House of Commons

Science and Technology Committee

The Census and social science

Third Report of Session 2012–13

Report, together with formal minutes, oral and written evidence

Additional written evidence is contained in Volume II, available on the Committee website at www.parliament.uk/science

Ordered by the House of Commons to be printed 12 September 2012

HC 322
[Incorporating HC 1666–i to -iii, Session 2010–12]
Published on 21 September 2012
by authority of the House of Commons
London: The Stationery Office Limited
£0.00

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

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.

Contents

Report

Summary

We undertook an inquiry into the proposed changes to the census not to consider whether there should or should not be another census but to consider the use of the data from the census by the Government, whether there were elements of the census that would be irreplaceable by other means and if the business of Government would be seriously impacted if census data was lost or changed.

We are committed to the principle of the Government basing its policy on evidence and social science based evidence needs to be just as robust as that from any other science. Social science in Government is strategically coordinated by a heads of analysis group and social scientists in Government departments are supported by a cross departmental resource based in the Treasury. We were concerned however that there was no clear voice, like the Chief Scientific Advisor provides in other sciences, to influence at a Ministerial level. We are more concerned about the lack of a Minister who could answer for Government as a whole whether social science provision was adequate and whether the data from the census and other sources was fit for purpose.

We found that there were many other social science surveys that would appear to overlap with the census. We were informed that these surveys could not replace the census and indeed depended on census data as a source of calibration. This dependence would appear to be due to the fact that there is no other survey that can approach the breadth of the census. We did hear of good use being made of non-census surveys to provide equivalent data in a number of areas in a much shorter timescale than can be achieved through the use of a decadal census.

There are a number of other key qualities of the census could not be replaced by other means. Census data provides a snapshot of the whole country at a moment in time which is invaluable to historians and to detect trends in the recent past; it also allows comparisons to be made of different areas in the country more accurately. The census also provides a means to recruit to longitudinal studies which, we are convinced, are the envy of social scientists in other countries. A wide variety of organisations and local historians told us that they depend on the availability of census data as they could not afford to finance any equivalent studies.

The key disadvantages of the census would seem to be the timeliness of the data. Census data is always at least two years out of date and up to twelve years just prior to a census day. In areas where the population changes rapidly census data would rapidly become non-representational. There is also a concern that the very presence of such an obvious dataset means that new and innovative solutions are not sought for social science as they can lean on census data whether or not it is the best solution.

We have concerns that social science could suffer if the census was to be discontinued without serious consideration as to how this data would be replaced. We have been told that surveys and administrative data can be used effectively but we have concerns that this would not easily scale to a nationwide coverage. There would need to be a serious consideration of how administrative data could be brought to a national standard to allow

it to be more easily used as a replacement for census data. A particular concern would be recruitment to existing longitudinal studies.

However, we anticipate that the absence of a census would also potentially stimulate a considerable amount of innovation in social science and examination of how to produce social data of an equivalent standard, but to much quicker timescales, than the current census data.

1 Introduction

Early Census activity in England

1. The ominous title of the Domesday Book often seems to obscure the fact that it was a simple census activity. The website of the Office for National Statistics (ONS) describes the Domesday book of 1086 as the "first thorough survey of England", adding:

> The Domesday Book paints a very detailed picture of life in Norman England. So in these terms it can be thought of as our first census. But unlike the modern census, it did not provide an accurate count of the people living in England then. Also unlike the modern census, Domesday's purpose was to establish the ownership of assets, so that owners could be taxed on these possessions.[1]

Thus the first recognisable census in England provides some insight into what was important in that society: the King wished to know what tax revenues might be raised from the populace. Arguably, every subsequent census has revealed what the government of the day thought it important to know at that time.

2. The incentive for the modern census was not money but the fear of overpopulation. This fear was fuelled by the publication in 1798 of Thomas Malthus's *Essay on the principle of population*, which suggested that population growth would soon outstrip supplies of food and other resources. Concerned at this alarmist view of the future, people began to see the need for a census.[2] Parliament passed the Census Act in 1800 and the first official census of England and Wales was held on 10 March 1801. Since then there has been a census every ten years except for 1941, during the Second World War.

3. There have been constant changes to the questions asked in the census with more and more detail required to provide greater 'granularity'[3] for a population that is probably subject to greater and faster change than ever before.

4. There is also a hugely different social science context where academics, charities and commercial organisations are all collecting and using social data to better serve their communities and customers. Digital recording of that information also increases the potential to share, compare and broaden the formal and informal data available about the population of the UK.

Beyond 2011

5. The Treasury Select Committee report *Counting the Population* in May 2008 recommended that:

[1] "Early census taking in England and Wales", *Office for National Statistics*, http://www.ons.gov.uk/ons/guide-method/census/2011/census-history/early-census-taking-in-England-and-Wales/index.html

[2] "The modern census", *Office for National Statistics*, http://www.ons.gov.uk/ons/guide-method/census/2011/census-history/the-modern-census/index.html

[3] Granularity is used to indicate how small a particular social group might be represented within the data. Looking at Constituencies is a useful level of geographic granularity but does not show nuances. Granularity at a postcode level is a much finer geographic granularity. There are many ways to group people and as the groupings get more detailed then it is referred to as the granularity becoming finer.

> the Statistics Authority set strategic objectives to ensure that the data gathered throughout the UK can be used to produce annual population statistics that are of a quality that will enable the 2011 Census to be the last census in the UK where the population is counted through the collection of census forms.[4]

6. Subsequently, in May 2010 Sir Michael Scholar, Chair of the UK Statistics Authority, wrote to the Minister for the Cabinet Office:

> As a Board we have been concerned about the increasing costs and difficulties of traditional Census-taking. We have therefore already instructed the ONS to work urgently on the alternatives, with the intention that the 2011 Census will be the last of its kind.[5]

The Beyond 2011 Programme was formally established in April 2011 to consider the alternatives to running a census in 2021. The ONS says: "Close collaboration is in place with the devolved administrations in Scotland and Northern Ireland to ensure that the obligation to produce consistent UK statistics is met".[6] *Beyond 2011* will report its findings in 2014.

Our inquiry

7. We decided to undertake a short inquiry to examine the potential impact of the ending of the census on social science research; we anticipate that our recommendations will feed into the ONS *Beyond 2011* consultation. We announced our inquiry on 9 November 2011 and issued a call for evidence based on the following terms of reference:

- How do social scientists use Census data?
- What impact will the ending of the Census have on social science research?
- What alternatives to the Census would provide population and socio-demographic data of equivalent or higher quality?
- What other existing sources of population and socio-demographic data could be improved upon?

8. We received 41 submissions in response to our call. We also held three evidence sessions during which we took oral evidence from four panels of witnesses:

i. On 7 December 2011 we took evidence from: Professor David Blane, Deputy Director, ESRC International Centre for Life Course Studies, Professor Heather Joshi, President, Society for Lifecourse and Longitudinal Studies, and Professor Leslie Mayhew, City University.

[4] Treasury Select Committee , Eleventh Report of Session 2007-08, *Counting the population*, HC 183

[5] Quoted in "Background to Beyond 2011", *Office for National Statistics*, 1 December 2011 www.ons.gov.uk/ons/about-ons/what-we-do/programmes---projects/beyond-2011/background-to-beyond-2011/index.html

[6] *Ibid*.

ii. On 14 December 2011 we took evidence from two panels. First: Professor Tim Allen, Local Government Association, Aleks Collingwood, Joseph Rowntree Foundation, Professor David Martin, Royal Statistical Society, and Professor Philip Rees, Royal Geographical Society; followed by Adrian Alsop, Director of Research and International Strategy, Economic and Social Research Council, Jeremy Neathey, Deputy Director of Policy, Economic and Social Research Council, Glen Watson, Census Director, Office for National Statistics, and Peter Benton, Deputy Director, Office for National Statistics.

iii. On 18 January 2012 we took evidence from the joint Heads of the Government Social Research service: Jenny Dibden, Department for Work and Pensions and Richard Bartholomew, Department for Education.

9. We would like to thank those who responded to our call for evidence, especially those who provided oral evidence. We are grateful to John Pullinger, Librarian of the House of Commons, for his helpful insight and commentary.

10. The census as an operation is not the subject of our inquiry. The Public Administration Select Committee scrutinises both the Office for National Statistics and the UK Statistics Authority and is expected to publish its own findings on the operation of the census. Issues relating to the conduct of the census, the questions asked in the census and the value of that data to the UK Government are core to the remit of that Committee. Our interest is in the conduct of social science and the potential impact of significant changes to, or indeed a discontinuation of, the census.

11. Our report begins by looking at the use of social data across society. We then consider the value of the census to those who use social data and some of the limitations of that source. Finally, we consider the alternative sources of data and what it would be important to preserve if there was never to be another census.

2 How the census is used currently

Social science in central government

12. We have consistently argued for evidence-based policy, and social science has an obvious role to play in providing evidence to plan future government spending. It is important that evidence from the social science services be properly considered when Ministers are making spending decisions. Richard Bartholomew, Joint Head of the Government Social Research service, gave an example where social science had played a role in recent decisions made by Government:

> For example, the pupil premium in my Department is certainly influenced by the concern about social mobility and improving the chances of children from disadvantaged backgrounds through a specific set of interventions. Clearly, you need a number of interventions with different age groups for the purposes of social mobility, but it is crucial in terms of the arguments as to why one should invest both in early years and the pupil premium.[7]

13. A cross-departmental social science resource is provided by the Government Economic Service (GES) and the Government Social Research service (GSR), which are responsible for giving evidence-based advice to support the rationale, objectives, appraisal, monitoring, evaluation and feedback to support effective policy making and delivery.[8] The Government Economic and Social Research Team provides professional support and leadership for social researchers and economists across all government departments.

14. Jenny Dibden from the Department for Work and Pensions explained to us that a Heads of Analysis group, on which sit all the heads of profession and the Government Chief Scientific Advisor, currently ensures strategic co-ordination of social science throughout central government.[9] Richard Bartholomew of the Department for Education highlighted the need for social science to be integrated into each department's needs "if we are to have an influence in making sure the evidence is available to make informed policy decisions".[10] We were told that within departments, all the various heads of professions, departmental directors of analysis and chief scientific advisers worked together. Jenny Dibden concluded: "That collective working and focus on issues within but also across Departments means we have a serious amount of input in deciding what the agenda is and securing it".[11]

15. The provision of social science advice is distinct from how Government receives other forms of scientific advice. Almost every Government department has a Chief Scientific Advisor and these are grouped together under Professor Sir John Beddington, the Government's Chief Scientific Advisor. There was a Government Chief Social Scientist but

[7] Q 141

[8] Q 119

[9] Q 103

[10] Q 109

[11] Q 103 [Jenny Dibden]

the holder retired in 2010 and his functions were split between the two civil servants already mentioned, Richard Bartholomew from the Department for Education and Jenny Dibden from the Department for Work and Pensions.

16. In 2011, the Science and Technology Committee of the House of Lords made two recommendations about leadership on social science in the Government:

> We recommend therefore that, at the earliest opportunity, the Government appoint a Chief Social Scientist (CSS) who reports to the GCSA and is an independent expert in social science research to ensure the provision of robust and independent social scientific advice.
>
> We further recommend that the Government consider whether existing mechanisms for the provision of social scientific advice, in particular advice on behavioural science, are fit for purpose. This should include consideration of how departmental CSAs and social scientists within departments can best work together to provide up to date social scientific advice to support evidence-based behaviour change interventions.[12]

The Government's response to that committee said: "The Government will therefore give careful consideration to the idea of appointing a Chief Social Scientist, which will involve weighing up the potential benefits against any potential costs".[13] In this inquiry we have found it difficult to establish who is responsible for taking a broader view of social science and its use in government. There is no obvious direct Ministerial responsibility for social science and therefore no 'customer' to decide where the benefits of the census, or changes to it, might lie or what might constitute best value for government as a whole.

17. **We are content that the structure of social science research in Government is organised in an effective manner to provide the information required by Ministers in planning departmental spending. We are, however, not convinced about the ability of social scientist advisors to influence Ministers when departmental considerations conflict with those of Government as a whole.**

18. **We recommend that the Government give a senior Minister direct responsibility for social science who would take a cross-Government view in Ministerial discussions, respond to Parliamentary questions and reports from Select Committees. We consider this essential, especially in the event that greater departmental co-operation is required to source data in the event of the census being discontinued.**

Use of census by public bodies

19. The UK supports social science research through the Economic and Social Research Council (ESRC). The ESRC told us that it had invested extensively in a programme to support data access services, expert support and training, research and development to

[12] House of Lords, *Behaviour Change*, Second Report of the Science and Technology Committee, Session 2010–12, HL Paper 179, paras 4.23–4.24

[13] Cabinet Office, *Government Response to the Science and Technology Select Committee Report on Behaviour Change*, 15 September 2011, www.parliament.uk/documents/lords-committees/science-technology/behaviourchange/BCGovernmentResponse.pdf

underpin social science research using UK census data. It argued: "This investment indicates the widespread and enduring importance of the census datasets as a key source for understanding British population and society. At present, over 20,000 registered users in UK Higher and Further Education take advantage of these services".[14]

20. The ESRC provided us with information on the top ten surveys (including the census) used by social scientists; those surveys asterisked are collected by the ESRC, otherwise the data is collected by the ONS or other government departments:

1. Census data
2. Labour Force Survey
3. Health Survey
4. British Household Panel Survey*
5. British Social Attitudes Survey*
6. 1970 British Cohort Study*
7. General Household Survey
8. British Crime Survey
9. National Child Development Study*
10. Workplace Employee Relations Survey*

21. The ESRC also funds the National Data Strategy, which is intended to identify major gaps or weaknesses in the current range of data resources, not only in terms of the limitations that lack of appropriate data may place on areas for possible research, but also in relation to technical problems about data access and data security and ethical issues surrounding the collection, curation, preservation and/or reuse of particular types of data.[15] The ESRC's second such strategy document covers the period 2009–2012, and a major part of it concerns making better use of the data that is already collected. **It was surprising that we did not hear more about this strategy in our oral or written evidence as it would seem central to the long term provision of data for social science, especially if the census is to be discontinued.**

22. Professor Philip Rees of the Royal Geographical Society emphasised the importance of census data in government planning:

> The census has a vital role in many of the resource allocation formulae that central Government Departments use. For instance, in allocating the NHS budget, the Department of Health has to have reliable information on the number of patients and the potential number of patients in future—the number in [Primary Care Trusts] currently and in future, and in clinical commissioning groups, and perhaps even individual practices or individual patients. It needs that sort of flow of very accurate information, and if you don't have it, you misallocate large chunks of Government spending.[16]

[14] Ev 37

[15] "UK Strategy for Data Resources for Social and Economic Research", *UK Data Forum*, 2011 www.esrc.ac.uk/funding-and-guidance/tools-and-resources/research-resources/data-services/NDS/index.aspx

[16] Q 48

23. Uses for census data include:

> funding formulae used to allocate central government resources to other organisations such as the devolved administrations (via the Barnett formula), local authorities and health bodies.
>
> policy development and planning by central, local and regional government, in areas including housing, transport, employment and health.
>
> democratic engagement: the Boundary Commission takes account of population change to reshape constituency boundaries. MEP representation is calculated using population figures based on the census.[17]

24. Glen Watson, Office for National Statistics director for the 2011 census for England and Wales, pointed out that the ONS had recently carried out a cost benefit analysis of the census in respect of the money spent and the money saved by use of the data:

> The estimated economic value of conducting the census far exceeded that £480 million [cost]. The business case prepared for the Government and considered by the Treasury in an earlier spending review estimated that the economic value of the census to the UK was probably in excess of £1 billion. We did not try to map every single case and every single use of census data that has been made, because the number was starting to get so large; we thought that we had probably gone far enough.[18]

Use of the census by industry

25. Much of our core infrastructure is now dependent on private investment for maintenance and development. Indeed a key rationale for privatisation was that private companies would be more able than Government to finance investments in necessary infrastructure. We had anticipated that the census would be a useful source of information upon which the future needs of water, energy and telecommunications infrastructure might be based. We had therefore anticipated that there would be concerns about the potential loss of this data source on the part of, for example, water and energy companies. However, despite central trade associations being made aware of the inquiry, none submitted evidence to us.

26. There are many large commercial datasets, for example the information from store loyalty cards. However, we were told it was doubtful whether the data would be valuable outside the context in which it was collected:

> Tesco rely on their data for marketing and they are a very successful company. Presumably, it is very data-driven, and, presumably, they do get something right for their purposes, but the words "for their purposes" are quite important. [...] The question is not whether they are very good data sets [...] but they would not substitute for basic demographic data consistently and transparently collected to a

[17] House of Commons Library Standard Note, *Preparations for the 2011 Census*, 9 March 2010 www.parliament.uk/briefing-papers/SN05230

[18] Q 72

standard somewhere by the public sector, or triggered and managed by the public sector.[19]

27. We were surprised that our call for evidence received no response from any industrial or commercial organisations that might be expected to show an interest in population statistics for business-planning purposes.

28. **Utilities presumably consider their investment decisions against projections made on the basis of local authority structural plans. However, as we note below, local authorities depend, at least in part, on the national census to test the validity of their own projections. We recommend that the Government investigate the potential for sharing anonymised social data collected by utilities.**

[19] Q 65 [Professor Allen]

3 Value of the census

29. Census data play a significant role in the research of most of the social scientists from whom we received evidence. In the words of the Institute for Fiscal Studies, the census is "the key source of information on population statistics".[20] A number of organisations highlighted the value of continuing to collect census data for:

- the very high coverage achieved due to the legal obligation to respond;[21]
- the longitudinal studies that have been taking place providing detailed information about the impact of changes across UK society;[22]
- the level of detail achieved despite the wide geographical coverage;[23]
- the fact that the census is a source of data independent of official information (ie information gathered by public bodies for their own specific purposes);[24]
- the increasing difficulty of comparing data from different authorities;[25] and
- the fact it provides a reference point based on a sample approaching the whole UK population.[26]

We will briefly consider each of these in turn.

National coverage

30. Householders are statutorily required to complete a census form on behalf of those sharing their property, and to do so accurately. The Office for National Statistics expends considerable effort in ensuring that as many people as possible comply with this requirement. Of all the sets of social data listed in paragraph 20 above, only the census has a sample size significantly approaching the whole UK population. Population studies are often judged on their sample size: as size increases, potential bias due to sampling decreases.

31. The value of a nationally co-ordinated data-collection system was expounded by Professor Martin of the Royal Statistical Society:

> My personal view is that the independence of the census as a nationally consistent exercise is one of the great strengths that any replacement system would have to have. It would be exceedingly difficult to mandate organisations of different sizes and

[20] Ev w19

[21] For example, Ev w38

[22] For example, Ev w22, para 5

[23] For example, Ev w15

[24] Ev w3

[25] Ev w16

[26] Ev w28

> shapes, with different biases inherent in their populations, to produce something that you knew was using the same methodology in every place.[27]

A key advantage of the current system is that it produces a national dataset, at a fixed moment in time, that is consistent across the whole of the UK.

