁴⁷) and for Christian and Muslim Black Africans to be differentiated. There is, too, a third tier of data, in addition to the 'pan-ethnicities' and the embedded cultural backgrounds, in the form of discrete categories used to code the free-text responses. While these are subject to selective attribution and do not provide accurate counts in the census, other administrative systems have used them to collect ethnicity data.

The problem of colour categories

The problem of using colour terms like 'black' and 'white' in census ethnicity classifications has been debated since the earliest field trials in the late 1970s. Such terms are perceived to be

linked with skin colour and other markers of phenotype and to have become racialised, although having a range of meanings for census respondents. Given a shift in 2001 to a conceptual base of 'cultural background' and recognition of racisms beyond colour, it can be argued that the time has come to abandon such labels in official classifications^{48,49}. The evidence base, however, shows that these colour terms have powerful acuity in directing people through the classification, for example, by filtering out 'white' sub-Saharan Africans who are numerically important amongst those with origins in Zimbabwe and South Africa. Similarly, 'black' differentiates those whose ancestral origins lie in sub-Saharan Africa from African Indians. The Scottish census agency took the bold step of testing a classification in 2006 that eliminated colour labels (substituting 'European' for 'White' and 'African or Caribbean' for 'Black')⁵⁰: however, post-test evaluations - including the linking of 2001 Census and 2006 Test individual records - showed that some respondents misunderstood the categories, there being a movement from 'White' to the 'African' categories⁵¹.

The current evidence indicates that the terms 'black' and 'white' are still needed to ensure accurate capture. While there are some voices of opposition (mainly from within the academy and from some African organisations)⁵², cognitive research indicates that the term 'black' is one which most community members find acceptable and frequently use spontaneously as a self-descriptor⁵³. Moreover, some public health professionals argue that the language of colour is needed to set white privilege against black disadvantage⁵⁴.

Discussion

The census continues to be an important source for public health specialists engaged in work on health protection, health improvement, and health inequalities issues for a number of reasons. The categorisation and classifications are the outcome of robust regimes of testing, resulting in their rapid adoption as the gold standard measure across most official ethnicity data collection (such as that by hospitals, GP practices, and the social care sector) and

monitoring. This facilitates comparisons across different administrative systems, provides denominator data for the calculation of disease burden rates/ratios (for example, more than half of the 17 ethnicity indicators in a Chief Medical Officer public health report used census denominator data⁴⁵), and enables population trend data to be derived. Moreover, since the enumerations aim to capture the entire population, census statistics provide the best opportunity for examining the characteristics of small ethnic minority populations or ethnicity at the level of small geographies. The latter is especially important for strategic public health aimed at driving health improvement locally and at the community level. An additional value added by census data is the wealth of contemporaneous data collected on poverty and disadvantage, work on health inequalities having established that deprived areas face the greatest challenges and where improving the health of the poorest fastest has been prioritised (the approach of ‘proportionate universalism’)⁵⁵.

Over the three censuses the public health utility of the categorisation has increased substantially, especially in addressing concealed heterogeneity through subdivision and scope to cross-tabulate by religion, language, and national identity. However, ‘Black African’ remains too broad a category to be useful, encompassing well-established communities like Nigerians and those extremely disadvantaged like Somalis and Congolese⁵⁶. The numerous national origins of ‘black Africans’ and complex intersections with religion, language, and migration history defy an easy subdivision, but a regional breakdown is feasible as is an African free text option (used in Scotland in 2011). The national origin categories of ‘Nigerian’ and ‘Somali’ may now merit predesignated category status in surveys in England and Wales by virtue of their size and, for Somalis, their socio-economic outlier status (their candidacy being argued in parliament)⁵⁷. The naming of ‘African’ categories may remove the need for a ‘black’ referent and thereby contribute to improved harmonisation of black African data across Britain.

