



House of Commons  
Science and Technology  
Committee

---

**The Census and social  
science**

---

**Third Report of Session 2012–13**

***Volume II***

*Additional written evidence*

*Ordered by the House of Commons  
to be published 7 December 2011, 14 December 2011, and  
18 January 2012*

## Science and Technology Committee

The Science and Technology Committee is appointed by the House of Commons to examine the expenditure, administration and policy of the Government Office for Science and associated public bodies.

### Current membership

Andrew Miller (*Labour, Ellesmere Port and Neston*) (*Chair*)  
Caroline Dinenage (*Conservative, Gosport*)  
Jim Dowd (*Labour, Lewisham West and Pinge*)  
Gareth Johnson (*Conservative, Dartford*)  
Stephen Metcalfe (*Conservative, South Basildon and East Thurrock*)  
Stephen Mosley (*Conservative, City of Chester*)  
Pamela Nash (*Labour, Airdrie and Shotts*)  
Sarah Newton (*Conservative, Truro and Falmouth*)  
Graham Stringer (*Labour, Blackley and Broughton*)  
Hywel Williams (*Plaid Cymru, Arfon*)  
Roger Williams (*Liberal Democrat, Brecon and Radnorshire*)

The following members were also members of the committee during the parliament:

Gavin Barwell (*Conservative, Croydon Central*)  
Gregg McClymont (*Labour, Cumbernauld, Kilsyth and Kirkintilloch East*)  
Stephen McPartland (*Conservative, Stevenage*)  
David Morris (*Conservative, Morecambe and Lunesdale*)  
Jonathan Reynolds (*Labour/Co-operative, Stalybridge and Hyde*)

### Powers

The Committee is one of the departmental Select Committees, the powers of which are set out in House of Commons Standing Orders, principally in SO No.152. These are available on the Internet via [www.parliament.uk](http://www.parliament.uk)

### Publications

The Reports and evidence of the Committee are published by The Stationery Office by Order of the House. All publications of the Committee (including press notices) are on the Internet at <http://www.parliament.uk/science>. A list of reports from the Committee in this Parliament is included at the back of this volume.

The Reports of the Committee, the formal minutes relating to that report, oral evidence taken and some or all written evidence are available in printed volume(s). Additional written evidence may be published on the internet only.

### Committee staff

The current staff of the Committee are: Dr Stephen McGinness (Clerk); Jessica Montgomery (Second Clerk); Xameerah Malik (Senior Committee Specialist); Darren Hackett (Senior Committee Assistant); Julie Storey (Committee Assistant); Henry Ayi-Hyde (Committee Office Assistant); and Nick Davies (Media Officer).

### Contacts

All correspondence should be addressed to the Clerk of the Science and Technology Committee, Committee Office, 7 Millbank, London SW1P 3JA. The telephone number for general inquiries is: 020 7219 2793; the Committee's e-mail address is: [scitechcom@parliament.uk](mailto:scitechcom@parliament.uk).

# List of additional written evidence

---

	<i>Page</i>
1 Julie Selwyn (Census 01)	Ev w1
2 Dr Eldin Fahmy (Census 02)	Ev w1
3 Professor Ceri Peach Emeritus (Census 03)	Ev w3
4 Dr Peter King (Census 04)	Ev w4
5 C J Morris (Census 05)	Ev w5
6 John Stillwell and Oliver Duke-Williams, School of Geography, University of Leeds (Census 07)	Ev w7
7 David Owen (Census 08)	Ev w11
8 TNS-BMRB (Census 09)	Ev w13
9 Mike Hogan (Census 11)	Ev w14
10 Raj Bhopal (Census 12)	Ev w14
11 David Truswell (Census 13)	Ev w14
12 Dr Julie Fish (Census 14)	Ev w15
13 TWRI Policy and Research (Census 15)	Ev w15
14 Dr James Kirkbride (Census 16)	Ev w17
15 Institute for Fiscal Studies (Census 17)	Ev w19
16 Dr Stephen Patterson (Census 18)	Ev w20
17 Professor Edward Higgs (Census 19)	Ev w21
18 Centre for Longitudinal Study Information and User Support (Census 20)	Ev w22
19 Dr Nicola Shelton (Census 21)	Ev w23
20 British Society for Population Studies (Census 22)	Ev w24
21 Dr Jennifer Mindell (Census 23)	Ev w27
22 NatCen (National Centre for Social Research) (Census 24)	Ev w29
23 Tees Valley Unlimited (Census 25)	Ev w31
24 CURDS (Centre for Urban and Regional Development Studies) (Census 27)	Ev w33
25 Welsh Language Board (Census 28)	Ev w34
26 The British Library (Census 29)	Ev w35
27 British Academy (Census 31)	Ev w38
28 The Association of Business Schools (Census 32)	Ev w41
29 Institute for Jewish Policy Research and the Board of Deputies of British Jews (Census 34)	Ev w41
30 Suffolk County Council (Census 35)	Ev w45
31 The Salvation Army (Census 37)	Ev w47
32 Equality Commission for Northern Ireland (Census 40)	Ev w48

# Written evidence

---

## Written evidence submitted by Julie Selwyn (Census 01)

1. I am the Director of the Hadley Centre for Adoption and Foster Care Studies, School for Policy Studies, University of Bristol. I have been a researcher for over 20 years in the field of alternate care for children unable to live with their birth parents.

2. My research team has used the Census data to calculate the number of children growing up in the care of relatives (kinship care). The data has also been used to estimate the growth in this form of care and the characteristics of those carers and children.

3. Our analysis of the 2001 Population Census found that the number of children in such arrangements had more than doubled since the 1991. In 2001, one in every 77 children in the UK was living with a relative and many urban areas had a much higher prevalence (eg London Borough of Newham one in four children). We also found that children's carers were often elderly, in poor health/disabled, and living in poverty. The patterns of kinship care also differed by UK country.

4. UNICEF has recently described the growth in kinship care as a "crisis", caused in the UK by parental drug/alcohol misuse. We expect the 2011 census data to show another large jump in the numbers of children being cared for by relatives.

5. Without a Census it will no longer be possible to monitor these changing patterns of family care. Although the numbers of children in kinship care are significant, they become invisible in other types of surveys. They are a hard to reach group, especially carers and children from minority ethnic groups.

6. Government policy (eg reduction in child poverty, supporting families) often assumes that "family" means parental care. But Census data is showing that increasingly neither parent is present.

7. Without population data being available, this group of carers are invisible and their needs are not considered when changes to policy/welfare reforms are being considered by government. For example, a grandparent kinship carer (age 75 years) is extremely unlikely to be able to find work to lift the family out of poverty.

8. Without this data, Local Authorities will find it hard to plan services at a local level.

### *Research referred to:*

Nandy *et al* (2011). Spotlight on kinship Care. Exec summary available at [www.bristol.ac.uk/hadley](http://www.bristol.ac.uk/hadley)

### *Julie Selwyn*

Reader and Director of the Hadley Centre for Adoption and Foster Care Studies, School for Policy Studies, University of Bristol

10 November 2011

---

## Written evidence submitted by Dr Eldin Fahmy (Census 02)

1. UK census tract data are an invaluable source in facilitating understanding of social distinctions and social divisions at a small area level. They are also essential in ensuring that relevant area-based policy interventions are effectively targeted. Using synthetic estimation approaches my colleagues and I have generated, for the first time, methodologically consistent estimates of poverty and wealth at a small area level over the 1971–2001 period (Dorling *et al*, 2007; Fahmy *et al*, 2011). This work was well received and widely quoted within the UK print and broadcast media and was quoted verbatim by the then UK Prime Minister, Gordon Brown within the House of Commons. This research demonstrated the growth in spatial inequalities in wealth over this period and is therefore essential in informing policies intended to redress spatial inequalities in socio-economic outcomes.

2. This work extends earlier analyses conducted by these authors into the distribution of deprivation at a small-area level which have informed the development and targeting of area-based regeneration initiatives (Cemlyn *et al*, 2002; Gordon & Fahmy, 2004), and the targeting of public health interventions (Fahmy & Gordon, 2003). Using a comparable synthetic estimation approach, the authors have developed small-area estimates of fuel poverty in England for the then Department of Trade and Industry (Gordon *et al*, 2006; Fahmy *et al*, 2011), and in Wales for the Welsh Assembly Government (Gordon & Fahmy, 2008). Data of this kind are essential in the effective targeting of area-based initiatives to tackle fuel poverty and was recognised as such by the then Parliamentary Under Secretary of State for Energy, Lord Truscott who referred to this work as "indispensable in helping understand which areas of England are worst affected by fuel poverty".

3. However, the decision to axe the decennial population census beyond 2011 will substantially undermine the capacity of UK social science to analyse and understand social processes at a small-area scale. This will have serious consequences for the capacity of policy makers and service providers to effectively target area-based policies and services at those areas with greatest need. A number of alternatives to the traditional decennial population census have been proposed including a rolling census or short-form census. Both of these

may be capable of providing robust socio-economic and demographic data at a spatial scale appropriate to the needs of research users interested in understanding societal change and its implications for the effective delivery of social policies and social welfare (eg in welfare, health and well being, education, employment, etc). However, it remains unclear to what extent a rolling census would be able to adequately address and adjust for the consequences of change over time in estimates arising from this methodology. Whilst a short-form census allows for the continued collection of “key” socio-demographic data and the subsequent modelling of additional information based on a sample of more detailed questionnaires, the reduction in data quality associated with this approach is likely to be very substantial since the model fit in many areas is likely to be uncertain and (since the census is not designed for this purpose) quite imprecise.

4. However, the implications for the social science community and for research users of other alternatives to a more limited census (such as the use of sample survey and administrative data sources) are likely to be even more serious. Sample surveys may provide valid and robust socio-economic and demographic data for the purposes of estimation at regional level and local authority levels which can inform national policy and provision. However, by their nature (and even allowing for the combining of data from multiple sources) they are incapable of providing the research community, policy makers and service providers with the kind of detailed spatial disaggregation necessary to inform the delivery of services and policies at a small-area level. The development of techniques for combining data from various survey sources in order to create “synthetic” households, whilst undoubtedly an important innovation, will substantially increase measurement error and undermine the validity of inferences drawn on this basis at a small-area level.

5. Clearly, administrative data have important applications in understanding and responding to social, economic and demographic change and will continue to have an important role in informing policy and service delivery. However, it is also vitally important to acknowledge that such data are collected for very different purposes to those associated with the objectives of scientific enquiry and that they reflect the perspectives, priorities, practices and criteria associated with policy development and policy implementation. As such their scientific validity, the extent to which they genuinely measure what we think they are measuring, is in many cases highly questionable with regard to the objectives of scientific enquiry.

6. In conclusion, no other source provides the research community with robust and precise socio-economic and demographic data which can be disaggregated to a high degree of spatial resolution. Whilst it would be highly desirable to supplement existing decennial census data collection with more extensive mid-term estimates than is currently the case (eg based upon combining sample survey data sources), the decision to move away altogether from the “gold standard” provided by the UK census of population will have serious consequences for the UK social science communities capacity to understand social distinctions and social divisions and their effects at a small-area level—and for policy makers and service providers to respond to them through the promotion of appropriate area-based initiatives which promote social equity.

REFERENCES (available on request):

Dorling, D, Rigby, J, Wheeler, B, Ballas, D, Thomas, B, Fahmy, E. & Gordon, D (2007). *The Changing Prospects of Place*. Bristol: Policy Press/Joseph Rowntree Foundation. (ISBN 958 1861349958).

Fahmy E, Gordon D, Patsios D (2011). Predicting Fuel Poverty at a Small-Area Level in England. *Energy Policy*, 39(7): 4370–4377 (ISSN 0301 4215).

Fahmy E, Gordon D, Dorling D, Rigby J, Wheeler B, (2011). Poverty and Place in Britain, 1968–99 *Environment and Planning A*, 43(3) 594–617 (ISSN 1472 3409).

Gordon, D & Fahmy, E (2008). *A small area fuel poverty indicator for Wales*. Cardiff: Welsh Assembly Government.

Gordon, D, Fahmy, E & Patsios, D (2007). *Updating the Fuel Poverty Indicator for England*. Report to the Department for Trade and Industry. See also: <http://www.fuelpovertyindicator.org.uk/>

Gordon, D & Fahmy, E (2004). *Poverty and Neighbourhood Renewal in West Cornwall: A 2001 Census Update*. Final Report to West Cornwall Local Strategic Partnership.

Cemlyn, S, Fahmy, E & Gordon, D (2002). *Poverty and Neighbourhood Renewal in West Cornwall*. Final Report to West Cornwall Local Strategic Partnership.

Dr Eldin Fahmy  
School for Policy Studies, University of Bristol

10 November 2011

---

## Written evidence submitted by Professor Ceri Peach Emeritus (Census 03)

### PREFACE

1. The decision to abandon decennial censuses after the publication of the 2011 census is a cultural, academic and statistical disaster. Interestingly, for your committee, it could have a major and negative impact on Members of parliament as their numbers are reduced and their existing constituencies are dismembered. British censuses have been taken every 10 years since 1801. Hitler in 1941 is the only person to have stopped the regular cycle. The census is much more than a count of heads, though that function remains important in the allocation of funds to local authorities, voters to constituencies, assistance to regions facing particular social or economic problems.

### RACE ETHNICITY, RELIGION, MIGRATION, SEGREGATION AND GHETTOISATION

2. I am Emeritus professor of Social Geography at Oxford University. I have a vested interest, for the future of my country, that this amazing resource for our national good survives. My research deals with ethnicity, immigration, segregation and ghettoisation of ethno religious groups (Muslims in particular, but Catholic/Protestant relations in N Ireland as well). The British censuses as they have developed over the years since 1951 have delivered data on these topics which are the envy of other European countries. They are also of major importance in the understanding and governance of our country.

3. France for example, which is supposed to have the largest number of Muslims in Western Europe, has no way of verifying whether that is the case. There is no question on race, ethnicity or religion in the French census and the French suffer from that lack. If one traces back the origin of the estimates of five million Muslims in France, they turn out to be the fiat of a minister of the interior, (M Pasqua or M Sarkozy most recently) but without any demographic data to support them. Michelle Tribalat, a dedicated researcher at INED the National Demographic Institute searched through a huge sample of Census returns to locate likely Muslim names, the best available surrogate for such a calculation and produced an estimate of three million Muslims as opposed to the five millions cited by the Ministers. She was summoned by Sarkozy, then minister of the interior, and ordered suppress the number as “unacceptable”. In Britain in 2001, however, we could look up the census to see 1.6 millions of our own. The French however can only speculate, estimate and guess. As their cities banlieus (suburbs) turn into the American Black Inner city, they have no easy way of charting the change.

4. If, in future members of parliament wish to follow the French example and have critical data invented for them by Ministers you will find yourselves in quicksand.

5. The British 1991 census was the first to introduce a question on ethnicity. It greatly clarified the socioeconomic conditions of the minority ethnic populations in this country. It quantified the educational, occupational, domestic housing conditions of the population. Although discrimination against minorities had long been recognised as a general phenomenon, census data made it possible to quantify the extent to which discrimination was acting.

6. This was done through the work of Anthony Heath, Oxford University’s Professor of Sociology Heath calculated the degree of “ethnic penalty” for each ethnic group. An ethnic penalty is the degree to which members of a minority group with educational qualification equal to those of their white peers fail to acquire occupations of the same level. His technique was to use whites as the reference group, take samples from the census of ethnic and white groups with the same educational qualifications and estimate the proportion of those equally educationally qualified minority and white, who achieved jobs in the salariat (jobs that were well paid, had an incremental salary and pension, for example). He not only found the expected systematic ethnic penalty but some unlikely cases as well. For example, Indians as a whole had a socio economic profile as high as or higher than whites. However, although they performed well, they performed less well than they should have done relative to their qualifications. Despite doing well, it is clear that they were nevertheless suffering an ethnic penalty. At the other end of the scale, using avoidance of unemployment as the criterion, Heath found that Bangladeshis with no qualifications performed better than they might have been expected to do.

### THE LONGITUDINAL STUDY

7. One of the great advantages of the census has been its development of longitudinal studies. This vital research facility is doomed by the threatened demise of the census itself. To explain, the census is a snapshot gives a picture of the population on a particular date each year. We measure change by examining the differences between the populations on 10 year gaps. However, changes are not solely the result of the same people 10 years later: people are born; people die people emigrate and immigrate. It is theoretically possible to have identical numbers of people present at two census dates but for the populations to have totally different ethnic compositions. What we need therefore is a way of tracking what changes take place for a sample of individuals whose individual characteristics can be tracked from census to census.

8. The Longitudinal study selects a largish sample of individuals born on the same date of the year. These individuals’ census forms are retrieved after each census to establish the changes taking places of a constant panel (panel is of course refreshed a little at each census, to compensate for deaths births and migration.) The longitudinal survey produces a different picture from that of the comparison of two national. Censuses as an

example, the occupational profile of Caribbean men in London shows rather little change from 1991 to 2001, but longitudinal study shows upward mobility in jobs. The Longitudinal survey also shows a suburbanisation of the Caribbean population.

#### THE SAMPLE OF ANONYMISED RECORDS (SARs)

9. A final example of what will be lost with the demise of the Census is the Sample of Anonymised Records. Sars allow one to have very detailed analysis of family structures and relationships without compromising the identity of individuals in such households. It allows detailed examination of society in which the variables that are sought can be specified to a degree that sample surveys by polling agencies cannot be achieved.

10. The bottom line in this argument is that a continual nation census which has been run since 1801 is a national treasure that cannot be replaced by an analysis of Sainsbury loyalty cards. There is a moral her for the preservation of the census from the Climate change debate. The key data on climate change come from ocean sea temperatures. As it happens, techniques of measurements have been changed over time but without overlaying the techniques by which data were captured. Originally a wooden bucket was dropped over the side of a sailing ship pulled onto deck, a thermometer placed in it and the temperature recorded Later, canvas buckets were used and ships' sides were higher out of the water and both the height and the bucket material affected the temperatures which recorded. At each advance in ships and buckets, the conditions under which the temperature was measures was different but no controls were taken so that the measures using the old system were not overlapped with the new. It thus became impossible to establish to what extent changes in water temperature were due to warming or cooling of the oceans or the process of measurement. The experience of measuring small but critical temperature differences over a long period of time, should be a warning against dismantling the long continuous data collections of our population censuses.

*Professor Ceri Peach Emeritus*

MA D Phil Professor of Social Geography Oxford University and Visiting Professor Institute for Social Change University of Manchester

*11 November 2011*

---

#### Written evidence submitted by Dr Peter King (Census 04)

I am an economic historian, though mainly concerned with periods before there was a census. I have also been engaged on behalf of an NGO in commenting on planning documents at a local level.

I believe that it is important to continue having censuses, because they provide an objective basis for providing future estimates for planning and government expenditure.

It is of course true that there are many other sources of information available, but a great many of these are estimates based on samples. It is in the nature of such an estimate that it is only accurate if the sample is a representative one. Since no sample will ever be perfectly representative, any estimate will almost inevitably be imperfect. If a biased sample is used repeatedly, there is liable to be a drift in the estimate away from reality. It is thus important that the "clock" should be returned to zero periodically, by having an objective survey, such as is provided by having a census.

Examples can no doubt be multiplied of where public policy has been poor because it has been based on poor estimates:

- Government funding for Derry/Londonderry was based on the census data, but significant parts of the city were not surveyed by the census, because the IRA imposed a boycott on it. This was no doubt for reasons that were sound in their minds, but were counter-productive to the welfare of their supporters.
- Slough BC has repeatedly complained in recent years that it was getting inadequate support from government funding. It alleged that it had suffered disproportionately from migration from eastern Europe, but that estimates made by ONS had failed to pick this up.
- In my own parish, the Local Education Authority looked at ONS data on the birth-rate in the parish and suggested from this that the Primary School would suffer from falling rolls. In doing so, they ignored that (as an affluent area with good schools) the parish is always under great pressure from in-migration from the adjacent major urban areas. This often consists of parents who have already started a family while living elsewhere. What actually happened was that the Primary School had to be expanded by making provision for a third stream. Even with that the school is always full, and is liable to have to contract the extent of its catchment.

The following will illustrate the tendency for estimates gradually to drift towards error:

- The pre-census estimate for 2001 proved to have a significant error in it, because ONS (or a predecessor) had underestimated emigration, largely by the professional and managerial class. If there had been no census in 2001 to correct this error, the error would have been compounded to an extent that ultimately the estimates would have drifted so far from reality that they would be useless.

- Almost everyone is registered with a GP, but,
  - people registering with a new GP without their medical number are assigned a new one. When their medical records are transferred, the necessary cross-reference is made. This will mean that the number of people appearing to be registered will slightly exceed the number of patients. No doubt that is an error for which a correction could be made.
  - However, it is likely that the medical records of those emigrating will not be transferred. Unless this error can be accurately be estimated, it is likely that the number of medical records will exceed the population and that there will be a continual drift in this.
- Similarly, National Insurance Records are an unsatisfactory source for the size of the population, because it is notorious that one class of benefit cheats defraud the system by creating multiple identities. This is an unquantifiable problem.

As a historian, I observe colleagues obtaining useful historical data from the census. In the period before enumeration records survive, historical data on occupations for example is only scantily available, from directories, parish registers, and other less systematic sources. The result is that we know much less of that period. If the 2021 census is missed, not only will data not be available for government planning, but in 2121, it will not become available to historians.

*Dr Peter King*

*12 November 2011*

---

### Written evidence submitted by C J Morris (Census 05)

The Committee has sought written submissions on the following matters:

#### *How do social scientists use Census data?*

1. Social scientists (including those employed by government, either directly or on contract) use Census data for two main reasons.
2. Firstly, the Census allows a quick and definitive statement of the numbers and percentages in some defined group of interest eg the number of car-owning households by socio-economic group in Wedmore, Somerset.
3. Although available only once a decade, this figure is at least available. To compile the data from scratch would be virtually impossible by reference to official statistics—government (local or central) most likely have no estimate of how many households there are in Wedmore, much less their socio-economic status, and while the VLA might perhaps, given time, be able to work out the number of cars registered in Wedmore, they have no idea of the socio-economic status of the owners and wouldn't tell the researcher if they did.
4. Such a figure, once compiled, would, in any case, be a dead-end whereas the Census would allow further cross-reference to other factors such as households containing someone with limiting long-term sickness. The Census is thus an invaluable starting point for analysing almost any social issue where a relevant question has been asked in the questionnaire, even if the follow-up involves using some other technique or approach.
5. Secondly, it is possible to subject Census data to more detailed analysis. This may be done at aggregate level or (through anonymised data sets such as those made available by the Census and the various UK longitudinal studies—England & Wales, Scotland and Northern Ireland) at individual level. The relationships found are very robust, being based on large numbers, and they allow the analysis to sidestep the problem that not everyone living a poor area is necessarily poor (or as the Duke of Wellington famously put it “Being born in a stable does not make one a horse”). Areal effects can thus be distinguished from personal effects.
6. Variables of particular interest include age, sex, occupation, education, social status, car ownership, religion, ethnic group, health, much as in the Census. This needs to be analysed at both individual and household level eg a household containing both Catholics and Protestants is of particular social interest in a Northern Ireland context (for GB, read “from different ethnic groups”). In fact, there is much greater demand for variables to be included in the Census than there is space to accommodate them, even discounting income which was of considerable interest but proved unacceptable (and probably less than ideally accurate). Not all of these variables are necessarily of particular relevance to those assembling administrative databases.

#### *What impact will the ending of the Census have on social science research?*

7. The ending of the Census would more or less require social scientists to depend solely on survey data, their own or others. Although at present, they carry out their own surveys to supplement the Census, asking questions or pinpointing groups that the Census does not, the Census provides a benchmark against which to test the results. It is completely impractical, however, for any researcher to carry out a survey even remotely approaching the scale and coverage of the census.
8. Government administrative databases are rarely available to social researchers outside government (and they are not always made available to those within government, either). Computer systems designed solely to

facilitate immediate administrative needs (though they do not always achieve even this); reluctance to capture data not directly and demonstrably relevant to the primary function of the system; and of course, the need (real or perceived) for confidentiality all combine to make these a limited asset for research.