Longitudinal studies/the advantage of comparisons over time

32. A key advantage of the census raised several times was that it provided a broad sample from which appropriate candidates could be selected to continue valuable longitudinal studies. Professor Blane, Deputy Director of the ESRC International Centre for Life Course Studies, considered that longitudinal studies required a large data context to be properly resourced:

> Britain remains the envy of the world in the richness of its longitudinal datasets. [The] 1946 birth cohort [...] have been tracked right across their lives and they are now in their 60s—my age. When you track people over that length of time, obviously, people drop out. What you need is a population count that tells you how representative the people left in these Longitudinal Studies are and what the directions of a selective bias might be. I do not see any other alternative than the census for that purpose.[28]

33. The British Library noted:

> 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').[29]

Organisations such as the British Society for Population Studies raised the issue of examining trends over long period to see societal changes and future trends, information that—they argued—is currently available in reliable form solely through the census:

> 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?'[30]

---

[27] Q 57

[28] Q 15

[29] Ev w37

[30] Ev w24

Detailed local information

34. Many submissions to our inquiry raised the value of the census in providing detailed information about local areas.[31] We were told that local authorities required the census to provide:

> 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.[32]

35. The Joseph Rowntree Foundation emphasised the ease with which the data could be used to compare different geographical areas:

> It is unique because it covers everyone at the same time and asks the same core questions everywhere. This makes it possible to compare different parts of the UK, Local Authorities and even smaller area statistics on many of the range of variables collected.[33]

36. Of particular interest was the information on household composition: the Royal Statistical Society pointed to the ability to "create bespoke cross tabulations of census variables and to analyse household structures".[34] We were told that the ability to look at fine detail was made possible due to the immense sample size of the census as a whole. John Stillwell and Oliver Duke-Williams of Leeds University said:

> 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.[35]

37. The British Library highlighted the use of census data by a range of 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

[31] For example, Ev w31

[32] Ev w31

[33] Ev 53

[34] Ev 42

[35] Ev w8

populations and socio-economic conditions in small areas. This information is then used to identify need, plan services and support funding bids and campaigns.[36]

38. The British Library also told us that the census had the potential to "provide greater representation of marginalised groups" as the near-universal coverage ensures that they are counted.[37] Evidence from the Salvation Army supported this; it uses census data to provide the organisation with "a foundation of what a community consists of".[38] This was not an uncommon perception among witnesses.

Reliance on official information/difficulties of combining other sources

39. The Joseph Rowntree Foundation pointed out a number of surveys carried out in the UK:

- The Labour Force Survey (LFS),
- The Longitudinal Study of Young People in England (LSYPE)
- Households Below Average Income (HBAI)
- The Annual Population Survey (APS)
- The Family Expenditure Survey (FES)
- Health Survey for England (HSE)
- Families and Children Study (FCS)
- Growing Up in Scotland (GUS)
- Millennium Cohort Study (MCS)[39]

These surveys are conducted mainly to provide specific information necessary for the formulation of particular policies. Concerns raised about such sources, in comparison to the census, include inadequate detail,[40] a lesser geographical coverage[41] or a lack of comparators to enable the aggregation of information in order to produce studies of greater depth.[42] Moreover, the provision of census data is something that has been done for so long that there is less suspicion among the populace about the motives for data collection and greater willingness to provide data, which is likely to mean that census data is more representative and more accurate than is obtainable from smaller-scale, more focused surveys.[43]

---

[36] Ev w37

[37] *Ibid.*

[38] Ev w47

[39] Ev 55

[40] For example, Q 61

[41] For example, Q 53

[42] For example, Q 62

[43] Q 47 [Professor Allen]

Census data as reference point

40. The large dataset produced by the census is valuable in that it can be used to adjust other surveys and studies which may, due to their more voluntary nature or because the participants do not entirely trust the purposes to which the information may be put, be less accurate. The British Academy argued for the value of census data as a central reference point against which other social data might be judged.[44]

41. The Joseph Rowntree Foundation described the role of the census as a lynchpin between other datasets:

> the census data enriches many other datasets by linking with them and therefore really enhancing the analysis possibilities. For example, the ONS Longitudinal Study (LS) is a data set comprising linked census and event records for 1% of the population of England and Wales (about 500,000 people at any one census). It was set up in 1974 to address problems with the adequacy of occupational mortality data, and the lack of longitudinal fertility data, but since then it has been used to address a wide range of other research questions.[45]

Disadvantages of reliance on census data

42. However, the fact that the census plays such a key role as a research tool may have a disadvantage. Professor Les Mayhew, Cass Business School, suggested that the census had a stifling impact on research in social sciences and that discontinuing it 'would lead to a period of huge innovation in the research community'.[46] This was not a perspective shared by the Economic and Social Research Council (ESRC) or the ONS.[47]

43. **We appreciate the central role that census data has played for social scientists. The main reasons for its importance for academia are datasets that can be used across many years, continuing longitudinal studies and a central reference point with which other data may be compared. However, Professor Mayhew's concern that social scientists may turn to the census data simply because it is there rather than because it provides the best data source for the research in hand was not convincing. However, we do recommend that the ESRC ensures that, among those researchers it funds, there is no over-reliance on census data to the exclusion of more appropriate data sources, or such use that stifles the development of innovative means of gathering or utilising social data.**

Timeliness

44. The problem with the census most frequently mentioned to us was that census data may quickly go out of date due to the gaps between census-taking, exacerbated by the lengthy data processing after the data-gathering exercise.[48] The advantages of the census—

---

[44] Ev w38 Check

[45] Ev 53

[46] Q 9

[47] Q 80

[48] *Ibid.*

size and comprehensiveness—are diminished when the data is not utilised until almost two years later, especially in areas such as London, where the population is highly mobile and changeable.

45. Professor Mayhew strongly argued that policy decisions needed to be made from data that is both current and accurate.[49] The Office for National Statistics itself questioned the value of a census that takes place only every ten years of a population that is increasingly mobile: "This can be a significant issue in areas experiencing rapid population change, or when the importance of a particular socio-demographic topic suddenly changes in response to new or emerging Government policies and priorities."[50]

46. Glen Watson, ONS, told us that it was not possible to expedite the publication of census data and that they had received representations that publishing raw data followed by a later, corrected, version caused confusion.[51] We were also told that the Government was not provided with early access to the data when it would be more up to date.[52]

47. **We strongly consider that there is a need for more up to date information than the census provides. We urge the Office for National Statistics during their 'Beyond 2011' deliberations to ensure that whatever solution they propose provides greater access to current accurate data.**

The impact of the census on the ONS

48. The normal working resource budget of the ONS has been around £140 million a year.[53] The running of the census, once every ten years requires the addition of thousands of staff and a budget of approximately £500 million. It is unlikely that the addition of such a task to any organisation would not interfere with the smooth running of the day-to-day tasks of that organisation. This is exacerbated by the high profile of the census operation.

49. One indicator of the impact is that census data takes so long to become available, the process occupies almost two years during which time those ONS employees are fully occupied on census work as opposed to the tasks they undertake eight years out of every ten. Glen Watson of the ONS outlined the timeline of the census:

> Census day was 27 March [2010], as everyone knows. The census field work finished in June, after we had spent a couple of months sending people out knocking on doors, reminding people of their obligations and chasing up non-responses. The capture and processing of all that information at our large processing facility in Manchester started at about the same time.
>
> [...]

[49] Ref

[50] Ev 49

[51] Q 90 [Glen Watson]

[52] Q 114 [Jenny Dibden]

[53] ONS website, 20 October 2010, www.ons.gov.uk/ons/media-centre/statements/older-statements/spending-review-oct-2010/spending-review-2010-announcement.html

> We are now in the phase where we are doing further statistical processing of that data within the Office for National Statistics, and we are embarking on a process of quality assuring all the results.
>
> [...]
>
> It is an incredibly complex process, and it will be next summer [2012] before we are in a position to get out the first set of results.[54]

50. **We have some concerns that the operation of the census may unduly impact on the day to day operations of the ONS. We recommend that when considering how to provide constant, accurate data to Government, the ONS devise how this might be accomplished in a way that will be less intrusive to the operation of their day to day organisation.**

[54] Q 89

4 A future without a census?

Existing alternatives to the census

51. As already indicated, the Government and other public sector bodies carry out a number of surveys of areas such as employment, health, household incomes and young people.[55] The *National Data Strategy* recognises the potential value of such 'administrative data' for research purposes.[56] The strategy outlines the steps necessary for this data, along with commercial[57] and tracking data,[58] to be useful for research. It states: "The [Economic and Social Research Council] will investigate the appetite for a collaborative programme of research involving both private and public sector organisations in the use of transactions data—information collected and retained by organisations in the conduct of their business".[59]

52. We earlier noted the difficulties that may arise in using data collected for one clear (commercial) purpose to provide information for other purposes (see paragraph 26). Our witnesses were sharply divided about the potential for, and the degree of difficulty of, combining different existing administrative databases to produce information comparable to that derived from the census. The importance of getting this right is thrown into relief when one considers, for example, that, "crudely, 75% of local authority funding is centrally funded with allocation based to a significant degree on population".[60]

53. Professor Les Mayhew, Cass Business School, explained that he was able to carry out surveys that were more timely and more accurate on a small scale than the census by combining data sets gathered by local providers of health, local authority and other services:[61]

> There is no dataset that completely covers all the population or is 100% reliable, so you have to combine them in some way. We link the population to property registers. We have a set of rules by which you can confirm or not confirm, based on whether they are on more than one dataset and other rules, which I could explain.
>
> Because I think I probably have as much experience as anybody in using all these datasets, I want to put on record the fact that you can look at household composition using administrative data by linking people to their addresses.[62]

54. He explained that:

---

[55] Ev 52

[56] "UK Strategy for Data Resources for Social and Economic Research", *UK Data Forum*, 2011 www.esrc.ac.uk/funding-and-guidance/tools-and-resources/research-resources/data-services/NDS/index.aspx

[57] Data generated by organisations that operate on a 'for-profit' basis, for example, loyalty cards.

[58] Data generated by watching traffic, real or virtual, for example CCTV or visitors to a web page.

[59] "UK Strategy for Data Resources for Social and Economic Research", section 6.4, *UK Data Forum*, 2011 www.esrc.ac.uk/funding-and-guidance/tools-and-resources/research-resources/data-services/NDS/index.aspx

[60] Ev 58

[61] Ev 57

[62] Q 14

> The data base produced may be contrasted with information available from the Census
>
> - The results are timelier than the Census which can take several years to process before being released and be more than 12 years out of date before it is refreshed. Based on the 2011 Census basic population aggregates will not be released before July 2012 and more detailed data will not follow in some cases until much later.
>
> - On Census day this year we obtained and processed snapshots of local administrative data covering the six Olympic Boroughs in London who commissioned the work. Fully geo-referenced databases down to household level by age and sex were completed and handed over to local authorities inside six months.
>
> - The databases contain much greater granularity than is possible using Census data. Each database is able to produce statistics for any size or shape of geographical area or administrative unit and unlike the Census is not constrained by any pre-determined geographical boundaries.[63]

55. While acknowledging the existence and importance of alternative sources of data,[64] other witnesses raised a number of concerns about the robustness of such administrative data, and the ease of comparison of data from different sources. For example, Professor David Blane commented:

> civil servants always laugh and say, 'You have seen nothing until you have faced the problems in administrative data.' A real ace like Professor Mayhew, who is motivated and skilled, can cut through a lot of problems, but I worry that most things are going to be done by people doing routine work. They are not going to be as motivated and skilled as Professor Mayhew. The potential for introducing error into the data is enormous.[65]

56. The British Library was particularly concerned that the loss of the census would lead to the fragmentation of information about the UK population:

> 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.[66]

57. Dr Eldin Fahmy, from the School for Policy Studies in the University of Bristol, believed that the decision to axe the decennial population census after 2011 would "substantially undermine the capacity of UK social science to analyse and understand

---

[63] Ev 57

[64] For example Ev 42

[65] Q 18 [Professor Blane]

[66] Ev w37

social processes at a small scale".[67] TWRI Policy and Research suggested 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".[68]

58. Moreover, other sources of data have restrictions on their use or general availability: we were told, for example, that the Labour Force Survey Annual Local Area Data Series and the Unitary Authority/Local Authority (UA/LA series) had been withdrawn on instruction from Office for National Statistics (ONS), due to confidentiality issues.[69] The Joseph Rowntree Foundation explained that, although these databases had always been anonymised when released for analysis by outside bodies to ensure that users could not identify any respondent with the information given, advances in technology and software had made it easier to link survey records to either other survey files or other administrative or commercial databases. The ONS had therefore concluded: "Although the risk for most respondents is very small, there remains a risk of identification for people with unusual combinations of personal circumstances. Thus the release outside the central government statistical services of social survey databases with small area identifiers, alongside a national database with detailed coding, has now been ceased".[70]

59. **Administrative data is often collected without consideration of potential wider application and use, thus often is only fit for a single purpose. There is reluctance on the part of research scientists and government social scientists to utilise it for other purposes due to the difficulties in its reuse. The existing *National Data Strategy* should provide good practice and guidance on expanding the number of uses and the longevity of data collected at public expense.**

60. **We also recommend that the ONS seek to remove bureaucratic burdens currently hindering the broader use of data. Too often the Data Protection Act is used as an excuse for not reusing data collected at considerable cost to the public purse. We consider it entirely possible that data could be collected in a way to facilitate better public administration that would not contravene the principles of the Data Protection Act. We would like the Government to indicate how it plans to more broadly use data from sources, such as the Labour Force Survey, as part of their response to this report.**

61. The Local Government Association indicated that they would like to see better and more available data in a number of areas:

Health data on the ageing population and disability

The availability of data on income and taxation

Second jobs

Broadband take-up

[67] Ev w1

[68] Ev w15

[69] Ev 55

[70] Ev 55–56

International migrants.[71]

This would suggest that currently neither the census nor other administrative datasets are adequate in these areas.

62. Even while arguing that expanding the system of utilising local administrative data adopted by his company to cover the country would cost roughly one tenth or less of the present cost of the Census, enable results to be available with six months and the exercise repeated at more regular intervals, Professor Mayhew acknowledged: "There is always scope to improve, and a population register would probably enable further improvements"; and "there is a demand out there from the policy community which is not quite the same as the requirements of the academic community".[72]

63. **We consider Professor Mayhew's evidence as confirmation that there is a credible alternative to the census for the purposes of local government. However, we note that local government are not the only users of census data, and—because of their ad hoc nature—Professor Mayhew's surveys would not substitute for the census in terms of being able to derive a snapshot of the whole nation at one time, with very widespread coverage (because of the mandatory nature of the census process) and the ability to make direct comparisons over time. The academic community would clearly lose more than the public sector by the ending of the census.**

64. **Furthermore, we are concerned that there would need to be a level of expertise not currently widely available amongst organisations collecting data in order to achieve results comparable with those obtained from census data. We recommend that the Government use the time until the next census is due to ensure administrative data is better able to supplement or replace census data. This will require a considerable investment, and possibly the production of a list of approved providers for local authorities, health bodies, etc, to ensure that the data produced is both robust and comparable across authority boundaries and devolved administrations across the whole of the UK.**

65. **Francis Maude, the Minister for the Cabinet Office and Paymaster General, wrote to us that "while cost is a driver, the real issue is ensuring that the best possible approach is taken". We are not persuaded that local and frequent surveys could provide an adequate substitute for census data despite the potential advantage of providing more up-to-date information unless they were designed and implemented to a high standard. We are therefore not convinced therefore that the use of administrative data would be a cheaper option over a ten year census cycle.**

66. **However, if standards could be set to facilitate integration with administrative sources, we consider it possible that obligations could be imposed on privatised utilities to produce and provide government with access to useful social data.**

[71] Ev 60

[72] Ev 58 and Q 14

Concerns about the future availability and reliability of administrative sources

67. The Royal Statistical Society had concerns that the exclusive use of administrative data had another problem:

> the overall data infrastructure would become more dependent on the policies of a range of government departments and organizations whose primary objective is not data collection. From a social science perspective this creates a risk that data series may change in unplanned ways and that comparable datasets through time may be difficult to achieve.[73]

68. In a future without a census there would be a need to ensure that social data was not compromised by policy decisions taken at local or national levels. The Royal Statistical Society suggested to us that government departments supplying information to the statistical system replacing the census should be required to consult with the National Statistician before implementing major changes to their data collections.[74] We think this suggestion has considerable merit. **We recommend that the ONS, if they decide to discontinue the census, should consider how administrative data might be collected over a sustained period without falling hostage to political considerations.**

69. **In this context, there is a particular problem in relation to ensuring the robustness of longitudinal studies by providing a benchmark against which the representative nature of the surviving cohort may be measured. Soundly-based longitudinal studies are a particular strength of the UK at present, and are vital in particular in relation to research into health and educational outcomes. We expect the ONS to pay particular attention to ensuring that any alternatives to the census enable the continuance of such studies.**

70. Professor Ceri Peach warned that using a multitude of unplanned, diverse surveys and studies to replace a monolithic census where surveying methods are standardised and changes planned in advance posed the risk that slight changes in sampling methods could lead to unnoticed influences on the data outputs.[75]

71. Academic witnesses stated that such a risk was mitigated by the existence of a central reference point.[76] This is currently provided by the census which, by its simple existence, influences the data collected in surveys and how that data is categorised. Many other surveys and studies employ terminology and definitions used in the census due to the obvious benefit of being more compatible with the data in the census. In the absence of the census, disparate surveys could become increasingly incompatible.[77] Professor Martin of the Royal Statistical Society told us that it would be "exceedingly difficult to mandate organisations of different sizes and shapes, with different biases inherent in their

[73] Ev 42

[74] *Ibid.*

[75] Ev w3

[76] Q 18

[77] Q 112

populations, to produce something that you knew was using the same methodology in every place".[78]

72. We asked whether the ESRC should set standards by which data collected by ESRC-funded research should be reported, but the academics we questioned rejected this.[79] **However, the National Data Strategy exists and we judge that this strategy would provide a vehicle through which greater coherence of data collection, both administrative and research, could be achieved in future.**

73. Concerns were also raised that because the census is used as a standard against which academic studies are compared to test the accuracy of the sampling methods used,[80] its loss might render a number of other studies less useful. The Joseph Rowntree Foundation suggested:

> The main problem however with [the new UK Household Longitudinal Study, Understanding Society (USoc)] as an alternative to the census, is that it's longitudinal, so over time because of wave on wave attrition (where the number of respondents decreases over time due to loss to follow up, emigrating or death) it becomes less representative of the population. Although it is large and robust, when you get down to small area statistics which census produces, then USoc is not appropriate.[81]

74. When looking at a central reference the key consideration was held to be something that reliably and consistently linked individuals to addresses.[82] Professor Mayhew told us that address databases did exist but that to make the whole process more efficient "the first thing that could be done is to link all administrative records to [a Unique Property Reference Number], and that would make a huge difference to the quality of the data and the processing of the data in future".[83]

75. There were problems raised with respect to the use and combination of administrative data. Professor Blane was worried about the sensitivities of the British public to cross referencing government collected data:

> I am aware that within the civil service there are problems of linking data. The example I know about is with the ONS Longitudinal Study where, for about 20 years, there has been talk of linking in people's benefit records from the Department for Work and Pensions. Some years it is on and some years it is off. There is a big problem about the Data Protection Act and whether a civil service department will release data to another civil service department because of the implications under the Data Protection Act. It could be that the culture in Britain is different from that in

[78] Q 57

[79] Q 42

[80] For example, Ev w28

[81] Ev 55

[82] Q 29

[83] Q 30 [Professor Mayhew]

> Scandinavia-that this relatively legitimised linkage in Scandinavia is foreign to the culture of Britain and that it would not work.[84]

76. Professor Joshi, President of the Society for Lifecourse and Longitudinal Studies, raised what might be a more pertinent point; that while addresses may be unique, "people may not be uniquely associated with addresses".[85] Any central database would have to ensure that it caught those people, like "visitors, second homes and children moving between parents", if it was to be as useful as the census.[86] Several current sources of data, such as the Child Benefit database,[87] the Local Land and Property Gazetteer[88] and the General Practice Research Database,[89] were mentioned as useful during our oral evidence sessions but none of them had all of the key features of the census.