For some public health purposes the 'Mixed' categorisation in England and Wales may be too limited as it conceals the important 'White and Chinese' mix and omits an explicit 'minority mixed' category⁵⁸. Further, the use in the England and Wales 2011 question of a single (rather than duplex) free text box leaves scarcely enough space to write in one group, the duplex box in 2001 - retained in Scotland - propelling form-fillers to name two groups. The evidence of testing indicates that predesignated mixed categories still produce better quality data than multi-ticking^{51,59}. The subdivision of the 'White' category is suboptimal, 'British' being included in a tick box option rather than part of the overarching label as with 'Black' and 'Asian', thereby excluding from Britishness those who tick 'Irish' or 'Gypsy or Irish Traveller'. Both census agencies appear to be relying on the national identity question to break down White into 'home country' origins, even though the concept is different.

In conclusion, census data on ethnicity is key to public health research and practice and ethnicity continues to be one of the most important epidemiological variables. There is an extensive body of evidence to demonstrate the utility of ethnicity categorisation in the 1991 and 2001 censuses for public health, including analyses of limiting long-term illness⁴⁵, determinants of health indicators, denominator data for rates/ratios, and for record linkage.

However, the poor stability, short-term or over a decade, of some of the census ethnic categories, including mixed, black African, black Caribbean, and UK-born Indians, has important implications for the use of this data in public health. While arguments exist for abandoning colour categories, their acuity in sorting responses into appropriate categories needs to be acknowledged. The 2011 census adds new categories (Arab and Gypsy or Irish Traveller) and, in Scotland only, Polish and an open response African category. Questions on national identity and language will be asked for the first time, the latter providing a measure of 'linguistic isolation'⁶⁰ at the household level to inform debates on social exclusion and the data needed to plan interpreting, translation, and English for Speakers of Other Languages (ESOL) services. Some problems remain with categorisation: the undivided 'black African'

category in England and Wales, the limited subdivision of 'white', and 'mixed' categorisation that is now suboptimal.

Ethical Approval

Nothing to declare

Competing Interests

Nothing to declare

Funding

Nothing to declare

References

1. Bhopal R, Fischbacher CM, Steiner M, Chalmers J, Pover C, Jamieson J, Knowles D. Ethnicity and health in Scotland: can we fill the information gap? Edinburgh: Public Health Sciences, University of Edinburgh; 2006.
2. Ellison GTH. 'Population profiling' and public health risk: When and how should we use race/ethnicity? *Critical Public Health* 2005; 15(1): 65-74.
3. Bartlett A, Fiander M. Census categories of ethnic group are limited. *British Medical Journal* 1995; 310: 332.
4. Bradby H. Describing ethnicity in health research. *Ethnicity and Health* 2003; 8: 5-13.
5. Burton J, Nandi A, Platt L. Measuring ethnicity: challenges and opportunities for survey research. *Ethnic and Racial Studies* 2010; 33(8): 1332-1349.
6. Witzig R. The medicalization of race: scientific legitimization of a flawed concept. *Annals of Internal Medicine* 1996; 125(8): 675-9.
7. McKenzie K. Ethnicity in psychiatric epidemiology. *British Journal of Psychiatry* 1998; 172(3): 278.