9. This is compounded by the fact that many government contacts with individuals at an address (possibly all save those dealing with benefit claims) disregard most if not all other individuals at that address. This loses the key social information on household structure, which would have to be imputed from some other source, probably a survey, or compiled by attempting record matching (and so far as that goes, is the Ms Smith recorded as living at the same address as Mr Jones, a previous occupant, a previous partner of Mr Jones, a current partner of Mr Jones or indeed now the Ms Jones recorded at that address?).

10. The absence of Census data to set a context results in a narrowing of perspective. This is extremely relevant to integrating social science research with policy making. Whilst qualitative research and small surveys may well demonstrate the existence of situations and relationships worthy of policy intervention, it is the scale of these which likely to be the determining factor in policy-making. Small problems attract less help than large problems.

*What alternatives to the Census would provide population and socio-demographic data of equivalent or higher quality?*

11. The current proposal to abolish the Census seems based on a rather naive view of the accuracy of data sources. Just because a computer record exists does not mean that it is accurate or indeed accessible. Any data source, the Census included, is subject to errors, and it is unrealistic to imagine that there is some panacea approach to data collection and data-matching that will yield error-free results. Different methods will yield different errors and different levels of error, so that it is simply a matter of deciding which is least unacceptable.

12. There are currently no existing reliable alternatives to the Census, and it is open to question whether an acceptable alternative could be constructed.

13. No survey can provide data of comparable quality. Apart from all the problems of non-response and the like:

- (a) Small surveys (ie up to an achieved sample size of say 25,000 or so) can provide very good data on large groups, but for smaller groups (under 10%) of the population, the effective sample is not large overall, and quickly becomes inadequate when disaggregated by eg region or sex or age.
- (b) Large surveys quickly become very expensive and yield ever decreasing benefits for increasing sample size. It is necessary to quadruple the sample, in order to halve the statistical error of the survey.
- (c) Very large surveys (eg the 1966 10% Census of Population in Great Britain) carry all the disadvantages of a survey, and are very much like a Census in the set-up costs. It was for that reason that the Northern Ireland government carried out a full census in 1966, calculating that the cost savings from a 10% census were not worth the reduction in data quality.

14. Some of the bigger existing surveys produce estimates expressed in total numbers eg X million in poverty or Y million unemployed. This should not, however be taken as an alternative to the Census. The survey results have been subjected to a complicated grossing process dependent on midyear population estimates that derive ultimately from the most recent Census. They are not, therefore, independent sources.

15. Theoretically, a population register holding a common identity number and basic demographic information could be cross-matched with one or more other data sources where the identity number was held, to produce an *ad hoc* database which would fulfil the purposes of a census. This, however, would depend on:

- (a) The establishment and maintenance of a universal population register—the Government has abandoned the identity card project for a variety of good reasons, and it would require a U-turn of policy to produce a register.
- (b) The modification of existing databases to capture and hold the common identity number.
- (c) The accuracy and timeliness of existing databases—many databases are well out of date, because of the infrequency of interaction with the subject and the failure to update them even when changes are notified. Address changes are a crucial issue, since there is a high level of mobility among the poorest members of the community.
- (d) Exclusion of data which although validly held on the subsidiary database, has no place on the compiled database eg lock-up garages are a valid part of the Northern Ireland Valuation List, but form no part of the habitable dwellings which might be occupied by the NI population. It will be noted that there might be justified resistance to adding a common identity number to property details—the owner and the occupier can be different people.
- (e) The implementation of a massive computerised data-matching capability—current evidence eg the NHS experience of integrated databases, suggests that Government lacks this capability and will continue to do so for the foreseeable future.

16. Such an *ad hoc* database would of course have to be set up every time that a research project was required, involving substantial set-up costs each time. The Census database, once set up, is extremely cheap to analyse in response to *ad hoc* needs. This suggests that the most likely approach would be a kind of “ersatz Census” where once a decade, a database would be set up by data matching and used in a similar way to the Census thereafter.

*What other existing sources of population and socio-demographic data could be improved upon?*

17. All of the existing sources of population and socio-demographic data could be improved, but not perfected, if government were prepared to spend the additional money required. This does not mean that they could be employed as a substitute for the Census.

18. When the 1981 Census of Population in Northern Ireland was the target of a moderately effective campaign of deliberate non-enumeration, it proved possible to arrive at a robust estimate of the level of non-enumeration, using alternative sources of information (as evidenced by the good match between the mid-year estimate and Census in 1991). The alternative sources would have been useless as a substitute for the Census.

19. The alternative sources (largely the school Census and past midyear estimates) were partial in coverage, lacking in the range of information and household data required, not aligned to the Census date and sometimes out of date. These problems would be encountered in any alignment of alternative data sources.

20. The Committee may wish to recommend that Office for National Statistics consult with Northern Ireland Statistics and Research Agency, as Northern Ireland has some past experience in dealing with seriously flawed Census returns.

#### DECLARATION OF INTEREST

21. I am a social researcher and statistician, working as Ulaidh Research Consultancy since my retirement from the NI civil service, where I worked as a statistician for 31 years.

*C J Morris*  
Ulaidh Research Consulting

17 November 2011

---

#### **Written evidence submitted by John Stillwell and Oliver Duke-Williams, School of Geography, University of Leeds (Census 07)**

#### DECLARATION OF INTEREST

The authors are both academic researchers funded by the Economic and Social Research (ESRC) Census Programme to provide users in the UK academic community (academic staff, research staff and students) with online access to Census and related interaction data via a Web-based Interface to Census Interaction Data (WICID) maintained by the Centre for Interaction Data Estimation and Research (CIDER). Both authors have longstanding research interests in migration and commuting behaviour, extensive experience in handling interaction data sets and publications in a range of international social science journals. The submission is therefore focused on the outputs of the Census generally referred to as “Interaction Data”.

#### *(1) How do social scientists use census data?*

1. Moving house is an event that most people experience at some stage during their lifetimes whereas going to work or to study is part of daily life for the majority of the pre-retirement population living in the UK. Censuses provide extremely valuable information about volumes and compositional structures of both these types of human behaviour including details of the origin and destination locations involved.

2. Data from the 2001 Census show us that migration and commuting are remarkably important phenomena in contemporary times, as they have been throughout modern history. Over 6.2 million people in the UK changed their place of usual residence in the 12 month period before 29 April 2001, representing around 10% of the total population of the UK. In addition, a further 467,000 immigrants arrived from outside the UK in the same period. In comparison, over 28.3 million people were recorded by the 2001 Census as commuting to work on a daily basis, around two fifths of the total population, although 2.1 million flows involved people whose home and workplace locations were the same.

3. Data on migration and commuting are available from various census products including aggregate statistics (various sets of tables), Special Migration Statistics (SMS), Special Workplace/Transport Statistics (SWS, STS), cross-sectional micro data sets and longitudinal studies. The SMS and SWS/STS, known collectively as the “interaction data”, are particularly important because they provide information about directional flows between origin and destination areas. This means that, depending on the geographical scale, the data sets may involve very large yet sparsely populated matrices.

4. Both migration and commuting have transformed societies and influenced how settlements and landscapes have developed over time. The role of the Census is providing the data to understand and analyse these transformations in the UK has been fundamental. In the 1970s, research attention turned from suburbanisation to new processes of population redistribution such as counterurbanisation, the movement of individuals and families away from major metropolitan centres and down the urban hierarchy to increasingly rural areas, as documented by a number of geographers.

5. Whilst the process of suburbanisation usually involves breadwinners whose place of work before and after migration remains the same, counterurbanisation includes those migrating away from the cities, severing their commuting to work ties and adopting a different style of living. In some cases, these individuals are those deciding to work locally or from home; in other cases, they involve those seeking unconventional lifestyles; the pioneers of the counterurbanisation movements were actually those reaching retirement age who, on becoming economically inactive, no longer required to maintain ties with a place of work.

6. Research based on migration data from the last three censuses indicates how the major losses from the big conurbations have continued in the 21st century, causing policy makers to fear implications that urban neighbourhoods are being abandoned whilst rural areas are continuing to come under increasing house-building pressures. However, research on the composition of migration streams has revealed important differences in sub-groups of the population with those in the high mobility age ranges, the late teens and twenties, showing patterns of movement away from rural areas and into cities that run counter to the net flows of families and the elderly in the opposite direction.

7. In addition to analysing the spatial variations in the role internal migration *vis a vis* the natural change component of demographic restructuring, migration research has increasingly focused on immigration flows and the ethnic minority communities that have established since the initial waves of newcomers arrived from the colonies in the post-war period. The addition of an ethnic question to the Census in 1991 has allowed studies of ethnic populations in the UK to flourish and the last decade has seen the emergence of a body of studies on ethnic migration—suggesting that, contrary to popular opinion that ethnic minorities in British cities are becoming more segregated, there is considerable evidence to indicate processes of spatial deconcentration are taking place, particularly in London, where the majority of ethnic minority populations are located.

8. Recent studies of commuting behaviour have demonstrated variations in patterns of mobility by demographic characteristics but also by people in different occupations or socio-economic categories. The results of the Census question on the mode of travel to work is particularly well used by researchers as it is by transport planners seeking to formulate solutions to the critical problems of transport congestion that afflict almost every city and small town across the country at certain peak hours of the day. Geographers have used commuting flows to construct self-contained labour market areas that reflect functional characteristics and the Census SWS have been used to define the official Travel-to-Work Areas (TTWAs) used by various Government Departments.

9. Research studies based on census interaction data have been carried out by researchers from various academic disciplines using a range of analysis methodologies and modelling techniques. Alongside the empirical analysts interested in flow composition and pattern identification, there are regional scientists or quantitative geographers who seek to model interaction behaviour, in aggregate or individual form, using statistical or mathematical modelling. This type of research has tended to focus either on deterministic approaches that seek to establish the causes of interaction behaviour or on predictive approaches that result in very practical applications: a new road is recommended that will reduce commuting congestion or new housing is created in response to the pressure of demand from potential in-migrants.

10. Internal and international migration statistics from the Census play a critically important role in the production of annual population and household estimates which underpin the provision of housing and services. They are used in the estimation of disaggregated flows of migrants (eg flows between regions by ethnic group or economic activity based on administrative sources), they are used by researchers to validate alternative sets of data from administrative or survey sources, and they are used as input data for projection models that generate better estimates of future populations.

11. The following text provides a number of case studies of the use of migration and commuting data, as well as information about the UK Census interaction data sets:

Stillwell, J, Duke-Williams, O and Dennett, A. (eds) (2010). *Technologies for Migration and Commuting Analysis Spatial Interaction Data Applications*, IGI Global, Hershey, pp 357.

(2) *What impact will the ending of the Census have on social science research?*

12. The migration and commuting data that are derived from the Census have a number of strengths. These include the spatial detail with which results are published, the comprehensiveness of coverage and the level of attribute detail (the socio-demographic characteristics of the persons or households involved) attached to each flow. In addition, they share the general authority of the Census as a high quality source of reliable data.

13. The ending of the Census will therefore have a pronounced impact on research on migration and commuting behaviour from a geographical perspective because it is very unlikely that there would be any alternative mechanism or combination of methods that would provide the attribute detail for small areas such

as output areas or super output areas across the whole country. That said, it should be acknowledged that the interaction data from the 1991 and 2001 Censuses were rendered considerably less useful by the suppression and small cell adjustment methods (SCAM) that were imposed on the data in both years in order to minimise the risk of disclosure and ensure the data remained confidential. The effects of SCAM effectively destroyed the utility of the flow matrices at output area level for England and Wales in 2001 and it is heartening to note that ONS will not be using this method of statistical disclosure control to adjust the data in 2011.

14. Claims have been made by ONS that the 2011 Census has been very successful in terms of the response and coverage. The accuracy of the results remains to be seen but if this is the case, then these results will form the basis of many research projects into the future—since it will be the last comprehensive count of the nation’s population disaggregated by such a large number of socio-demographic characteristics. There is no other data source that indicates migration flows by ethnicity, religion or health status, for example, and there is no equivalent source of information on commuting flows by demographic attributes, occupation or mode of travel. We can expect to see the data from the last Census being accessed and used by social science researchers with greater intensity and for a longer period of time than any previous Census.

15. However, the lack of another census will also act as a driver for the search for and exploitation of other existing administrative data sources as well as the stimulus for new data sets to be collected through mini-censuses or surveys and the improvement of existing surveys (such as the International Passenger Survey). It will encourage much greater use of the existing cohort and longitudinal studies and will drive the effort to link data from different sources together to provide “linked data”. A raft of new data sets are being released by Government Departments and other agencies that will be of enormous interest to the social science community if agreements can be reached for the data suppliers to provide social science researchers with access to these data sets. The “Beyond 2011” Programme is a good example of a context in which new and valuable data sets are being assembled and used by ONS. The Economic and Social Research Council must take a pro-active role in reaching agreements with data suppliers to disseminate these data sets, taking advantage of the open government licensing arrangements that are now in place.

*(3) What alternatives to the Census would provide population and socio-demographic data of equivalent or higher quality?*

16. In the case of migration, there are a number of alternative sources which can or could provide useful data. These are arguably of “higher quality” in terms of their timeliness and extent, but they are of less utility than Census data in a number of other respects, including coverage and level of detail.

17. Data from NHS administrative sources have been made available for the period since the mid-1970s. The structure and nature of these data have changed over the course of time, reflecting changes in recording and data processing technologies, as well as changes in the administrative framework of the NHS. The current “NHS derived” data series are based on comparisons of patient registers at different points in time, with patients being identified as migrants if their postcode has changed between the start and the end of the observation period.

18. The Patient Register data are currently published on an annual basis, using a mid-year to mid-year reporting period. A variety of tabular summaries are produced detailing characteristics of patients involved in various types of move; for the purpose of comparison with the Census flow data there is one table of interest, which shows the number of persons migrating from specific origins to specific destinations, using a reporting geography of local authority areas in England and Wales. The table shows the total number of persons moving between (or within) locations, rounded to the nearest 10.

19. These data are of great importance in monitoring and/or estimating migration levels within the UK in the intercensal period. There are some known shortcomings, primarily that certain population groups—young adult males in particular—are less likely to notify their GP of a change of address, and thus will not be captured as migrants; however this weakness is understood, and is common to many other administrative sources. However, the data only show the number of persons or, at best, the number of persons by age and sex. They do not include additional socio-demographic information of the sort common in Census data, such as individuals’ religion or ethnic groups, their marital or family status, their economic status or occupation.

20. International migration data are generated from a number of sources, including the International Passenger Survey, and from the National Insurance Number database. Immigration is a politically sensitive issue, and all data sources including the Census share—to a greater or lesser extent—the weakness that they are avoidable by those who wish to avoid enumeration.

21. The International Passenger Survey (IPS) is a sample survey which includes visitors (both long and short term) to the UK, as well as UK citizens leaving the UK (both long and short term). Unlike Census data, flows published from the IPS are disaggregated by visitors’ reasons for entering the UK (holiday, business etc); however the amount of socio-demographic information is very limited and the sample size remains relatively small.

22. The Department for Work and Pensions publish annual aggregates (on a fiscal year basis) of the numbers of new National Insurance Numbers (NINo) allocated to foreign adults entering the UK to work. These are disaggregated by nationality (as opposed to origin country). These are disaggregated by local authority of

residence. Information is also published regarding the age and sex of persons involved, but only at a national level. The NINo data are a valuable source of immigration information, but are limited in their socio-demographic characteristics.

23. Alternative sources of commuting data are very limited; this is an area of acute concern to those who have used the Census commuting data for research. The National Travel Survey (NTS) is collected by the Department for Transport, and has been running on an annual basis since 1988. It is a sample survey, and in 2009 collected data from over 8,000 households. There is very limited spatial information in the NTS, with disaggregation by Government Office Region or “area type” only. Even at this very coarse scale, there is no disaggregation by origin, destination and mode of travel. The essential reason for this is the small sample size, which does not permit the publication of detailed disaggregated flows.

24. The Department for Transport also publish annual average daily traffic flow data, for traffic passing fixed points on motorways and A roads. These provide important information about traffic levels, but cannot provide information about either the origins or destinations of trips, or about the individuals making those trips.

*(4) What other existing sources of population and socio-demographic data could be improved upon?*

25. The NHS Patient Register data are captured at the postcode scale, and thus could potentially be published at more detailed geographic scales than is presently the case. There is only limited scope to include additional socio-demographic information about migrants; however we would welcome the publication of aggregate observations by age and sex on a geographically detailed basis.

26. The lack of socio-demographic data in the Patient Register limits its scope for use in social science research: disaggregation of internal migration data by characteristics such as ethnicity or family status often reveals significantly different patterns to those suggested by counts of total persons alone.

27. The Schools Census—known as the Pupil Level Annual School Census (PLASC) prior to 2007—records a variety of information about pupils in state education in England. In addition to academic performance information, the data also include student’s home addresses. Comparisons of these over time can be used to generate migration matrices for schoolchildren. Access to data from the Schools Census is currently limited, especially and necessarily to those fields—such as home address postcode—which act as personally identifying. However, it would be possible for aggregate observations to be produced showing the migration of school children.

28. The Schools Census also has the potential to generate a regular series of origin-destination matrices disaggregated by pupils’ modes of transport (walking, bus, car passenger, etc) to school. Data of this kind were produced as part of the output of the 2001 Census in Scotland (and will again be produced from the 2011 Census), but differences in questionnaire wording has not permitted the production of these data elsewhere in the UK. These data would thus deliver information in England and Wales that are not available from the Census.

29. The Higher Education Statistics Agency (HESA) record information for students at Higher Education Institutes, including the postcode of the student’s address (usually, their parental address) prior to entry to an HEI. These could be used to produce migration matrices showing the movement of students to universities and colleges.

30. There are no obvious alternative sources of data on commuting. The most widely used attribute of the Census commuting data is the mode of transport to work, and it is this information which tends to be missing from candidate data.

31. The Annual Survey of Hours and Earnings (ASHE) includes information about employees’ workplace and home postcodes, but it does not include any socio-demographic information about the individual, nor does it include information about the usual mode of transport to work.

32. ASHE is captured through a questionnaire sent to employers (rather than employees) and thus there is no scope in the current collection model to collect any additional personal data, such as mode of transport. Even if this were to be the case, the small sample size (1%) means that no detailed spatial disaggregation could be made.

*John Stillwell and Oliver Duke-Williams*  
School of Geography, University of Leeds

22 November 2011

---

## Written evidence submitted by David Owen (Census 08)

### THE CENSUS AND SOCIAL SCIENCE

#### 1. *How do social scientists use Census data?*

1. The Census of Population is the most comprehensive source of data on the demography, economic circumstances and housing of people living in the UK. It achieves a vastly higher response rate than any other social survey and unlike all the administrative sources, permits social, demographic, economic and housing variables to be cross-classified. It is the most detailed source of data on internal migration and commuting.

2. UK social scientists have benefited greatly from the purchase of the whole of the electronic outputs of the Census of Population since 1981 via ESRC/JISC and latterly Treasury funding. The Treasury purchase has permitted a vast increase in the use of Census data in the private and voluntary sectors as well as the public sector and promoted a great improvement in knowledge of the characteristics of the population, more widely spread across society.

3. The disciplines which make most use of the Census are probably Geography and other closely-related subjects with a spatial perspective, such as Town Planning and Regional Science. However, spatial analysis and hence use of the Census is increasing in Economics and Sociology.

4. Geographers use Census data to analyse the characteristics of localities at a range of spatial scales. A wide range of analyses have been undertaken, including exploration of socio-economic differentials, the geography of ethnic composition, geographical patterns of demographic change and migration and the creation of classifications of areas. The latter have been developed by the commercial sector to create powerful geo-demographic classification systems used in credit scoring, insurance and marketing and for predicting individual consumer behaviour. The Census has been an important source of denominators in the creation of indicators of local disadvantage using more regularly up-dated information sources (indeed it is one of the only sources of information on the economic activity of the population at the local level, which is required in calculating small-area unemployment rates). It has also been extensively used as contextual background data in qualitative research on local areas.

5. The analytical tools applied to the Census include regression models, cluster analysis, spatial statistics and the application of Geographical Information Systems to create profiles of areas within distance bands of physical features and within travel time bands. Mapping and graphical display have been important ways of presenting Census data. Census “flow” data (eg journey-to-work data) has been used to delineate functional urban regions and city regions, which identify the hierarchy of cities and towns. This is not apparent from any other source of information. Small-area Census data has also been used to identify the geographical “segregation” of ethnic groups and explore the factors behind this. The detailed geography of the Census is used to link Census of Population data with a wide range of other data sets to enrich the analysis undertaken.

6. The Census of Population is under-used by UK social scientists. The reasons for this are three-fold: (i) the data is collected only once every ten years and hence the potential for time-series data analysis is very limited; (ii) most of the data is in the form of pre-designed cross-tabulations for geographical areas and hence most analysis involves exploring ecological associations; (iii) other data sources exist which are collected more regularly, contain more variables and are available at the level of individuals (ie ONS sample social surveys such as the Labour Force Survey).

7. On the other hand, the recent UPTAP Initiative funded by ESRC centred around the award of (mostly) small grants for the secondary analysis of quantitative data sets yielded a large amount of research activity around the analysis of the 1991 and 2001 Censuses and yielded a number of new data sets which open new possibilities for analysis, created by linking the Census with other data sources.

8. The availability of microdata from the Census (in the form of the Samples of Anonymised Records) has permitted more in-depth analysis of Census data because the relationships explored are not constrained by the pre-specified tables used for delivering the bulk of the census output. Regression models can identify the separate influence of geographical location and socio-economic characteristics on a range of phenomena ranging from labour market participation to health.

9. The Longitudinal Studies in England and Wales, Scotland and Northern Ireland provide a powerful tool for tracing the evolution of individuals and families over time and have yielded new insights.

#### 2. *What impact will the ending of the Census have on social science research?*

1. The impact of the ending of the Census of Population on UK social science research is entirely related to what will replace it.

2. The Census is unique in the combination of geographical detail and information about individuals and households which it yields. The 2011 Census includes new questions which allow more detailed analysis of the migrant population, identity, the impact of student populations and the phenomenon of weekly commuting.

3. If the alternative to the Census yields the same range of data, for the same geographical detail and at greater frequency, then the data environment for social scientists will have been improved by abandoning the Census.

4. However, it is more likely that the replacement will be a series of partial products—such as enhanced population estimates, more localised data on housing, economic activity and health. It will be difficult to link these together and the problem of only having ecological associations between variables and not direct causal associations to analyse will be intensified.

*3. What alternatives to the Census would provide population and socio-demographic data of equivalent or higher quality?*

1. The UK is hampered (both in terms of research and administration) by the lack of a population register. Without a comprehensive register with a system for updating when individuals move residence, there is no definitive measure of the population and migration is estimated from data sources which were not intended for this purpose. Population registers are not ideal, because they cannot achieve complete coverage and it cannot be ensured that every instance of migration is recorded. Additionally, obtaining socio-economic information for the people on the register still requires large-scale surveys and/or the linkage of official data bases using the individual registration code number (as occurs in Scandinavia).

2. In the UK, the linking of administrative and commercial data bases to yield Census-like data (which has been suggested by Francis Maude) is difficult because no unique personal identifier is used in the collection of data.

3. The nearest equivalent is the National Insurance number which is allocated to nearly everybody for purposes of collecting tax and payment of benefits and is allocated to new migrants when they seek work. The DWP maintains the “Lifetime Labour Market Database” which can trace individuals over time and across space. This could be used to identify families and their dynamics. However, the NI number has problems—it is not necessarily unique and people leaving the country do not have to notify anybody that they have done so, so there will be numbers which are inactive or used fraudulently. Though the postcode provides location information, there is no method for linking information about the characteristics of the dwelling an individual is living in, unless the Land Registry and Electoral Register also require NI numbers. These sources would also have to be regularly updated with information about the dwelling in which an individual lived in order to yield data on overcrowding and the presence of amenities (eg central heating, exclusive use of bathroom).