77. **There is a danger that, if the census is not repeated, there will be no equivalent large-scale collection of trusted data that can be used to correct smaller surveys. We are convinced of the need to have a national reference point that other datasets might use as a benchmark for their own parameters. We recommend that the ONS consider how this might be achieved in the absence of a census; it appears to us that making an existing dataset better would be more advantageous than adding a new one.**

78. We were also concerned to establish the degree to which government departments will be able to continue to collect data on the scale which they have done recently. As our predecessor Committee found,[90] it is difficult to track research and development (R&D) spend by government departments accurately over time using published statistics in the R&D scoreboard. Nor do the statistics readily show what proportion of R&D is social science related. There have been concerns, however, that government research and development will see a significant cutback given the stringent budgetary cuts faced by government departments. In 2007, Lord Sainsbury's review of research spending in government[91] recommended better identification of and protection for departmental R&D budgets but the Campaign for Science and Engineering in the UK considers that, while the Government accepted this recommendation, there has been little progress.[92] The joint heads of the Government Social Research Service (Jenny Dibden and Richard Bartholomew) told us that they did not believe that social science was being unduly affected by the budget cuts within departments:

> Heads of Analysis undertook an exercise, which Sir John Beddington has spoken to in the House of Lords Science and Technology Committee, that looked at how

---

[84] Q 30 [Professor Blane]

[85] Q 34

[86] Q 34

[87] Q 39

[88] Q 29

[89] Q 18

[90] Science and Technology Committee, Sixth Report of Session 2009–10, *The impact of spending cuts on science and scientific research*, HC 335–I

[91] HM Treasury, October 2007, *The Race to the Top: A Review of Government's Science and Innovation Policies*. Recommendation 8.4

[92] "Government departmental R&D spending", *Campaign for Science and Engineering in the UK*, http://sciencecampaign.org.uk/?p=7144

> Departments were dealing with the spending review settlement specifically in relation to analysts and what they spent on the research budget. The conclusion that we reached on the early returns was that analysis was not being disproportionately affected, which was important. There were indications that in some Departments spending would be preserved, or potentially increased, and in other Departments there would be reductions.[93]

79. **The Government spends significant sums on R&D, though it is not clear what percentage of this relates to social science. However, we accept the assurance of the joint heads of the Government Social Research Service that spending is, in the main, being protected even in this difficult economic climate. That same climate makes it even more vital that the Government ensures such expenditure, firmly based on evidence, achieves the maximum benefit possible.**

80. We also have concerns about the future availability of data to the third sector, especially volunteer and community-based organisations with limited resources. Any move away from the provision of data free of charge under the Neighbourhood Statistics service would clearly have a detrimental effect on their ability to continue to provide services needed locally. **We are not convinced of the value of government collecting data simply because it has happened in the past and we consider that the responsibility to maintain particular datasets should rest with those bodies most interested in the dataset. However, we recognise the difficulties that local charities and support groups may have in accessing information if the census is discontinued. We regard it as essential that the Government recognise these needs and confirm that appropriate steps would be taken to ensure these groups to have continued free access to whatever alternative data is gathered and shared by public bodies in order to avoid detriment to the valuable local services provided by them. We anticipate a central data repository from which all publicly funded social data, not subject to legal or commercial restrictions, would be made available.**

Future developments

81. The majority of submissions to the inquiry were more concerned about the loss of the census as a source of social data than with anticipating new data sources. Despite this, various resources were proffered as potential sources of census-style information, such as a Swedish-style population register,[94] expanding the data associated with the National Insurance number (for example, associating the address or information on characteristics of the dwelling where an individual is living),[95] and sharing government administrative databases more widely.[96] None of these options was seen as a ready replacement and some effort would be necessary to make them easier to use, more comprehensive or more tightly associated with individuals over time.

[93] Q 120 [Jenny Dibden]

[94] Ev w16

[95] Ev w12

[96] For example, Ev 35

82. Professor Blane told us that the collection of data really depended on the principle of informed consent. In Sweden, every citizen has a unique ID that appears on every official document relating to that citizen, making cross referencing of information very easy. This is balanced by the fact that the use of the data is controlled through an ethics committee and the system is subject to renewal through a referendum every ten years to ensure that the state still has a mandate for that level of data co-ordination by government.[97] Our witnesses doubted that, in a larger and more heterogeneous society such as the UK, this degree of "information swapping" would be as acceptable, despite the ready access to computing power making the sharing of even the largest datasets more feasible.[98] In this context, we note the difficult data protection issues already being raised. **Therefore, although we do not rule out the development of new sources of data in the future, we consider it would be wrong to discontinue the census simply hoping that new developments will provide a solution to the gaps caused by the loss of census data. ONS must be sure that the tools used to collect data will be adequate. One key concern is that we have not identified any dataset that will really enable social scientists/historians to follow individuals over time. Most public sector data-gathering is focused on the size of specified groups, rather than details of individuals, and private sector databases (such as those for loyalty cards) cover only parts of the population and are of little relevance for many of the economically excluded and the poorer sections of society.**

83. **We are convinced that the social science benefits of the census are valuable and that they outweigh the financial costs. However, we are also convinced that there remain significant benefits to be gained in terms of improving the consistency, currency and availability of administrative data to government planners. Although we put forward these conclusions to assist in the ONS's 'Beyond 2011' project, we consider it essential that the Government not only retain access to the breadth and quality of data it collects but seeks to improve its currency and consistency.**

84. **The regular conduct of a census in the UK has provided Government and social scientists with an almost unique dataset with which to examine the changing nature of UK society over the past 200 years. We consider that good evidence-based social policy is founded on such data and that the Government needs to ensure future access to high quality social data.**

[97] Q 3

[98] For example Q 30 [Professor Blane] and Q 133 [Richard Bartholomew]

Conclusions and recommendations

Social science in Government

1. We are content that the structure of social science research in Government is organised in an effective manner to provide the information required by Ministers in planning departmental spending. We are, however, not convinced about the ability of social scientist advisors to influence Ministers when departmental considerations conflict with those of Government as a whole. (Paragraph 17)

2. We recommend that the Government give a senior Minister direct responsibility for social science who would take a cross-Government view in Ministerial discussions, respond to Parliamentary questions and reports from Select Committees. We consider this essential, especially in the event that greater departmental co-operation is required to source data in the event of the census being discontinued. (Paragraph 18)

3. The ESRC's second [National Strategy] document covers the period 2009–2012, and a major part of it concerns making better use of the data that is already collected. It was surprising that we did not hear more about this strategy in our oral or written evidence as it would seem central to the long term provision of data for social science, especially if the census is to be discontinued. (Paragraph 21)

4. Utilities presumably consider their investment decisions against projections made on the basis of local authority structural plans. However, as we note below, local authorities depend, at least in part, on the national census to test the validity of their own projections. We recommend that the Government investigate the potential for sharing anonymised social data collected by utilities. (Paragraph 28)

Importance of census data

5. We appreciate the central role that census data has played for social scientists. The main reasons for its importance for academia are datasets that can be used across many years, continuing longitudinal studies and a central reference point with which other data may be compared. However, Professor Mayhew's concern that social scientists may turn to the census data simply because it is there rather than because it provides the best data source for the research in hand was not convincing. However, we do recommend that the ESRC ensures that, among those researchers it funds, there is no over-reliance on census data to the exclusion of more appropriate data sources, or such use that stifles the development of innovative means of gathering or utilising social data. (Paragraph 43)

6. We strongly consider that there is a need for more up to date information than the census provides. We urge the Office for National Statistics during their 'Beyond 2011' deliberations to ensure that whatever solution they propose provides greater access to current accurate data. (Paragraph 47)

7. We have some concerns that the operation of the census may unduly impact on the day to day operations of the ONS. We recommend that when considering how to provide constant, accurate data to Government, the ONS devise how this might be accomplished in a way that will be less intrusive to the operation of their day to day organisation. (Paragraph 50)

Potential use of other administrative data

8. Administrative data is often collected without consideration of potential wider application and use, thus often is only fit for a single purpose. There is reluctance on the part of research scientists and government social scientists to utilise it for other purposes due to the difficulties in its reuse. The existing National Data Strategy should provide good practice and guidance on expanding the number of uses and the longevity of data collected at public expense. (Paragraph 59)

9. We also recommend that the ONS seek to remove bureaucratic burdens currently hindering the broader use of data. Too often the Data Protection Act is used as an excuse for not reusing data collected at considerable cost to the public purse. We consider it entirely possible that data could be collected in a way to facilitate better public administration that would not contravene the principles of the Data Protection Act. We would like the Government to indicate how it plans to more broadly use data from sources, such as the Labour Force Survey, as part of their response to this report. (Paragraph 60)

10. We consider Professor Mayhew's evidence as confirmation that there is a credible alternative to the census for the purposes of local government. However, we note that local government are not the only users of census data, and—because of their ad hoc nature—Professor Mayhew's surveys would not substitute for the census in terms of being able to derive a snapshot of the whole nation at one time, with very widespread coverage (because of the mandatory nature of the census process) and the ability to make direct comparisons over time. The academic community would clearly lose more than the public sector by the ending of the census. (Paragraph 63)

11. Furthermore, we are concerned that there would need to be a level of expertise not currently widely available amongst organisations collecting data in order to achieve results comparable with those obtained from census data. We recommend that the Government use the time until the next census is due to ensure administrative data is better able to supplement or replace census data. This will require a considerable investment, and possibly the production of a list of approved providers for local authorities, health bodies, etc, to ensure that the data produced is both robust and comparable across authority boundaries and devolved administrations across the whole of the UK. (Paragraph 64)

12. Francis Maude, the Minister for the Cabinet Office and Paymaster General, wrote to us that "while cost is a driver, the real issue is ensuring that the best possible approach is taken". We are not persuaded that local and frequent surveys could provide an adequate substitute for census data despite the potential advantage of providing more up-to-date information unless they were designed and implemented to a high standard. We are therefore not convinced therefore that the use of

administrative data would be a cheaper option over a ten year census cycle. (Paragraph 65)

Consideration of change or discontinuation of the census

13. However, if standards could be set to facilitate integration with administrative sources, we consider it possible that obligations could be imposed on privatised utilities to produce and provide government with access to useful social data. (Paragraph 66)

14. We recommend that the ONS, if they decide to discontinue the census, should consider how administrative data might be collected over a sustained period without falling hostage to political considerations. (Paragraph 68)

15. In this context, there is a particular problem in relation to ensuring the robustness of longitudinal studies by providing a benchmark against which the representative nature of the surviving cohort may be measured. Soundly-based longitudinal studies are a particular strength of the UK at present, and are vital in particular in relation to research into health and educational outcomes. We expect the ONS to pay particular attention to ensuring that any alternatives to the census enable the continuance of such studies. (Paragraph 69)

16. However, the National Data Strategy exists and we judge that this strategy would provide a vehicle through which greater coherence of data collection, both administrative and research, could be achieved in future. (Paragraph 72)

17. There is a danger that, if the census is not repeated, there will be no equivalent large-scale collection of trusted data that can be used to correct smaller surveys. We are convinced of the need to have a national reference point that other datasets might use as a benchmark for their own parameters. We recommend that the ONS consider how this might be achieved in the absence of a census; it appears to us that making an existing dataset better would be more advantageous than adding a new one. (Paragraph 77)

18. The Government spends significant sums on R&D, though it is not clear what percentage of this relates to social science. However, we accept the assurance of the joint heads of the Government Social Research Service that spending is, in the main, being protected even in this difficult economic climate. That same climate makes it even more vital that the Government ensures such expenditure, firmly based on evidence, achieves the maximum benefit possible. (Paragraph 79)

19. We are not convinced of the value of government collecting data simply because it has happened in the past and we consider that the responsibility to maintain particular datasets should rest with those bodies most interested in the dataset. However, we recognise the difficulties that local charities and support groups may have in accessing information if the census is discontinued. We regard it as essential that the Government recognise these needs and confirm that appropriate steps would be taken to ensure these groups to have continued free access to whatever alternative data is gathered and shared by public bodies in order to avoid detriment to the valuable local services provided by them. We anticipate a central data

repository from which all publicly funded social data, not subject to legal or commercial restrictions, would be made available. (Paragraph 80)

20. Therefore, although we do not rule out the development of new sources of data in the future, we consider it would be wrong to discontinue the census simply hoping that new developments will provide a solution to the gaps caused by the loss of census data. ONS must be sure that the tools used to collect data will be adequate. One key concern is that we have not identified any dataset that will really enable social scientists/historians to follow individuals over time. Most public sector data-gathering is focused on the size of specified groups, rather than details of individuals, and private sector databases (such as those for loyalty cards) cover only parts of the population and are of little relevance for many of the economically excluded and the poorer sections of society. (Paragraph 82)

21. We are convinced that the social science benefits of the census are valuable and that they outweigh the financial costs. However, we are also convinced that there remain significant benefits to be gained in terms of improving the consistency, currency and availability of administrative data to government planners. Although we put forward these conclusions to assist in the ONS's 'Beyond 2011' project, we consider it essential that the Government not only retain access to the breadth and quality of data it collects but seeks to improve its currency and consistency. (Paragraph 83)

22. The regular conduct of a census in the UK has provided Government and social scientists with an almost unique dataset with which to examine the changing nature of UK society over the past 200 years. We consider that good evidence-based social policy is founded on such data and that the Government needs to ensure future access to high quality social data. (Paragraph 84)

Formal Minutes

Wednesday 12 September 2012

Members present:

Andrew Miller, in the Chair

Ms Caroline Dinenage	Mr Jim Dowd
Mr Gareth Johnson	Mr Stephen Metcalfe
Mr Stephen Mosley	Ms Pamela Nash
Mr Roger Williams	

Draft Report (*The Census and social science*), proposed by the Chair, brought up and read.

Ordered, That the draft Report be read a second time, paragraph by paragraph.

Paragraphs 1 to 84 read and agreed to.

Summary agreed to.

Resolved, That the Report be the Third Report of the Committee to the House.

Ordered, That the Chair make the Report to the House.

Ordered, That embargoed copies of the Report be made available, in accordance with the provisions of Standing Order No. 134.

[Adjourned till Wednesday 17 October at 9.00 am

Witnesses

Wednesday 7 December 2011

Page

Professor David Blane, Deputy Director, ESRC International Centre for Life Course Studies, **Professor Heather Joshi**, President, Society for Lifecourse and Longitudinal Studies, and **Professor Leslie Mayhew**, City University — Ev 1

Wednesday 14 December 2011

Professor Tim Allen, Local Government Association, **Aleks Collingwood**, Joseph Rowntree Foundation, **Professor David Martin**, Royal Statistical Society, and **Professor Philip Rees**, Royal Geographical Society — Ev 10

Adrian Alsop, Director of Research and International Strategy, Economic and Social Research Council, **Jeremy Neathey**, Deputy Director of Policy, Economic and Social Research Council, **Glen Watson**, Census Director, Office for National Statistics, and **Peter Benton**, Deputy Director, Office for National Statistics — Ev 16

Wednesday 18 January 2012

Richard Bartholomew, Joint Head of the Government Social Research Service, Department for Education, and **Jenny Dibden**, Joint Head of the Government Social Research Service, Department for Work and Pensions — Ev 23

List of printed written evidence

List of additional written evidence

(published in Volume II on the Committee's website www.parliament.uk/science)

List of Reports from the Committee during the current Parliament

The reference number of the Government's response to each Report is printed in brackets after the HC printing number.

Session 2012–13

First Special Report	Science in the Met Office: Government Response to the Committee's Thirteenth Report of Session 2010–12	HC 162
First Report	Devil's bargain? Energy risks and the public	HC 428
Second Report	Pre-appointment hearing with the Government's preferred candidate for Chair of the Medical Research Council	HC 510–I
Second Special Report	Engineering in government: follow-up to the 2009 report on Engineering: turning ideas into reality: Government Response to the Committee's Fifteenth Report of Session 2010–12	HC 511

Session 2010–12

First Special Report	The Legacy Report: Government Response to the Committee's Ninth Report of Session 2009–10	HC 370
First Report	The Reviews into the University of East Anglia's Climatic Research Unit's E-mails	HC 444 (HC 496)
Second Report	Technology and Innovation Centres	HC 618 (HC 1041)
Third Report	Scientific advice and evidence in emergencies	HC 498 (HC 1042 and HC 1139)
Second Special Report	The Reviews into the University of East Anglia's Climatic Research Unit's E-mails: Government Response to the Committee's First Report of Session 2010–12	HC 496
Fourth Report	Astronomy and Particle Physics	HC 806 (HC 1425)
Fifth Report	Strategically important metals	HC 726 (HC 1479)
Third Special Report	Technology and Innovation Centres: Government Response to the Committee's Second Report of Session 2010–12	HC 1041
Fourth Special Report	Scientific advice and evidence in emergencies: Government Response to the Committee's Third Report of Session 2010–12	HC 1042
Sixth Report	UK Centre for Medical Research and Innovation (UKCMRI)	HC 727 (HC 1475)
Fifth Special Report	Bioengineering: Government Response to the Committee's Seventh Report of 2009–10	HC 1138
Sixth Special Report	Scientific advice and evidence in emergencies: Supplementary Government Response to the Committee's Third Report of Session 2010–12	HC 1139
Seventh Report	The Forensic Science Service	HC 855 (Cm 8215)

Seventh Special Report	Astronomy and Particle Physics: Government and Science and Technology Facilities Council Response to the Committee's Fourth Report of Session 2010–12	HC 1425
Eighth Report	Peer review in scientific publications	HC 856 (HC 1535)
Eighth Special Report	UK Centre for Medical Research and Innovation (UKCMRI): Government Response to the Committee's Sixth Report of session 2010–12	HC 1475
Ninth Report	Practical experiments in school science lessons and science field trips	HC 1060–I (HC 1655)
Ninth Special Report	Strategically important metals: Government Response to the Committee's Fifth Report of Session 2010–12	HC 1479
Tenth Special Report	Peer review in scientific publications: Government and Research Councils UK Responses to the Committee's Eighth Report of Session 2010–12	HC 1535
Tenth Report	Pre-appointment hearing with the Government's preferred candidate for Chair of the Technology Strategy Board	HC 1539–I
Eleventh Special Report	Practical experiments in school science lessons and science field trips: Government and Ofqual Responses to the Committee's Ninth Report of Session 2010–12	HC 1655
Eleventh Report	Alcohol guidelines	HC 1536 (Cm 8329)
Twelfth Report	Malware and cyber crime	HC 1537 (Cm 8328)
Thirteenth Report	Science in the Met Office	HC 1538
Fourteenth Report	Pre-appointment hearing with the Government's preferred candidate for Chair of the Engineering and Physical Sciences Research Council	HC 1871–I
Fifteenth Report	Engineering in government: follow-up to the 2009 report on Engineering: turning ideas into reality	HC 1667 (HC 511, Session 2012–13)

Oral evidence

Taken before the Science and Technology Committee

on Wednesday 7 December 2011

Members present:

Andrew Miller (Chair)

Stephen Metcalfe	Pamela Nash
David Morris	Roger Williams
Stephen Mosley	

Examination of Witnesses

Witnesses: **Professor David Blane**, Deputy Director, ESRC International Centre for Life Course Studies, **Professor Heather Joshi**, President, Society for Lifecourse and Longitudinal Studies, and **Professor Leslie Mayhew**, City University, gave evidence.