8. Mortimer L, White A. Ethnic Group Question: Findings from focus group discussions. London: ONS Social Survey Division, 1996.
9. Pringle M, Rothera I. Practicality of recording patient ethnicity data in general practice: descriptive intervention study and attitude survey. *British Medical Journal* 1996; 312: 1080-2.
10. Aspinall PJ, Song M, Hashem F. The ethnic options of 'mixed race' people in Britain. ESRC Research Grant RES-000-23-1507. End of Award Report, 2008.
11. ONS. A guide to comparing 1991 and 2001 Census ethnic group data. London: ONS: 2006.
12. Modood T, Beishon S, Virdee S. Changing Ethnic Identities. London: Policy Studies Institute, 1994.
13. ONS. Ethnic group statistics: A guide for the collection and classification of ethnicity data. London: ONS, 2003.
14. Jenkins R. Social Identity. London: Routledge, 1996.
15. Harris M, Consorte JG, Lang J, and Byrne B. Who are the Whites? Imposed Census categories and the racial demography of Brazil. *Social Forces* 1993; 72(2): 451-462.
16. US Office of Management and Budget. Federal Register 7 September 1997; Part II: 36873-36946.
17. Aspinall PJ. The conceptual basis of ethnic group terminology and classifications. *Soc Sci Med* 1997; 45(5): 689-698.
18. Salt J. International Migration and the United Kingdom. Report of the United Kingdom SOPEMI Correspondent to the OECD, 2008. Paris: OECD; 2008.
19. Mackintosh M. Inter-ethnic unions. Census update 2005/12. London: Greater London Authority (Data Management and Analysis Group); 2005.
20. Aspinall PJ. 'Mixed race', 'mixed origins' or what? Generic terminology for the multiple racial/ethnic group population. *Anthropology Today* 2009; 25(2): 3-8.
21. Aspinall PJ. Operationalising the collection of ethnicity data in studies of the sociology of health and illness. *Sociology of Health and Illness* 2001; 23(6): 829-862.

22. Vertovec S. Superdiversity and its implications. *Ethnic and Racial Studies* 2007; 30(6): 1024-54.
23. ONS. Information Paper. Deciding which tick-boxes to add to the ethnic group question in the England and Wales Census. London: ONS; 2009 (March).
24. OPCS/GRO(S). 1991 Census. Ethnic Group and Country of Birth. Great Britain. Vol. 2. London: HMSO; 1993.
25. OPCS & GRO(S). 1991 Census. Definitions. Great Britain. London: HMSO; 1992.
26. ONS. 2011 Census Special Populations Advisory Group. Paper SPAG(05)02. Taking forward the 2001 Census Community Liaison Programme: Draft Proposals and Timetable. London: ONS; 2005 (November).
27. ONS. The One Number Census – an estimate of the whole population. Available at: <http://www.statistics.gov.uk/census2001/onc.asp>. Last accessed 20th July 2010.
28. Tucker C, McKay R, Kojetin B, Harrison R, de la Puente M, Stinson L, and Robison E. Testing Methods of Collecting Racial and Ethnic Information: Results of the Current Population Survey Supplement on Race and Ethnicity. Statistical Note No. 40. Washington: Bureau of Labor Statistics; 1996.
29. Ndofor-Tah, C., Hickson, F., Weatherburn, P., Amamoo, N.A., Majekodunmi, Y., Reid, D., Robinson, F., Sanyu-Sseruma, W., Zulu, A. Capital assets: a community research intervention by The African Forum in Redbridge and Waltham Forest. 2000. London: Sigma Research; 2000.
30. ESDS Question Bank. People, Families and Communities Survey, 2001 (qn. R7). Available at: <http://survey.net.ac.uk/>. Last accessed 20th July 2010.
31. NHS Information Authority. Data Set Change Notice 16/2001. Data Standards: Characteristics. Birmingham: NHS Information Authority; 2001.
32. Aspinall PJ. The New 2001 Census Question Set on Cultural Characteristics: is it useful for the monitoring of the health status of people from ethnic groups in Britain? *Ethnicity and Health* 2000; 5(1): 33-40.