4. A database built around linking the tax, benefit and other systems does have the potential to yield much more regular information on travel-to-work than the Census. However, in order for systems like this to yield information on ethnicity and other identity variables (such as religion), each person with a NI number would have to be required to report this information at each contact with the government, since these can change over time.

5. Retailers and credit scoring companies hold large amounts of geographically detailed information on the spending patterns and use of financial products of individuals. However, most of the transactions captured would not involve the same individual identifiers used in contact with the public sector and matching would therefore probably depend upon the accurate recording of residential address. There is also the problem of how transactions for multiple credit cards and store cards held by individuals would be combined, especially when personal details recorded differed slightly.

6. While record linkage could yield data on a more frequent basis than the Census of Population, it is not clear that the quality would be equivalent or better. This is because the collection of data is based on administrative processes in which there is less control over the quality and consistency of the information collected. In some instances, information is provided by an individual and in others is provided by an official on behalf of an individual.

*4. What other existing sources of population and socio-demographic data could be improved upon?*

1. The other sources of population and socio-demographic data in the UK are very limited and should all be improved, whether the Census of Population continues or not. The most reliable element of the UK population data is the registration of births and deaths which are very accurate and form the basis of the annual estimates of population made by the ONS and the Registrars-General of Scotland and Northern Ireland. It would be possible to add more questions about the parents or the deceased. Attempts have been made to add an ethnic group classification to birth registration, but this is difficult because the concept of ethnicity adopted by official statistics is a social construct and cannot be applied to a new-born child. Recording each parent’s ethnicity might be preferable, but this does not necessarily indicate the ethnicity of the child and may not be possible where the father is absent.

2. Most of the information on the socio-economic characteristics of the population between Censuses is derived from ONS social surveys. The main drawback of these sources is the small sample size, which means that data for small geographical areas cannot be generated without potentially risking disclosure of data for identifiable individuals. Much larger sample sizes would improve this situation, but the costs involved in data collection would be substantial.

3. All surveys suffer similar drawbacks to the Census in not asking all the questions that are potentially of interest to data analysts while becoming regarded as intrusive by a section of the population which will not co-operate. Hence all social surveys are experiencing declining participation rates and are biased in (often unknown) ways. The 2011 Census made great improvements, but this involved substantial expenditure to achieve. However, it was undoubtedly cheaper in terms of expenditure per head than each wave of a social survey.

#### DECLARATION OF INTERESTS

I have no commercial interest in the Census or connection with suppliers of Census data. Much of my research work is highly dependent upon the availability of Census of Population data.

*David Owen*  
University of Warwick

24 November 2011

---

### Written evidence submitted by TNS-BMRB (Census 09)

#### BEYOND 2011: AFTER THE CENSUS

##### SUBMISSION:

##### 1. *How do social scientists use Census data?*

We work in the field of social surveys and use census data in four main ways: (i) to assess and correct for non-response bias in survey samples, (ii) to pre-classify neighbourhoods and ensure that each sample of neighbourhoods is as representative as possible, (iii) as additional context data added to interview data, and (iv) to assist in estimating key survey statistics at the neighbourhood level (direct estimates from survey samples are unreliable at this very granular level).

##### 2. *What impact will the ending of the Census have on social science research?*

##### 3. *What alternatives to the Census would provide population and socio-demographic data of equivalent or higher quality?*

The answer to these questions will depend entirely on what it is replaced with. All survey data would be weakened if neighbourhood level data was unavailable but the census is not the only possible source of neighbourhood data. Administrative data of various types is already available at local levels but there are problems with it. Firstly, there is a lack of full linkage between these sources as well as an inconsistent approach to updates. Secondly, there are considerable difficulties in accessing the data in a usable form. This means that administrative data is considerably more awkward to use than census data and, for the most part, less reliable (even if more up-to-date). Furthermore, administrative data is inevitably purpose-specific. There is no overarching data strategy to ensure that key evidential gaps are filled.

However, it is not impossible to reduce these problems to a manageable level. There is no census in the Netherlands but there *is* a reliable database of citizens based on a range of linked administrative data sources. Neighbourhood level data can be extracted from this and used widely and the database is also used to draw samples for government and academic-sponsored surveys. In Britain, the uninformative, if comprehensive, postal address file is used as the sample frame for high quality general population surveys. A reliable administrative database would represent an improvement, especially with regard to correcting survey non-response bias.

Census micro-data is currently only available on a 5% basis and only for specific purposes. This is rarely used in survey research because (a) it is impossible to link specific address-level data to survey sample data (for reasons of confidentiality and due to the 1-in-20 nature of the sample), and (b) even a 5% sample (c 1 million households) is too small for micro-level analysis of specific neighbourhoods. Unless a way can be found that allows micro-level analysis of specific neighbourhoods the absence of the 5% micro dataset will not greatly affect survey research.

If the census is absent, it would be valuable if the census 2011 figures could be updated in a reliable way, even if that means replacing actual data with part-modelled data. The Annual Population Survey—as well as many other surveys—provides reliable national and regional level updates of census 2011 data (albeit with a greater level of uncertainty due to the much higher level of non-response than attends the census) but not at sub-regional level.

However, the APS data could be *modelled* at the neighbourhood level given a strong enough base of administrative data. For our primary purpose—the accurate classification of neighbourhoods—modelled estimates of this type would suffice in place of real data, although there is a risk that these estimates will become more and more unreliable as time passes without the decennial anchoring role of the census.

Finally, one possible option is to make the Annual Population Survey mandatory as the census is at present. That would go some way to reducing non-response bias for national and regional estimates. Increasing the size of the APS might also allow for reliable updates at local authority level too (currently estimates are only considered reliable if three-years of data are aggregated together). Modelled estimates would then be required only for neighbourhoods rather than for larger geographies.

*Joel Williams*  
Head of Survey Methods  
TNS-BMRB

*24 November 2011*

---

#### **Written evidence submitted by Mike Hogan (Census 11)**

##### **MAINTAIN THE CENSUS**

In essence, the census tells us important things about the changing face of our society.

We will only really know how important and informative it when it is gone.

So lets not axe it.

The demographics of the UK a re changing more rapidly and in complex and unpredictable ways.

The census is more important than ever.

*Mike Hogan*  
Integrated Commissioning Manager  
Adult Mental Health

*29 November 2011*

---

#### **Written evidence submitted by Raj Bhopal (Census 12)**

Colleagues, without the census, epidemiological research and surveillance will be seriously dented, and much of it will become untrustworthy for lack of reliable denominators for the calculation of disease and death rates. Public health activity, both surveillance and evaluation of policy, will be undermined. We will also find it very difficult to undertake reliable linkage studies that can inform us on population health, with appropriate adjustment for potential confounding factors. The only serious alternative is to have a population register and that is likely to be even more expensive than the Census. Minority populations will be disadvantaged, as information on them tends to be particularly sparse, and the census is an essential resource.

At something like a few pounds a head the census is very good value, especially when compared to other forms of social survey, which costs about £100 per head. I am hopeful that your consultation will lead to a reappraisal of the general view that the census is no longer necessary.

*Raj Bhopal, R S Bhopal, Bruce and John Usher*  
Professor of Public Health  
Edinburgh Ethnicity and Health Research Group  
Centre for Population Health Sciences  
University of Edinburgh

*29 November 2011*

---

#### **Written evidence submitted by David Truswell (Census 13)**

##### **CENSUS—IMPORTANCE OF POPULATION DATA IN SOCIAL SCIENCE AND HEALTH CARE**

While there are well known limitations to the accuracy of Census Data it still remains one of the internationally best regarded studies of detailed population demographics. I have used Census data adjusted by ONS published on annual projected changes in my work in London on mental health related access issues for programmes of work with a national profile.

From 2005 to 2008 I led a programme of work on improving mental health seevices for people from Black and Minority Ethnic backgrounds. This was the Focussed Implementation Site Programme that was part of the DH Delivering Race equality Action Plan. I worked in a mental health trust that provided services across five boroughs in West London.

Then from 2009 to 2011 I was the lead for a pan-London programme of work for Commissioning Support for London supporting commissioners in improving services for people living in London in line with the National Dementia Strategy.

I have found Census data an invaluable support to cross-reference and either confirm or challenge other information sources in support of improved services, particularly with improving access for Black and Minority Ethnic Groups. One specific areas that the census has been useful for me has been, comparisons between electoral ward level census data, GP referral rates to mental health services of people from Black and minority ethnic groups and mental health ward admission for the same population. I was able to demonstrate that in Brent and Westminster the mental health services were not admitting disproportionately high numbers of these populations. A summary of this work has been published and I have presented it at both national and international conferences.

Another area of my work has been the underestimation of the proportion of people with dementia from Black and minority ethnic populations in Central of Census data, and dementia prevalence projections. This work has recently been published and I will be presenting it at a national conference in the next few days.

These findings have significant impact on planning of services and investment decisions by commissioners and are an essential contribution to a reliable evidence base for these decisions.

*David Truswell*

Senior Project Manager Central and North West London NHS Foundation Trust

29 November 2011

---

#### Written evidence submitted by Dr Julie Fish (Census 14)

The census provides valuable information that is not available from other sources. It collects baseline data that I draw upon in my research, teaching and writing.

*Dr Julie Fish*

Reader in Social Work and Health Inequalities  
De Montfort University

29 November 2011

---

#### Written evidence submitted by TWRI Policy and Research (Census 15)

1. TWRI Policy and Research (TWRI) appreciates the opportunity to write to the Science and Technology Committee about the value of the Census to Social Scientists. We do this in the context of a number of years<sup>1</sup> experience analysing, manipulating<sup>2</sup> and subsequently presenting statistics for Tyne & Wear and parts thereof.

TWRI was established in 1986, following the closure of the Tyne & Wear county council, to support the five Tyne & Wear district councils through the assembly and analysis of statistical information. From 1 October 2011, TWRI became a private company which aims to continue providing and maintaining a range of socio-economic, demographic and community safety research and information services.

We would wish to note that, whilst no census is perfect, the absence of a 2021 census will be extremely detrimental to the understanding of changes in society in the decade 2011 to 2021.

#### 2. *How do Social Scientists use Census data?*

In summary, we have used the Census to undertake *trend* and *multivariate* analyses.

This is possible because the Census is the one and only source of data collected consistently across the country at a sub-district (ie at an output area)<sup>3</sup> level. In the past this would have been what was known as an enumeration district, a small area of contiguous addresses covered by a Census enumerator.

#### 3. *Analysis of trends across a sub-area of a district*

Because the Census is collected at a small area level, we can aggregate the data up to any area. Where appropriate, this can include appropriate apportioning and manipulating. So if, for example, Local Authority wards or Parliamentary boundaries change, we can by going back to the original small area data, manipulate it and aggregate up the old data to new boundaries.

Hence we can establish trends, from the past, in particular indicators, for new local authority wards and Parliamentary constituencies.

---

<sup>1</sup> One of us was first involved in understanding and adjusting the 1961 Census to the 1974 Ward boundaries following on from the 1973 Local Government re-organisation. He has worked on all subsequent censuses, including the 1966 sample census; thus he has had over thirty-five years working with census material. Another has over 30 years and another over 20 years experience, again including multivariate analyses.

<sup>2</sup> By *manipulating* the census results we mean a) *adding* cells to create new variables or new areas, b) *subtracting* cells, where this is better than adding, c) creating indicators [eg percentage of the population aged under 16, or over retirement age; percentage of the population/households without access to a car] and d) apportioning to create measures for boundary changes [eg new constituencies and wards], this enabling local, [eg within district], trends to be detected.

<sup>3</sup> An output area is a group of full postcode addresses of between 40 and 125 households.

So, for example, we can establish “household car ownership” rates, “average household size” or “percentage aged under 16”, from previous censuses for new wards. This helps in understanding the nature of a particular ward and is useful for politicians and service providers.

#### 4. *Multivariate analysis of small area Census data*

Again, because the Census is collected at a small area level, we can calculate a number of social and demographic indicators at that level. For example, we can calculate “unemployment rates” for males and for females, “percentage long-term sick” and “percentage ethnic minorities”, “car ownership rates” and “percentage owner-occupiers”. No other apportioning or manipulating is required. We can then input the indicators into a multivariate Factor analysis.

5. Using Factor analysis, we can determine the importance of each of these indicators in the underlying Factor, (or Factors if there is more than one) which best describes the data. Generally speaking, from the Census, we find that the first factor can be described as a “poverty” measure.

Some of the indicators used in the analysis, may turn out to be unimportant. That would mean that that indicator was measuring something else. This is important to learn.

We can then calculate Factor scores for each small area. We can then see which of the small areas has a low (or high) score on each of the Factors. This will have policy implications for intervention to reduce the impact of poverty.

[We can take two (or more) census small area output data sets and undertake a Factor analysis on both sets together.]

In summary without small area data collected through the Census in a consistent way across districts this kind of analysis is not possible.

#### 6. *What impact will the ending of the Census have on social science research?*

It follows from what was said in the previous paragraphs that unless Districts and/or others invest in their own data collection, these kinds of analyses will be impossible to undertake, leading to a lack of understanding of poverty and other measures, at the small area level.

It would also be impossible to get the same kinds of analyses done between different authorities.

#### 7. *What alternatives to the Census would provide population socio-demographic data of equivalent or higher quality?*

In theory, it would be possible to use administrative data to undertake these kinds of analyses. In practice, however, changes in definition make this next to impossible.

For example, the administrative definition of unemployed has changed a number of times during the last 30 years. This would mean that whilst we can monitor unemployment at a ward level, when there are changes in the definition this leads to discontinuities in the analyses. Any relative change in unemployment rate could have arisen because of the change in definition.

Similarly, there have been changes in those allowed to have access to council tax benefit; this means that this would not be a suitable means of comparing sub-areas of districts over time.

Additionally there are three problems with administrative data namely confidentiality, uniformity in collection and non-comprehensive coverage of the data. People do not always register a house move with a GP, so that this information would be inaccurate. The Census, on the other hand, counts people at their actual address on Census day.

#### 8. *What other existing sources of population and socio-demographic data could be improved upon?*

It is difficult to say that any source would be adequate. To enable analyses to be undertaken, from administrative data, it would be necessary make sure that there is no change in definition. This is, as society changes, impossible.

An alternative to the Census would be a population register, as in Sweden. However, this would only give Population estimates by age and sex, household size and type. It would be unable to give measures of car ownership, tenure, disability, ethnicity and other socio-demographic indicators provided by the Census.

#### 9. *Summary*

It is our view that the Census is irreplaceable. Any attempt to do the above kinds of analyses, without it, will fall well short of the desired level of accuracy that can be provided by undertaking a 100% population census.

TWRI

29 November 2011

## Written evidence submitted by Dr James Kirkbride (Census 16)

RE: CENSUS AND BEYOND 2011

I am writing to share my views of how Census data is currently used in the research that I conduct. My specialism is epidemiology, and in particular, psychiatric epidemiology. While I will naturally draw upon my own experiences of how the Census data is used for epidemiological research into psychiatric disorders, much of what I say will apply to epidemiologists and public health practitioners interested in a range of other population health issues. To structure my comments, I have used the question headings you provided for this call. I, as other academic colleagues I work with, are grateful that you are taking on board the views of those engaged in social science and public health research in moving forward from the Census.

### 1. How do social scientists use Census data?

In my field we use Census data to provide an accurate estimate of the population who may be “at-risk” of developing a disorder in a given catchment area. This catchment area may be geographically focussed ie on just a single Borough, or may be broad ie an estimate of the population at the national level. This data is primarily used as the “denominator” number which goes into calculating estimates of the incidence (number of new cases) and prevalence (number of new and existing cases) of a particular disorder. Estimates of incidence and prevalence are vital for several reasons related to population health. These include:

- (i) accurate health service planning;
- (ii) clues for aetiological (ie causal) research;
- (iii) information to develop potential public health campaigns (prevention and health promotion);
- (iv) Health economics—to estimate the burden of disorder(s) and potential associated costs in a given population.

Of these, it is the second (ii) with which I have most direct experience. By calculating incidence rates of disorders in specific populations (and communities) we can look at variation in those rates along several factors which might be of interest to us. These include looking at how incidence rates of disorder vary geographically, or vary by other factors such as age, sex, ethnicity or occupation. Identifying variation in incidence rates can provide clues about the aetiology (or cause) of a disorder, which can advance scientific understanding, reduce the burden of disorder (through prevention strategies which follow) and reduce the direct and indirect economic costs to health services and society associated with disease. The census provides essential underlying denominator data, stratified according to such factors, to enable us to estimate these incidence rates. One of the unique aspects of the census here is that the data can be stratified according to more than one variable at a range of spatial geographies. By having this information, we are able to know how many people in a given geographical location belong to each group we are interested in—for example, we can split the population by age, sex and ethnicity. By knowing this, we can estimate the number of cases of a given disorder which occur in particular groups of people, allowing for the estimation of incidence in different groups. By comparing incidence in different groups, we get valuable information about the burden and impact of disease in different communities. For example, we have used census data in this way to show that men have greater rates of psychotic disorder than women at young ages, but that this pattern reverses in mid-life (Kirkbride *et al*, 2006), and we have also shown that black and minority ethnic groups have higher rates of psychotic disorder than their white British counterparts (Coid *et al*, 2008, Fearon *et al*, 2006, Kirkbride *et al*, 2008b); vital public health information and important in understanding the causes of such disorder. Furthermore, because census estimates of the population can be stratified across several factors at once (age, sex, ethnicity, etc...) this allows us to build this complexity into our models of health and disease, controlling for the effect of one variable upon another, to eliminate confounding—the chance that an association between disease and a risk factor is really due to another factor. For example, in our 2008 paper, we used census data in East London to show that the incidence rate of schizophrenia remained elevated for ethnic minority groups, after taking into account possible differences between these populations and the white British group in terms of age, sex or social class, which could otherwise have explained this association.

*Accurate estimates of the population at-risk—provided for by the census from the national level to the level of output area—are vital for understanding variation in the incidence and prevalence of a variety of diseases, which can then be important for aetiology (causal research), health services planning, health economics and public health. Whatever supersedes the census will need to provide accurate estimates of population data, stratifiable according to several factors, down to fine spatial (ie geographical) scales.*

### 2. What impact will the ending of the Census have on social science research?

Without an adequate replacement for the census we will not be able to conduct the type of research outlined above. This will have major impacts on our ability to detect variation in the rates of disease, geographically and according to different sociodemographic groups, meaning that models of service provision will be less accurate, it will become harder to investigate possible aetiological associations between sociodemographic factors and disease, and harder to forecast (predict) which groups are at greatest risk of developing illness. These consequences will have downstream effects for health economics.

A particular aspect of my social science research—and that of many others—which will be affected is our ability to study geographical variation in the incidence of health (well-being) and disease. Currently population estimates from the census for different geographical aggregations (particularly the super output area and the electoral ward) allow us to estimate rates of disorders in small communities. If these rates vary, we can discover whether any social or physical factor in our environments varies with them, which has possible implications for causal understanding. For example, in psychiatric epidemiology we have used census data to show that the incidence of schizophrenia varies across neighbourhoods (Kirkbride *et al*, 2007, Kirkbride *et al*, 2006), and that this variation is associated with certain characteristics of neighbourhoods (Kirkbride *et al*, 2008a), including deprivation and social capital (more urban, more deprived, less socially cohesive neighbourhoods have more schizophrenia). Interestingly, these findings have motivated neuroscientists to study differences in the brains of people living in rural and urban environments, finding that those people in urban environments have a greater reaction to stressful events. Thus, the epidemiology (facilitated by the census) provided a testable hypothesis for neuroscientists, with the results suggesting that urban environments may alter the way we process information in our brain, and very tentatively, govern our risk of developing certain psychiatric disorders.

*This is the kind of translational research which is threatened (and which will be lost to overseas competitors) if the census ceases to exist, or if suitable equivalent data is not made available in its place.*

### *3. What alternatives to the Census would provide population and socio-demographic data of equivalent or higher quality?*

The census provides valuable, accurate estimates of the population. As with any statistics there is a degree of error in their estimation, but for the kind of social science literature I have outlined, they represent the gold standard, with currently available UK alternatives lagging some way behind in scope, scale and precision. The alternatives which would provide equivalent or better data for population health research of this kind can be separated into (i) research-led initiatives or (ii) government-led initiatives:

- (i) Research-led initiatives would include certain study designs—like birth cohorts—which follow-up a group of people born at the same time, with repeated assessments at various stages over the life course, for the duration of their lives. Examples in the UK exist, such as the 1946 or 1958 Medical Research Council birth cohorts. The advantages of such studies is that they provide prospective clues for causal research of health and disease (ie the risk factor was measured before the person became ill), providing great power to detect causal associations. However, such studies are costly, have issues with loss to follow-up, take a long time (several decades) and even the largest are small relative to the rate of certain disorders. For example, the 1946 birth cohort contained only ~5000 singletons in the sample. For disorders such as schizophrenia, what we know from epidemiology is that this would yield only roughly 50 cases over the entire life time of the participant members; too small for meaningful population research. Thus, I believe such options, while powerful in their ability to detect associations between disease and risk factor, do not provide adequate population data for the type of research currently facilitated by the census.
- (ii) Government-led initiatives—one possibility, although it shamefully does not currently exist in the UK—is to develop national registers capable of linking up a range of government databases on the population to provide extremely valuable, prospective, nationally-collected linked datasets for the study of health and disease. Such registers are commonplace in Scandinavian countries where databases are linked through people's social security number. Such data is extremely powerful to detect possible associations between disease and risk factors and provides population level coverage. However, such databases no doubt have implications for how data is shared between government agencies, and the fiasco over ID cards illustrates the difficulties in setting up such a registry system, though it would be a valuable—and in some ways more powerful—method of analysing health data in the population.

### *4. What other existing sources of population and socio-demographic data could be improved upon?*

Unsure. There is nothing currently available in my eyes which comes close to providing the rigour and precision of the census. Random small sampling with extrapolation to the remainder of the population is a possibility, but samples will rarely be representative, even if participation if selected is made compulsory. Another option would be to link up existing government databases for such research, but this has massive cost, logistical and moral challenges. At best, the decision to scrap the census is very disappointing for social science research, and at worst, this could set back some areas of social research by several decades.

I am very grateful to the Committee for giving researchers the opportunity to express their views on this matter. It is of grave concern that this resource should be superseded with something capable of permitting (and ideally improving) upon the type of population-based research made possible by the Census. The estimation of incidence and prevalence rates (of any health outcome) will only be as good as the denominator data from which those data are ascertained.

I have no conflicts of interest to declare.

I am a Wellcome Trust research fellow based at the University of Cambridge.

## REFERENCES

Coid, J W, Kirkbride, J B, Barker, D, Cowden, F, Stamps, R, Yang, M & Jones, P B (2008). Raised incidence rates of all psychoses among migrant groups: findings from the East London first episode psychosis study. *Arch Gen Psych* 65, 1250–1258.

Fearon, P, Kirkbride, J B, Morgan, C, Dazzan, P, Morgan, K, Lloyd, T, Hutchinson, G, Tarrant, J, Lun Alan Fung, W, Holloway, J, Mallett, R, Harrison, G, Leff, J, Jones, P B & Murray, R M (2006). Incidence of schizophrenia and other psychoses in ethnic minority groups: results from the MRC AESOP Study. *Psychological Medicine* 36, 1541–50.

Kirkbride, J, Boydell, J, Ploubidis, G, Morgan, C, Dazzan, P, McKenzie, K, Murray, R & Jones, P (2008a). Testing the association between the incidence of schizophrenia and social capital in an urban area. *Psychological Medicine* 38, 1083–94.

Kirkbride, J B, Barker, D, Cowden, F, Stamps, R, Yang, M, Jones, P B & Coid, J W (2008b). Psychoses, ethnicity and socio-economic status. *British Journal of Psychiatry* 193, 18–24.

Kirkbride, J B, Fearon, P, Morgan, C, Dazzan, P, Morgan, K, Murray, R M & Jones, P B (2007). Neighbourhood variation in the incidence of psychotic disorders in Southeast London. *Social Psychiatry and Psychiatric Epidemiology* 42, 438–445.