Q1 Chair: I welcome the three of you to this morning's session. Would you be kind enough to introduce yourselves, for the record?
Professor Blane: I am David Blane from Imperial College, London.
Professor Joshi: I am Heather Joshi from the Institute of Education.
Professor Mayhew: I am Les Mayhew from Cass Business School.

Q2 Chair: Thank you very much indeed. As you know, we are looking into the census in some detail. From your perspective, is the census the "key source of information on population statistics" or is it just one of many?
Professor Mayhew: No. It is one of many. The other sources are largely administrative. They have to be combined and used in certain ways to produce something that is equivalent. I could go into detail, but, in my experience, there are enough other sources out there from which you can replicate more frequently, more accurately and in more detail what is available in the census.

Q3 Chair: Is that view broadly held?
Professor Joshi: My answer would be slightly different. The main census is part of a structure of sources of evidence for social and demographic research. A census underpins the rest, which are mostly sample surveys.
Professor Blane: I do quite a lot of research in the rest of Europe and I am aware that there are many other ways of doing these sorts of things. I find that on the mainland a lot of people are still quite envious of this country with its census. The problems with alternatives are quite formidable, not least in terms of informed consent. The country that has the best alternative is probably Sweden and the Scandinavian model, where each citizen has a unique identification number which appears on all Government records. In theory, you can link every record the state holds on any individual. To get access to these records, you have to pass very stringent ethical committee approval.

Q4 Chair: Does that approval extend to academics?
Professor Blane: Yes.

Q5 Chair: It is not just Government organisations.
Professor Blane: No. It is a major research tool. Most importantly, it requires a referendum every 10 years or so, so that there is population support for this sort of record linkage.
Chair: That is interesting.
Professor Blane: The most recent one in Sweden, which was a very close call, was essentially on the function of these linked registers for medical research.

Q6 Chair: Historically, as I understand it, the reason for having a census is to help manage increases in population and develop structures to match change. Does that remain true today or are there other reasons for having a census?
Professor Mayhew: I can speak from a personal standpoint. I use the census very little these days because I find that alternatives are much better, timelier and more accurate. I work a lot in local authorities with healthcare providers and the census is regarded as out of date and not fit for purpose for their needs: commissioning local services, managing complex budgets and identifying unmet need. All those things require timely, detailed and accurate information which the census cannot provide in the form that they need.
In areas like London, in the last 10 years, there have been absolutely huge changes in the number of people and in the ethnic make-up of the population, yet some people are still using census data. A good example I came across recently was looking at the prevalence of TB cases in the African population. The population denominator that the public health specialists are using is hopelessly wrong, and the outcome of that kind of analysis can be very misleading, misdirect resources and so on. In the areas where I work the case is quite clear. In that particular role, the census does not fulfil the need.

Q7 Chair: There is, in that particular instance, a significant difference in the available data coming from the census versus stuff that is policy driven—in this case your work on TB.

Professor Mayhew: It was not my work on TB, but, yes, there are so many local issues for which you require good quality population information.

Q8 Chair: Stemming from that, if the other witnesses could consider it as well, are there similar examples elsewhere?

Professor Joshi: Once the 2011 census data are available, they won't be out of date or not very out of date, and they will refresh the baseline information that is used for all kinds of things: for example, the ethnic composition of local areas. That census will also produce evidence that is difficult to collect locally about the flows of people from place to place, from residence to workplace, commuting across boundaries and moving house across boundaries. That evidence does not become available very often, but it is an important source for knowing how the population is not only distributing itself residentially but how it is moving between different types of households. It also covers people not living in households, in communal establishments, to a greater extent than many other sources.

Professor Blane: A major problem with the census as a research tool is that it takes place only once every 10 years, so you don't really know what has happened in between. You have to rid yourself of the idea that you are ever going to get a perfect dataset. You have to see how different types of datasets fit together. We have annual panel studies that can fill in between the 10 years of the census. The census gives us a secure block every 10 years. It is that in which all the other datasets are situated, in my experience.

Q9 David Morris: What would be your major concerns if census data were no longer available? Is there something that you could do that would no longer be possible?

Professor Mayhew: If the census was not available, it would be hugely beneficial to research because it would lead to a period of huge innovation in the research community as they would learn to use other sources and other datasets, which would become more available. That would create a completely different research environment. You are not completely comparing like for like.

However, there are some things in the census, for example, on caring responsibilities, a question on religion and a few other questions that cannot be replicated easily from administrative data. On the other hand, administrative data contain so much more information than is available in the census. That is what I mean by creating a new research environment and atmosphere that would lead, in my view, to huge innovation.

On the issue of turnover and change, in two months we completed a six-borough study of the Olympic boroughs in London. The information on the population was given to ONS to help validate, as a benchmark, some of their census estimates. This information was very detailed. It was on a personal level by age, sex and ethnicity. It followed on from previous snapshots of the same local authorities in the last two to three years. So we were able to look at turnover change and the flows into and out of those boroughs. There are changes of up to 20% per year in population. If you are measuring something that is 10 years old—you can be using data that are 12 years old with the census because it takes a couple of years to publish—you could be way out. You could be misdirecting resources and all sorts of other things. I cannot speak for the academic use of those sources, but at the local level, for the management of local resources, localism, the unmet need, the creation of health and wellbeing boards, much better data of population intelligence are essential.

Q10 David Morris: You are saying that this form of census could be outmoded and simply collecting this kind of data could be more expensive and more costly.

Professor Mayhew: It is not quite like for like, but we think you can count the population every year for about one tenth of the cost of the census. Statistics Finland, which has used registers for a number of years, calculates that their population intelligence is 35 times cheaper than the census that it replaced per head of population on the basis of equivalent data[1]. The case is clear, but there will be gaps. The targeted use of surveys can fill many of those gaps and extend your intelligence. For example, you might be interested not only in a person's religion but their propensity to give up smoking or something like that. This is what goes on. You can take that sort of data, link it to the administrative data and extend the range and usefulness in that way. You have to take that step. That is what I am saying.

Q11 Chair: Do the other witnesses wish to comment?

Professor Joshi: I would not be quite so optimistic, although I don't really know what sort of database might be able to replace the census. As to the function of acting as a benchmark for sources of information that are not readily obtainable from administrative sources, neither are perfect, but the census tells you who is living with whom, which the administrative sources do not. It tells you about unusual combinations of people living together like husbands and wives who are more than 20 years of age apart, for example. You will never detect those reliably in a sample survey. It is something that is interesting from the research point of view and also from the policy point of view.

Q12 Chair: But they would exist inside public datasets of some sort.

Professor Joshi: If the datasets link those two individuals, yes.

Q13 Chair: Just reflecting back on Professor Blane's observations about the Scandinavian approach, presumably, they address issues like this by drilling down into separate datasets and putting them together, but having some legal protection to ensure that data are anonymised properly and so on. Is that how it works in Sweden?

Professor Blane: I recently led a comparative study of Finland, Italy and the UK. Finland and Italy both

[1] Note by witness: Because it is produced annually the volume of statistics produced is also 10 times greater than previously.

have linked registers, and they are green with envy that we have the ONS Longitudinal Study, which is based on census, linking people across the four censuses. The Finns and Italians, anyhow, seem to envy our census.

Q14 Roger Williams: The Chairman has covered some of the ground I was going to cover. Do I detect a little bit of a difference between academics, on the one side, and people who want to deliver policy, on the other side? One prefers the census. I am not quite sure about the comments that Professor Mayhew made. Is this digging down into data that is there already, or is it commissioning new work outside the census?

Professor Mayhew: No; it is data that are there already. There is no dataset that completely covers all the population or is 100% reliable, so you have to combine them in some way. We link the population to property registers. We have a set of rules by which you can confirm or not confirm individuals, based on whether they are on more than one dataset and other rules, which I could explain.

Because I think I probably have as much experience as anybody in using all these datasets, I want to put on record the fact that you can look at household composition using administrative data by linking people to their addresses. You can look at the household demographics. In fact, we have a classification system of eight household types that are generated out of a subset of 120 types. They include things like single-parent families, older people living alone, three-generational families and so on. It is surprising what you can do with the existing data. There is always scope to improve, and a population register would probably enable further improvements. You are right, in a sense, that there is a demand out there from the policy community which is not quite the same as the requirements of the academic community, particularly those who have used the census for many years and have built up their research history agenda on that basis. It is different, but I think it is much better.

Q15 Roger Williams: Is there anything that either of you would like to add?

Professor Joshi: There are uses of census data for research—as you say, I have different requirements—to combine, with some certainty, a lot more information about people and the people they live with, how long the states that they have been observed in have gone on for, and whether they were in the same place doing the same sort of work 10, 30 or 40 years ago, which the Longitudinal Study can tell you. As Professor Blane said, it is one of the major longitudinal research resources that this country has built up. Much use has been made of it, and much more would be made of it if it continued to accumulate.

Professor Blane: My use of the census is much more limited than that of my colleagues. I use it in the context of the British longitudinal studies. I do not know whether the members of the Committee are aware of it, but Britain remains the envy of the world in the richness of its longitudinal datasets. For example, if you take the 1946 birth cohort, these people have been tracked right across their lives and they are now in their 60s—my age. When you track people over that length of time, obviously, people drop out. What you need is a population count that tells you how representative the people left in these Longitudinal Studies are and what the directions of a selective bias might be. I do not see any other alternative than the census for that purpose, and that is the purpose I use it for.

Q16 Roger Williams: Does the fact that the census exists in any way remove an incentive for people to look at novel and new ways of analysing data?

Professor Mayhew: It probably does, for a combination of reasons. One is that people may lack knowledge about the availability of the administrative data that exist, and it takes time to learn that. Secondly, there are the different interpretations of whether or not you can have access to that data. In that area, there is scope for a lot of improvements. Hypothetically, you could be involved in a study that requires administrative data from five different data owners, which means five different sets of negotiations, getting ethical approval and governance, depending on what the research is for. That can be extremely off-putting. I would like to see some changes to that where, perhaps, the individual researchers are more regulated or licensed to use that data in a trusted and secure way, rather than having to go through all of those barriers. That is a barrier to change. Of course, the census is freely available. If you are used to using it, you will tend to fall back on it. If you can reduce those barriers and change the environment further, then that would be helpful.

Q17 Roger Williams: I guess that one of the virtues of the census is that it is a complete piece of work or as complete as we can get, anyway. Therefore, there is little of a sampling element in it. If it was replaced by a sampling technique or different sampling techniques, slight changes in how those samples were selected or decided upon could have quite profound results on the work that was being done. In the way forward that you are suggesting, is that a weakness?

Professor Mayhew: You seem to be saying, if I may say so, that the census is some kind of gold standard, but it is not, because there have been difficulties with the census in the past, such as low response rates and imputation in areas of London with a 70% to 80% response rate. From my point of view, those potential levels of accuracy count against the census.

Suppose you were designing a survey today and you wanted to have an ethnically representative sample of your population, you used the 2001 census and you were doing it in London. You would get a very misleading sampling framework if you adopted that approach. It is not a straightforward question. It is not one or the other. You have to look at it on its merits. There is no easy answer.

Q18 Chair: If we went down the road of some alternative method of collecting data, in some parts of the publicly collected data there is already an element of compulsion. You are supposed to register to vote,

but we know that people do not, for example. That is one source of data about people. In the absence of a census, would it require there to be a great deal more compulsion on the data provider to ensure that there was accuracy in datasets?

Professor Mayhew: We have been working with administrative data for 10 years and we have noticed steady improvements in the quality, because there are now British standards for addressing or referencing records in datasets. That is noticeable, but you can always take that further. When you are using them, you have to look at how they are put together and how they are maintained in order to form a judgment in terms of whether they are suitable for the purpose for which you want to use them. There have been huge improvements over time. Something like this would lead to further improvements.

Professor Joshi: Administrative datasets that are being put together are not national. Even the four administrations of the UK generate their data in different ways at the moment. There is a question of how you harmonise the standards of collection and the content. Some places will, undoubtedly, have more problems and/or better quality data than others.

Let me come back to this question about innovation, which Professor Mayhew mentioned. In the past 15 years there has been tremendous innovation in the way that social scientists analyse data with the IT and statistical resources available to them. There has been a response from the ONS to enable social scientists to use micro data in a secure way which makes it much more useful scientifically but, nevertheless, preserves the confidentiality which is quite rightly assured to census informants. Doubtless this structure will continue to evolve, but, if it is evolving to preserve the different sorts of confidentiality of different sorts of informants in a hybrid dataset, that is a challenge which might well be met, but it needs to be thought about.

Professor Blane: If you are looking at alternatives, you want to pay great attention to issues of the reliability of the information and its completeness. I use two datasets of administrative data that come from the national health service. They are the Hospital Episode Statistics and the General Practice Research Database. The Hospital Episode Statistics are very difficult to use because people get classified differently. The General Practice Research Database is almost unusable because general practitioners will not register the diagnostic categories in a consistent way.

My colleagues who have been civil servants tell me that academics are used to cleaning research databases, but the civil servants always laugh and say, "You have seen nothing until you have faced the problems in administrative data." A real ace like Professor Mayhew, who is motivated and skilled, can cut through a lot of problems, but I worry that most things are going to be done by people doing routine work. They are not going to be as motivated and skilled as Professor Mayhew. The potential for introducing error into the data is enormous.

Q19 Pamela Nash: We have talked a lot this morning about the problems with the census as it currently stands. Could each of you reflect on the usefulness of the census, particularly in planning ahead for public service provision?

Professor Mayhew: I have tried to comment a little on that already. The fact that it can be up to 10 or even 12 years old before it is replaced is a major issue because of population change. The entire landscape has changed in some areas, particularly in London, where we do a lot of work. Also, the spatial granularity of it is too coarse to aggregate it to drill down and answer some of the questions that health commissioners and local authorities have, whether it is about estates, brownfield sites or something like that. The way in which the data are delivered to you does not get you as far as you want to be. What they want is something much more flexible so that they can flex age groups, time periods and geographical areas that they are interested in, whether they are neighbourhoods, school catchments or whatever. The demand for this level of information has been part of a trend over the last 10 years since we have been involved in it, but it is increasing with concepts like localism and the democratic issues of whether all people who are eligible to vote are actually there. All those things add to the pressure on getting better intelligence to support local decision making. I do not think that the census really meets, or will ever meet, that kind of agenda.

Q20 Pamela Nash: Do any of you have anything, additionally, positive to say about the census and the uses that we still have for it?

Professor Blane: I have two points. I have already mentioned the thing that I use it for most, which is to make best use of Britain's investment in terms of tens and hundreds of millions of pounds and 60 or 70 years of research effort in the longitudinal datasets. We need something like a census to take account of inevitable attrition from these longitudinal datasets. Secondly, one study, the ONS Longitudinal Study, links the same individuals in the censuses of 1971, 1981, 1991, 2001 and 2011, which is a unique dataset because it takes Britain from the end of the post-war settlement in 1971, through the industrialisation in 1981 and the feminisation of the work force in 1991, to globalisation in 2001. You can look at these large-scale social processes and their impact on individuals' lives. For me, the ONS Longitudinal Study is not the major issue, but it would be a definite loss.

Professor Joshi: For the study of what is going on in local areas, the census is an element; it is not always perfect or complete, especially in London. All sources of information put together would improve policy makers' understanding about what is going on and researchers' understanding of the process. I do not think that the census is perfect, but I do not think it is irrelevant. For understanding less local issues like trends in the labour force, trends in the family or projecting what the birth rate is going to be, the census is a complete treasure trove of evidence about the social fabric on which many other data sources build. It is like a tapestry; you can embroider on it, but the framework that the census gives you is the basis.

Q21 Pamela Nash: The census still has support as being a very useful historical document, but would you agree that that is the bigger use of the census information or does it still have a role in planning for the future?

Professor Joshi: Certainly, it would have a role for historians in the future. If the census series of evidence came to a halt, there would be dismay by future historians, but I am not sure that that is your prime concern now. It has laid down this fantastic document of how Britain has changed over more than a century. To know where you are in the picture of long-term change should be of importance to policy makers as well as to social scientists.

Q22 Pamela Nash: It seems that population changes much more quickly than the census is able to capture at 10-year intervals at the moment. Do you think there is a possibility of changing to a smaller census which would be more frequent?

Professor Mayhew: There is a possibility, but I am not sure that it would be valued by local authorities and people at that level, because a sample is a sample and you cannot generalise easily what is happening in London with what is happening in Cornwall or other parts of the country. There is an argument for smaller-scale surveys which ask questions about, perhaps, income, journey to work, commuting and that sort of thing. I wonder whether some of those kinds of issues could be piggy-backed on to existing national surveys, like the Labour Force Survey and other surveys. A more creative use of existing surveys, coupled with more use of administrative data, will get us to a point where we can ask the next question of whether we need a population register or something like that, which is being proposed, but which is a much greater step, in my view. A carefully calibrated mix and use of surveys of administrative data actually gets you to a better place[2].

Q23 Pamela Nash: Is there anything either of you would like to add?

Professor Blane: Yes. There is the new Understanding Society annual panel study. It is 40,000 households and 100,000 individuals and is a random sample across Britain. On an annual basis, that will fill in the gaps between the 10-year censuses. We already have the research infrastructure in place to look at the short-term changes.

Q24 Pamela Nash: Do you agree that technology could be used to improve the cost-effectiveness and efficiency to continue the current census as it stands?