33. Finney N, Simpson L *'Sleepwalking to segregation'?* *Challenging myths about race and migration*. Bristol: The Policy Press; 2009.
34. Berthoud R. Defining ethnic groups: origin or identity? *Patterns of Prejudice* 1998; 32(2): 53-63.
35. Morning A. Ethnic classification in global perspective: a Cross-National Survey of the 2000 census round. *Popul Res Policy Rev* 2008; 27: 239-72.
36. McKay. *Cognitive Research in Reducing Nonsampling Error in the Current Population Survey Supplement on Race and Ethnicity*. Washington: Bureau of Labor Statistics & Bureau of the Census; 1996.
- Available at: <http://www.bls.census.gov/cps/racethn/1995/statcanp.htm>. Last accessed 20th July 2010.
37. Modood T, Berthoud R, Lakey J, Nazroo J, Smith P, Virdee S, Beishon S. *Ethnic Minorities in Britain: Diversity and Disadvantage*. London: Policy Studies Institute; 1997.
38. Rainford L. *Census Testing Programme: Report on the Ethnic Group and Religion Question Test Carried out in March 1997*. London: Office for National Statistics; 1997.
39. OPCS/GRO(S). *1991 Census General Report*. Great Britain. London: HMSO; 1995.
40. Platt L, Simpson L, & Akinwale B. Stability and change in ethnic groups in England and Wales. *Population Trends* 2005; 121: 35-46.
41. Simpson L., Akinwale B. Quantifying stability and change in ethnic group. *Journal of Official Statistics* 2007; 23(2): 185-208.
42. Georghiou T, Thorlby R. Quality of ethnicity coding in Hospital Episode Statistics: beyond completeness. Paper presented at Healthcare Commission seminar 'Measuring What Matters: measuring ethnicity in health data', 12th February 2007. London, King's Fund, 2007.
43. Bhopal R. *Concepts of Epidemiology*. Oxford and New York: Oxford University Press; 2002.
44. Aspinall PJ & Chinouya M. Is the standardised term 'Black African' useful in demographic and health research in the United Kingdom? *Ethnicity and Health* 2008; 13(3): 183-202.

45. Fitzpatrick J, Jacobson B, Aspinall P. Indications of Public Health in the English Regions. Vol. 4: Ethnicity & Health. London: APHO, 2005.
46. This was as a result of advice given by Professor R Bhopal, University of Edinburgh.
47. Nazroo JY. *The Health of Britain's Ethnic Minorities*. London: Policy Studies Institute, 1997.
48. Aspinall, P.J., 2007. Is it time to abandon colour categories for ethnic groups? *Journal of Epidemiology and Community Health*, 61(2), 91.
49. Agyemang, C., 2007 (29 January). Abandonment of colour categories for ethnic groups is long overdue (online response). *Journal of Epidemiology and Community Health*: Available at: http://jech.bmj.com/content/61/2/91.extract/reply#jech_el_1309. Last accessed 20th July 2010.
50. General Register Office for Scotland. 2006 Census Test Forms. Edinburgh: GRO(S); 2005. Available at: <http://www.gro-scotland.gov.uk/files/2006-census-test-form.pdf>. Last accessed 20th July 2010.
51. Scottish Government and General Register Office for Scotland. *Scotland's New Official Ethnicity Classification*. Edinburgh: Scottish Government & GRO(S); 2008.
52. Ligali Organisation. *Race to Ethnicity: The Classification of African Identity in Britain*. London: The Ligali Organisation; 2005.
53. Homes, A., Murray, L. *Cognitive Question Testing Scotland's Census Ethnicity Classification*. Research Findings No. 2/2008. Edinburgh: Scottish Government Social Research; 2008.
54. Krieger N, Williams D, Zierler S. "Whiting Out" White Privilege Will Not Advance the Study of How Racism Harms Health. *American Journal of Public Health* 1999; 89(5): 782-785.
55. Marmot Review. *Fair Society, Healthy Lives*. Strategic Review of Health Inequalities in England post-2010. London: Marmot Review, 2010.

56. Aspinall PJ and Mitton L. The Migration History, Demography, and Socio-Economic Position of the Somali Community in Britain. In *Somalia: Economic, Political and Social Issues*. Ed. Elias P. Hoffman. New York: Nova Science Publishers; 2010.
57. UK Parliament. House of Commons Hansard. Written answer of 14 July 2006. Col.2116W.
58. Aspinall PJ. Concepts, terminology and classifications for the “mixed” ethnic or racial group in the United Kingdom. *J Epidemiol Community Health* 2010; 64: 557-560.
59. ONS. Information paper. Recommended questions for the 2009 census rehearsal and 2011 census. London: ONS; 2008.
60. Siegel P, Martin E, Bruno R. *Language Use and Linguistic Isolation: Historical Data and Methodological Issues*. Washington: US Census Bureau; 2001.