Kirkbride, J B, Fearon, P, Morgan, C, Dazzan, P, Morgan, K, Tarrant, J, Lloyd, T, Holloway, J, Hutchinson, G, Leff, J P, Mallett, R M, Harrison, G L, Murray, R M & Jones, P B (2006). Heterogeneity in Incidence Rates of Schizophrenia and Other Psychotic Syndromes: Findings From the 3-Center AESOP Study. *Archives of General Psychiatry* 63, 250–258.

*Dr James Kirkbride PhD (Cantab), MSc, D.LSHTM, BA (Hons)*

*29 November 2011*

---

**Written evidence submitted by the Institute for Fiscal Studies (Census 17)**

DECLARATION OF INTEREST: NIL

This is an ESRC funded investment. The views and statements expressed are those of the authors and do not necessarily reflect the views of the ESRC.

*How do social scientists use Census data?*

1. There are a number of ways in which researchers at IFS use Census data.
2. The census is the key source of information on population statistics and this information is very important in their own right. Information on how a population is ageing for example is key to identifying where policy needs to take into account a changing population structure and to plan how to spend resources.
3. Population statistics are also important in work that uses data on sub-samples of the population. Cross-sectional and longitudinal samples of the population are inevitably subject to biases in the type of individual or household that responds. In order to generalise results to the whole population, one needs to appropriately weight the data to take account of non-response. These weights are calculated using information about population totals and the structure of population taken from the Census.
4. Population total are also used to “gross up” data to work out the *number* of people or households affected by a particular policy for example.
5. Information on socio-economic characteristics is available from the Census at a very disaggregated geographical level. This information can be used and combined with other data and will give a very accurate picture of the extent of any deprivation in any local area. For example we have used information on various indicators of socio-economic status at the output area level and merged it into administrative data we have on schools and universities.

*What impact will the ending of the Census have on social science research?*

6. The most likely negative impact is that researchers will have to pay for alternative sources of data.

*What alternatives to the Census would provide population and socio-demographic data of equivalent or higher quality?*

7. While there other sources of information such as ACORN or MOSAIC data which contains information about socio-economic status as a very disaggregated geographical level, the census is the only freely available data.

*What other existing sources of population and socio-demographic data could be improved upon?*

8. No comment.

Institute for Fiscal Studies

29 November 2011

---

**Written evidence submitted by Dr Stephen Patterson (Census 18)**

RESPONSE TO INQUIRY BY PARLIAMENTARY SCIENCE AND TECHNOLOGY COMMITTEE INTO FUTURE OF CENSUS OF ENGLAND AND WALES

1. I have received a copy of the request for submissions to the parliamentary inquiry on impact of ending the Census on social science research. You may be interested in the following comments.

2. There is a long tradition of conducting decennial Censuses in the UK. The first official Census of England and Wales was conducted in 1801. It would be a loss if this tradition came to an end.

3. In the modern day the task of conducting the UK Censuses is complex. Through long experience the Office for National Statistics has developed the expertise to conduct the Census of England and Wales.

4. If the Census is to be replaced, it needs to be determined that alternative sources for all Census data items really are available and accessible.

5. From our point of view in the Public Health Department of NHS Suffolk, the Census is one of the main sources of data for reporting on the health of the local population. Census topics that provide particularly useful data include:

- rebased mid-year population estimates by age and sex;
- population estimates by ethnic group;
- limiting long-term illness;
- carers;
- household and family composition;
- lone parents;
- housing and tenure;
- communal establishment populations;
- qualifications;
- occupation and socioeconomic classification;
- religion;
- car ownership.

6. These data are widely used in public health, in health needs assessments, Annual Reports of the Director of Public Health, ward profiles and other reports.

7. Census data are available from the level of Output Area to the full range of higher geographies. A similar geographical range of data may not be available from alternative sources.

8. It is also useful to be able to access this whole range of data from a single clearly organised source, eg ONS Neighbourhood Statistics, the main ONS website (for 2011 Census data, when it becomes available), SASPAC, InSite or C2001 Census Data Manager.

9. ONS has developed the expertise and infrastructure to conduct the Census. The amount of information requested in the Census is not large and it is not a great burden on the public to complete the forms, as is shown by the high response rates for the 2011 Census. If anything, this foundation should be built upon, with additional questions added to the Census in future years. More questions could be added on lifestyle, including smoking, for example.

I hope these comments are of interest.

*Dr Stephen Patterson*  
Epidemiologist  
Department of Public Health  
NHS Suffolk/Suffolk County Council

29 November 2011

---

---

**Written evidence submitted by Professor Edward Higgs (Census 19)**

1. As a British social and economic historian I have been using the raw household returns from the historic British censuses for the 35 years. I am also one of the co-researchers currently working on the Integrated Census Microdata (I-CeM) Project, a three year project funded by the Economic and Social Research Council (ESRC) to create a single integrated dataset of the British census returns, 1851 to 1911, for academic research purposes. This involves receiving data from a commercial partner, standardizing it, and coding and classifying it to make it comparable with data from modern censuses and longitudinal surveys. We hope to link the people in the censuses across time, and with modern surveys, so that we can study subjects such as social and physical mobility, the life courses of the disabled, the intergenerational experience of migrant families, and so on.

2. As a historian of the nineteenth century the original, individual-level census returns are a vital source because they locate everyone spatially, socially (by family and occupational strata), and temporally. The latter is very important because the census is a *de facto* rather than a *de jure* enumeration—it is a snapshot on one night of the year, and so avoids double counting.

3. The existence of names and addresses that identify individuals is also of great importance because they allow historians to link the census returns with other nominal records so enriching our understanding of individual circumstances in the past.

4. As a social scientist interested in following families and individuals across time it is also important to have data that is comparable. In England and Wales census data has been created by a single body, the General Register Office (now part of the Office for National Statistics) for the past 170 years, and this has ensured a degree of continuity in the information collected, and in the definitions used when collecting that data. For example, the modern definition of what constitutes a family was developed in the nineteenth century. When definitions of terms and variables have changed, the census-taking authorities have been scrupulous in explaining those changes. For example, in some Victorian censuses women were told not to put down an occupation if they worked at home, except where the home was a farm or a shop, when they were assumed to be working in the family business. But in other censuses this instruction was not given. We need to be able to reconstruct these instructions so that we know what the data means.

5. As a historian and social scientist, I am not particularly wedded to the continuation of the decennial census, as long as what replaces it can provide comparable information. That would mean:

- (a) data collected at one point in time;
- (b) data which is collected using consistent definitions, which are consistent over time, or if there are changes in definitions, these we can reconstruct; and
- (c) data which gives names and addresses.

6. In practice, I am not at all sure that alternative means of data collection to the census could meet such criteria. Thus:

- (a) I cannot see how one could guarantee that data collected from a multitude of differing sources would not relate to people at different dates, and so be open to double counting. How do we know that Edward Higgs at one address is not the same as another Edward Higgs recorded at another address some time later? With the abolition of the ID card scheme of the last government, we do not have a common identifier to tie people down.
- (b) In such circumstances I cannot see how one can ensure that data collected in different ways, and by differing agencies, is strictly comparable. Do they give the same instructions to respondents when asking questions and collecting data? We need to remember the old data processing adage—GIGO—Garbage in Garbage Out.
- (c) Will the data collected contain the names and addresses of identifiable people? I cannot see how information from commercial, or non-census government, databases could be used for census purposes without breaking the provisions of the Data Protection Acts. Certainly, I cannot see businesses being willing to hand over the data they have collected from their customers for marketing purposes, and I believe that Tesco has already refused to hand over the data collected via its store loyalty card scheme.

7. In these circumstances I cannot but think that a shift away from a *de facto* census every ten years will seriously undermine the usefulness of the data collected for longitudinal, historical analysis.

*Professor Edward Higgs*  
Department of History  
University of Essex

28 November 2011

---

### Written evidence submitted by Centre for Longitudinal Study Information and User Support (Census 20)

1. Responses are sought to the question: “How do social scientists use Census data?” One of the census data resources for social scientists is the subject of this submission.

#### DECLARATION OF INTEREST

2. This submission is being made by Professor Emily Grundy and staff of the Centre for Longitudinal Study Information and User Support (CeLSIUS) at the London School of Hygiene & Tropical Medicine <http://www.celsius.lshtm.ac.uk/>. CeLSIUS has been funded by the Economic and Social Research Council since December 2001 to promote and facilitate the research use of the ONS Longitudinal Study by UK academic researchers, both staff and students, and this submission is made on their behalf. Additionally Emily Grundy has made extensive use of data from the LS in her own research.

#### *How do social scientists use census data?*

3. This response relates to how social (and other) scientists use the census data incorporated in the Office for National Statistics Longitudinal Study (LS). The LS was established in the 1970s largely to address the problem of inadequacies in the data available to analyse mortality differentials in the population of England and Wales and biases resulting from the fact that denominator (population) data came from the Census and numerator (death records) data from vital registration. The initial sample was drawn from the 1971 Census on the basis of birthday, in order to facilitate linkage. All those born on four birthdays per year were included giving a sample amounting to just over 1% of the population of England and Wales. The study has been maintained through the addition of new births and immigrants with the same birth date, and includes individual level data from four censuses (1971, 1981, 1991 and 2001) as well as linked information on births, deaths and cancer registrations. Linkage of data from the 2011 Census is currently underway and will be available to the social science research community in 2013. Access to anonymised data for research purposes is permitted under various strictly adhered-to conditions which include only allowing access to microdata in an ONS secure data laboratory.

#### *Advantages of the LS*

4. The LS is representative of the whole population of England and Wales, including those in non-private households (such as nursing and residential care homes, prisons etc), includes all age groups and includes census information about other people in sample members’ households at each census which provides additional opportunities for examining intergenerational continuities and changes. The “width” of the sample in terms of size means it is possible to study relatively small groups, such as members of particular ethnic minority groups or older people resident in institutional settings (a group excluded altogether from most surveys). The “depth” of the study over time makes it increasingly valuable for research including a life course or intergenerational perspective, for example, it is possible to examine characteristics of adults aged 31–46 in 2001 by attributes of the parents they lived with as children in 1971. A further strong advantage of the LS is minimal bias due to non-response or attrition, as census coverage is good and rates of linkage high.

5. Since the LS comprises all persons born on four days of the year, the sampling fraction is approximately 1.05% and sampling bias is almost nil. At each census the LS sample numbers over 500,000 and the full database currently includes well over 1 million people. This is by far the largest longitudinal dataset in the UK; it allows analysis of small areas (well below local authority level), particular ethnic groups and specific occupational groups. These are not possible with any other longitudinal dataset currently available.

#### *Research based on the LS*

6. The current list of publications based on LS research includes over a hundred books, more sections of books, and approximately 350 articles in journals—see list at <http://www.celsius.lshtm.ac.uk/publications.html>. These publications present results from research on a wide range of topics including, for example, studies of associations between unemployment and mortality; effects of marriage and marital history on health and mortality; ethnic variations in social mobility; trends and differences in fertility patterns; consequences of lone parenthood; social class differences in cancer survival; links between geographic and social mobility; influences on household structures among older people and many others. Current projects, listed at <http://www.celsius.lshtm.ac.uk/projects.html>, are also wide ranging and include many studies of inter relationships between ethnicity, location and labour market or demographic behaviour; many studies focussed on understanding differences in health and mortality; and many examining effects of regional or local characteristics and change in specific locations (for example, effects of in and out migration on the population of Cornwall).

#### *What impact will the ending of the Census have on social science research?*

7. Most of the research carried out using the LS could not be undertaken using other existing sources. Indeed, no national, cross-sectional dataset offers the same combination of sufficient numbers for detailed work with repeated measures spanning several decades, and for this reason the LS is used for cross-sectional as well

as longitudinal analysis. The older birth cohort studies, for example, provide more information on long periods of study members' life courses (the LS now spans 40 years) but are not large enough to examine small population sub groups. Even large scale surveys do not provide the potential for looking at sub regional differences or reorganising geographies in the way that the LS permits.

*What alternatives to the Census would provide population and socio-demographic data of equivalent or higher quality?*

8. No other source would compensate for the absence of census data in the LS, but some parts of the shortfall might be addressed by more linkage of records from other sources. The Scottish Longitudinal Study (SLS) and the Longitudinal Study of Northern Ireland (NILS) were both created much more recently than the LS and used the design and working of the LS as a template. However, unlike the LS, the two newer studies have steadily expanded the range of existing records which are added to the dataset (or, in some cases, can be linked to it at an individual level for each project that requires these data).

9. Records available in the Scottish or Northern Ireland Longitudinal Studies and not in the LS include:

- marriages;
- hospital admissions and discharges (including day hospitals and psychiatric hospitals) in Scotland;
- school census data, including exclusion records, in Scotland;
- migration data within Northern Ireland;
- valuation and Lands Agency data for Northern Ireland, also value of property;
- records of the prescription of drugs in Northern Ireland;
- records of breast cancer screening in Northern Ireland; and
- births and stillbirths to fathers who are sample members—the LS only has mothers; (data for fathers was linked 1971–74 but was discontinued due to poor linkage rates).

10. There is a legal gateway for similar linkages in England & Wales but to date little progress has been made in using this to enhance opportunities for social science research.

Centre for Longitudinal Study Information and User Support

November 2011

---

### Written evidence submitted by Dr Nicola Shelton (Census 21)

#### DECLARATION OF INTERESTS

The author Dr Nicola Shelton is a health and population geographer in the Department of Epidemiology and Public Health at UCL. The author has used UK Census data for health geography and health surveillance purposes.

#### 1. *How do social scientists use Census data?*

In the context of health geography and health surveillance the Census is used for a variety of purposes. Census data on health have been asked variously since 1991 and have included caring for others since 2001. Health outcome data including cancer registrations and mortality and hospital admissions and discharges are also linked to the Census through the various Longitudinal Studies.

The Census is also used to provide sampling frameworks from which sample surveys are drawn and to create reference populations against which sample survey data can be age standardised for comparison between eg geographical regions and socio-economic groups. Census data is also used to develop indices that allow small area estimation for health statistics such as smoking and obesity prevalence and for measuring health inequalities.

#### 2. *What impact will the ending of the Census have on social science research?*

The national and regional and socio-economic and demographic variations in health and care will be lost as will the wealth of analysis of health variations from the longitudinal studies.

Because the UK does not have a population register there will be no alternative sampling framework other than that of the electoral register which includes data for adults only or the NHS Patient Register data which includes only registered patients. Developing sampling frameworks for sample surveys of socio-economic and geographical and ethnic health inequalities will be very difficult especially for children and immigrants and for those groups less likely to be registered with a GP or to keep that registration up to date eg young men. Age standardisation would be problematic as it would be based on increasingly inaccurate population estimates over time.

3. *What alternatives to the Census would provide population and socio-demographic data of equivalent or higher quality?*

A comprehensive population (identity) register linked to vital events and migrations would deliver the population and socio-demographic data of equivalent or higher quality, but may be a more costly than continuing with the Census in 2021 and beyond if it were to include the full range of variables covered in the Census, and would take some decades to develop. It would not provide continuity for the Longitudinal Studies.

4. *What other existing sources of population and socio-demographic data could be improved upon?*

The NHS Patient Register data could be made available as a sample framework of individuals. This would provide also an opportunity to sample individuals and demographic characteristics and particular health conditions subject to disclosure control.

The lack of socio-demographic data in the Patient Register limits its scope for use in social science research: disaggregation of characteristics such as ethnicity often reveal significantly different patterns of health inequality to those from the total population.

The electoral register could be extended to include all ages as a basic population register, but again would not provide the socio-economic and demographic detail of the Census.

*Dr Nicola Shelton*

Head of Health and Social Surveys Research Group  
Epidemiology and Public Health, UCL

30 November 2011

---

**Written evidence submitted by the British Society for Population Studies (Census 22)**

DECLARATION OF INTEREST

The British Society for Population Studies (BSPS) comprises persons with a scientific interest in the study of human populations. Its main objectives are to further the scientific study of biological, economic, historical, medical, social and other disciplines connected with human populations and to contribute to the public awareness of them. Our four hundred BSPS members include the public sector (including central and local government and health services), academia and business. They have a wealth of experience in empirical research, with much of it drawing on Censuses in the UK and other countries but also much based on data from other population surveys and from administrative sources not designed primarily for demographic analysis.

SUMMARISING POINTS

1. Good data is the key priority of those who currently rely on Census data for information about the size, distribution, characteristics and welfare of the UK population: good in relation to population coverage, geographical detail, accuracy of information, and range of topics. While many users would prefer greater frequency than once every 10 years, this should not be at the expense of quality on any of these criteria. At the same time, there are some aspects covered by the Census that change relatively slowly, for which 10 years is a satisfactory time interval.

2 The increasing ubiquity of geographically coded data means that the need for small area data will continue to grow inexorably. Any move away from providing such data could not be more detrimentally timed.

3. Much valuable policy-relevant research relies on the high level of consistency between decennial Censuses to understand change over time and this would be lost by shifting to data collected in different ways.

4. Most administrative data sources have a very limited number of variables on personal characteristics, so the challenge of replacing the Census is to link the administrative sources together for individuals and households.

5. Key policy concerns including energy use and social segregation cannot be supported without local analysis of the characteristics of commuting and migration which only the Census provides at present (mode of transport to work; socio-economic and ethnic background of migrants within the UK).

6. For researchers, the key question is "Is there an alternative data source with the range, detail and quality of information that the Census currently provides?" and not "Which is the best alternative, given that Census will be replaced?"

[1] *How do social scientists use Census data?*

7. BSPS members use UK Census of Population data for three main purposes: (1) to provide descriptive contextual information on populations of interest; (2) for secondary data analysis, often using advanced quantitative methods, to address questions which improve our understanding of UK society and how it is

evolving; and (3) as a means of benchmarking other surveys to ensure that they represent the total population of an area.

8. The most traditional form of Census outputs, dating back to the first Census in 1801, comprises what has become known as the Area Tables (AT). These provide data for geographical areas ranging in scale from the UK level through the 4-country and regional level down to the very local scale, allowing understanding of neighbourhoods of around 100 households or more. The parts of the AT most heavily used by BSPS members are the “multivariate tables”, in which the answers to one Census question are cross-tabulated with the answers to one or more other questions, such as in the table of age by sex by whether living in a household or in some form of institutional accommodation such as a hall of residence or nursing home. The “univariate tables”, which present the results for each Census question separately, are less widely used in social science research, because the latter is primarily concerned with the way in which the population’s various individual and household characteristics relate to each other, as for instance between ethnicity and the Census measures of health, deprivation and well-being.

9. BSPS members are also major users of the Origin-Destination Statistics (ODS) which allow the study of how people move between places, most notably migration and commuting but also (though for Scotland only) travel to place of study. While the Area Tables include some of the data on these topics in summary form, the ODS are highly valued because they show the flows between each geographical area and every other one separately, with each between-place flow broken down by a set of population characteristics and (for data on travel to work and to school) also by mode of transport. Applications of these data include studying whether migration is leading to greater mixing or segregation of people by aspects such as age, ethnicity, employment and occupation, how much occupational groups vary in the distances over which they access work, how reliant each place of employment or residence is on its workers travelling by car as opposed to forms of transport with a lower carbon footprint and—by comparison with the results of previous Censuses—the extent to which these have altered over time. In addition, users are eagerly awaiting new “mover” data sets from the 2011 Census on the whereabouts of people away from home on Census night and on secondary residences.

10. The suite of datasets known collectively as the Samples of Anonymised Records (SAR) have greatly extended the value of the Census since 1991. Besides allowing users to readily extract alternative cross-tabulations, it also permits them to undertake micro-level modelling. The huge benefits of this approach are well rehearsed in the document “A business case for microdata files from the 2011 Census” ([www.ccsr.ac.uk/sars/2011/documents/businesscase.pdf](http://www.ccsr.ac.uk/sars/2011/documents/businesscase.pdf)), which has prompted the Census agencies to repeat this exercise for the third Census in a row.

11. Even more powerful for social science research is the ability of users to track anonymised individuals through the UK’s three Longitudinal Studies (LS), based on linking their Census records together but also through adding information from other sources. The LS for England and Wales was set up after the 1971 Census with the particular aim of using the Census data on people’s characteristics to help understand differences between them in their life expectancy and proneness to cancer and other notifiable diseases, with data on the latter being linked to their Census records. Analyses have been both retrospective in nature (as in seeing what conditions preceded death or illness) and prospective (seeing what changes in work or housing happened after an event such as birth of a child or widowhood). Following the major insights achieved in this work, similar studies have been set up in Scotland and Northern Ireland.

12. Finally, a significant number of BSPS members have looked back over longer historical periods. Much of their work has examined long-term trends in places’ size and characteristics by comparing the Area Tables of past Censuses, helped by Censuses including many of the same or similar questions. For Censuses taken more than 100 years ago, it is permitted to access the data for named individuals, providing added richness to these analyses as well as proving extremely popular amongst those wishing to trace their ancestry. It should be noted that, for more recent Censuses, great care has been taken to ensure that no individual persons can be recognised from the data made available to researchers, this being achieved primarily by controlling the detail allowed in microdata releases but also including record swapping and other disclosure control measures in the Area Tables and Origin-Destination Statistics.

[2] *What impact will the ending of the Census have on social science research?*

13. Impact is dependent on the nature of any replacement, which means that our answer to this question should be read in conjunction with those to the two following questions. As the Census is unique in the combination of geographical detail and information about individuals and households which it yields, its abandonment without any form of replacement (ie with researchers needing to rely on existing alternatives) would have devastating impacts. At the same time, the severity of the impact would vary between the major areas of Census usage outlined above in answer to the first question, as set out below.

14. Much (but by no means all) of the information provided by the univariate sections of the Area Tables could be proxied by data drawn from existing surveys and administrative sources, though as currently in the Neighbourhood Statistics Service (NeSS) data holdings the population coverage, accuracy and geographical detail would not be as high as in the Census and would vary between items.

15. The majority of crosstabulations provided by the multivariate sections (i.e. the majority) of the Area Tables could not be generated from existing alternative sources, except at the broadest geographical scales (eg

whole UK, regions) where sample sizes are large enough, because to get meaningful findings crosstabulations have to be based on the individual persons or households rather than by comparing the univariate area aggregates.

16. Migration is monitored between Censuses, but the international migration data are based on very small samples with correspondingly high error and lack of geographical detail and the within-UK migration data record only sex and age and do not provide information on moves within local and unitary authority areas, so most applications of the relevant ODS would not be possible. For commuting, there appears to be no viable alternative source.

17. The value of the SARs over any other survey or administrative dataset lies in its combination of relatively large sample size (allowing a fair degree of detail about small groups and areas), complete coverage of the population (including—unlike many sample surveys—people living in communal establishments) and a wide range of characteristics for each person and household (which can all be crosstabulated against each other at the individual level).

18. The UK's three Longitudinal Studies grow ever more valuable as data are added from each successive Census so as to cover an increasing length of people's lives, and it is extremely unlikely that the records for all these LS members could be updated from non-Census sources on a consistent and comparable basis for more than a very few of the current topics, leading to steadily reducing value over time from a final Census point in 2011.

19. Longer-term historical research by the social scientists of the future will be rendered less viable if the Census is replaced by data collected in a variety of ways which themselves may change over time.

20. In this context, it is worth noting that the demands placed on the Census, driven principally by central government but also coming from the data needs of local government, academia and the private sector, have steadily increased over recent decades, and this in spite of the development of new surveys, longitudinal cohort studies and administrative data releases. Most recently, the 2011 Census included several new questions which allow more detailed analysis of the migrant population, self-ascribed identity, the impact of student populations and the phenomenon of weekly commuting.

21. The impact on social science research and on the intelligence which it provides to policy makers and practitioners in both public and private sectors therefore depends very much on whether new data sources can be developed or existing sources can be accessed and manipulated in novel ways.

[3] *What alternatives to the Census would provide population and socio-demographic data of equivalent or higher quality?*

22. The short answer is that currently there are none. A replacement for the Census would need to provide the combination of small area detail and accuracy that make the Census the one essential data resource.