Professor Mayhew: You can always improve it to some extent, and there have been huge improvements this time round with the address database and other things that are used. Ultimately, I do not think it is the right model for collecting these kinds of data. You are never going to be as efficient as the Scandinavians have managed to be and what I think is possible today. While you can improve it, I question whether it is the right model for collecting that kind of data.

[2] Note by witness: However, national surveys are poorer at addressing local issues.

Professor Joshi: There is the technology of analysis as well as the technology of collection to be thought about. That will change. It will have to change when there is room for innovation and improvement there. Going back to your question about having a smaller census, there is one model where you would have a rolling census. Every year it would be in some local authorities but not in others. I would not like to hazard a guess about whether that would save any of the cost, but it might make the whole thing slightly more up to date if you updated the local estimates with all the other sources of information there are locally, and every 10 years you replace it and have a census, but not necessarily in a particular year. It may be more cost-effective, but I am not sure.

Professor Blane: There is a sea change in people's preparedness to answer public surveys which, undoubtedly, is the problem with the census. If people knew that every time they filled in a tax return, they went to their doctor or registered or insured their motor car the information was going to be used for an invasion of their privacy, you are going to get the same problems with administrative data, which might be of greater concern to Government Departments. We just have to face the fact that we live in a changing world. I do not think that technology is going to solve the problem that people are less deferential.

Q25 Chair: In your earlier remarks about the Longitudinal Studies, you were talking about yourself as a researcher using historic data, and then Professor Joshi referred to historians of tomorrow. You and I will have filled in about the same number of census forms, I am guessing. The one thing that I recall from them is that every one has been different. What is the definition of the core dataset that is mission critical for the kind of Longitudinal Studies that you are interested in and future historians would be dismayed about if it were not collected?

Professor Blane: Professor Joshi will give you a better answer to that question. A problem with a survey is that everybody wants more. It is the responsibility of the person leading it to say what you cannot include. The census does a pretty good job of including core information. In recent years, when they have added new questions, by and large they have been very fruitful in terms of our understanding of social processes. The question that was added in 2001 about informal caring, "Are you looking after someone who is ill?", or what have you, has been really important in terms of understanding work-life balance. In 1991 there was a question that was added about limiting long-standing illness, which has been very important in terms of understanding levels of functional disability in the population, the relationship with people who are categorised as permanently sick within the labour force and so on.

Professor Mayhew: But some of that data is available in existing surveys: for example, the Health Survey for England.

Chair: Stephen wants to take this a bit further.

Q26 Stephen Mosley: Thank you, Chair. From what you are saying, pretty much all the data that are available in the census are actually available in other

data sources, but, from where I am sitting now, these data sources seem to be spread across a wide range of Government institutions and organisations. The one advantage with a census is that it is in one place. You can go to the ONS website and go down to the lowest super output area, wherever it is, and see all the data. The problem that I can see in the future is that you will have all this data in different places. That might be okay for a Government institution which has millions of pounds to spend on sorting it all out, getting access to the data and producing it in a useful format, but, for the researcher in a university or a social historian, how are they going to gather all of these pieces of data together? Can you see some way how all of this data can be brought together so that it is in an easy-to-use format to give us the advantages that we can only get with the census?

Professor Mayhew: It is a question of management, transition and vision. Having a core academic dataset that mimics or extends what is already available in the census is something that we ought to be thinking about. You are absolutely right that there are different data owners. The owners of the data that we use are the local authority and the primary care trusts. That is sufficient to count the population in considerable detail at the level that they require. With all these other add-ons and so on, you may have to have a rolling series of surveys, like the existing surveys, but you need the vision that will pull down that data and get it into a form that can be used.

Q27 Stephen Mosley: Does that vision exist at the moment?

Professor Mayhew: No, it needs to happen.

Q28 Stephen Mosley: Thank you. That is something we can work on in our report. We are talking about the historical record, which is something I am particularly interested in. I think it is great that you can look at the 1901 census results and see what your great-grandfather was doing, where he lived and all those kinds of things. Would you envisage in future a snapshot of this data being taken on one day every 10 years, having it as a snapshot and putting it back for historical purposes, or do you imagine that the whole thing would just be a rolling process? It would be constantly updated and in 100 years' time all of this data will be on a CD somewhere, or whatever they will use at that time, but there would be no actual snapshot?

Professor Mayhew: My colleagues may be more expert than I am on this. Genealogists, for example, probably go back to the original registration records that are held in Southport to do a lot of the tracing work that you are reflecting. Beyond that, I cannot comment. It is something we might need to look into to see whether that exists in administrative records already: i.e. the births and deaths registration system. Maybe Heather would like to comment.

Professor Joshi: You would need to devise some form of database which could be interrogated and linked by users in the future as well as the more immediate policy users, once you had put things together. I am not saying that it is impossible but it is quite a challenge if you are going to put people's names on it. You could think of maintaining the Longitudinal Study database, but it would be quite a challenge to collect records of those people if they are not going to reappear in the census.

Q29 Stephen Mosley: Coming back to something you said earlier, you mentioned that there is no database of individuals, essentially, or database of addresses. Do you need something like that to act as a key to the database?

Professor Mayhew: Addresses are key. There is an address database for the whole country. In fact, there are two or three. I am not an expert on this. The one I use is called the Local Land and Property Gazetteer, which is a geo-reference address of every property in an area which links through into council tax and so on. You assign individuals to addresses within that controlled framework where you validate people's current address and so forth. Going back to the original question, has that answered it?

Q30 Stephen Mosley: Yes. So you have this gazetteer of all the addresses. Do you then go in and cross-reference the stuff yourself, or is it a central cross-reference?

Professor Mayhew: Some local authorities have reached the stage where they allocate what is called a UPRN—Unique Property Reference Number—to every record, but many have not reached that stage. The first step we use is that we assign people's records to a UPRN and then we proceed from there. If you wanted to think in terms of making this whole process more efficient, the first thing that could be done is to link all administrative records to the UPRN, and that would make a huge difference to the quality of the data and the processing of the data in future.

Professor Blane: I am a bit peripheral to this issue, but I am aware that within the civil service there are problems of linking data. The example I know about is with the ONS Longitudinal Study where, for about 20 years, there has been talk of linking in people's benefit records from the Department for Work and Pensions. Some years it is on and some years it is off. There is a big problem about the Data Protection Act and whether a civil service department will release data to another civil service department because of the implications under the Data Protection Act. It could be that the culture in Britain is different from that in Scandinavia—that this relatively legitimised linkage in Scandinavia is foreign to the culture of Britain and that it would not work.

Q31 Chair: That is nothing to do with the Act. That is how the data are collected, is it not? If, on collecting the data, the Benefits Agency, for example, said, "This information is going to be made available to the HMRC", that is all they have to do to clear their back. It is not the Act that stops it.

Professor Blane: It is the legal advice of the DWP, as I understand it, that you cannot link to the ONS Longitudinal Study without informed consent.

Q32 Chair: What is the evidence that informed consent would be withheld by honest people?

Professor Blane: Within a medical context, people often do. That is all I can say. They feel that it is their private business.

Professor Mayhew: There is no doubt that there is sensitive data and there is not so sensitive data, and health records are an example of sensitive information. Core information for counting the population is everyday low-level information. There is already a perception in the population that Government already merge these records, and they are surprised to be told that that does not actually happen and all this information exists in different departmental silos across Government. They cannot understand why that information is not—

Q33 Chair: Every MP who has dealt with a Child Support Agency case will be told, whether by the custodial parent or the non-custodial parent, that they are surprised that this data cannot be pulled together.

Professor Mayhew: You are absolutely right. All these complexities come together at the level of children's social services. That is why we get all these problems because everybody hides behind confidentiality. These are all, ultimately, bureaucratic issues. They are counter to common sense, really, in many cases. That ought to be looked at.

The point I wanted to make is that there is a key difference between the use of administrative information for statistical purposes and the use of information about the individual. Very often these things get confused. Section 33 of the Data Protection Act states that we can use administrative information for research purposes, but it needs to be made absolutely clear, from a research point of view, for managing local resources and managing the resources of the country, that that is a legitimate use and it does not have any of the normally associated dangers.

Q34 Chair: Just before I ask Stephen Metcalfe to come in, your previous answer to Stephen Mosley about linking people to a unique property number would instantly create a universal electoral register that is accurate.

Professor Mayhew: It is funny you should say that because we have just done some work in one part of London looking at precisely that issue. It would make a huge difference, yes, although this particular local authority already links to addresses. The problem it has is that, when it comes to update the register every year, it get lots of people who do not respond. Something like 37% of its electorate is not registered. So it is not the UPRN that is causing the problem. It is just the flow of information issue and the turnover of the population.

Professor Joshi: I would like to say something about that. Addresses may be unique but people may not be uniquely associated with addresses, as the census finds visitors, second homes and children moving between parents. It is a good start but it is not perfect. We get evidence from various sources, administrative and the census, about where the adjustments have to be made on the edges.

I would also like to say two things about getting consent from people who are getting benefits. First, it does not get informed consent from people who have had benefits in the past, so you would not be able to link that information, and, secondly, you would not be able to get information about people who do not collect benefits, people in that age group or demographic group, who are not unemployed or are not collecting carer's allowance, for example, so that you can see what the rates of claiming these benefits are. You need some other information on the population that is not a beneficiary in order to make best use of the administrative information.

Q35 Stephen Metcalfe: I think we established this morning that there is quite a large divide between the needs of academic research and those of public policy research. Therefore, the question is whether the Government should be funding those two tracks separately. Should it be funding the collection of social science research as a public good and then funding policy research separately? Do you have views on that?

Professor Mayhew: The priority to have good quality core population information is absolutely key. Whatever else you decide, academic research should sit on top of that. That is why I said earlier that we need a vision of how we are going build the research needs on top of that core information. That means looking at access to administrative data, dealing with that issue to get the core right, and then looking at what needs to change over and above that by looking at the existing surveys or other arrangements. We need a vision to make that transition. That is the way things will go in the future. The population will always drive the agenda, whether it is about migration or simply counting the number of people living in an area. All of those things are key.

Professor Blane: I never thought of the issue before you asked the question. While Professor Mayhew was speaking, I was imagining the debates that are going to go on between the Medical Research Council and the Economic and Social Research Council about what proportion each research council should contribute towards the research component of the decennial census. I could imagine that this would go on for ever. It is probably easier just to recognise that it is a multi-purpose survey and funded centrally, as it is.

Q36 Stephen Metcalfe: Stick with what we have got is really what you are saying.

Professor Blane: Yes. We must continue to modify and adapt it to new circumstances. I think split funding might be a bit complicated.

Q37 Stephen Metcalfe: Do you want to add anything to that, Professor Joshi?

Professor Joshi: I certainly agree that history reveals that, when you try to support big data resources for research across research councils, it is another challenge. The micro datasets are public goods. They would mostly be used by academics, but they will be informing policy makers as well as science. It is probably best to keep their funding integrated, as David suggests.

Q38 Stephen Metcalfe: I have a couple of final questions. Assuming we keep the existing arrangements, you mentioned earlier about tracking individuals through that. Is the census, as it stands, the best way or the main way of doing that, or are there viable alternatives of tracking individuals over a long period of time?
Professor Blane: The alternatives are the annual panel studies. I am always full of admiration for the people who agree to be part of an annual panel study, because a researcher comes to them once every year of their life and asks them questions. I have forgotten your question and was getting carried away with the thought of this.

Q39 Stephen Metcalfe: The first is the extent that it is necessary to track individuals, and, secondly, is the census the best way of doing it?
Professor Blane: The annual panel studies have small numbers. If you want to look at a small group like Chinese people in Britain, there are too few to be able to get any statistical power, so you do not know whether it is just an odd person or whether they are representative of Chinese people in Britain, whereas in the decennial census you have the whole population. In the ONS Longitudinal Study, which is a 1% sample, you have half a million people. You could look at Chinese people in Britain through the census or the ONS Longitudinal Study, but you could not in one of the smaller-scale studies.
Professor Joshi: We also have the birth cohort studies. A new one is just about to start. All these longitudinal data resources complement each other because they have different features and different periodicities, and the longitudinal sample of the census does form a standard to which the others can be linked. The others elaborate detail, but the census and its longitudinal sample give the long-term rates of change. They do not give you the year-on-year changes. They give you decade-on-decade changes, for which you have to wait for a terribly long time if you are collecting data prospectively.
Professor Mayhew: One administrative route that we have used is to combine NHS numbers, which are with you all your life, with address level information—UPRNs—and you can track changes of address and movements into and out of areas. We have done it at a population level of up to half a million, but, if you were to do that at the national level, you would probably be looking at two databases—the Child Benefit database and the DWP database. If you could populate those data with UPRNs, you would be able to monitor statistically movement whenever and wherever you wanted, subject to the accuracy of the underlying information, which should be about 95% to 98% accurate.

Q40 Stephen Metcalfe: In the absence of a census, if we were to change the way we collected information and moved to alternative models, would you all suggest that that data be made available publicly—not necessarily the details of the individuals but the headline numbers?
Professor Mayhew: Although we put together datasets from individual local data when it is handed over to users, it is completely anonymised and you cannot identify any person or individual in it. At some stage in the production process, it is inevitable that you are going to be handling that kind of information.

Q41 Stephen Metcalfe: You are handing it over to the user. Should it be publicised but not published, effectively?
Professor Mayhew: The work we have done has been published on the web by local authorities, yes. I know quite a lot of examples of that.

Q42 Stephen Metcalfe: A lot of social scientists are employed by the taxpayer. All three of you touched upon standardising or regularising the way in which data are collected and the way in which they are put on to systems. Do you think the ESRC should impose a standardisation requirement when it is funding social research?
Professor Mayhew: The university of Essex runs the data archive and it tries to maintain standards, but I do not think that they are universally adhered to. In principle, of course, you are right, but thinking of the ESRC as the ultimate authority on this is a bit tricky.
Professor Blane: The only standardisation that the ESRC would claim is top quality research which can take many forms.
Professor Joshi: In effect, the way that the census classifies variables, without any compulsion, does form guidance for other people collecting data. If you collect data on ethnic groups, it is useful to follow the standard ONS classification. There is a problem also of harmonising with other countries.

Q43 Chair: Presumably, just looking at the other side of the coin, in relation to data that are held by Government, which are accessible to researchers, there has been some minor move in the right direction, but it would help tremendously if there was a consistent format in which Government held data about us.
Professor Mayhew: And a consistent policy on the release of that data for statistical purposes.

Q44 Chair: The day before yesterday we had the debate about health data. It was clear from that that there needs to be a public dialogue and an agreement or a contract between citizen and state that has some longevity associated with it.
Professor Mayhew: Yes. Medical records and that type of information is in a different category to basic standard administrative information about where you live and whether you are on benefits and that sort of thing. You are absolutely right that having a standard approach across Government is important. I do not think that we are far off that. In principle, they will use common identifiers. There is not complete consensus on this. Some use the NHS numbers and some use a child benefit number, or whatever that is, or an NI number. You can create what are called "Look up tables" to link those identifiers together and they could then be used for statistical purposes. It may even be done at the moment.

Q45 Chair: The one that always amuses me is the CSA, which uses a reference number that is longer than the number of people on the planet. There you are.

Professor Blane: You have to qualify the idea that medical records are in a separate category. If you look at the things that people can be blackmailed for, few of them are to do with their health. Psychiatric illnesses would be an exception. The things that people are blackmailed for are much more about their criminal past and so on. There are many areas of state data which are highly sensitive.

Professor Mayhew: I agree that criminal data are not in the same category.

Chair: Can I thank you for your contributions this morning? It has been very informative. There are some challenging issues hidden under all of this. Thank you very much indeed.

Wednesday 14 December 2011

Members present:

Andrew Miller (Chair)

Stephen Metcalfe
Stephen Mosley
Pamela Nash
Graham Stringer
Roger Williams

Examination of Witnesses

Witnesses: **Professor Tim Allen**, Local Government Association, **Aleks Collingwood**, Joseph Rowntree Foundation, **Professor David Martin**, Royal Statistical Society, and **Professor Philip Rees**, Royal Geographical Society, gave evidence.

Q46 Chair: Good morning, ladies and gentlemen. Thank you for coming here this morning. Just for the record, perhaps you would introduce yourselves.
Professor Allen: Good morning. I am Professor Tim Allen, representing the Local Government Association.
Aleks Collingwood: Good morning. I am Aleks Collingwood from the Joseph Rowntree Foundation.
Professor Martin: I am Professor David Martin, representing the Royal Statistical Society.
Professor Rees: I am Professor Philip Rees, representing the Royal Geographical Society.

Q47 Chair: Thank you. We have a few fairly straightforward questions. As you know, we are trying to understand more about the census process. We have been told that the British population is becoming more reticent to publish information to Government while at the same time becoming more profligate about publishing personal information here, there and everywhere. You have only to look at facebook sites to understand that. Does that suggest that the Government should step back from purchasing data from sources that they commission, and manage with the data that people freely put in the public domain? To put it slightly differently, does the census have a unique selling point that makes it special and different from other sources?
Professor Allen: I guess that trust is a much wider issue; as we know, the public's willingness or not to offer data is part and parcel of that wider trust in institutions and in public policy and services. The bottom line for us is that it clearly behoves the public sector to gather only the data that it really needs for useful, sensible purposes for which it can account to the public. But at present there does not appear to be an adequate substitute for the basic demographic data that the census provides. That does not mean that we could not think about other solutions, but at present the census is the bedrock on which our demographic work is predicated. Of course, that leads to some very practical things that the public do care about, like how many school places you need in Leeds, London, Bradford or wherever.
Aleks Collingwood: The census is not perfect but we think that it is essential, its key role being the allocation of resources. Mainly, we need an affirmed population count. The ONS produces mid-year estimates, but by the time it comes to the next census in 10 years' time there will be a slight discontinuity between the mid-year estimates and the overall population count. The main thing that we get from the census, which we would not get from other sources, is the ability to break things down into small areas of statistics for local area information.
Professor Martin: There are two elements to your question. It is quite telling that a major element of what seems to have been a successful publicity campaign around the 2011 census was that of explaining to the public the purposes to which the data would be put. Explaining the impacts on health care, education, planning and transportation was very valuable in justifying the data collection. My concern is that we may need alternative sources; we may have difficulty in doing another census so well, but there needs to be some framework on which the same justifications can be clearly given. Stopping the census without having an adequate strategy for how to explain and justify the replacement would be a worst possible case, because we would be relying on fragmentary evidence without being able to justify what it was being used for. That would undermine people's motivation to ensure that they were included.
Professor Rees: Your question goes to the heart of democracy. Here in Parliament, you represent the whole people. If you moved from a legally enforceable instrument like the census or an equivalent replacement, you would be relying on voluntary information from interested parties, and that means that your democratic role would be reduced. You would be ignoring all those people who are not vocal enough or educated enough to make a proper representation of their views or characteristics.

Q48 Graham Stringer: Every 10 years, the Government spend about £500 million to collect lots of data. If the Government did not do so, what would you pay to get statistical information? Is there anything that you would pay for, if the census was not available?
Professor Rees: The census has a vital role in many of the resource allocation formulae that central Government Departments use. For instance, in allocating the NHS budget, the Department of Health has to have reliable information on the number of patients and the potential number of patients in future—the number in PCTs currently and in future, and in clinical commissioning groups, and perhaps even individual practices or individual patients. It needs that sort of flow of very accurate information,

and if you don't have it, you misallocate large chunks of Government spending.