23. The nearest alternative is the Annual Population Survey (APS), but much of the specialist Census research for policy makers and others simply could not be done at all with the APS due to it not providing robust data for local authorities and smaller areas. The value of the APS is also diminished by it not attempting full coverage of institutional accommodation. Other official surveys have lower sampling fractions and/or less complete coverage of the population.

24. Countries which have abandoned the Census now rely on a comprehensive population register with a system for updating when individuals move residence. By themselves population registers are not ideal, because it cannot be ensured that every instance of migration is recorded. Additionally, obtaining socio-economic information for the people on the register requires linkage to data bases using a unique code number for each person (as occurs in Scandinavia) and/or large-scale surveys.

25. As regards the linking of data bases, this is difficult in the UK because no single unique personal identifier is used in the original data collection. The National Insurance (NI) number could be used to identify families and their dynamics. However, the NI number is not necessarily unique and people leaving the country do not have to notify anybody that they have done so, so there will be numbers which are inactive or used fraudulently. Also, there is no method for linking information about the characteristics of the dwelling an individual is living in (tenure, type, conditions), unless other sources such as the Land Registry and Electoral Register were also required to record NI numbers and introduced ways of regularly updating their information about the dwellings.

26. For commuting data, there is no viable alternative to the Census at present. An experiment was undertaken by ONS to see whether APS could provide the necessary data on commuting for updating official Travel-to-Work Areas (TTWAs) and it was found that this was very far from true. If the Census is not continued this updating will be impossible, leaving policy-makers and many others without a resource they have had for many decades (in fact, Eurostat is currently discussing the need for TTWA-type definitions in all EU member countries). The Annual Survey of Hours and Earnings (ASHE) has some potential, but only if its sampling fraction could be raised from 1% to at least 10%. Moreover, ASHE does not collect mode of transport, which is a key policy interest these days in terms of the climate change agenda, nor a sufficient range of personal characteristics to better understand trends in commuting behaviour.

27. A cautionary note also needs to be sounded about the pitfalls of relying on a variety of surveys and administrative sources. These include the fact that, unlike the Census which is specifically tasked to provide a 100% population count and answer a wide range of questions, most of these other sources have no such remit. For one thing, there is normally less control over the quality and consistency of the information collected. For another, if an alternative was built around one or more key administrative datasets, there would need to be legislation—and no doubt funding—to ensure that the relevant data collection was continued even if the body in charge of the source had no further need of such data or was abolished. Hopefully, lessons have been learnt from the effect of NHS restructuring on migration data availability and the decision to withdraw child benefit from wealthier families next year, as well as the recent failure of Royal Mail to play ball with address registers.

28. If a population register is selected as a replacement for the Census, then its worth must be properly validated against a full Census enumeration, so the Census should not be withdrawn until the alternative is shown to be of a sufficiently high standard for both policy making and research. It took Sweden some 20 years to bring its register-based data system up to the required standard, and ONS's Beyond 2011 briefing is suggesting a similar time frame. The latter assumes considerable investment of resources in dataset development over and above the funding needed for continuing the Census to provide this benchmarking.

30. If some form of population register is adopted as an alternative to the Census or as a supplement to a reduced-length Census, then it is vital that regular downloads of the data are stored in a safe but user-friendly setting in order to enable studies of change over time and to permit historical research by future generations.

[4] *What other existing sources of population and socio-demographic data could be improved upon?*

31. The other sources of population and socio-demographic data in the UK are limited and each has its own weaknesses as well as strengths, so improvements should be sought in all of them, whether the Census continues or not.

32. The most reliable element of the UK population data is the registration of births and deaths which is very accurate and underpins the annual estimates of population made by the ONS and the Registrars-General for Scotland and Northern Ireland. It would be possible to add more questions about the parents or the deceased.

33. Most of the information on the socio-economic characteristics of the population between Censuses is derived from ONS social surveys. The main drawback of these surveys is their small sample size, which means that reliable and non-disclosive data for smaller geographical areas and sparsely populated rural areas cannot be generated. Much larger sample sizes would improve this situation, but the costs involved in data collection would be substantial.

34. All surveys suffer similar drawbacks to the Census in not asking all the questions that are potentially of interest to data analysts while becoming regarded as intrusive by a section of the population which will not co-operate. Hence all social surveys are experiencing declining participation rates and are biased in (often unknown) ways. The 2011 Census made great improvements in response, but this involved substantial expenditure to achieve. However, it was undoubtedly cheaper in terms of expenditure per head than each wave of a social survey and it has yet to be shown that a Census is more expensive over a decade than an alternative that provides a similar or better intelligence.

British Society for Population Studies

30 November 2011

---

### Written evidence submitted by Dr Jennifer Mindell (Census 23)

#### DECLARATION OF INTEREST

Dr Mindell is an academic public health physician who regularly uses Census data. She and her team at UCL work on the Health Survey for England, the UK-wide National Diet and Nutrition Survey, and the Scottish Health Survey, each of which uses Census data for sampling and for weighting the data to increase the representativeness of the data, and for age-standardisation.

#### *How do social scientists use Census data?*

1. When conducting surveys on a sample of the population, social scientists make extensive use of Census data, to ensure that their results are representative and hence generalizable to the population as a whole. This enables social scientists and others to gross up percentages to give population numbers, and to use the results for planning policies and spending. The census data are used when selecting the sample, assessing the representativeness of the participants, and when weighting the survey estimates to increase representativeness, as well as to judge the extent of bias through differential non-response.

2. When conducting household based surveys, social scientists often use a stratified sample to better ensure a representative sample. Using strata ensures that the sample chosen will closely resemble the population along the stratifiers chosen. For example, that the proportion of households sampled in areas of low, medium and high proportions of minority ethnic groups, match the proportions of households within the target population.

Census data on areas are often used for this purpose. Other important stratifiers include, for example, socio-economic position (eg the proportion of households with a head of household in a non-manual occupation).

3. Survey data are often weighted to correct for imbalances in the *responding* sample, whether the data are collected by household interview, telephone, internet, or by post. Such imbalances may occur as a result of the selection criteria (for example, a survey that places a limit of two adults per household to be included in the survey will therefore under-represent adults in three or four adult households, compared with one and three adult households). More commonly, such imbalances may occur as a result of response bias, whereby certain types of people are more likely to be contacted and more likely to agree to take part. To detect, and correct these imbalances, population estimates provided by ONS on the basis of the Census are used. As a result, the survey estimates, and any policy or spending decisions based upon them, are more accurate.

4. Due to cost constraints, surveys are limited in size, which then limits the size of geographical area that they can represent accurately. In large scale household surveys, the nine former government office regions or 10 strategic health authority areas are commonly the smallest geographical area presented in the results. Using accurate population figures and socioeconomic characteristics of smaller areas, available from the census, researchers can use synthetic estimation to “fill in the gaps” and provide small area estimates, which enable more detailed planning in smaller local areas.

5. An additional key use of the Census is to provide the definitive population structure by age and sex required for age-standardisation. This is undertaken to enable comparisons between groups that are not confounded by age.

6. Surveys face similar limits on the number of categories that their estimates may represent along other variables, such as income (often analysed by quintiles), deprivation level (often presented in tertiles or quintiles), education level and so on, and the census data offers the opportunity of synthetic estimation in the same way.

7. Census data are of course used to provide population level estimates of many variables, and these too can be used by social scientists conducting other surveys. For example, health surveys using the question “how is your general health” can compare their estimates with those generated in the census. This may illuminate other biases in who responds to surveys where participation is voluntary, even after adjustment for demographic and socio-economic factors, and also how responses to the same question may differ depending on its context within a survey labelled “health” and a census gathering general information.

*What impact will the ending of the Census have on social science research?*

8. Social scientists working on surveys with a sampled population will no longer be able to sample and weight their surveys to the same degree of accuracy and representativeness. Less confidence may therefore be attached to the results given, and policy and funding decisions may have a less robust platform of evidence.

9. Age-standardisation and other forms of adjustment for age will become increasingly inaccurate, resulting in misleading findings in social science research. While mid-year population estimates are used, the degree of re-estimation of these when a subsequent Census is published demonstrates how crucial the Census is for accurate population statistics: even with recording of vital events and of inward and outward migration, these estimates become increasingly inaccurate as time passes since the previous Census.

*What alternatives to the Census would provide population and socio-demographic data of equivalent or higher quality?*

10. The unique value of the census is its coverage levels of the population, and its bringing together basic information on several core topics (demographics, population movement, work, travel, health, education). Individual surveys cover samples of the population, and suffer greater levels of non-response. In addition, individual surveys focus on different topics. Administrative sources of data, such as post-office address files, health records, tax records, each cover only certain sections of the population (those registered with a GP, those with income) or only certain pieces of information (the post-office address files for example only enumerate households, not individuals, and are subject to errors such as empty properties being counted, or properties with more than one household).

In short, we believe that a composite of other sources of data would be unlikely to have the same accuracy and completeness that the census offers.

*What other existing sources of population and socio-demographic data could be improved upon?*

11. It is difficult to think of other sources of data that could replace the Census. It has been proposed that the Census is unnecessary because of other sources of routine data and bespoke surveys. This completely misses the mark:

- (a) First, no other data source covers the entire population (excluding, admittedly, rough sleepers). All other data sources are limited to certain sub-groups, such as school aged children (the Pupil Level School Census); hospital records (people attending hospital); GP data (those registered with a GP— and individual level data are available from only a portion of GP practices);

- (b) Secondly, it is not that the Census is unnecessary because of other data sources but on the contrary other data sources are useful only because of the information supplied by the Census that forms the base on which all other population level data depends, as described above for demographic and socio-economic data;
- (c) The Census provides data that are simply not available through other means. For example, the Index of Multiple Deprivation (IMD) 2007 comprises some information that is updated frequently, such as unemployment figures, but others that are based on Census data, such as housing without central heating, overcrowding, and adult skills levels.

12. The Census would be less necessary if the UK introduced an individual-level population register, as is held in most other European countries. A population register that contained the name, sex, date of birth and address of all residents and that was updated annually would fulfil some, but not all, of the essential functions of the Census. However, this may not be considered politically acceptable in the UK.

*Dr Jennifer Mindell*  
Clinical Senior Lecturer  
Research Department of Epidemiology & Public Health  
UCL (University College London)

29 November 2011

---

### Written evidence submitted by NatCen (National Centre for Social Research) (Census 24)

#### INTRODUCTION

1. This note sets out a response from NatCen (National Centre for Social Research) to a request from the Science and Technology Committee for submissions on the Census and Social Science.

2. NatCen is Britain's leading independent social research organisation. We are a not-for-profit business, driven only by our mission to ensure that policy making is informed by a rigorous scientific understanding of a wide range of social issues; to lead on research innovation; and to share our knowledge with the wider research and policy communities.

3. The remainder of this note addresses the four questions posed by the Committee.

#### *How do social scientists use Census data?*

4. Census data are used in a variety of ways by social scientists. Of key interest to our organisation is the use of census data in the design, conduct and analysis of UK population surveys. We would like to draw the Committee's attention to five key uses:

- (a) Census data are used when designing surveys of particular sub-groups that are recorded by the Census. An example of this is surveys of BME groups (described as "ethnic boosts"), for which Census data are used to identify relatively small geographical areas (eg postcode sectors, wards, MSOAs and LSOAs) within which higher proportions of residents from BME groups live. This means that fieldwork can be concentrated in areas which are likely to generate a larger number of interviews per issued address, which makes the fieldwork much more cost efficient.
- (b) In addition to being used for over-sampling specific sub-groups for social surveys, Census data are also used for sample stratification; in other words to select a sample in such a way that it is representative of the population for geographical measures such as region. If other measures that are related to the topic of the survey are also used to stratify (order) a sample before selection, this can help to improve the precision of survey estimates and minimise the risk of selecting a sample that is not representative of the population. Most large-scale social surveys use Census data when selecting the sample. For example, the National Travel Survey (commissioned by the Department for Transport) is stratified by Census estimates of the proportion of households that own or have a car available for use, and the population density of the area derived from Census population counts.
- (c) Census data are used in analyses to investigate the impact of the area in which someone lives on other measures collected from social surveys. NatCen regularly receives requests for Census variables to be added to our survey datasets to provide context and to enable particular research questions to be explored. A recent example involved adding data on religious composition of local areas to British Social Attitudes data in order to explore the influence of this on individual religious beliefs, values and attitudes.
- (d) Synthetic estimation is an approach that has been used to obtain area-level modelled estimates for a range of measures, for example, to estimate health lifestyle measures such as the prevalence of smoking or binge drinking. These are generated by observing the relationship between individual-level survey estimates (eg for smoking status) and accurate external area-level measures, a large number of which are obtained from the Census. This relationship is then used to predict the prevalence of smoking at the area-level. The more accurate and varied the area-level measures available, the more accurate the predicted estimates.

- (e) Finally, census data are used to weight survey data to make them representative of the population of interest. Most large-scale social surveys use mid-year population estimates for calibration weighting to make them representative by age, sex and Government Office Region. This is an essential step that reduces potential sample bias caused by any differential non-response between age/sex groups and across regions. Area-level Census data (such as population density, proportions of ethnic minorities etc) are also used in non-response modelling to help reduce non-response bias. And as survey response rates are likely to remain a challenge, trust and confidence in survey estimates will depend heavily on the quality of the population data being used for weighting.

*What impact will the ending of the Census have on social science research?*

5. If accurate data on ethnicity were not available, then it would not be possible to carry out ethnic boosts for surveys without them being prohibitively expensive. No other source of population data is available for the relatively small geographic areas with the required level of accuracy.

6. Ending the Census will impact on how samples are stratified and weighted, potentially increasing sampling error for surveys. It will mean that the stratification variables will need to be changed for most of the UK's large-scale social surveys. It would also be likely to compromise the accuracy of mid-year population estimates used for calibration weighting, as the time lapse since the last accurate population count increases.

7. Census estimates are a key component of synthetic estimation and so ending the Census would impact on the range and accuracy of modelled estimates that could be produced. It would also reduce the range and accuracy of area-level estimates that could be included when modelling the impact of area-level characteristics on individual measures.

*What alternatives to the Census would provide population and socio-demographic data of equivalent or higher quality?*

8. A rolling large annual survey could be used to replace the Census. The sample size each year would need to be much larger than the current Integrated Household Survey (IHS) and if the IHS model was used it would be necessary to expand it to include the parts of the population that are not covered at the moment (for example, people living in institutions).

9. However, data from such a source would not be as accurate as Census data. All surveys are subject to non-response error. The Census has a response rate of over 95%. The response rate for the IHS, for example, was 65% for the 2009–10 survey. For this to be a feasible option, participation in the alternative, large-scale survey would have to be made compulsory.

10. The risk is that, in the absence of the Census, population data would become increasingly out of date as the years pass. This will make it increasingly difficult for social surveys to provide accurate estimates and explanations of social phenomena, as the information bedrock on which they are sampled, weighted and analysed becomes weaker. Therefore, in our view, even a very large survey based method would not provide the same quality information as the Census.

11. An alternative method would be a population registration system against which other data sources, such as administrative records and survey data, could be matched. However, substantial improvements would have to be made in the scope, completeness and quality of administrative records in order to create a reliable population registration system. Also issues related to public acceptability, data security and data access would have to be addressed.

*What other existing sources of population and socio-demographic data could be improved upon?*

12. The IHS could become the main source of population and socio-demographic data, but the annual sample size of 420,000 would have to be increased to be able to obtain estimates that were precise enough for analysis purposes for small areas (eg MSOAs), even if several years worth of data were combined.

13. The sample size could be increased by including the core module in other large-scale social surveys so that the data could be combined. This would require collaboration between ONS, and other survey commissioners or survey organisations. However, the limitations of such an alternative, raised in paragraphs 7–10 above, would still be valid.

14. As noted above, the scope, completeness and quality of administrative records and survey data would have to be improved if these data sources, in combination with a population register, are to replace the UK Census.

---

## DECLARATION OF INTEREST

15. The key interest to note is that NatCen is a key provider of social survey data to government—conducting studies like the Health Survey for England for the Information Centre for Health and Social Care, the Family Resources Survey for the Department for Work and Pensions and the National Travel Survey for the Department for Transport.

*Kevin Pickering*  
 Head of Statistics  
 NatCen (National Centre for Social Research)

30 November 2011

---

## Written evidence submitted by Tees Valley Unlimited (Census 25)

1. Tees Valley Unlimited is the Local Enterprise Partnership covering the boroughs of Darlington, Hartlepool, Middlesbrough, Redcar & Cleveland and Stockton-on-Tees. In addition to our role in understanding and regenerating the local economy, we also work in partnership with our borough councils on the Census, and provide research, expertise and analysis of the Census and other sources of socio-economic information to them.

2. The fundamental requirements of the Census by Local Authorities are:

- Accurate local authority population estimates as a key element of financial settlements.
- Accurate estimates of particular populations eg children of school age and pre school age, the number of elderly, the number suffering from ill health.
- Detailed information on local areas to inform needs and priorities, and to find out the individual needs of particular areas. The sorts of information that local authorities find useful at local level include single parent households, old people living alone, housing data (including overcrowding), unemployment, qualifications of residents, health issues, and employment and commuting data.

3. The Censuses have provided an important role in providing this information, and the scope and detail available in the Census has steadily increased to match demand and ability of new technology to process it. Even when the Census process has been flawed, the data produced by the Census was extremely valuable. Until recently, the Census was almost the only source of data for local areas. Whilst some of ONS' alternative proposals involve some form of a Census but carried out on a different basis, others anticipate the replacement of the Census by the use of alternative, pre existing data sources. Given our reliance on the Census, we therefore view any proposals to cease undertaking the Census with concern—if replacement arrangements provide less data and / or data of inferior quality, this will have a negative impact on Local Authorities.

## POTENTIAL BENEFITS OF REPLACING THE CENSUS

4. We are not blind to the possibilities of benefits arising from alternative arrangements. One of the drawbacks of the Census is that data is only published every 10 years, and the extensive processing of the data means that it normally takes around two years for each set of data to be published. This means that Census data becomes out of date very rapidly, and the time between Censuses is long, making monitoring of progress over short periods using Census data effectively impossible. The undertaking of more frequent surveys, or use of administrative data on a regular basis could give us more frequent results, with smaller timelags. In addition, use of other data sources means that there is also the potential for publication of data not currently collected by the Census such as data on income and taxation, and second jobs.

## CHALLENGES FOR ALTERNATIVE DATA SOURCES TO THE CENSUS

5. Administrative datasets are complex entities, compiled for their own specific purpose and not primarily intended to supply statistics. They are liable to changes through time as the rules and procedures of the underlying process evolve. There are likely to be inconsistencies between data sources, in definitions, in coverage and completeness. Some datasets under-record populations (eg the electoral register) whilst others over-record (eg GP Patient lists). Data on benefits is a useful source of data, but take-up rates will vary and will never be 100%. Ensuring consistency between data sources would be very important for reliable results, but this would be very difficult and time consuming.

6. There may be a greater reliance on sample surveys, though these have many problems, especially of poor response rates. To make matters worse, these vary widely between different socio-economic groups. Furthermore, sample surveys inevitably have a degree of error in their results, which becomes compounded as smaller areas and smaller populations are examined.

7. ONS has experimented with producing modelled outputs for various indicators for small geographies, where the data is not normally collected (eg Average Household income estimates for small areas). In this technique, the data can be collected by a relatively small survey, and the results for small areas estimated by their socio-economic characteristics. This sort of estimate can on occasion be useful, but they are reliant on

accurate socio-economic data used in the modelling process. Without accurate base data (eg from the Census), such modelled estimates will be even more problematic.

#### RESEARCH CAPABILITIES AT RISK

8. ONS are of course yet to make firm proposals as to what arrangements could replace the current Census, so we cannot evaluate specifically the impact of the new proposals on the work of local authorities. However, the following paragraphs list key outputs from the Census (as currently produced), which would appear to be difficult to replicate by other means. The loss of any of these would be a major concern and would limit research that can be performed.

9. *Data for very small areas*: The most unique ability of the Census is to provide detailed information for small areas—not just for standard geographies like Wards or LSOAs, but for even smaller areas (called Output Areas in 2011) which can be combined to create data for neighbourhoods or other custom areas, and used to produce data for revised administrative areas when boundaries change. Furthermore, the Census can provide detailed information for each output area, not just a few counts. The only reason the Census can supply this level of detail is its near 100% sample size. Apart from the Census, we are only aware of a tiny number of data sources currently available for Output Areas, and these only give a few pieces of information for each area. We feel that the level of detailed information from the Census is invaluable, but will be at risk under proposals to replace the Census, since it is hard to see how this level of detail can be replicated by other means.

10. *Commuting and Migration Patterns*: The Census currently provides the only practical and reliable set of data on patterns of commuting and migration, and the only source of data on this at all for small areas. This data is obviously important to study travel patterns and for transport planning, and contains valuable information on mode of travel and type of job as well as the journey itself. ONS did release some alternative data on commuting from the Annual Population Survey, but this was only available at Local Authority level, and the sample size was insufficient to provide reliable estimates for smaller flows. It is not easy to think of a replacement source of local commuting data, and so this valuable resource could be at risk if the Census is replaced.

11. *Sample of Anonymised*: Records and Longitudinal Study data sets are further unique features of the Census, where sets of sample records can be released to researchers allowing more detailed analysis and social, economic and health characteristics to be tracked from one Census to the next. These are specialist areas but are valuable resources, which are likely to be at risk if there are no further Censuses.

12. *Data on Ethnicity*—there are still no alternative sources of data to the Census on ethnicity below local authority level.

13. *Household Composition*: We are not aware of any other datasets that can provide such detailed information on households and their composition. Since families structures are becoming more complex and diverse, retaining a good source of information to understand this issue will be important.

#### CONCLUSIONS

14. The Census is a vital source of data for local authorities both for its role in providing accurate population counts and to provide detailed data for small areas. We feel that the current format of the Census should not be abandoned until suitable alternative(s) can be defined that provide as much of the current functionality as possible and are proven to be effective.

15. There is a strong case for running a 2021 Census in parallel with the proposed alternative solutions to ensure consistency between the results and to prove the effectiveness of the new procedures.

16. If administrative records are to be used as an alternative source of data, there needs to be significant work undertaken in order to enable them to provide the consistent and reliable data that is needed.

17. There are many important areas of the Census where alternative solutions would appear to be very difficult to achieve, notably the provision of detailed information for very small areas and commuting and migration patterns. The loss of this data would have a significant impact on much research work, especially by local authorities.

18. Whilst the Census has a substantial cost associated with it (approximately £500 million for the 2011 Census), ONS are optimistic that the Census will be of significantly better quality than in 2001 and it seems likely that the money spent will produce a very high quality, detailed, reliable set of data. This can be compared with the recent Regional Fire Control Centre scheme where approximately the same amount of money spent produced nothing. The costs, work needed and difficulties involved of producing alternative solutions to the Census should not be underestimated. We feel that a careful analysis of the costs and benefits of any proposed replacement solutions should be undertaken, and that these should be compared to those associated with continuing to hold a Census.

**Written evidence submitted by CURDS (Centre for Urban and Regional Development Studies)  
(Census 27)**

2021 POPULATION CENSUS

DECLARATION OF INTEREST

0. CURDS has undertaken academic and policy-relevant research for 35 years, and the datasets from successive Censuses underpinned much of its geographical analysis. Some of this work was recently part of our contribution to the Spatial Economics Research Centre funded by two government departments plus the national social science research council. In this long-standing contribution to public debate and academic knowledge of CURDS research into spatial dynamics, far less would have been possible without the Census.

[1] *How do social scientists use census data?*

1. CURDS researchers use UK Census of Population data both directly in secondary data analysis to reveal the spatial patterns and processes at work in the British economy and society, and also indirectly to “benchmark” our own primary research (ie to validate the coverage by our survey of the total population of an area or group of particular interest). In this respect, research by CURDS is perhaps “representative” of that carried out by many social and spatial scientists. *The following paragraphs focus on more CURDS-specific considerations so as not to duplicate the inputs from the learned societies to which CURDS actively contributes.*

2. CURDS researchers use the nearest alternative to the Census—datasets from the Annual Population Survey (APS)—but much of the specialist research for policy makers and others simply could not be done at all with this due to it not providing robust data at a sufficiently detailed area level. The irreplaceability of the Census across a wide range of CURDS research stems from its multiple and inter-linked value for:

- analysing neighbourhood effects (eg on child poverty), due to its local detail.
- highlighting the situation of minority groups, due to its “sample” nearing 100%.
- separating multiple influences on outcomes, due to its wide subject coverage.
- recognising the role of mobility in social and economic inequalities, due to its unique provision of fine-grain data on migration and commuting.

[2] *What impact will the ending of the Census have on social science research?*

3. For almost 30 years CURDS has defined Travel-to-Work Areas for the government by analysing each new Census commuting dataset, and now has begun research for Eurostat exploring the possibility of a “European standard method” for defining labour market areas; it would be incredible if the world-leading British research in this field could not be conducted in Britain in future, but this would be the case if the Census was discontinued.

4. CURDS research which relies on Census commuting data (among other datasets) made a prominent contribution to the recent urban and regional policy developments calling for economic and housing policy delivery at the functional economic area (FEA) scale that, for example, is invoked in relation to Local Economic Partnerships. As yet there is no established way to identify these economic “places” in practice, although there is a consensus that the analysis of the Census commuting data is the single most important way forward to such definitions. Most importantly, the patterns observed by such analyses reveal change in the economic self-sufficiency of areas, with the consequence that these analyses need repeating to update understanding: this will not be possible if the key Census datasets no longer exist.

5. With a more academic emphasis, CURDS has been a major user of the Census Origin-Destination Statistics which include data on migration as well as commuting. These datasets are irreplaceable in measuring the flows between each geographical area and every other one in the country, right down to the neighbourhood scale. CURDS research has, for example, shown how migration can alter the social and economic profile of areas through, for example, “brain drain” trends due to the northern cities being unable to retain the graduates from their universities.

6. It should always be remembered that the data collected in 2011 and any subsequent Census will be used for decades or even centuries into the future, just as the data from past Censuses are prized by researchers (along with the ever growing numbers of family historians). If there had been no Census in 2011 it would not be possible now or later to “go back and fill that gap” in knowledge: we have a duty to future generations to collect data on the present, just as we rely on the data collection carried out in the past by our forbears.

[3] *What alternatives to the Census would provide population and socio-demographic data of equivalent or higher quality?*

7. The idea that there are effective alternatives to the Census is wrong:

- It is sensible to test some candidate alternatives, but this must be done without the expectation that at least one can replace the Census: the test should be “can an alternative provide the data the Census currently provides” and not “which is the best alternative, given that Census will be replaced”.

- None of the proposed alternatives will provide the combination of small area detail and accuracy that make Census data the one essential data resource: the nearest alternative is APS but much specialist research for policy-makers and others simply could not be done at all with this due to it not providing small area data.
- What makes the Census data absolutely irreplaceable is not the small area detail alone but its combination with a rich range of variables, all of which have been exhaustively evaluated to ensure they are the most important for policy-makers and other users: none of the alternative sources would provide data that can answer such a wide range of key questions about communities (whether these are communities defined in terms of neighbourhoods using small area data, or communities defined by dimensions such as religion, which means for example being unable to answer many questions on the equalities agenda).
- An experiment was undertaken by ONS to see whether APS could provide the necessary data on commuting for updating official Travel-to-Work Areas (TTWAs) and it was found that this was very far from true: if the Census is not continued this updating will be impossible which will leaving policy-makers and many others without a resource they have had for many decades (in fact Eurostat is currently discussing the need for such definitions in all countries).
- The increasing ubiquity of GIS means that the need for small area data will continue to grow inexorably so any move away from providing such data could not be more bizarrely or detrimentally timed.
- Much valuable policy-relevant research relies on the high level of consistency between decennial Censuses to understand change over time and this would be lost by shifting to data collected in different ways; a specialist case of this risk is the increasingly valuable Longitudinal Survey whose linkage of Census records provides a unique resource whose value increases with each decade.

8. The only really robust alternative to a Census is provided by a register system such as those in Scandinavian countries where such systems are so familiar and trusted that there is also little public concern at considerable linkage between these systems and many other datasets. However there are considerable public qualms about such systems in the UK (and in fact there is no single unique personal identifier used across official datasets).

[4] *What other existing sources of population and socio-demographic data could be improved upon?*

9. The key limitation of alternative data sources such as ONS social surveys is their sample size, preventing data being made available for smaller geographical areas. The value of these sources derives from them providing “top up” information between Census years so that aggregate trends can be identified. Reflecting on the way that such surveys were designed to complement the robust comprehensive audit of the national population that the Census provides periodically, it can be seen that the arguments which are made for dropping the Census are wrong:

- “Census data “out-of-date” much of the time”—but most analyses for which Census is the only plausible data source are not of rapidly changing trends.
- “Census is expensive”—averaged over 10 years it is not expensive, and then divided by its myriad uses makes it very low cost: almost certainly the options for replacement will cost at least as much over the decade and will add little by being more frequently available (see above) while being of far less use (see below) so overall of much less cost effective.
- “Census information is increasingly difficult to collect”—in fact the signals from the 2011 Census are there was less of a problem than in 2001.

*Prof Mike Coombes*  
on behalf of CURDS (Centre for Urban and Regional Development Studies)  
Newcastle University

30 November 2011

---

**Written evidence submitted by the Welsh Language Board (Census 28)**

**SUMMARY**

The Census of Population is the only source of information providing full details of the geographic distribution of the Welsh language skills of the population. Sociolinguistic research into Welsh based on samples generally requires Census information concerning the Welsh language for the sample design process and subsequently at the results stage to weight the sample correctly to the population estimates. We do not believe there is currently any other source of information concerning the distribution of Welsh language skills amongst the population which could be utilised for these functions. A version of a rolling Census could provide a viable alternative.

---

#### ABOUT THE WELSH LANGUAGE BOARD

1. The Welsh Language Board was established by the Welsh Language Act 1993. The Act was "...to establish a Board having the function of promoting and facilitating the use of the Welsh language, to provide for the preparation by public bodies of schemes giving effect to the principle that in the conduct of public business and the administration of justice in Wales the English and Welsh languages should be treated on a basis of equality, to make further provision relating to the Welsh language...".

2. The Welsh Language (Wales) Measure 2011 will, with effect from April 2012, abolish the Welsh Language Board and establish the Welsh Language Commissioner.

#### DECLARATION OF INTERESTS

3. As part of its function to promote and facilitate the use of the Welsh language the Board has undertaken and commissioned research and funded PhD studentships.

4. The Welsh Language Board was represented by one of its officials on the Census Advisory Group for Wales.

#### *How do social scientists use Census data?*

5. Questions about Welsh have been included in the Census since 1891. The distribution of the population able to speak Welsh (and more recently those with other skills in Welsh) has in itself been a recurring theme of research by geographers through the decades.

6. Sociolinguistic research (including attitudinal and behavioural research into Welsh) involving the drawing of samples often relies on Census information, frequently both at the sample design stage and subsequently at the results stage. At the sample design stage, Census data permits sample sizes to be determined efficiently to cover the section of interest of the total population. At the results stage, Census data permits the results from the sample to be weighted appropriately so they may be interpreted as reflective of the population. At both stages information about the complete population is required.

#### *What impact will the ending of the Census have on social science research?*

7. In some respects it could be argued that without the Census information society would have a poorer understanding of the distribution of Welsh within the population than since 1891. Sociolinguistic research could continue but interpretation of the results would be harder as their relationship to the actual population would be unclear.

#### *What alternatives to the Census would provide population and socio-demographic data of equivalent or higher quality?*

8. Instead of a decennial census covering the whole population at one point of time, a rolling census, covering the whole population in different areas at different points of time might be a viable alternative. As this would ensure that information about Welsh would be updated in some regular pattern this could even be preferable to what some might think would be the more obvious alternative, ie a population register, including details of an individual's skills in Welsh. (The Welsh Language Board is not expressing support for any alternative by noting these two different approaches). The main point is that some means of obtaining complete coverage of the population with some regularity is needed. A sample survey, or a "census" in which participation is voluntary rather than compulsory, could not provide data of equivalent or higher quality.

#### *What other existing sources of population and socio-demographic data could be improved upon?*

9. With the exception of data on school pupils, very few sources provide data concerning an individual's Welsh language skills. Accordingly many sources could be improved upon.

Welsh Language Board

30 November 2011

---

#### **Written evidence submitted by The British Library (Census 29)**

##### 1. INTRODUCTION

The British Library welcomes the opportunity to contribute our views to the enquiry looking at the impact of ending the Census of Population on social science research.

By way of introduction, The British Library is the national library of the United Kingdom and one of the world's greatest research libraries. It supports the UK's research infrastructure, serving business and industry, researchers, academics and students, world-wide, as well as in the UK. We receive a copy of every publication produced in the UK and Ireland via legal deposit; our collection includes well over 150 million items, in most known languages and grows by approximately three million items a year.

The British Library is passionate about providing both physical and digital access to world-class information where and when people need it. Over 16,000 people use the collections of the British Library each day (on site and online). In the course of a year six million searches are generated by the British Library online catalogue and nearly 400,000 people visit the Reading rooms. We have a specific department for Social Sciences, which works to support and inspire the commissioners, producers and users of social scientific research publications and material, including research monographs, reports, theses, conference proceedings, journals and statistical publications. We also acquire research-level material from around the world, again including statistical information.

Our collections include the UK's Census of Population publications from inception, as well as an extensive collection of official publications, statistical reports covering social and economic trends, incomes data, economic information and material from the major international organisations such as the Economic Community, OECD, World Bank, United Nations and so on. All of this material is used by our readers, who include academic researchers (from undergraduates to postgraduates, post-doctoral fellows and up to professorial levels); government and third sector researchers; and researchers from consultancy companies and other private sector organisations.

As well as supporting the social science research community, we are committed to enhancing the public's understanding of social science. We have an active events programme, including a series entitled Myths & Realities, which looks at major social and economic policy issues. Topics have included crime, migration, welfare and education and many of the speakers have used Census data as supporting evidence.

From March to June 2011 we held an exhibition on the Census at the Library: *Census and Society: why everyone counts*. This included material from the late 18th century to the present day looking at the scope and power of the Census to inform public policy and underpin decision-making. Its content ranged from very early maps of employment distribution, to sermons extolling the virtue of completing the census to detailed information of the distribution of different ethnicities. It was visited by almost 50,000 people.

Alongside the exhibition we held a number of conferences and workshops, covering the potential use of the census in research and the value of cohort and longitudinal data (for the research community); on the use and abuse of statistics (for a public audience), and on family history and a historical perspective—broken down by age, sex and religion (also for a public audience). These were all very well-attended.

## 2. DECLARATION OF INTERESTS

No formal interests in the process of running, analysing or producing the Census.

## 3. HOW DO SOCIAL SCIENTISTS USE CENSUS DATA?

Social scientists make extensive and multi-faceted use of data provided by the Census. It is the one, single dataset that provides evidence on the characteristics of whole population on a regular basis, across time, and geography. It's coverage of the whole population provides one major benefit for social scientists: the ability to carry out detailed and, where appropriate, comparative analysis of small areas and neighbourhoods—essential for local service planning, but also providing information on groups often under-represented in even large-scale sample surveys. The Census allows analysis and understanding of the shape of the population on a geographical, regular, and comparative basis; it facilitates understanding of the way household composition, including socio-economic status, age, education, ethnicity, religion, employment patterns, health, travel to work, children and more, all change over time. No other source provides such comprehensive data on the population, and plays such a key role in the development and implementation of both local and central government policy.

Even where there are concerns about representation and under-counting, the Census has the ability to provide more reliable evidence than any ad hoc survey could ever do. In fact, the Census is used as a benchmark against which the validity of samples drawn for ad hoc surveys can be measured. While some of the information gathered by the Census can be found via administrative data (which, there is no doubt, should be used to greater effect), there is no substitute to collecting the data consistently and simultaneously from the population as a whole, at the same time. The Census is unique, and provides a unique historical perspective of an evolving population.

Some examples of specific uses of the Census follow.

### 3.1 Local Authorities

Census data are used by local authorities as the basis for local planning decisions about requirements for housing, infrastructure development and services. The Census is also the basis of allocation of central government funding to local authorities. Neil Storer at a meeting of the Royal Statistical Society in October 2009 gave an example of how the Census is used in Camden:

*“The Census is used for tracking the distribution of the population of the borough, mapping key variations within it (eg overcrowding, spread of students across the borough) and profiling areas at a very detailed level. Census information, mashed with local data, is then used to target scarce resources on local areas and communities. (It provides): affordable commissioned tables; national*

*comparisons and rankings; and essential evidence base for creation of the local development framework, the housing strategy, and the economic development strategy.*

*Between 2001 and 2009 Camden received 2,199 information requests, of which 37% were answered using Census data. Census information has been used to support: funding applications, area-based initiatives such as Sure Start, policy development, eg distribution of green spaces, and service delivery, eg defining catchment areas for libraries and post offices.”*

### 3.2 Third Sector Organisations

Census data made available free of charge via the Neighbourhood Statistics service is relied on by voluntary and community organisations, self-help groups and other bodies active at local levels to provide them with a detailed picture of local populations and socio-economic conditions in small areas. This information is then used to identify need, plan services and support funding bids and campaigns.

### 3.3 Social historians

Historic census data is used in combination with other sources such as the Registrar-General’s Reports by historical demographers and social historians to:

- Map changes in population spread over time.
- Track changes in housing, household composition and employment over time.

Census data can be combined with results from other surveys to present demographic change alongside attitudinal change. Examples include family and household composition, the change from manufacturing to service economy, journeys to work, changing household composition, changing housing conditions, and the use of domestic servants (including the more modern, “help”).

The Online Historical Population Reports website—*histpop*—is an online resource of almost 200,000 pages of all the published population reports created by the Registrars-General of (and its predecessors for) England and Wales and Scotland for the period 1801–1920, including all Census Reports for the period 1801–1937, along with ancillary archival material from The National Archives, plus critical essays contextualising much of the material. The whole resource illustrates the changing demographic structure of Britain and Ireland over this period, and for the first time researchers are able to search and browse the entire collection of published pre-World War Two Census and registration material for the British Isles.

## 4. WHAT IMPACT WILL THE ENDING OF THE CENSUS HAVE ON SOCIAL SCIENCE RESEARCH?

As noted earlier, Census data are more complete than other survey data as they capture the whole population at a given point in time rather than based on sampling. One of the Census’s great strengths has been its regularity and coverage. Administrative data cannot substitute for census data as they do not capture the whole population; not everyone in the population uses the NHS or claims social security benefits. Hard-to-find groups (including those who do not want to be found) tend not to use services, nor appear on formal, official registers. They often do not have access to bank accounts, or credit; children may not attend school and so on. Signifiers of age, household composition, even address, may all be recorded differently across different administrative organisations. So, the Census has the potential to provide a greater representation of marginalised groups. Census data thus form a reliable baseline or framework which validates, and is supplemented and enriched by, survey and administrative data.

In sum, incomplete administrative and sample-based survey data cannot:

- substitute for complete and impartial census data;
- provide micro-level neighbourhood data required by local authorities and third sector organisations; and
- support analysis of long-term trends due to methodological changes.

## 5. WHAT ALTERNATIVES TO THE CENSUS WOULD PROVIDE POPULATION AND SOCIO-DEMOGRAPHIC DATA OF EQUIVALENT OR HIGHER QUALITY?

User needs would have to be met by a combination of sources including survey data and administrative data collected during the normal operation of central and local government eg for education, taxation, payment of benefits, electoral registration, council tax records, local municipal records, births deaths and marriage records, records collected for use in the health service and so on. These are complex datasets, with different coding schemes, units of measurement, geographies and identifiers. We suggest, therefore, that we are some considerable distance from a situation where these sources could replace the unique level of detail, impartiality and completeness of the Census. It is unlikely that administrative data would provide the extensive coverage of the population it obtains as certain groups may not be registered for or access services. It is also a matter of concern that some major surveys are now being discontinued, eg the National Citizenship Survey and the General Lifestyle Survey, leading to a significant loss of primary research data for analysis.

Conducting surveys in many countries—and certainly in Britain—is proving harder as society changes. Issues that need to be addressed if these are to provide data of equivalent quality to the Census include overcoming the reluctance to fill in surveys, public acceptability as public concern about confidentiality, data security/protection and the “surveillance state” has grown, data processing, storage and data linkage methodologies, estimation methods, and standard concepts/definitions. Examples range from the proposed use of “biometric” data on passports and ID cards, to the vast databases built up from consumer loyalty cards. Most recently, stories of large-scale data loss have commanded news headlines across days at a time. The loss of public confidence in the intentions and competence of those collecting data carries the risk of impacting directly on the amount and quality of data recorded, and the ability of government to make informed decisions based on accurate, reliable statistics.

#### 6. WHAT OTHER EXISTING SOURCES OF POPULATION AND SOCIO-DEMOGRAPHIC DATA COULD BE IMPROVED UPON?

Moves to improve upon existing sources of population and socio-demographic data would include:

- Reinstating the major large-scale social surveys as noted above.
- Instituting additional surveys to cover marginalised groups such as homeless people and asylum seekers. Whatever replaces the Census will need to ensure that “even the most overlooked people in our society are represented so the allocation of public services can be more effectively geared towards the most needy”.
- Joining up administrative data to ensure that there are no gaps, and supporting efforts to enable government departments to share data more effectively.
- Ensuring that data is made freely available to researchers outside of higher education through Internet sites such as [www.data.gov.uk](http://www.data.gov.uk). It is crucial that access is maintained and links are kept up to date and not broken. Clicking a few other links from [data.gov.uk](http://data.gov.uk), it appears everything from ONS now sends you to a dead link.
- Development of Geographic Information Systems (GIS) and other tools for data manipulation.

The British Library

30 November 2011

---

### Written evidence submitted by the British Academy (Census 31)

#### INTRODUCTION

1. The British Academy welcomes the opportunity to submit evidence to the House of Commons Science and Technology Select Committee inquiry on the census and social science. The Census has been an important and highly useful resource for social scientists and historians, enabling them to carry out research, analyse policy development and evaluation social behaviour and mobility. It is vital that, should the Census cease to operate in its current format, its replacement continues to meet the needs of social science research. It may also be worth remembering that the Census has long-run value and can be expected to be used by many future generations of historians and social scientists who wish to understand how, and why, society has changed.

2. Along with other members of the UK Strategic Forum for the Social Sciences,<sup>4</sup> we were concerned when we heard about plans to end the Census in its current form last year. Sir Adam Roberts, KCMG, President of the British Academy wrote to the Right Honourable Francis Maude MP on 30 July 2010 urging that the Census 2011 went ahead with sufficient support to ensure its success. We were clearly very pleased to receive confirmation from Mr Maude in August 2010 that the Census 2011 was to go ahead (as it did in March this year), but we remain worried about how this extremely valuable work will continue in the future.

#### *The uses of census data*

3. The Census in its current form provides social science researchers with an extremely large sample size, and arguably almost complete coverage of the UK population (because of the legal obligation to respond to the Census). The advantage of this is that it in turn means that social scientists can look at particular sub-groups of the population where there may be specific issues which may require certain policy responses (minority groups such as travellers is one example). This is hugely valuable as it is often not possible to obtain sufficiently large sample sizes from other sources such as the Annual Population Survey (which is derived from the Labour Force Survey, the Welsh Labour Force Survey, the Scottish Labour Force Survey and the English Local Labour Force Survey).

4. Social scientists use Census data for a number of important additional purposes. The data enable them to assess the representativeness of sample survey data that they might use in their own research. Thus, the Census

---

<sup>4</sup> The Forum brings together a number of funders, commissioners, advocates and users of social sciences to facilitate discussion about areas of shared interest twice-yearly and is hosted by the British Academy. The Forum was established in 2001 to provide an arena to enable the major issues and challenges facing the social sciences to be debated at the highest level, and to ensure the needs of the social sciences are addressed and tackled.

data can be used to provide a yardstick against which to judge the accuracy of the demographic profiles of sample surveys, or to weight sample survey data to ensure they are representative of the presumed national demographic.

5. Social scientists can also use the Census in statistical analysis, by using aggregate level (especially geographical) data contained within it, including region, constituency, ward or super output areas (SOA).<sup>5</sup> These data can be used as controls that assist social scientists in properly estimating the relationship between individual-level variables, as predictors of individual-level variables or as predictors of the variations in the relationships between and / or among attitudes and behaviours.

6. Historians also find the Census extremely valuable, as they are able to make excellent use of the long runs of comparable data to study, among other things, patterns of employment and demography over time.

#### *The impact on social science research of ending the Census in its current form*

7. Without the data produced by the Census in its current form, social scientists will lose an extremely effective yardstick for judging the representativeness (or otherwise) of their own sample surveys. The overall response rate to the 2001 Census was 98%.<sup>6</sup> We do not yet know the response rates for the 2011 survey but it is reasonable to assume it will remain high, given the legal obligation to complete. This response rate is much higher than that in the best of the other large-scale government surveys, such as the Annual Population Survey and the Labour Force Surveys on which it is based. It is also important to note that response rates for surveys have been falling, and will no doubt continue to do so. For example, the response rate to the Labour Force Survey fell from 80.6% in the third quarter of 1997 to 71.2% in the second quarter of 2006.<sup>7</sup> This suggests that any alternative to the Census that is reliant on surveys may be prone to increasing bias over time.

8. The ending of the Census in its current form would have extremely serious implications. We noted in paragraph 3 the value of the Census in providing data about particular small groups; these data could well be less robust in the future without the Census. Other uses are also important. For example, Census data are used for estimating social composition of constituencies or smaller geographical areas—which is not possible to do from Annual Population Surveys (where data below the level of Unitary Authority is not released). There is a major concern in the social science community about the reluctance of government departments to release data with detailed geographical information because of potential identification issues—the data obtained from the Census enable us to overcome this particular issue (we return to this point in paragraph 10).

9. It is true that Census estimates (for example, for constituencies) can rapidly become out of date, hence the need for Office for National Statistics to find ways of estimating mid-Census period adjustments. However, this will only be worthwhile if suitable alternative instruments are put in place to provide information about the changing demographics of the UK public. That is, to replicate (as far as is possible) the functions enabled by Census data discussed in paragraphs 3–6.

#### *Alternatives to the Census to provide population and socio-demographic data of equivalent or higher quality*

10. We note in paragraphs 2–5 important functions of the Census and some particularly valuable uses for the data provided made by social science researchers and other academics. Any alternative would need to continue to fulfil these functions.

11. To be even more specific, regular targeted surveys (eg approximately every five years) could provide a more in-depth study of important population groups (we have already mentioned travellers, but data about young Black men are also less robust than for other groups). Such groups are relatively small, and so sources such as the Annual Population Survey will not provide sufficient numbers of respondents for robust analysis. Even more importantly, such groups tend to exhibit high non-response to standard surveys so that survey-based analysis may generate misleading conclusions. Particular arrangements will be needed in order to be confident that government departments, social scientists and other interested parties can be sure that robust data are available. It is very unlikely that administrative data would be able to satisfy this need.

12. The Census is almost certainly less prone to response bias than even the best government and academic surveys. It is true that this is still an issue for the Census (again as examples, coverage of certain disaffected groups like young Black males, or of undocumented residents is less strong) and so any alternative would need a lot of work to identify response bias and to adjust for any such bias.

<sup>5</sup> See the ONS website for more on SOAs <http://www.neighbourhood.statistics.gov.uk/dissemination/Info.do;jessionid=C9XqTKsC0p1qCDhhvffJgZP5xC1mQLvbzWMh5jTVz71GgSL06thG!-857689500!1321888994656?m=0&s=1321888994656&enc=1&page=aboutneighbourhood/geography/superoutputareas/soa-intro.htm&njs=true&nsc=true&nsvg=false&nswid=1280> [accessed 29 November 2011]

<sup>6</sup> Data from <http://www.ons.gov.uk/ons/guide-method/census/census-2001/data-and-products/quality-of-the-census-data/response-rates/summary/index.html> [accessed 24 November 2011]

<sup>7</sup> Labour Force Survey User Guide—Volume 1: Background and Methodology, Office of National Statistics, table 5.3, p33 available from [www.ons.gov.uk](http://www.ons.gov.uk)

13. It is not simply a case of alternative surveys improving the way they identify and adjust for response bias<sup>8</sup> (for which we mean both bias in survey response rates and on a question-specific basis within surveys). There would almost certainly need to be supplementary investigations, targeted at particular groups that are believed to be at high risk of non-response, in order to compare their profile with those of respondents to particular surveys (for example, the Labour Force Survey). These supplementary investigations might be expensive, and probably much less definitive than the census in the absence of a legal requirement to respond.