Q49 Graham Stringer: The 2001 census had to be adjusted for a number of inner-city areas because it was flawed. Would it not have been better to have used those sources that were eventually used in Westminster, Manchester and one or two other places, rather than spending all that money nationally? Would it not be a more accurate way of getting what I agree is vital information so that we can plan our public services?

Professor Allen: That is a challenging question to answer, because you are not dealing simply with either spending or not spending. There is a cost attached to whichever way you try to access the data. For me, the question is what level of accuracy you need, and therefore need to pay for, and how best and most effectively you can reach that data. From our point of view at the moment, we are very open-minded about examining alternatives to the census, but at present local approaches still rely on the collection of and access to data, and the local approaches, which I understand you heard about last week, are still based on the public sector collection of data. The question is not whether you spend money or don't spend money. The question is how, most wisely, you operate a statistical system that gets you to the data that you really need—the core demographic data.

Forgive me, but that is not the entire answer to your question, because trying to tease through the cost of the alternatives is quite challenging. From the LGA's point of view, we don't rule out the possibility of alternatives. For example, you might be able to systemise the use of things such as electoral registers and the base data that the public sector collects for other purposes, and you might, out of that, be able to create population figures that might be more up to date. However, before you went down that route, you would first obviously want to know what it might cost, but secondly you would also want to know what the weaknesses of those systems were. At present, none of those alternative systems are necessarily fit for purpose, but lots of people are beginning to think about how you might develop systems that would be fit for purpose.

Q50 Chair: Putting it simply, if electoral registration data were to be utilised, one would have to seriously invest in it to get it accurate, because it is not accurate at the moment.

Professor Allen: Yes.

Chair: That is quite an indictment of our electoral system, but that is another story.

Professor Allen: Yes.

Q51 Graham Stringer: May I turn the question round the other way? We accept that there are different ways of getting the core information that we need for planning our public services. Is there anything that could be dropped from the census? Many more questions were asked in 2011 than were asked in 1911. Could money be saved by dropping some of the extra questions, and what questions would they be?

Professor Martin: I don't think you would save money by dropping questions within the census frame, because most of the cost and effort of conducting the census as it stands is in designing the enumeration process. The removal of one or two questions would be quite marginal on the overall effectiveness of the data collection exercise. Much more fundamental is the extent to which we are able successfully to enumerate everybody—delivering a form to every household, getting a form back from every household, and being confident that that is what we have achieved. There are elements of the census—for example, the large coverage survey, which is designed to work out how well we have covered all those people—which could be very valuable if you were evaluating another frame. In fact, if you moved to a system that used alternative data sources, some of the elements, such as the coverage survey to work out how well we are doing and what the biases are, would be ones that you would want to retain. Within the context of the census, the specifics of the questions asked are not really the ones that determine the success or failure of the whole venture.

Professor Rees: The other point is this. Roughly speaking, last time it cost £1 million a question, in terms of the add-on costs of the questionnaire. If you took off £40 million for the 40 questions, you would still be left with the £310 million needed to develop the infrastructure for administering the process[1]. I don't think the size of the questionnaire is moot; it has to be kept short and straightforward because everyone has to answer it. That is the real constraint on the questionnaire.

Aleks Collingwood: Essentially, I agree, but fundamentally, the population count is the most important thing about the census. One of the worries is the impact on data quality overall of all the other major surveys, as most of them are sampled from postcode sectors from the census. That would definitely affect all the other surveys. The census—the overall population count and the local area statistics—is also used, for instance, for ethnic boost samples for other huge surveys such as Understanding Society, which is one of the biggest of its kind. Their ethnic boost sample is from that. Without having this overall population count by these small areas, it would be a hugely expensive operation to work out where the minority ethnic groups are.

Q52 Stephen Metcalfe: If you were starting from scratch, if you had a blank slate, what would your individual visions be for data collection? Forget the history—we want to collect all the data that we think is important; how are we going to do that?

Professor Rees: Just to pick up on the point made by my colleague from the Joseph Rowntree Foundation, to get the population spine right is the critical thing that National Statistics have to do. I recommend that Parliament look again at something first proposed in 1753—namely, a population register. It did not pass the House of Commons then, but although you have had some experience with a national identity register,

[1] *Note by Witness:* These figures are indicative and refer to the 2001 Census. The cost of the 2011 Census is reported by Glen Watson, Director of the Census, in response to Q72.

I think that a population register that had legal force and was comprehensive across the whole population would be good. It would be something that people had to do whenever they moved house, wanted a benefit, wanted to pay taxes or whatever; they would have to record their current address. That would give a really good anchor, year by year, month by month if necessary, for any additional population data that you wished to create through surveys or tapping into administrative data sources.

Q53 Chair: That would be predicated not on a 1753 version but on a modern version of data sharing across Government Departments to make any sense of it, would it not?

Professor Rees: That is the solution with a virtual population register, where you try to match individual records through various probabilistic and other techniques, so that, in the end, when you look at all the Government data sets, you are pretty sure that you have about 98% of the population. You will never get 100%, because there will still be a small gap, with people somehow slipping out of the different registers.

Professor Martin: If I had a completely clean sheet, I would probably run a census that was an unproblematic one. The failing is that the reality of delivering the census suffers, at the opposite extreme, from Professor Rees's suggestion in that population compliance is what makes neither of them work. If I were enumerating a real population, I would probably seek to match the sources that we have in an integrated way, rather than trusting that a number from one place could possibly be matched to an address from somewhere else, and that if we got them all together we would have a reasonable estimate of a local population. I would seriously design a system that put those things together systematically, and I would put some effort into bringing about consistency between sources and the completeness of sources.

Aleks Collingwood: I would highlight how important transparency should be. Right from the start, we should be much more open with the general public about what we are doing and what the main reasons are for a census.

Q54 Stephen Metcalfe: You would stick with the census, even if you had a blank sheet of paper and could design however you wanted to collect this data.

Aleks Collingwood: Ideally, yes.

Q55 Stephen Metcalfe: As is.

Aleks Collingwood: Yes, as is. The thing about the population register is that, in those countries where it works successfully, it is a lot more accepted and supported by the general public than here.

Professor Allen: I would probably reflect the comments that have been made. Nirvana would be some form of population registration system and getting back to trust and public acceptability. I don't think that, currently, the public would necessarily welcome the step of introducing a register. To be successful, there would need to be a linkage to the way in which other public sector data on individuals is collected. There should be a clear public service element to it; you should be able to demonstrate to the public that by going through the registration process you are making certain bureaucratic public service processes simpler for them. In other words, your register has to be multifunctional; it is not just counting people but it is about them gaining a benefit in access to services. Culturally, clearly, to shift to a registration approach would be quite challenging, even if you could demonstrate quite vigorously that it had that element of public benefit attached to it, because, as you identified, clearly there is public resistance to providing data to the public sector.

Q56 Stephen Metcalfe: Whichever system we move to or stay with, are you saying that the Government have a key role in that? It is bigger than just setting standards asking people to collect data on their behalf. It has a role in that. Is that a fair summary?

Professor Rees: Absolutely.

Aleks Collingwood: Yes.

Professor Martin: One of the very important features is that there is no other organisation than the Government which can set up a system that has demonstrable national standing and independence which is not somehow funding the system because they want a particular piece of data or because of a particular regional interest. However those tensions exist, if you try to put things together without that integrated and completely national solution, people will always have possible grounds for challenging the numbers.

Q57 Stephen Metcalfe: Rather than mandating each local authority, setting a standard, and saying, "You do it", which you don't think would work, it would have to come from the governmental level.

Professor Martin: My personal view is that the independence of the census as a nationally consistent exercise is one of the great strengths that any replacement system would have to have. It would be exceedingly difficult to mandate organisations of different sizes and shapes, with different biases inherent in their populations, to produce something that you knew was using the same methodology in every place.

Professor Rees: Another point about making the census or census equivalent local is that you would be handing over the business of producing statistics to bodies that had an incentive to produce particular numbers. They would know that, in the resource allocation formulae from central Government Departments, a large measure depends on the population. I would not recommend that kind of step at all.

Q58 Chair: Professor Allen is going to tell us that none of his members would do that.

Professor Allen: I am sure our members would not wish to falsify figures, but you would need, for reasons well articulated, a system that is transparent and nationally consistent. It is not impossible to do what you have suggested. For example, we have moved pretty well toward that for property location data for addressing. It is not impossible, but it is hard work.

That brings us back to the cost question. Forgetting any of the other questions, is it more cost-effective to do it once, do it properly, and do it as a national exercise, albeit with a clear engagement about why you want to do it and therefore what data you want to collect; or do you want to disaggregate it so that it becomes a data collection exercise by a whole lot of different organisations, with all the safeguards and management that that would require? I don't have a figure in mind for the cost benefit, but you would need to through that process.

Q59 Chair: Is it the case that, whichever way you go, the starting point is very hard? Even the Post Office address file, which local authorities rely on so heavily, is full of inaccuracies. It is bound to be in a sense because it is a snapshot in time which is not likely to be accurate six months later.

Professor Allen: That is true of the Post Office's address file. As for the way in which addresses are collected via local authorities, they have a statutory duty to create the address at the point when a new development is about to come into being. There is a system that feeds that data consistently into a national machine; it is a joint operation between the Local Government Association and Ordnance Survey. I would not claim 100% accuracy; indeed, the census went through a process of double double-checking all that; it was quite a good process because it reflected the data back to local government, saying, "Let's improve on that." It is not 100% accurate.

Q60 Chair: Is it now moving towards an accurate property register?

Professor Allen: Yes, but it is not 100% perfect. It is not the Post Office address file; it is a much bigger database than that, in fact. I can explain the technicalities, but I am not sure that you want to go into the technicalities right now.

Q61 Graham Stringer: If we are looking at using administrative data either as an alternative or a supplement, should standards be set for its collection? If so what should those standards be?

Professor Martin: Yes, although you would have to approach it, as I have just hinted, from the mindset of designing a system that produces the required statistics, because the purposes of the various administrative sources are very different. That is our dilemma with them at the moment. For instance, the purpose of patient registration with general practitioners is extremely different from the purpose of delivering benefits to benefit claimants. The biases in those systems are not amenable to the simple fix of saying, "Record the data in this way." They are recording different data, which is their primary purpose.

It is not a simple matter of telling people to use a consistent format. That is important, but any system would have to understand the inherent differences of who is likely to be counted in those systems, as none of them will be complete. We can converge towards the concept of when someone is resident, what constitutes a household, and thinking about whether we want more consistent ethnicity recording one of those lists, but we cannot make them perform the same function when they are designed for something different. The research challenge that ONS is starting to grapple with in the Beyond 2011 programme is to do with recognising the fact that combining those lists means putting things together that are not exactly the same. You cannot completely standardise them, except perhaps at the technical level, but definitionally they are doing different things. That is where the research challenge lies.

Q62 Graham Stringer: I sort of asked a question—I tried not to—about standardisation, but I don't really quite understand what it would mean. Perhaps you could explain to me what I meant when I asked the question.

Professor Rees: One of the problems with administrative data systems is that they capture different populations. There has to be a very clear understanding of the concept of population or property register or whatever that each system is capturing. They are not alternatives to the definition used in the census. One thing that would improve the situation would be a requirement that, when central Government Departments develop or change their administrative data systems, and there is in place a census replacement, they consult with National Statistics and the UK Statistics Authority so that they make their changes in collaboration with statistical needs. I point to the failure of the Home Office to do this when developing the electronic UK borders system, which was meant to capture people coming into the country. To my knowledge, there was no consultation with National Statistics that would have made those trip-counts—those journey counts—useful for counting international migrants. There has to be joined-up government in any census replacement that uses administrative data.

Professor Martin: May I try to answer your question more concisely, with an example? We could take two different administrative registers currently in existence in two Government Departments; both may record an individual's address but in quite different ways. It is a substantial challenge, but a technical one, to use the standard address list that Professor Allen referred to and encourage all those organisations to record addresses in a standard way such that we do not repeatedly have to rematch them. It is a technical standard. The fact that the NHS register could include someone as resident because they are registered on a GP's list, and therefore use a local service in a particular place, is a very different definition of residence from what we might get if we tried to trace someone from a national insurance number. The numbers are there for different purposes, and the definitional difference is one that the system has to handle. We cannot make them the same, but the technical standards have a long way to go in order to make life easier. I am trying to make the distinction on what standardisation might mean.

Aleks Collingwood: Another of the problems with administrative data is that it can be totally filled up and collected by, for instance, a member of the hospital staff, and subjective things like ethnicity may be guessed. They are not necessarily filled in by the

individual, and it will have a big effect on data quality if we do not have some firm rules on how it is collected.

Q63 Graham Stringer: When Members of Parliament were concerned about the size and accuracy of the electoral register, they became convinced that the better data set was held by Experian, the people who hold data sets for the big supermarkets and analyse who shops where and what they buy. Do you think that those data sets should be made available for public use?

Professor Rees: That depends on the terms and conditions that the private sector organisations negotiate with the Government.

Q64 Graham Stringer: It is partly whether we legislate for it as well, isn't it? We could do that sort of thing here. We could change the rules.

Professor Rees: It would be part and parcel of it for organisations to sell you the information if they did not want to make it available—if they felt that it was not in their interests. Speaking as a social scientist and a geographer, the difficulty with all private sector information is that it looks very good when you talk to them, but if you say, "Show me your methodology; show me your peer-reviewed paper and report, which has been refereed and signed off," they don't want to do that. They keep it completely secret. You have to guess how they have assembled their supposedly good statistics. On that basis, I don't use private sector information.

Q65 Graham Stringer: Can you guess whether they are doing it to high standards? They just don't want to tell you because there is commercial value in it. Or are they not doing it to high standards?

Professor Rees: We don't know: they keep it secret. If you look at medical research, which is the gold standard, there are very high standards in publication and transparency. However, half of the trials funded by drug companies have to be thrown out because they are inadequate. That is a case of private organisations producing information and having to meet high transparency standards but failing. Forcing people like Experian to reveal all their marketing data would be a step too far.

The other point is that marketing companies are interested in the top end of the socio-economic spectrum; they are not interested in the bottom end. They are satisfied if they have all the people with disposable income to spend in the shops, but they are quite happy to miss out of their results all those who are short of the readies.

Professor Allen: I would add that, while I would entirely agree that these are not necessarily peer-reviewed, Tesco rely on their data for marketing and they are a very successful company. Presumably, it is very data-driven, and, presumably, they do get something right for their purposes, but the words "for their purposes" are quite important. As my colleague said, if we were to compel it to publish, the question is the degree to which the data might be appropriate, given the challenges described earlier in relation to the use of other sources of administrative data in the public sector because Tesco's sample will be skewed to Tesco's particular market niche. It will focus on that, although it might want to expand to other markets and therefore use other data, but it will still be a particular cut and a particular view of the data and the information. The question is not whether they are very good data sets—I have no reason to doubt that they serve their particular purpose, and many of us would love to have a look at them; they would be very interesting—but they would not substitute for basic demographic data consistently and transparently collected to a standard somewhere by the public sector, or triggered and managed by the public sector.

Q66 Pamela Nash: I take you back to Graham's previous question. Each of you highlighted problems with us using other forms of data collection from public bodies, but there are other countries that seem to be successfully doing that. The example we have been looking at is Sweden. In your opinion, is Sweden doing this successfully, or are there major problems? Do you foresee a time when the UK will not need a census or anything similar, or do you believe that the census provides something unique that can never be replaced?

Professor Rees: Sweden has some of the best social statistics in the world. I am told that at every conference by my Swedish colleagues. They start with a population register and a unique personal identification number, and citizens are used to supplying that to all the administrative registers. The task of merging the administrative registers at the individual record level is fairly straightforward. We are a long way from that. Take the names and organisations you are representing in this list. For this call, there are several mistakes in names and affiliations, so you can see how difficult it is to match lists without a unique personal identification number. For instance, it refers to my colleague Alex Benton, who in fact is Peter Benton, and so on.

Professor Martin: It would be possible to create a good system using linked lists, but that is exactly the core of your question. It will not happen, because it will not evolve to a position in which the lists are sufficient to replace the census. That has to be an active design decision. The reason it works, as you heard, is, first, that there is a common identifier; a step towards that in our world would be to use standard addressing. Some of those technical standards would help us a little, but they would not necessarily resolve the fact that we need to know whether we have the same person; also, there is a level of population acceptance, which has taken a long time to develop. If you investigate how those countries reached that point, you will see that they did not do a census and then decide halfway through not to do the next one; they developed a replacement system, in some cases over decadal time scales. We obviously have better technologies and can make changes faster, but nevertheless we cannot switch off one system and use the other one the next day. When you look at how well they work, you can see they are the product of a very long period of development and we can learn in parts from that, but it would not happen automatically.

Q67 Pamela Nash: We all have a national insurance number. I appreciate that this is not its primary objective, but each person is given one at birth and it is used by a variety of Government Departments. How far away are we from that being used as a unique identifier?

Professor Rees: When you reach 16 that is the case, I understand.

Q68 Pamela Nash: You are given your card when you are 16, but you are allocated the number much earlier.

Professor Rees: So you miss out children aged 0 to 15[2]. The national health register is probably the most comprehensive, because it includes children, but again, it misses people out. If you put them all together, you are probably covering virtually all the population if you can match them. That is the critical research task which I know National Statistics is engaged in. There is lots of social science experience in this matching of record level. But, in the end, unless you have a unique personal identifier on each database, it is still only a probability that a person in this register is the same as the person that has almost the same characteristics in another register.

Aleks Collingwood: I am not sure how the national insurance number works for people who have migrated into the country, but that is not the main point that I wished to make. The main point is that the most important reason for the census is that it is vital for allocating services overall. The only possible alternative would be a central population register to get the overall population count, as far as I see it. In places like Sweden, as my colleague said, it is a lot more accepted and supported, but I don't know whether that would be more acceptable here than the census; I think that it would be more beneficial to keep going ahead with the census.

Professor Allen: My response is the same as that of my colleagues. With the Swedish experience, it took 20 or 30 years to get where they are; and then there is the cultural dimension—but never say never. I think it is feasible for us to move to that system, but I would worry if we said that we don't want another census in 10 years on the basis that we are going to achieve the alternative. I would want to see the evidence for that being achievable. There are implications not just for the practicalities of the data standards and all the technical stuff, but also for the cultural shift in how we operate through public sector departments, agencies and so on to get us to the necessary state where we could effectively substitute a different system, making it work and making it cost-effective. That is quite challenging. If you take a long enough horizon and a sensible costing, I am sure that it is possible.

[2] *Note by Witness:* A Child Reference Number (CRN) is created on DWP's Customer Information System for every child for whom child benefit is claimed at the point of award. The number is used primarily for the administration of child support. This number is then registered on HMRC systems as a National Insurance Number at age 15 years 9 months and is then notified to the young person. DWP allocates NI numbers to adults, usually migrant workers when required for employment or benefit purposes.