14. The Beyond 2011 Programme suggests a number of valuable alternatives to the Census in its current form. It may be possible to achieve high-quality population and socio-demographic data by using a shorter census plus relatively frequent sample surveys to update changing demographic patterns between censuses. It may also be possible to achieve high-quality population and socio-demographic data using the administrative data collected by government in a combination with sample surveys. However, it is not clear that either surveys or administrative data on their own or in combination would meet the needs for authoritative data to provide a yardstick for evaluating survey data (given that administrative data may well be subject to unknown biases), or for analysing small groups.

15. We therefore see value in a “rolling census” which, for example, undertook a legally compulsory census on a rolling basis in different geographic areas each year, perhaps revisiting more frequently areas that have faster rates of change/turnover. A census could then be carried out in a different county or borough each year. This might have the advantage of a census plus the benefit of sampling—but in this case, the ONS would be sampling areas for census treatment (stratified sampling). This may also be a less costly option than the Census in its current form.

16. In discussing the use of administrative data, we think it relevant to alert the Committee to numerous current concerns in the academic community about the availability of these data. Social scientists are increasingly finding problems with obtaining access to both administrative and research data held by government departments. The Government Social Research Unit and the departmental Heads of Analysis have promoted good practice, but there still seems to be a lack of clear, consistent guidelines to cover all departments. Instead, there seem to be many obstacles to social scientists obtaining data to support them in their research. The Committee referred to these issues in its report in July this year on peer review<sup>9</sup> as did the British Academy’s response to the Cabinet Office’s consultation on open data in October.<sup>10</sup>

#### *Improving on existing sources of population and socio-demographic data*

17. As we noted above, data are already available in government departments. An improvement would be to secure, as standard, the open release of all data (suitably anonymised) to archives that grant access to social science researchers and other academics. Protocols for access and use would need to be established, but this would be an improvement on the current situation, and vital if social scientists no longer have the data produced by the Census in its current form.

18. Data on health and education inputs and outcomes that were comparable on a geographical basis (for example by region, constituency, ward, SOA) would be enormously useful for analysing the quality and consequences of policy decisions.

19. Regularly collected data on a wide range of social, political and economic attitudes and behaviours would be very useful in informing policymakers about the consequences of their policies. The cancellation of the Citizenship Survey means that there is no longer a vehicle for informing policy-makers about the consequences (if any) of current “Big Society” policies, to take one example. This means, therefore, that the ONS should look at the whole array of government-funded data collection and not treat the Census in isolation. If other relevant surveys are also being cancelled, the argument for maintaining the Census (or establishing something very similar such as a rolling Census) surely becomes stronger.

#### SUMMARY

20. The Census in its current form has enormous value: to government, to other policy organisations, and to historians and social science researchers. Any changes to the Census must ensure that, by and large, the data produced are still able to provide users with valid, reliable and usable information. If changes to the Census result in efficiency savings, then consideration should be given to investing in new targeted research that can investigate some of the issues around non-response and small population groups. Consideration also needs to be given to the issue of bias in non-Census alternatives such as sample surveys and administrative data. The aim must be to ensure that whatever replaces the Census is at least as effective as the current system.

<sup>8</sup> There is an important point about adjusting for response bias. While one can in principle use weighting in order to adjust, this assumes that there are no interaction effects with non-response. That is, weighting makes the assumption that young blacks who did respond are similar to the ones who did not. These estimates can then be grossed up. However, they may very well not be similar—for example non respondents may be more likely to be undocumented migrants (eg among Black respondents only 5% may be undocumented, but among non-respondents the figure might be 20%, so grossing up using weights could be seriously inaccurate—not that the Census actually tries to ask about legal status of course, but this illustrates the point.)

<sup>9</sup> House of Commons Science and Technology Committee—Eight Report: Peer Review in Scientific Publications, 2011

<sup>10</sup> Available on the British Academy’s website at <http://www.britac.ac.uk/policy/policy-centre-reports.cfm> [accessed 21 November 2011]

The British Academy, established by Royal Charter in 1902, champions and supports the humanities and social sciences across the UK and internationally. It aims to inspire, recognise and support excellence and high achievement across the UK and internationally. As a Fellowship of over 900 UK scholars and social scientists, elected for their distinction in research, the Academy is an independent and self-governing organisation, in receipt of public funding. Views expressed in this submission are not necessarily shared by each individual Fellow.

The British Academy

30 November 2011

---

### Written evidence submitted by The Association of Business Schools (Census 32)

The Association of Business Schools is a membership organisation representing 125 UK business schools and related bodies. Our members teach one in seven students in UK Higher Education and employ over twelve thousand academic staff.

#### 1. *How do social scientists use census data?*

Within UK business schools, census data is used extensively in the research areas of Health Economics, Marketing/Consumer behaviour and Economics more generally.

Looking at Health Economics specifically, the use and allocation of healthcare resources is directly proportioned to age. Mortality rates, incidence and prevalence of common as well as rare conditions are a function of age, gender and ethnicity distribution of the population. Those data are only available through census at national level. The link between income and health is also extensively researched using census data.

#### 2. *What impact will the ending of the Census have on social science research?*

Research of the kind described above would be very difficult to undertake without census data. The impact on the evidence base for public policy making would be considerable.

#### 3. *What alternatives to the Census would provide population and socio-demographic data of equivalent or higher quality?*

The lack of census data can only be replaced with large cohort studies, which are lengthy, expensive and far less significant and broadly applicable than census data.

#### 4. *What other existing sources of population and socio-demographic data could be improved upon?*

See answer to question three.

If you have any questions about our response please get in touch.

Head of Policy Development  
The Association of Business Schools

30 November 2011

---

### Written evidence submitted by the Institute for Jewish Policy Research and the Board of Deputies of British Jews (Census 34)

#### ASSESSING THE POTENTIAL IMPACT OF THE TERMINATION OF THE NATIONAL CENSUS ON THE PLANNING AND PROVISION CAPABILITY OF BRITAIN'S JEWISH COMMUNITY

##### 1. INTRODUCTION

The purpose of this document is to present before the Science and Technology Select Committee, the shared position of the Institute for Jewish Policy Research (JPR) and the Board of Deputies of British Jews on the role of the census in research and policy planning carried out by these organizations on behalf of Britain's Jewish community.

- JPR is an independent research organization, consultancy and think-tank. It aims to advance the prospects of Jewish communities in Britain and across Europe by conducting research and developing policy in partnership with those best placed to influence Jewish life.
- The Board of Deputies of British Jews is the representative body of the British Jewish community. Its research unit carries responsibility for gathering fundamental community statistics that are widely used to inform policy throughout the UK.

Data collection and analytical work conducted by JPR and the Board of Deputies are strongly policy-oriented. The chief consumers of our work are communal planning bodies at a variety of levels, charitable organizations and private enterprises operating within, and for, the Jewish community. Census data on Jews

provide JPR and the Board of Deputies with detailed information on the demographic characteristics and socio-economic make-up of Jews in Britain. No other data source provides direct information of this kind. The principal reason for this is that religion is not captured by any publically available national register in the UK.

We maintain that, as data on religion are not routinely captured, all religious groups will be disadvantaged as a result of potential discontinuation of the census and the religion question contained within it. This is particularly important to religious minority groups which plan and provide group-specific services. In the context of Jewish community, the termination of the census, without an appropriate substitute, would see a return to the pre-2001 Census period, forcing complex, expensive and inaccurate methods to be utilised in order to gather data on the basic characteristics of the Jewish population.

We would further maintain that the gathering of data on religious groups should be in the government's interest. Given Britain's multicultural nature and some of the challenges that exist within and between religious minorities, it is surely essential to have access to data that provide a detailed view of the internal dynamics of each sub-population.

## 2. HOW DOES THE JEWISH COMMUNITY USE CENSUS DATA?

2.1 Both JPR and the Board of Deputies are engaged in large-scale collection of data on Britain's Jewish population. The tradition of data collection in the community dates back well over a century. The types of data routinely collected and analysed include:

- 2.1.1 Vital statistics on Jewish births, deaths, marriages and divorces (none of which are routinely or comprehensively collected elsewhere).
- 2.1.2 Social, economic and political attitudinal data of Jews (collected through community surveys and again not comprehensively captured elsewhere).
- 2.1.3 Data on Jewish faith school education, standards, take-up and attendance.
- 2.1.4 Data on synagogue membership.
- 2.1.5 Census data obtained from the Office for National Statistics (ONS) by means of standard and commissioned tables and through the Samples of Anonymized Records (SAR).

2.2 Data from the 2001 Census has become a principal element of the entire data "warehouse" of the Jewish community. Census data on Jews are important in their own right by providing JPR and the Board of Deputies with detailed information on the demographic and socio-economic characteristics of Jews in Britain.

2.3 Moreover, census data have come to form the backbone for the system of demographic accounting for Jews in the UK. Data from the census combined with data collected independently in the Jewish community (see 2.1) constitute the base for production of annual estimates of the size, structure and composition of the Jewish population in the UK, as well as of summary measures of fertility and mortality pertaining to this population. The census has dramatically increased the scope and accuracy of these types of demographic indicators.

2.4 Further, the census provides vital benchmark data that are used to calibrate community surveys, making them statistically more robust and valuable. In the absence of a full sampling frame for Britain's Jewish population that would enable random sampling, sampling processes for Jewish community surveys have to rely on non-random sampling or particular indirect methods developed for hidden populations. These methods are far from ideal, as they are costly, inaccurate and potentially result in biased samples. In this situation, it is especially important to have a source which provides a picture of the basic distributions of demographic, geographic and socio-economic characteristics of Jews, allowing for assessment of bias and weighting of community survey data, if necessary.

2.5 Below we provide the most pertinent examples of the use of census data by JPR and the Board of Deputies for communal needs. Further examples will be provided in the next section of this document:

- 2.5.1 Census-based age/sex distributions of Jewish population are used to estimate the current and future care needs of the Jewish elderly population, both in terms of provision of residential care facilities and service provision.
- 2.5.2 Census age/sex data, in combination with measures of fertility, are used to plan for the future provision of Jewish faith schools and child care facilities.
- 2.5.3 Census-based data on the age/sex structure and geographic distribution of the Jewish population are used by community providers such as community centres, kosher shops, charities and synagogues, many of whom have come to rely on census data for planning their activities and services.
- 2.5.4 Census data on self-reported health, presence of limiting disability, access to transport, employment status, number of persons per household, relationship status, mixed-religious households, and educational status, as well as comparisons of all these metrics with other communities in Britain, has become an important planning tool for a multitude of organizations working inside the community to provide relief and support to the economically deprived sectors of Jewish population, as well as other disadvantaged and marginal groups within the Jewish community.

### 3. *What impact will the termination of the census have on social scientific research in the Jewish community?*

3.1 It is our view that the termination of the census, without the provision of alternative data collection systems able to identify Jews in a meaningful and comprehensive way, will severely compromise the quality and scope of our knowledge of the Jewish population. Specifically, JPR and the Board of Deputies will no longer possess direct and detailed information on the demographic and socio-economic composition of Jewish population. Additionally, the continuation of demographic accounting (eg the periodic updating of a given population base with data on births and deaths) will have major systemic drawbacks in terms of accuracy and cost as compared with the census. The absence of census data will take the analytical capacity of JPR and the Board of Deputies back to a situation where simply establishing the size of Jewish community becomes a major research project in itself, and one which is ultimately unreliable and uncertain.

3.2 As a consequence, the planning and performance capacity of the Jewish communal, charitable and private sectors will ultimately be curtailed. British society as a whole, and the Jewish population within it, faces multiple challenges in relation to the provision of targeted services (such as education and care support). Access to reliable and timely data on the size and characteristics of the Jewish population is a pre-requisite of informed policy.

3.3 The termination of the census will seriously negatively impact on the following unique aspects of census data:

- 3.3.1 The census uniquely provides the Jewish community with data on ageing, and demonstrates it exhibits a considerably older age profile than the UK average. Demographically, this means that the proportion of persons at very advanced ages (eg 95+) among Jews is relatively high compared with other communities. From a planning point of view this reality means there is a high demand for Jewish care facilities and caring services, requiring for example special attention to the type and location of Jewish care homes.
- 3.3.2 The census uniquely provides the Jewish community with accurate data on fertility levels and the scope of “intermarriage” (Jews married to, or cohabiting with non-Jews). These have a clear bearing on the planning of Jewish educational provision for children and the assessment of potential demand for educational and other communal services of mixed-faith families.
- 3.3.3 The census uniquely provides the Jewish community with accurate data on its geographical distribution down to very small and detailed areas. This means all census metrics can be examined in vastly more detail than was ever possible with even the most comprehensive of surveys. Specifically, the 2001 Census for the first time indicated the very wide dispersion of Jewish people across Britain through all districts and local authorities.
- 3.3.4 The census uniquely provides the Jewish community with data on internal inequalities in terms of educational attainment, access to vehicles, deprivation and hidden disadvantage and, in particular, of child poverty, within the community.
- 3.3.5 The census uniquely provides the Jewish community with data on migration in terms of both international and national population movements. Again this is very important from a planning point of view and is unavailable elsewhere.
- 3.3.6 The census uniquely provides the Jewish community with baseline data which can be used to evaluate the representativeness of Jewish population surveys conducted within the community.

3.4 These brief examples of the additional insights the census has provided gives an indication of what is likely to be lost with the discontinuation of the census especially without a suitable replacement. Planning and performing activities of a multitude of Jewish organizations that rely on these unique insights will inevitably be curtailed, and the community will once again have to resort to informed guesswork and speculation about the nature and future composition of the community. Discontinuation of the census, therefore, represents a threat to good governance in the Jewish part of the third sector.

### 4. WHAT ALTERNATIVES TO THE CENSUS WOULD PROVIDE HIGH QUALITY POPULATION AND SOCIO-DEMOGRAPHIC DATA FOR THE JEWISH COMMUNITY?

4.1 The UK system of quantitative data collection about the population is one of the oldest and most advanced in the world. However, in contrast, for example, to Canada and Australia, the United Kingdom does not possess a strong tradition of collecting data on religion. Most important administrative datasets (such as the National Health Service Central Register, the Electoral Register, the vital statistics registration) do not include information about religion. National sample surveys that do ask about religion (such as the Labour Force Survey and the British Household Panel Survey) only locate a very small number of Jews—typically 5 per 1,000 respondents, which, whilst a true reflection of the community’s relative size, is insufficient for analysis and communal planning.

4.2 A useful change to the current state of affairs would be the initiation of systematic inclusion and publication of information on religion in administrative data files: the vital statistics registration, National Health Service Central Register, and the Cancer Registry being chief examples. Irrespective of this development, collection of information on religion and identification of Jews in principal administrative datasets will result in a significant analytical and practical gain for research and planning agencies in the Jewish sector.

4.3 A number of countries of northern Europe possess population registers, incorporating individual-level data about each citizen and/or resident of a country, and including characteristics such as date of birth or immigration, sex, ethnicity, religion, and place of residence. Typically, population registers rely on (a) the existence of unique identifiers for each individual; and (b) operation of subsystems of registration that “feed into” the register and keep it up to date. Systems of vital registration and border control, for example, are linked to the register, and any changes in personal status of individuals or their movement inside and across country borders are adequately reflected in the register.

4.4 A population register, with a well-developed and functioning updating system, could potentially present an alternative to the census. The experience of countries with the register is instructive, however. Not many countries with the register abandoned the census completely. The chief reason for this is that the functioning of the registration systems updating the register is often imperfect. This is especially true of registration of cross-border migration. Thus, the census continues to provide a way for benchmarking population size, structure and composition.

4.5 We maintain that the intention to discontinue the census should be accompanied by thinking about, and taking practical steps towards adequate alternatives. We identified a population register as such an alternative. It is important to stress that, bearing in mind the research and policy needs of the Jewish population, such a register should include information about individuals’ religion. In parallel, all systems “feeding into” the population register should include this information on religion too. While by no means impossible in theory, currently no such system (most importantly, the system of vital registration) contains this information.

## 5. CONCLUSION

5.1 The potential discontinuation of the national census is a cause of major concern to JPR and the Board of Deputies. A viable alternative to the census (such as a population register of the kind maintained in several Scandinavian countries) must include the collection of data on religion if the community is not to be put at a significant disadvantage in its data gathering capacity.

## 6. FURTHER READING

6.1 The best example of how the British Jewish community has utilized the vast amount of data the census has provided can be seen in the following report:

Graham D J, Schmool M, and Waterman S (2007). *Jews in Britain: a snapshot from the 2001 Census*. London: JPR / Institute for Jewish Policy Research.

6.2 To understand what the removal of the census means in terms of demographic data gathering see:

Graham D J (2011). “Enumerating Britain’s Jewish Population: reassessing the 2001 census in the context of one hundred years of indirect estimates”. *Jewish Journal of Sociology* Volume LIII..

Signatories to this document

*Lord Haskel*  
President, JPR

*Harold Paisner*  
Chairman, JPR

*Jon Benjamin*  
Chief Executive, Board of Deputies of British Jews

*Jonathan Boyd*  
Executive Director, JPR

*Dr Laura Staetsky*  
Senior Research Fellow, JPR

*Dr David Graham*  
Senior Research Fellow, JPR

*Marlena Schmool*  
Former Director, Community Research Unit, Board of Deputies of British Jews

*Daniel Vulkan*  
Research and Information Officer, Board of Deputies of British Jews  
Institute for Jewish Policy Research (JPR) and the Board of Deputies of British Jews

30 November 2011

---

---

## Written evidence submitted by the Suffolk County Council (Census 35)

### SUMMARY

1. Local Authorities make extensive use of census counts to assemble the evidence needed for formulating policy, service delivery and monitoring. The census has many advantages over other data sources as it spans the whole Country, is consistently collected, allows meaningful comparisons between areas and provides many complex variables for small geographical areas that can be tessellated to form customised areas.

2. Without the census counts the quality of the evidence underpinning policies would be less robust.

3. Administrative systems, surveys, modelling or linking of data could all be employed to provide counts of straight forward variables, but estimates of complicated variables or data for small areas could not be so readily provided. Consideration would also need to be given to academic researchers testing hypotheses using original anonymised records.

4. Many administrative systems such as Electoral Registers or registration of vital events could be extended to include other variables, but coordination is required to provide sufficient useful data. Legislation will be required to ensure records are comprehensive and fields completed.

#### 1. *How do social scientists use Census data?*

1.1 Local Authorities make extensive use of Census data despite not specifically being instructed to do so. Councils do not conduct social research using the actual records, preferring to rely instead upon others for this sort of guidance. Instead we use the census counts per geographical area as a key information source.

1.2 There are a number of statutory plans for which we clearly need population characteristics data, such as the Waste Plan, Minerals Plan, Joint Strategic Needs Assessment, Local Transport Plan, Local Development Frameworks and the Survey of the Area (under 2004 Act and proposed to continue in The Town and Country Planning (Local Planning) (England) Regulations [2012]).

1.3 In addition there are several statutory responsibilities for which a background context is desirable, if not essential. Suffolk County Council uses the census results to develop profiles of population subsets living or working in the County, or for specific geographical areas within the County eg a particular town. These profiles form the evidence base for anticipating need directly associated with a particular Council service, or by services provided by other organisations in the public sector eg public health, or suppliers or contractors over whom the Council has a controlling responsibility eg house builders. The greater understanding provided by these profiles facilitates the development of policy, with priorities or targets to address these needs. Finally delivery of these services needs to be planned in a coherent manner. Being able to monitor service effectiveness enables policies to be updated to meet changing circumstances or to adapt them to meet emerging needs and requirements.

1.4 We also use the census data indirectly as a denominator for the nationally collated indicators that County Councils are expected to assemble and within indices of deprivation.

1.5 The census results have two main advantages over many other data sources and it is these that the Council exploits fully to prepare its evidence:

#### (a) Census results are consistently available for small geographies

1.6 The Council needs data for small areas within Suffolk that are consistent with data for the County as a whole and with elsewhere in the Country. Such a suite of data makes it possible to both look at the area in question and put it in context; for example by comparing its characteristics with equivalent nearby areas by the simple expedient of ranking using a key variable or mapping values for this key variable across the County to identify “hotspots” or spatial patterns. The County average provides the benchmark comparison. At a County level, comparisons need to be made with elsewhere in the Country using similar techniques. All these comparisons enable a greater understanding of an area so as policies may be tailored accordingly.

1.9 The advantage of having data for small areas is that these constitute building blocks that form other larger more meaningful geographical areas, with data for the smaller areas simply added to give data for the larger area. Data from the 2011 Census will be provided for output areas that can usually be assembled to give the key administrative areas of Parishes and Towns, plus District Council wards.

1.10 The need for data for small areas is not going to diminish; in fact there is a focus on decentralisation and localism for which more data for small areas will be required. The Localism Bill stipulates that the organisations with which County Councils are legally obliged to cooperate need evidence. The Bill also enables

town and parish councils or new Neighbourhood Forums to draft neighbourhood Plans for which Local Authorities will be required to provide support and which Local Authorities will be legally required to adopt, the inference being that the Plans need to be properly evidenced. Suffolk County Council is encouraging the private sector or third sector organisations to take more responsibility for running some services but they will only be prepared to do so if a full business case itemising existing and potential client base is available.

(b) Census results summarise complex issues

1.11 Categorising the answers to one census question using the answers to another census question amplifies the response to each question by giving extra information. Users are therefore able to explicitly differentiate between separate subgroups to better understand their specific characteristics. The outcome is a suite of data that gives the user a better understanding than if the answers to each question were to be categorised independently. For example, knowing how many pensioners there are in an area or how many people live on their own is valuable, but cross tabulating gives a much richer dataset. The census has the added advantage that all the variables refer to one point in time rather than at different points in time.

2. *What Impact will the ending of the Census have on social science research?*

2.1 Although much of the use Local Authorities make of the census is not specified by law, many of the documents in which the data is used are legally required. Without the census, these would still have to be prepared but would be less well evidenced, with less clarity. We would therefore have less confidence in what they have to say.

3. *What alternatives to the census would provide population and socio-demographic data of equivalent or higher quality?*

3.1 It would be too expensive for Local Authorities to conduct commensurate sample surveys themselves. It is widely acknowledged around the world that people are proving less willing to complete such surveys, so the utility of conducting local surveys may be low. There have been experiments in the past involving enhancements to the electoral register, but these have not, by all accounts, yielded comprehensive coverage due to the obligatory nature of the additional questions. Whilst it is true Councils do have sophisticated administrative recording systems these are incomplete (since they refer to just the clients at the moment under current guidelines) so cannot give a full socio demographic picture of the area or a complete evidence base. Without a national effort, we would not be able to put our clients in context or make comparisons with other Authorities.

3.2 It is possible national databases could be used to yield basic counts of the population with particular characteristics, but these will have the same drawback as locally held administrative systems, namely that the records will only refer to clients. Cross tabulations of one variable against another would also not be so readily achievable. In particular, household characteristics such as household size and composition, both exceptionally valuable data, could not be categorised very readily using straight forward administrative systems. Record matching may be required to link data in one database with data from another to produce the more complex variables, but the weakness of this procedure is the quality of the matching and key used to perform the match. Such linked but anonymised records would be of value though to the academic researcher testing hypotheses. Other data, such as travel to work information can really only be reproduced by a survey, which if it is to produce reliable results for small areas, will need to have a high percentage coverage (like a census) or be repeated over a number of years to boost coverage. It may be possible to use modelling techniques to merge data from administrative sources or from surveys, or the two sources combined, to give approximate estimates but this procedure may still not be adequate for small geographic areas.

4. *What other existing sources of population and socio-demographic data could be improved upon?*

4.1 Some data sources could be easily enhanced to yield additional useful information; for example, vital event registrations could easily include a question about ethnicity information, but without a coherent and comprehensive work plan the results will be uncoordinated and piecemeal and may not be sufficient to yield the basic and necessary evidence required.

4.2 The Electoral Register could be extended to include those who cannot vote (to complete records) and with additional questions that are compulsory; in effect these would be annual censuses.

4.3 There are a number of issues that could arise if existing sources of population and socio-demographic information were to be used for generating data. Legislation may be required:

- to avoid changes to any such existing administrative systems; otherwise the data time series will be disrupted;
- to ensure records are fully completed correctly;
- to ensure records are kept up to date; and
- to enable individuals to inspect their personal details particularly those derived by matching and linking.

However if the administrative processes are onerous these will equate to a census form and not reduce the burden on the public.

Response prepared by:

*Mary Moore*  
Senior Business Development Analyst  
Business Development  
Suffolk County Council

30 November 2011

---

### Written evidence submitted by The Salvation Army (Census 37)

#### 1. *How do social scientists use Census data?*

Founded in East London by William Booth in 1865, The Salvation Army (The SA) is one of the largest, most diverse providers of social services in the United Kingdom. In the United Kingdom and Ireland, The SA has approximately 50,000 members, 4,000 employees and 1,500 Salvation Army officers (full-time ministers).<sup>11</sup>

The SA uses Census demographic data to understand the demographics of a community in which The SA is working, or is considering working. We use the data to assess deprivation, need and development opportunities. The SA produces demographic profiles which report on the population, health, employment, education, housing and deprivation of an area. The Census provides us with a foundation of what a community consists of. Census data directly informs what work The SA will commission and what work we will need to further investigate.

We use ONS population projections and estimates regularly to assess the future demographic of an area, eg ethnicity projections and age group projections. This helps to future proof any investment we may make and influences what activities we may offer to a particular community. Population projections are an important tool for effective action.

We use Census data to show, through longevity, how a community has changed over a period of time.

We predominantly use lower super output area and ward based geographies. We will also make data comparisons with local authority geographies. We do not use Government Office Regional data. The ability for the Census to report on small geographic areas robustly and with high statistical significance is invaluable for The Salvation Army. Many of our locations are inner city areas whose intricacies often get dwarfed by the urban area at large. The Census provides us with an intimate picture of small but unique areas.

In conjunction with Census data we use The Department of Work and Pensions (DWP) data to report on the health of a community and its employment status. We also report on migrants entering an area through the National Insurance registration. We also report on housing figures and school performances. This data is accessed via various sources, including the DWP tabulation tool, NOMIS, local councils, etc. These other sources of information are beneficial as they use the same geographic categories as the census. This enables us to use fixed data parameters when presenting information.

When presenting demographic information on a community, as mentioned above, we use multiple data sources but we make explicit that it is the Census data which forms the foundation of our research.

#### 2. *What impact will the ending of the Census have on social science research?*

We have three main concerns for the removal of the Census:

1. *Lack of statistical verification of other data sets:* the Census is a comprehensive data set that covers the whole of England, Northern Ireland, Scotland and Wales. We do not believe the same high level of statistical significance and coverage can be achieved by amalgamating other national/regional/local data sets. We believe the same geographic levels that are so consistently reported on now will be lost and, therefore, data will not be used to its fullest potential. Ultimately, the communities on which this data is used to make decisions on, will loose out.
2. *Loss of longevity:* it will be difficult and time consuming to ensure any historical comparisons will stand up to statistical rigour.
3. *Ineffectual integrated working:* any amalgamation of data sources will involve in-depth change management across huge organisations (DWP, NHS, DCLG, BIS, Department for Education etc). What guarantees does the charitable sector have that an equivalent integrated data system, like the Census, would be up and running by 2021? Considering the failures of the NHS' National Programme for IT—is this something the Government is confident that it can deliver on?

---

<sup>11</sup> The Salvation Army About Us, [http://www.salvationarmy.org.uk/uki/www\\_uki.nsf/vw-dynamic-arrays/8EA9C0B3D06D8814802578CC002EA25E?openDocument&charset=utf-8](http://www.salvationarmy.org.uk/uki/www_uki.nsf/vw-dynamic-arrays/8EA9C0B3D06D8814802578CC002EA25E?openDocument&charset=utf-8)

3. *What alternatives to the Census would provide population and socio-demographic data of equivalent or higher quality?*

We do not think there are any alternatives to the Census that would provide equivalent population and socio-demographic data. However, we do believe the data is out there and currently collected by different sources. The difficulty, as mentioned above, would be coordinating the data to provide a consistent national, regional and local output.

4. *What other existing sources of population and socio-demographic data could be improved upon?*

The level of migration data is poor. We are only able to report on migrants that have applied for a National Insurance number. We are not able to report on those that haven't or track the through flow of migrants within the UK. The categories of ethnicity as presented through current and previous Census collections are not particularly helpful in mapping the ethnicity of respondents. It is assumed that the UKBA collect data on country of origin and ethnicity in greater detail but how much of this data is publicly available to social researchers?

5. *Declaration of Interest*

Fleur Bragaglia, Researcher for The Salvation Army, was a Census Collector for the 2011 Census in the Barnet area. The Salvation Army responded to the Office of National Statistics Work Programme Consultation on 24 December 2010.

Research & Development Unit  
The Salvation Army

22 November 2011

---

**Written evidence submitted by the Equality Commission for Northern Ireland (Census 40)**

EXECUTIVE SUMMARY

1. The Equality Commission for Northern Ireland is concerned that the ending of the decennial Census of Population could have considerable implications for the understanding of equality related issues in Northern Ireland.

2. Although we recognise that the Office for National Statistics (ONS) "Beyond 2011 Programme" is centred on assessing alternative options for producing the population and socio-demographic data required in England and Wales, we wish to respond to the inquiry as changes in England and Wales will likely have considerable implications for Northern Ireland.

3. The ending of the Census will provide a number of significant challenges to social science research in the UK including the provision of robust statistics on equality-related demographics, small population equality subgroups and small area statistics.

4. In Northern Ireland, the availability of high quality local level data, (which can be reliably disaggregated by equality group) is central to the effective implementation of key pieces of equality legislation such as the Fair Employment and Treatment Order, and Section 75 of the Northern Ireland Act.

5. Surveys conducted using samples of the population, for example the Labour Force Survey, do not currently provide a robust alternative to Census based small area statistics in Northern Ireland, as small sample sizes often do not allow figures to be meaningfully disaggregated to consider equality characteristics at the local area level.

6. Administrative data in Northern Ireland would need to robustly record characteristics across a range of statutory equality (Section 75) grounds and be comprehensively cross-linked to provide any suitable alternative to the Census.

7. There are likely to be considerable cost implications of enhancing UK wide sample surveys (ie by increasing sample sizes) so as to provide robust small area estimates in Northern Ireland; and in comprehensively linking administrative records to form robust estimates of the population at the local area level.

INTRODUCTION

8. The Equality Commission for Northern Ireland ("the Commission") is an independent public body established under the Northern Ireland Act 1998.

9. The Commission is responsible for implementing the legislation on fair employment, sex discrimination and equal pay, race relations, disability and age. The Commission's remit also includes overseeing the statutory duties on public authorities to promote equality of opportunity and good relations under Section 75 of the Northern Ireland Act 1998 (Section 75) and the disability duties under the Disability Discrimination Act 1995.

10. The Commission, along with the Northern Ireland Human Rights Commission, has been designated under the United Nations Convention (UNCRC) as the independent mechanism tasked with promoting, protecting and monitoring implementation of the UNCRC in Northern Ireland.

11. The Commission welcomes the opportunity to provide evidence to the Science and Technology Select Committee Inquiry on the Census and social science, as we acknowledge that any changes to the Census methodology in England and Wales will have profound implications for the UK as a whole.

#### QUESTIONS POSED BY THE COMMITTEE

##### Question 1: *How do social scientists use Census data?*

12. The Census of the Population play a fundamental role in the provision of comprehensive and robust statistics essential for modern policy-making in Northern Ireland and is used extensively throughout the public, private and community and voluntary sectors. The detailed, reliable information provided on, for example, age distribution, spatial distribution, community background, national or ethnic origin etc. is crucially important for developing, targeting, delivering and evaluating policy and programmes and is pivotal in the planning and development of front-line service provision.

13. The Census of Population informs a range of equality relevant work in Northern Ireland—both via research undertaken by the Commission and in facilitating employers and service providers to fulfil their statutory obligations.

14. Northern Ireland has historically been at the forefront of developments in Equality legislation—including in the use of robust data to identify, monitor and challenge inequality. The Fair Employment and Treatment Order, and Section 75 of the Northern Ireland Act are two examples of where high quality, local level data provided by the Census is central to their effective implementation.

15. Article 55 of the Fair Employment and Treatment (Northern Ireland) Order 1998<sup>12</sup> requires employers to monitor the composition of their workforce and of those applying, appointed, leaving or being promoted (Article 52) and review their workforce composition at least once every three years (Article 55), “for the purposes of determining whether members of the Protestant and Roman Catholic communities are enjoying... fair participation” in their employment. This consideration involves the employer comparing the community composition of their workforce to the composition of the local labour supply. Robust, local level data is central to this comparison. The Commission considers that the Census of Population is the only available data source in Northern Ireland which provides such data at the local area level.

16. Recognising the importance of Equality of Opportunity and Good Relations between different groups in Northern Ireland, Section 75 of the Northern Ireland Act 1998<sup>13</sup> placed a statutory obligation on public authorities in carrying out their various functions relating to Northern Ireland, to have due regard to the need to promote equality of opportunity:

- between persons of different religious belief, political opinion, racial group, age, marital status or sexual orientation;
- between men and women generally;
- between persons with a disability and persons without; and
- between persons with dependants and persons without.

17. In addition, without prejudice to this obligation, Public Authorities are also required to have regard to the desirability of promoting good relations between persons of different religious belief, political opinion, and racial group.

18. In delivering these duties, public authorities use data to undertake audits of inequalities to identify key areas within their remit on which they might exert a positive influence. Public authorities also undertake screening and equality impact assessments as part of their policy development processes. Robust local area data is not only required to identify inequalities, but to track the subsequent impacts of any associated policies. The Census of Population is the only data source in Northern Ireland which can provide robust data, by equality ground, where it is required at the local area level.

19. Beyond statutory requirements, social science research using Census data plays a vital role in increasing the understanding of inequality and forms a significant basis for evidence based policy-making in Northern Ireland.

<sup>12</sup> The Fair Employment and Treatment (Northern Ireland) Order 1998 makes it unlawful to discriminate on the grounds of religious belief and/or political opinion in the fields of employment, the provision of goods, facilities and services, the sale or management of land or property and further and higher education. In addition to providing protection from discrimination, the legislation seeks to promote “equality of opportunity” (for those in, or seeking to be in employment or any occupation) and provides for “affirmative action” to secure “fair participation” in employment for members of the Protestant or Roman Catholic communities. <http://www.legislation.gov.uk/nisi/1998/3162/article/55/made>

<sup>13</sup> <http://www.legislation.gov.uk/ukpga/1998/47/section/75>

20. Over the years, the Commission (and its predecessor bodies) have used the Census of Population as a data source for key analyses (for example—Equality Commission (2006) *Census 2001: Limiting Long-term Illness in Northern Ireland*) or as part of the evidence base to inform flagship publications (eg Equality Commission (2007) *Key Inequalities in Northern Ireland*).<sup>14</sup>

21. The Commissions have also supported academic work, using the Census, to consider a range of key issues over time—for example, with regards to fair employment:

- Osborne, R D; Shuttleworth, I G; (eds) (2004) “Fair Employment in Northern Ireland: A Generation On”, Equality Commission / Blackstaff Press.
- Cormack, R J; Gallagher, A M; Osborne, R D (1993); “Fair Enough? Religion and the 1991 Population Census”; Fair Employment Commission.
- Gallagher, A M; Osborne, R D; Cormack, R J; (1994); “Fair Shares? Employment, Unemployment and Economic Activity”; Fair Employment Commission.

22. The Commission also draws on the work of others, for example:

- Shuttleworth, I; Lloyd, C; (2006) *Are Northern Ireland’s Two Communities Dividing?: Evidence from the Census of Population 1971–2001*, Shared Space, Northern Ireland Community Relations Council.

23. Often, a key defining characteristic in the decision to use the Census of Population is not only the precision that a census provides, but that sample surveys in Northern Ireland are often insufficient to allow a detailed consideration of patterns at the local (community) level. Census data has allowed the Commission to establish quantitative data for small population subgroups (for example, Irish Traveller, BME) and small geographical areas for which sample surveys cannot supply robust statistics.<sup>15</sup>

24. Census data has also played a crucial methodological role in equality-related Commission research. Census data has been used to benchmark and improve the quality of information from other data sources used by the Commission such as sample surveys. For example, census-based population estimates have been used to benchmark population-based quotas in surveys and weight data post-survey to ensure a representative sample of key equality groupings in Northern Ireland.<sup>16</sup>

25. The harmonisation of the Census across UK regions has allowed the Commission to establish cross-comparisons between equality groupings in Northern Ireland and other parts of the UK. It has also provided a reliable means of tracking change across social and economic indicators across time.

26. In summary, Census based research has been used to provide at least part of the evidence-base to allow the Commission to identify and describe key inequalities in Northern Ireland; to influence policy-making in Northern Ireland, to assist in the monitoring of fair participation in the labour market; and in the evaluation of legislation and policy in Northern Ireland. As noted above, robust local area data, disaggregated by equality grounds, is also central in allowing employers and public authorities to fulfil their legislative obligations.

*Question 2: What impact will the ending of the Census have on social science research?*

27. The ending of the Census will provide a number of significant challenges to social science research in the UK which have the potential to impact on Commission research given the small population in Northern Ireland. These challenges include the provision of robust demographics on equality grounds; the provision of robust data on small population equality subgroups and at small geographical levels; disaggregation of data on equality grounds, and benchmarking for sample surveys.

28. In general terms, population data on equality grounds plays a pivotal role in policy-making and is essential for monitoring policy implementation and outcomes. Policy without valid and reliable data is potentially costly and wasteful. The ending of the Census will potentially impact on the ability to obtain data on equality grounds as at present the Census is the most comprehensive means of capturing equality-related statistics.<sup>17</sup>

29. Sample surveys in Northern Ireland are often insufficient to allow a detailed consideration of patterns at the local (community) level. Sampling variability due to small sample size will impact on the ability to provide disaggregated data on equality grounds. In the context of population estimates in Northern Ireland this would mean that surveys would need to be larger and would require costly booster samples to ensure appropriate coverage and reliability.

30. The issue of sample size may have wider compliance implications, for example, Article 31 of the United Nations Convention requires that data on disability should be disaggregated to assess the implementation of the Convention and to identify and address barriers.<sup>18</sup> In addition, the Regulation of the European Parliament and the Council on Population and Housing Censuses requires Member States to provide data at the level of

<sup>14</sup> <http://www.equalityni.org/archive/pdf/Keyinequalities%28F%291107.pdf>

<sup>15</sup> See NISRA (2010). *The 2011 Census of Population in Northern Ireland: Proposals*. DFP: Belfast. Section 1.13 and 1.17.

<sup>16</sup> For example in ECNI (2008) *Equality Awareness Survey 2008*. ECNI: Belfast.

<sup>17</sup> Except information on sexual orientation, a question on which was still not included in the 2011 Census.

<sup>18</sup> United Nations Convention on the Rights of People with Disabilities (2007) Article 31.

Local Area Unit 2 (LAU2) equivalent to electoral ward level in Northern Ireland.<sup>19</sup> Currently this information can only be fully captured at this geographical level by the Census.<sup>20</sup>

31. The small size of the total population in Northern Ireland adds a further layer of complexity. The Labour Force Survey (LFS) User Guide notes that *“it is the nature of sampling variability that the smaller the group whose size is being estimated, or from which an estimate is being derived, the less precise that estimate is”*.<sup>21</sup> Due to the relatively small size of the Northern Ireland population, population estimates from surveys may be unreliable, in particular for small population subgroups (eg ethnic minorities) or small geographical areas as they may be subject to a high degree of sampling variability.

32. Concern with the quality and reliability of survey statistics in the absence of a Census have been raised by the Northern Ireland Statistics and Research Agency (NISRA). NISRA has stated that sample surveys cannot provide robust statistics for small population subgroups, such as ethnic minority populations<sup>22</sup> and at small geographical levels such as electoral ward.<sup>23</sup> In presenting the business case for the 2011 Census, NISRA was of the view that *“the Census was still the most authoritative source of information for a wide range of uses. It provides a snapshot of the country, with consistent and comparable information for small areas and sub-populations, and allows multivariate analyses that are not feasible using any other data source”*.<sup>24</sup>

33. The ending of the Census may also create challenges in the provision of statistics for persons living in communal establishments (CE) that may receive limited coverage in sample surveys, such as the LFS. In 2009, the ONS stated that *“the decennial Population Census is likely to remain the most reliable integrated source of CE population data”*.<sup>25</sup>

34. Other challenges in relation to the use of administrative and/or survey data include the consistency and stability of concepts used in survey and administrative data. The quality and robustness of equality data for the population is contingent on the consistency of concepts used between surveys. For example, research<sup>26</sup> has identified that the concept of disability used in a survey can impact on population estimates of disability. The challenge would be to ensure stability of concepts between surveys and administrative records used for the purpose of gathering census information and longitudinally across time.

35. In addition, inconsistencies between regional administrative records and surveys due to differences in the regional context have the potential to impact on the harmonisation of UK statistics and will need to be addressed. For example, the Northern Ireland version of the Family Resources Survey (FRS) and Labour Force Survey (LFS) use a different sample design and/or sampling frame than in Great Britain.<sup>27</sup>

36. There is a related danger that the ending of the Census may result in de-harmonisation of statistics between the UK regions and greater challenges in making robust inter-regional and/or whole UK comparisons. It is, therefore, crucial that continued close cooperation and joint working is carried out on Census policy between England and Wales and other UK regions to ensure harmonisation of any alternative arrangements.<sup>28</sup>

*Question 3: What alternatives to the Census would provide population and socio-demographic data of equivalent or higher quality?*

37. We offer a limited response to this question, in seeking to confine our response to areas within our core area of expertise.

38. We consider that the Census provides the most reliable source of population characteristics in the UK and is used to improve and quality assure information collected from other data sources such as sample surveys. Other sample surveys, for example the LFS, are comparatively less reliable, in that they provide estimates of population characteristics rather than exact measures.<sup>29</sup>

39. NISRA is of the view that *“without the Census, surveys would be less reliable or would need to be larger and more costly”* to ensure robustness.<sup>30</sup> Given the issues with sampling variability noted earlier, this may present challenges in relation to the small population of Northern Ireland.

40. In light of the particular challenges for Northern Ireland the Commission is of the view that should the UK government decide to discontinue the decennial Census the Northern Ireland Executive should be strongly encouraged to introduce a Northern Ireland Census. The Commission recognises that the issue of scale and

<sup>19</sup> See NISRA (2010). The 2011 Census of Population in Northern Ireland: Proposals. DFP: Belfast. Section 1.17.

<sup>20</sup> Ibid.

<sup>21</sup> ONS (2009). Labour Force Survey User guide—Volume 1: Background and Methodology. ONS: UK.

<sup>22</sup> NISRA (2004). The future provision of demographic statistics in Northern Ireland (towards the 2011 Census). Information paper. NISRA: Belfast.

<sup>23</sup> NISRA (2010). The 2011 Census of Population in Northern Ireland: Proposals. DFP: Belfast. Section 1.17.

<sup>24</sup> NISRA (2010). The 2011 Census of Population in Northern Ireland: Proposals. DFP: Belfast. Section 1.13.

<sup>25</sup> ONS (2009). Labour Force Survey User Guide—Volume 1: Background and Methodology. ONS: UK.

<sup>26</sup> MSA-Ferndale (2004). Review of disability information project for DFP NISRA. DFP: Belfast.

<sup>27</sup> Rafferty A (2010). Introduction to Complex Sample Design in UK Government Surveys. ESDS:UK.

<sup>28</sup> See NISRA (2010). The 2011 Census of Population in Northern Ireland: Proposals. DFP: Belfast. Section 1.16.

<sup>29</sup> Department of Enterprise, Trade and Investment. (2006). Northern Ireland Labour Force Survey: Performance and Quality Monitoring Report. DETINI: Belfast.

<sup>30</sup> See NISRA (2010). The 2011 Census of Population in Northern Ireland: Proposals. DFP: Belfast. Section 1.8.

cost of a Northern Ireland Census would be an important consideration, however, Ireland which has little more than double the population of Northern Ireland has a quinquennial census.

Question 4: *What other existing sources of population and socio-demographic data could be improved upon?*

41. Surveys conducted using samples of the population, for example the Labour Force Survey, do not currently provide a robust alternative to Census based small area statistics in Northern Ireland, as small sample sizes and associated sample error often does not allow these estimated figures to be disaggregated to consider equality characteristics at the local area level. In the context of robust population estimates in Northern Ireland this would mean that surveys would need to be larger and would require costly booster samples to ensure appropriate coverage and reliability.

42. Administrative records in Northern Ireland do not always collect information on all equality-grounds covered by Section 75. Where data is collected this may be confined to a limited number of grounds (eg age or gender) or data may be incomplete. This is because administrative records are by design focussed on their administrative purpose and thus may not meet wider user needs or may only provide a proxy for the underlying population.<sup>31</sup>

43. To present an alternative to the Census, administrative records, registers would need to be comprehensively linked and cross-tabulated<sup>32</sup> and capable of tracking change on a longitudinal basis. This may require a comprehensive population register and system of linkage, such as that used in Denmark, Finland, Norway and Sweden,<sup>33</sup> and such data linkages may require legislative change.<sup>34</sup> Research in Northern Ireland has however indicated that many such records are subject to gaps in the range or quality of data.<sup>35</sup> It is also likely that none of these would at present be able to provide the full range of equality demographics required by the Commission and provided for by the Census.

44. NISRA has indicated that the Census provides the basis for equality monitoring by government, in particular “census information on age, sex, ethnicity, religion and disability help to identify the extent and nature of disadvantage and to measure the success of equal opportunities policies”.<sup>36</sup>

45. The Commission is of the view that should an alternative to the Census include administrative records, this must include the recording of equality demographics on Section 75 grounds. This level of information is not only useful to the work of the Commission and equality-related social science research in general, but would be useful to public authorities in fulfilling their statutory duties under Section 75 of the Northern Ireland Act 1998.

46. The Commission acknowledges that administrative data cannot always be completely accessible due to issues of privacy and confidentiality. If administrative records were used to gather Census data the issue of data access for research whilst maintaining the confidentiality of individual records would need to be addressed.

47. Any decision to move away from a Census of Population must also recognise that there are likely to be considerable cost implications of enhancing UK wide sample surveys (ie by increasing sample sizes) so as to provide robust small area estimates in Northern Ireland; and in comprehensively linking administrative records to form robust estimates of the population at the local area level.

The Equality Commission for Northern Ireland

December 2011

---

---

<sup>31</sup> NISRA (2004) The future provision of demographic statistics in Northern Ireland (Towards the 2011 Census). DFP: Belfast.

<sup>32</sup> NISRA (2004) The future provision of demographic statistics in Northern Ireland (Towards the 2011 Census). DFP: Belfast.

<sup>33</sup> See Ralphs M and Tutton P (2011) Beyond 2011: International models for census taking: current processes and development. Office for National statistics (ONS): UK.

<sup>34</sup> See NISRA (2010). The 2011 Census of Population in Northern Ireland: Proposals. DFP: Belfast.

<sup>35</sup> Ijpelaar J, Marshall D, Paul S and Moylan K (2011) Quality Report / User Guide—Northern Ireland Population Estimates. NISRA Occasional Paper No. 32: NISRA.

<sup>36</sup> NISRA (2010). The 2011 Census of Population in Northern Ireland: Proposals. DFP: Belfast. Section 1.8.