Q69 Roger Williams: The structure of Government has probably changed a bit since 1753—I don't know whether for better or worse. As the structures change, the Government may need different sets of data and different information. Should that in some way determine the style of the census and how it collects that data?

Professor Martin: There are certain core demographic characteristics for which we have not seriously changed our need over a very long time scale. The unambiguity of those in the central system will remain.

One of the difficulties with the census is that, if you do it only once a decade, you have a long time window waiting for the data and only a small opportunity to change the questions. That is, potentially, one of the great strengths of a system that uses other sources, in that you can adapt on a more measured time scale. But there is also the danger, which must be recognised by anyone making a decision which will affect the future for a long time, that, if there were to be some other change in Government policy that changed an administrative source that was not under the control of the Statistics Authority, it would be very easy to undermine your administrative population system without having a census from which to go back and check. A major change in a benefits policy that took a lot of people out of the register would clearly undermine the system that you had built. It cuts both ways, but my general feeling is that working from sources that you can change more frequently is the right move.

Professor Rees: May I add to that? The options that ONS and others are looking at beyond using the traditional census all include an element of a big social survey. Because it is usually administered by interviewers, it can last much longer and have more questions than the regular census, and you can introduce new questions. There is room for introducing a lot more questions on health of the population than we ask—we ask only a couple in the census which are very basic—and on health-related lifestyles. We do that in current national surveys, but they are all too small to get at the local and small-area detail. But if you had good methodology that linked a good population spine built on administrative data—hopefully, a population register with a large survey—you would be able to keep the information up to date to suit the current needs of society and Government.

Q70 Roger Williams: Basically, the question revolves around whether the Government should be asking specific questions or gathering large databases that can be interrogated for specific purposes.

Professor Martin: It is the latter, because we are not very good at anticipating next year's specific questions. That is why the census remains so important. We could replace it with something else if it could answer those general questions, but designing a system that asks only specific questions is probably a short-term objective.

Professor Allen: The world changes very rapidly. For example, you are well aware that the recent census switched, although not entirely by any means, to the use of the internet and web-enabled technologies to

fill in the data. It seems to me that a bedrock of core demographic data underpins some very basic human needs around health, housing, mobility and the basic functions of government, which is to understand the nature of the population and to be able to respond to its needs. Those things may change, but the consistent factor in all of that is the need for the data that underpins those decisions to make them wise decisions. In that, I would echo my colleagues. Those requirements are surprisingly stable over time.

Aleks Collingwood: I agree that the most important part of the census is the basic demographics, with age, sex and the population count being the most fundamental. Although giving a picture of the whole population in health, education and everything is important, the basic demographics are the main key for allocating resources and so on. However, there are so many very robust national surveys that we can use for other things. For instance, the English Longitudinal Study of Ageing, the Millennium Cohort Study and the Understanding Society survey, which is the biggest of its kind, all cover a wealth of information. Perhaps we should look further into linking these studies to the census.

Professor Rees: Let me give an illustration of why we need a national instrument that is comprehensive. I cite education. We have an excellent school census national pupil database, incredible longitudinal information about pupil performance and so on. Why do we need education questions in a census or the attached survey? It is because the school census does not cover all of the school population. It leaves out all the private schools. That is a decision that this body made at some point in the past. If you want to improve the school census, you would have to extend the coverage. That is the essential purpose of having a national census-like instrument, so that you can cover everyone with those kinds of questions.

Chair: I thank the four of you. It has been an extremely useful session.

Examination of Witnesses

Witnesses: **Adrian Alsop**, Director of Research and International Strategy, Economic and Social Research Council, **Jeremy Neathey**, Deputy Director of Policy, Economic and Social Research Council, **Glen Watson**, Census Director, Office for National Statistics, and **Peter Benton**, Deputy Director, Office for National Statistics, gave evidence.

Q71 Chair: Good morning, gentlemen. Thank you for attending. Would you please introduce yourselves?

Adrian Alsop: I am Adrian Alsop, director of research and international strategy at the Economic and Social Research Council.

Jeremy Neathey: Good morning. I am Jeremy Neathey. I am deputy director of policy at the ESRC.

Glen Watson: Good morning, I am Glen Watson. I am the Office for National Statistics director for the 2011 census for England and Wales.

Peter Benton: I am Peter Benton, a deputy director of the Office for National Statistics. I am currently leading the Beyond 2011 programme.

Q72 Chair: Thank you very much. We have heard that the last census cost about £500 million. In the current economic climate, is that a justifiable expense?

Glen Watson: I believe that it is. The estimate of the final cost is £480 million. In real terms, as opposed to cash terms, and taking account of inflation and population growth, that is about 35% more expensive than the 2001 census. The estimated economic value of conducting the census far exceeded that £480 million. The business case prepared for the Government and considered by the Treasury in an earlier spending review estimated that the economic value of the census to the UK was probably in excess of £1 billion. We did not try to map every single case and every single use of census data that has been made, because the number was starting to get so large; we thought that we had probably gone far enough.

Jeremy Neathey: I would essentially agree with that. As you have heard all morning, the census is probably the most effective mechanism we currently have for measuring the demographic profile of the UK's population. Therefore, it is a critical resource for social science, and, despite the cost, the benefits to social science research are enormous. Again, as you have heard most of the morning, there isn't a reliable alternative to the census.

I pick out a few of the issues that have been brought to the Committee's attention. The level of geography is a key one; coverage is another. Those are key areas in which other data resources just cannot achieve the same level of coverage as the census. There is also the key question for social scientists of continuity over time and being able to measure change, which is the life blood of social science and important in all other aspects of the way in which Government work and so on. It may be expensive, but it is worth every penny. We are not averse to the idea of looking at alternative ways of doing it in future, but the cost of the last census was a worthy and worthwhile investment.

Q73 Chair: Let me put you on the spot. ESRC's budget for the current year is about £174 million, give or take.

Jeremy Neathey: Grant in aid, yes.

Chair: We are talking about a project costing the best part of £500 million. If I was able to magic a transfer of £500 million as an adjunct to your budget for the next 10-year cycle, what would you do with it? Would you choose to run a census, or would you do other things?

Jeremy Neathey: As I say, there is no reliable alternative at the moment. There are several things that we would consider doing, but I suspect that many of them are already in train through the Beyond 2011 programme, which you might hear more about in future. In a very modest way, we are working with ONS in looking at those and ensuring that the academic community is actively engaged in the

process of looking at the alternatives. You have heard some of the issues that surround trying to put an alternative in place. Those are the issues that need to be addressed head-on to see whether it is feasible to move beyond the current model. If you were to give me £500 million, I suspect that I would spend a good part of it in the same way as ONS is doing now, which is to develop an absolutely secure evidence base to take us in a direction away from the current model, with its associated costs. Essentially, we would need to build a similar evidence base to that being built by ONS if we are to transit out of the current arrangements.

Q74 Chair: If the census was stopped, you are arguing that it would be necessary to collect a lot of the data in another way—a national address register or whatever—but that other things would be necessary in a core data set. Would it be possible, adding them all together, to do it for a lesser sum, or would you end up reinventing the wheel in creating a census?
Jeremy Neathey: I suspect that, ultimately, you could do it for less. If you look at models in place outside the UK, I suspect that their investment in censuses is less, but I think that some are comparable. But the up-front investment in getting that right would still be significant, given some of the issues that have been raised about the definition of other data sources, their quality and comprehensiveness.
I shall not go over those issues again because you have heard them more eloquently articulated by others this morning, but they would still need to be addressed regardless of whether the funding was being directed through the ESRC or other sources. In any case, I suspect that the up-front investment would be significant. Over time, the cost profile would probably come down significantly once the systems were in place and could be deemed to be reliable and robust. On that journey, you would have to consider the costs and benefits of doing certain things, particularly on issues such as geography; the position of local geography would be a key factor among other issues. I suppose my short answer is that you probably could save money but it would take time. The experts here today on my right may take a different view.
Peter Benton: May I chip in at that point? The fundamental basis on which we will be evaluating the potential alternatives is by understanding the benefits that they provide for all the uses that people have articulated—the small area geography, the multivariate rich picture of the population—and setting against them the cost of the alternatives. There are very real questions about the alternative costs, how high the peak of that cost would be in setting up an alternative, and how low the cost would be in steady-state after getting over that hurdle. At this point, we do not know the answers to those questions. That is the purpose of the Beyond 2011 programme, and ONS is taking a good look at those questions.

Q75 Chair: Have you ever sat down with any of your customers in the various Government Departments and produced a cost-benefit analysis of the work you are doing?
Peter Benton: Yes. As Glen explained, in undertaking the 2011 census we looked carefully at the costs and benefits. We evaluated the benefits for the 2011 census primarily by looking at the process of resource allocation to the health service and local government, understanding the errors that might accrue if the population statistics were wrong and where the funding might move around. But there is a whole range of other uses. We did not actually value the financial benefits of all the small-area data that is used by social scientists, and one of the challenges is to understand how we might do that. We are looking to work with the social science community to put a value on some of those benefits so as to inform the decision.
Glen Watson: May I add one point? It is also important to look at the international context here. In constructing the case for the 2011 census, we looked long and hard at the cost of running censuses in other advanced western democracies. The UK census stacks up rather well compared with that of the US, which spends something like two or three times as much per head of population on its census, and a number of countries conduct a census every five years, including Australia, New Zealand and Canada. Consequently, over a 10-year cycle, they spend around twice as much per head of population as we do on a census. Although the cost has grown, and although it is right that we should look at a programme of alternatives for the future, it is worth keeping the international context in mind.

Q76 Chair: Mr Watson, it would be extremely helpful to have a paper on those international comparisons, if you have any data.
Glen Watson: Yes, we can follow up with that.

Q77 Stephen Metcalfe: Between them, your two organisations are the largest publicly funded collectors of social data. What should the balance be between state-gathered data and research-directed information? If you have finite resources, which side of the fence is more valuable? Where does the balance lie?
Glen Watson: In terms of constructing a census or a replacement for the traditional field enumeration census model, I would say that we have to take account of all uses of the information, whether it is state use by central Government or local government, or whether it is by the research community, the voluntary sector or the private and commercial sectors. Our job is to weigh up the strength of the case being made for different types of question, different types of data and different types of analysis, and to come up with something that optimises the value to the country as a whole. We don't prioritise one sector over another—state above research, or research over state—but try to come up with a balanced view that meets as many needs as possible within the budgetary constraints and the bounds of what is logistically possible and publicly acceptable.
Adrian Alsop: Indeed, they are complementary in many respects. We have a regular dialogue with ONS to ensure that that complementarity is made the most of both now and in future.
Jeremy Neathey: It is worth noting that the data currently produced by ONS and other Government

Departments is extensively used by social scientists. In some respects, the boundary line is not irrelevant—I would always like more money for ESRC to collect data—but the statistics from our UK data service show that, of the top 10 data sets used by social scientists which access that service, half are funded by ESRC and half are funded by ONS. There is a sense in which the collective is the important thing in terms of access to resources. We are working with ONS and other Government Departments in a number of ways to ensure coherence. We have something called the UK data forum, which brings ONS and other Government Departments together to talk about data issues and to hammer out common issues and problems; we also have a framework called the national data strategy, which tries to identify key data issues that are of interest to all organisations. We have a regular dialogue going on, which, in some respects, allows us to optimise the opportunities for social scientists by working with ONS rather than in opposition to it. We have a pretty comprehensive data infrastructure in place—it is one of the best in the world—but we could always improve on it. In answer to your original question, I think that we are in a complementary space here.

Q78 Stephen Metcalfe: The UK data strategy plays an active part in getting that balance and making sure that there are no gaps.

Jeremy Neathey: I think that it does, and I hope that ONS would agree. Certainly it is a framework that we use to drive forward our own strategy in working with ONS.

Q79 Stephen Metcalfe: You say that, potentially, we have the best system in the world for collecting and using data, but how well is data collected locally, and how does that co-ordinate with national statistics? Let me add a caveat to that. If data is collected locally and funded from the public purse, should it conform to national standards so that the data is consistent across the whole country?

Peter Benton: Local collections happen in all kinds of places for all kinds of reasons. People have their own reasons for undertaking those things, and the funding they have is for a particular purpose. It is not for ONS to say, "You must do it in this way." We can set standards and give advice on particular classifications that people might use, but the census and the information that it provides, and the information that we would need to provide in future, needs to provide the bedrock that lets others calibrate their local surveys to a national set of information in order to adjust for missing people and the bias that results. The national picture gives the bedrock that enables the local activities to function.

Jeremy Neathey: I agree that the census plays a critical role in framing how we go about the different forms of data collection that we fund. A colleague in the previous evidence session talked about one of our biggest surveys, which is the biggest in the world. It is the household panel survey called Understanding Society, which is spread over 40,000 households. But, in order to deal with issues such as ethnicity, we use the census as a frame to do the sampling, and to find out where we need to go to ensure that we get a representative spread. Going into local areas and doing analysis on where you put your boosts and how you organise the whole data collection strategy for major and expensive data collection exercises is guided by the census. In that sense, the census provides a frame, but it also provides an approach to standards on how to go about collecting data on a whole range of areas.

Glen Watson: It is not just in the academic or research community that that is true. That is also true in the private sector. I know that a lot of market research companies and other research organisations use the census results as a framework for designing and constructing their own samples and survey instruments that serve a whole host of different needs. It is true within central Government as well, so when ONS is designing a household survey—we do many such surveys—we typically use the census to provide the national framework for constructing the sample. When we try to take sample estimates from the social surveys and gross them up to something that is akin to a regional or national population, the census results are used to scale up those figures.

One more thing is related to this, which is the important part that the census plays in providing denominators for national rates, or regional or local rates. Whether it is a mortality rate or an unemployment rate, or some other rate, that is the sort of thing that we use and see in the press, and policy makers use on a daily basis. There is often a denominator sitting underneath that is derived from the census or from the population figures pulled from the census. It is quite a complex picture in terms of the spin-off uses of the census results.

Q80 Stephen Metcalfe: I have a final question. Listening to what you have said, I suspect that you will not agree, but Les Mayhew told us that the census as it is at the moment stifles social science innovation. Is that a fair comment, or do you disagree with that?

Adrian Alsop: We would disagree with that, as you might expect, because of the point about complementarity that I made. The existence of the census has not in any way prevented us from developing Understanding Society. With the birth cohort studies, we are in the process of adding social and biological data to the major data sets further to enrich their capability. While there may be other factors that bear on innovation in social science, the census does not in any way restrict our ability to innovate.

Glen Watson: My feeling is exactly the opposite. I believe that it actually encourages innovation and provides an incredibly rich treasure trove of data that other organisations can use and build on. That is true of social research instruments, but it is also true, for example, of the way that we will be exploiting the 2011 census results through the web, and the sort of web views of the data that we will be providing and the machine-to-machine transfer of small-area aggregate data that will be used by other website builders in the private, public and research sectors to power their own websites and mash together the census data with data from other sources. That is very

much the way that the world wide web is going, it is the future direction of the internet, and the census will be part of it.

Q81 Graham Stringer: I return to Andrew's initial questions about the value of the census. There was no census in 1941, and, arguably, between 1945 and 1949, there was the biggest change in the provision of public services in this country that there has ever been, with the health service, pensions and various other things, but as far as we can tell they were done pretty well. Does that not suggest that the census is less necessary than you are making out?
Glen Watson: I am not a historian. I was not around and I cannot comment in any detail on how social policy was developed in that era. I would imagine that, by the time the 1951 census was taken, it was eagerly awaited and was reporting on some incredible changes in society that had taken place over the 20-year period that possibly then sent people off in all sorts of different policy directions. It was such a different time that I find it quite hard to answer that question.

Q82 Graham Stringer: I am not surprised that you find it hard to answer the question, but it is an interesting point, is it not? We saw the planning of the health service, a change in the ownership of many industries that had previously been in the private sector, and changes in social security and pension provisions. As I said, they were done pretty well. In some ways, they were socially administered better than we do it now. The census data that would have been available at the time would have been 17 or 18 years out of date. Don't you think we have something to learn about that before we spend £500 million on the next census? Do you think it is a field worthy of study?
Glen Watson: It is always worth looking back to see how census results have been used and how statistics generally have underpinned the development of public policy. It is also worth considering changes to how policy was developed over that period, with a greater reliance on evidence and a greater expectation of transparency. We are living in an age when, if there is a significant policy change, people expect, and have a right to expect, to see the evidence and the data on which that change of direction is based. Perhaps it was less so in the 1940s.

Q83 Graham Stringer: A central issue in the previous evidence session and this one has been administrative data compared to census data. How usable do you believe administrative data is?
Peter Benton: There are clearly a lot of potential benefits in the existing administrative data sets. As you have heard, the NHS register covers a large number of people in the population, but not all, and the electoral roll covers a large number of people, but not all. There are challenges in coverage, but there are also significant challenges with definitions.
You have heard people talk about why the information is collected; it is there to support a particular purpose. Taking the example of the NHS register, somebody might live in Winchester and travel to London during the week, and they may choose to register with a doctor in London during the week because that is the convenient place for them to visit a doctor.

Q84 Graham Stringer: That applies to MPs.
Peter Benton: Okay. If you tried to count the population from those sources without understanding some of those definitional issues, you would get a wrong result. Some of our early analysis of those sources in the Beyond 2011 programme suggests that the NHS register in London over-estimates the population by 10% or 15%. In other parts of the country it is lower and more of an accurate reflection of the rural parts of the country where people do not move around so much. If you were to use it at face value, you would make wrong decisions. The challenge is to understand those patterns of coverage and those differences in definition, and build statistical models that enable us to produce accurate outputs on the definitions that we want to use. We have some ideas about that, and we are discussing our thoughts with others, but it will take a year or two to really understand whether we can do that.

Q85 Graham Stringer: Mr Watson has just made the point that we live in a different world from that of the 1940s. People want to see the evidence, and, if the evidence is not there, they will sometimes challenge things legally. Would the public accept the use of administrative sources, even with the statistical analysis and changes that you are talking about? What are the risks involved and would the public accept it?
Peter Benton: To an extent, they are unknown at the moment, but we know that people have strong views about the census. There are some who very strongly support it, particularly genealogists, who love to have a historical record. There were others at the time of doing the census who clearly said, "We don't think Government should be doing these sorts of things. It's an intrusion." The range of opinions is similar on the use of administrative data sets. Some people say that it is obvious that we should use it; it is more efficient for Government; but others say that it would be an intrusion of their privacy. That debate is probably in its infancy. People are not necessarily well-informed about data sets, how they are currently used in Government and how we might use them in future. As we start talking about some of the research that we might do in the Beyond 2011 programme, opinions may start to emerge and change.

Q86 Graham Stringer: Is the core of that debate about having contemporaneous information from administrative data sets against having a national data set that is comparable over hundreds of years? Is that the basis of the argument?
Peter Benton: There are many facets to the kind of information that people want that we need to consider. There is accuracy, there is the amount of small-area detail that people can get, and there is the multivariate analysis. We have heard all those things, but the idea of frequency is really important. The census currently provides information once every 10 years. It might be possible to move to a system where we could produce good statistics every year by using those

administrative data sets. The question that results is this. We may be able to do it more frequently, but can we do it as accurately, and can we get information down at the small geographic areas that people use? If we were to rely on a survey to plug some of the information gaps that we simply cannot get from administrative sources—for instance, information about carers—you might move to a survey to do that. If you had a survey, by definition, it would not give you detail down to small geography. There are some real choices and trade-offs on accuracy, frequency and geography, in particular, and that is the focus of the current consultation that we are running with users in the Beyond 2011 programme.

Q87 Graham Stringer: All the nationally collected administrative data sets require checking and correction. Which is more difficult to correct?

Peter Benton: At the moment, I would say that they are all difficult to correct. Do you mean which is more difficult in terms of people missing or being put in the wrong location?

Q88 Graham Stringer: I suppose I mean: which would take more resources and effort to check? Which would be more technically difficult?

Peter Benton: The challenge of knowing whether people are in the wrong location on a particular register is probably one of the larger ones, as I explained. It is not only about people who have two addresses. For instance, people may move house and don't need to see a doctor for two years, and therefore they don't tell a doctor that they have moved. So they are still recorded at a previous location. With the level of population churn that we have, that is a real issue. Trying to understand who has been correctly or erroneously included in different sources is something of a challenge, but it is not insurmountable. There are potential ways of dealing with it, but we need to do some research to see whether it is feasible, what level of accuracy it would achieve, and at what cost.

Q89 Pamela Nash: We are now in the middle of December 2011, yet we are still waiting for the first tranche of information to come from the census. Can any of you shed some light on why it takes such a long period of time to collate the information?

Glen Watson: Census day was 27 March, as everyone knows. The census field work finished in June, after we had spent a couple of months sending people out knocking on doors, reminding people of their obligations and chasing up non-responses. The capture and processing of all that information at our large processing facility in Manchester started at about the same time. The initial capture, and—what we call keying and coding—of some of the written-in responses on the census form has now completed, just about, very recently. We are now in the phase where we are doing further statistical processing of that data within the Office for National Statistics, and we are embarking on a process of quality assuring all the results.

It might be worth explaining that people tend to think of the census as a count, as a form that every household fills in and sends back, but that is only partly right. That was the case historically, but there are now other parts to the process. They include matching the census coverage survey records, which is a 1.5% sample of households in the country, with the returned census forms. That gives us an assessment of those households that must have been missed or did not fill in the form so that we have a basis to adjust the results and make sure our final published results are a best estimate of 100% of the population. That is quite a complex and labour-intensive process.

We also use administrative sources from a range of places, including council tax data, data on benefits and the school census data. We use all that information, not at an individual or personal level to link it, but in terms of the statistical counts at a small-area level so that we can confront the results of that against the emerging census results. We can then ask the question whether the results are right. Does this point to any anomalies? Do we need to investigate or adjust? The combination of the adjustment for the undercount, using the census coverage survey and all this quality assurance, is quite a substantial task. For example, to help you visualise it, we have a system that can produce something like 80,000 diagnostic charts to compare the census results with all of these other sources, to check area by area that the results are plausible and that they do not need to be adjusted for any anomalies. We also have to get data from many other institutional areas of the country, for example, from military bases, prisons and boarding schools. We get data from lots of these other sources to check against the institutional part of the census form completion that was being done earlier in the year.

It is an incredibly complex process, and it will be next summer before we are in a position to get out the first set of results. The more detailed results will follow after that. In fact, the more detailed results that start coming out later next year are often the ones that social scientists eagerly await when we issue the results, for example, through micro-data samples or linkage to the longitudinal study. These are the sorts of things that social scientists are waiting for. The first results to come out next summer will be the population estimates by local authority, age and sex. They are certainly eagerly awaited by those in Government who want to use the census results to update the next round of resource allocations to local areas.

Q90 Pamela Nash: I appreciate the amount of work that must go into providing analysis of that detail, but is there any way in which the raw information that has been collected can be published at an earlier stage?

Glen Watson: I am sorry, but can I be clear that I understand the question? Are you asking whether there is any way in which the raw information can be published in terms of the census counts rather than the results that take account of the adjustments that we make?

Pamela Nash: Yes.

Glen Watson: No, we don't do that, and we don't intend to. That is partly because it could mislead. We are sure that there are differential response rates in different parts of the country. We set some targets this

year for the response rate, and we are very pleased to say that it looks as if we have met them. For example, it seems that we have a better than 80% response rate to the census form in every single local authority, and we have achieved a national average of about 94%. Nationally, that is a bit better than was achieved in 2001; but in the local areas, where there were some problems in 2001, we have done substantially better this time. We have done a lot better in London, for example, where we had a lot of problems in 2001. If we were to publish counts without those adjustments, it would be misleading. Some parts of the country have a 98% response rate and some parts might have an 85% response rate, and that would lead people to the wrong conclusions.

Jeremy Neathey: There is a general acceptance within the social science community about the complexity of trying to get the numbers out, and accuracy over speed is probably what the social science community would want. Of course, we would like to get hold of the data as quickly as possible, but that constraint is acknowledged. It takes time to get the data that the social scientists really want—the micro-data and some of the longitudinal data out there—but there is a recognition that delivery of the data used for resource allocation modelling and so forth probably has to come first. There is an acceptance that that is the way it has to be, even though we want to get there quicker, if possible.

Peter Benton: In the past, we used to do exactly that. After the 1991 census we published raw counts, followed by adjustments for coverage. In the end, the users said, "It is confusing. Please don't do that next time. Just give us the best results that you can when you're ready."

Q91 Graham Stringer: The obvious question resulting from your answer to Pamela's question is how you know that you have returns of 85% when you are trying to do the original count.

Glen Watson: We have two mechanisms. First, as people have already said this morning, for the 2011 census we built a national address register; we brought together all the different address sources—from Royal Mail, local government, Ordnance Survey, utility companies and other places—into a single high-quality list of all the addresses. We know what proportion of returns we have had from those addresses. That is one measure.

Q92 Graham Stringer: So it is a percentage of the addresses.

Glen Watson: That is one indicator. It is a little more complicated than that, however, because it is possible that a lot of those addresses are vacant or not yet occupied. It might be possible that within an address a census return came back but two or three people were missed off it. All these things are possible, and it is the comparison between the census coverage survey and the census itself that allows us to make an estimate of the national response rate. There are methodological papers on this, and there is a very complex statistical methodology that leads us to those conclusions, but that is the basis of it.

Q93 Graham Stringer: When you are doing your check with the 1.5%, one of the things that concerned me about the 2011 census, among other things, was that, if you look at affluent areas, say Buckinghamshire, and compare them to poor inner-city areas such as in London or Manchester, the disability statistics are higher in the poorer areas. That may be true, but is it not what one would expect, and it might represent the literacy rate of people in Buckinghamshire compared to elsewhere? Do you correct or check that, as it would have a terrific impact on the funding of social services and health?

Glen Watson: We do a number of checks. The first thing is to try as hard as we can to get the response rate up in the inner city areas with traditionally hard-to-count populations; for instance, with young men and most ethnic minorities, we find historically that their response rates are lower. Our community engagement, and our publicity campaigns and marketing, was all very much focused on making sure that we got the message across to those areas, and that we did it in a nice, simple-to-understand way. For example, with our partners in local authorities, we ran something like 7,000 completion and awareness events up and down the country, trying to bring people in from the community and neighbourhoods, explaining what the census was about and how to do it. We had people who could speak the languages of the people who lived in those communities and, there and then, we could help and advise them on how to fill in the census form.

But we also do other things. For example, we do a quality survey, a small survey at the end of the census operation, where we go back out and conduct in-depth interviews with people to go through and check their understanding of the census questions and, through a conversation, to get the right answers from them; rather than it being self-completion, they have someone to advise them. It is a very small sample, but we bring those results back to head office and analyse them, and we can then compare them to the results of the census itself. That enables us to come up with estimates of how accurately and reliably the questions were answered. Part of the collection of material that we publish with the census results will be the response rates by areas and population groups, the extent to which questions were filled in or left blank, and information on how reliably we think the questions were filled in. I do not claim perfection in any of this, but we are very transparent, and we publish all of that information so that social scientists and other users can form their own opinions and draw their own conclusions about how good and reliable the data is.

Q94 Graham Stringer: Do you think that their perception about disability is correct?

Glen Watson: I know that disability is one of the barriers to being able to engage with the census process—I know that—and we have worked very hard with the Royal National Institute of Blind People, the Royal National Institute for Deaf People and various other groups to try and help people. Whether or not the returns from people with disabilities are of the same quality and reliability as others, we really need

to have been through the process of publishing the results to be able to analyse that.

Q95 Graham Stringer: My final question is this. The census is a huge one-off event every 10 years. We get better at most tasks if we do them twice. Sometimes, people are not up to the task. It is my view that Mr Cook was not up to the task in 2001; he did it very badly.
I am asking the question from both ends. If you are going to have a census every 10 years, do you think there should be an attempt to recruit people who are young enough to do it twice? Secondly, if it is clearly going wrong, how do you deal with somebody who is not up to the job? It is a one-off, rather like running the Olympics. If you do it only once, you either get it right or wrong. How do you deal with that from both sides?
Glen Watson: Some countries deal with it by doing a census every five years. That gives a lot more continuity from census to census in terms of the people, the systems and the processes, and they can incrementally improve and change the way things work. That is one way of dealing with it.
I cannot envisage a situation in this country where the Government would give us the money to do a census more frequently than every 10 years. If we stick with the 10-year model, whether or not there is a traditional census or some variant in the future, or whether we have a completely different model, if it carries on every 10 years, it is important to try and get the expertise from the last census, and as much of it as you can, and incorporate it into the process for the next time. For example, I was not involved in the 2001 census, but many of my senior managers and my team were. Quite a lot of people at headquarters designing and planning and working on it had the experience of the last census. In fact, some have had experience of the last two or three censuses, and we try to retain that expertise as far as possible.
I cannot comment on Mr Cook's experience, but my understanding is that when he came in to the Office for National Statistics most of the preparations had already been done. I came in to run this census in 2007, and you would be surprised how much of it was already decided and set out. It is a very long-term project, and the thinking and design takes place quite early on in the process.

Q96 Chair: In terms of changes in society, it was fairly obvious from an earlier conversation about mobility that the population has become considerably more mobile, and very rapidly so. Given that there is a 10-year gap between censuses, how can the design of the next census take into account changes in a society that is moving very fast? How do you go about it? How can you ensure me it is accurate?
Glen Watson: Between 2001 and 2011, we saw the population of the UK grow by more than 3 million, and two thirds of that was a direct or indirect result of migration. When doing the 2011 census, we knew that we had a very different population than we had 10 years ago.

Q97 Chair: It has also become more mobile, with people working away.
Glen Watson: Yes. The population is more diverse. There is more migration and more internal movement. There is more mobility, with more people living in different arrangements and family members living in different households, and there are more people working away from home. We addressed that in 2011 by adding some new questions and gathering new data to help us understand it; for example, you will be aware that for the first time there was a new question on second residences. That will allow us to publish population estimates that are not only on the usual residence basis—

Q98 Chair: That was a very poorly designed question, I have to say, for those of us who had to fill it in.
Glen Watson: Okay, but we tried very hard when designing these questions.

Q99 Chair: That was a good example of a question that was incredibly badly designed. If you have a second home, it should have said, "If you are living in a second home, you should be filling in both forms at the same time." It was all very vague. If you followed the questionnaire and said, "I'll be filling in this one today, and on Thursday I'll fill in the next one," you would have sent confusing data back to the system.
Glen Watson: We tried various alternatives and tested them on thousands of people. They tend to work very well in most cases, but there will always be examples and there will always be people for whom it did not work awfully well.

Q100 Chair: Perhaps I can't read. I found it incredibly difficult to follow the logic of that question.
Glen Watson: Okay. I am not sure what to say about that. The reason that we added that question, and the reason why it is important, is that we will be able to publish statistics not only on the usual residency basis but also on some other population bases. We will be able to publish estimates of working week populations. We will be able to publish estimates of students, whether they are in their term-time address or at home. For the first time, using the census, we will be able to publish estimates of short-term migrants who are in the country for only a few months.
That is part of the richness that we will be publishing, and social scientists will be able to use these different cuts of the data. The other reason why we add questions like that is because, if we do not, people will find excuses not to fill in the census. Unless we include something about visitors or second homes, people who are away from home on census night will find a reason not to fill it in at all, and it is quite important that we do not miss people.
Chair: Gentlemen, thank you very much indeed.

Wednesday 18 January 2012

Members present:

Andrew Miller (Chair)

Stephen Metcalfe
David Morris
Stephen Mosley
Pamela Nash
Graham Stringer

Examination of Witnesses

Witnesses: **Richard Bartholomew,** Joint Head of the Government Social Research Service, Department for Education, and **Jenny Dibden,** Joint Head of the Government Social Research Service, Department for Work and Pensions, gave evidence.

Q101 Chair: I welcome the two of you here. Thank you for coming. Would you kindly introduce yourselves?
Richard Bartholomew: I am Richard Bartholomew, joint head with Jenny Dibden of the Government Social Research Service representing social science across Government and externally.
Jenny Dibden: I am Jenny Dibden, joint head of the Government Social Research Service.

Q102 Chair: To help put this in context, perhaps each of you would explain your role as half of the Chief Social Scientist for Government.
Richard Bartholomew: We completely share the role. Jenny and I have worked together very closely over the years. We work together very well. It depends on the availability of particular things. We will tend to major on one issue and then the other will pick it up, but we often share responsibility. It is a completely shared role which gives us more input in terms of time and so forth. We are as one; so address us both, please.

Q103 Chair: We are interested in the use of social science by Government. Who is ultimately responsible for setting the parameters for what is required by Government, and whether the information they receive is adequate?
Jenny Dibden: We are from the Government Social Research Service, but there are five analytical professions in Government. We work very closely with the other analytical professions. There is a Heads of Analysis group to ensure strategic co-ordination, chaired by Sir Nicholas Macpherson at the Treasury, on which sit all the heads of profession. In terms of co-ordination and deciding the strategic direction for analytical work in Government, we work through the Heads of Analysis group. Sir John Beddington is also on it. Within Departments, all the various heads of professions, departmental directors of analysis and chief scientific advisers also work together. That collective working and focus on issues within but also across Departments means we have a serious amount of input in deciding what the agenda is and securing it.

Q104 Chair: There is a fuzzy group that is the customer; there is not a single customer.
Jenny Dibden: Within a Department, individual Ministers have responsibility for particular areas and the analytical community within it will support them. Depending on the issue, we would support them too. Across Government, it depends on what the issue is. If it was particularly pertinent to science—

Q105 Chair: Are the issues driven by Ministers or by you?
Jenny Dibden: It depends. We have an agenda. For example, one of the things you have been talking about in this inquiry is census data, but there are also other sorts of data. We have an agenda on data which we have been very keen to push for some time, but it fits very well with Government priorities—

Q106 Chair: But who has established that?
Richard Bartholomew: Ministers will have specific needs and policy officials will advise on them. Our role through the directors of analysis and chief scientific advisers is to take an overview and see where we need to push and look at cross-cutting issues. We have set up a number of groups to look at different cross-cutting issues, because we need to see the whole picture of what Government and the public need in terms of information, working also with the National Statistician. There is that overview as well as the specific needs of each Department expressed by Ministers and their senior policy advisers.

Q107 Chair: That includes the public need for data and, for example, the desire of universities to have access to it?
Richard Bartholomew: We are very active on access to data, and we have an active policy of putting our data into the archives. There is a public use of the data that individual Departments collect that is also an important use, and national statistics is very much about getting data to the public. The new approach to transparency is all about getting data out to local users so that they can use it. That is also part of our role.
Jenny Dibden: It is also important to remember that all the analytical work on which Government draw for reasons of policy strategy and delivery is not done just within the analytical community in Government; it is also done externally. It is either contracted out or we draw on the vast amount of work done in academia, often funded by the research councils. The sum total of input on the analytical side is both inside and outside Government, and it is important we make sure that the external community is able to do that work.
Richard Bartholomew: If I may give you a specific example of the new major birth cohort study funded through the Economic and Social Research Council

and Medical Research Council, recently I have been asked to be part of its governing board. It is clearly a study of huge benefit to the academic community in the future and ultimately the public, but one can also see very specific interests for departments, particularly my own, the Department for Education, in terms of looking at the development of children. That is a joint enterprise. We bring these things together by working alongside the academic community and the research councils to identify the areas of future data we need to pursue commonly.

Q108 Chair: From the outside, the structure of social science in Government seems to be confusing, to say the least. I take it from your response that there is not a single Minister who has co-ordinating responsibility. Is this because your unit has effectively been the most successful and broken down all the silos in Government or would it be better run in a different way?
Richard Bartholomew: David Willetts is the Minister for Science, including social science, so in terms of science funding he has broad public responsibility for that.

Q109 Chair: You do not report to him.
Richard Bartholomew: No, but we have a very strong network that reports through Sir Nick Macpherson to the Cabinet Secretary and the relevant Minister in each case. I do not think it would be that helpful. There are specific issues for each area of policy relating to social science, so it is quite difficult to say that that is all social science. That would probably marginalise the role of social science. We need to be very integrated into each department's needs if we are to have an influence in making sure the evidence is available to make informed policy decisions.

Q110 Chair: It seems a very sensible approach. Could you apply that to other aspects of science in Government? Do you have a diffused group like you?
Jenny Dibden: You regularly see Professor Sir John Beddington in his role as Chief Scientific Adviser, but he too works through a network. He has a network of chief scientific advisers in Departments, much like we have heads of profession. They work alongside the departmental directors of analysis and all the analytical heads of profession in Departments too. The way in which John is able to do his job and deal with the issues of the day is often through the individual chief scientific advisers but also because they meet as a network, much like the departmental directors of analysis.

Q111 David Morris: In your opinion, how central and important are the census data to Government Departments in their planning?
Jenny Dibden: We have spoken with a whole range of Departments about it. They have also been feeding into the Beyond 2011 consultation that ONS has been doing, much as we have. The view from the users we have consulted is that currently there is no obvious replacement for the census, but they are confident that ONS will do the work required to assess all the alternatives and look at the potential trade-offs we might have to make in terms of quality, speed of data and so on. At the moment it remains very important to Government for a number of reasons such as our ability to forecast, assess potential policy changes and those sorts of things.

Q112 David Morris: What do you think are the major advantages and disadvantages to the Government of collecting these data via the census? Could there be better ways or methods of collecting similar data? What would your opinions be on that?
Richard Bartholomew: The main advantage and the absolutely unique thing about the census, which would be very hard to replace, is the quality and comprehensiveness of local data. That is crucial for many Government decisions, particularly by local authorities but also in other ways—for example, in education in terms of pupil projections. It is that unique aspect that is very hard to replicate through other approaches such as surveys. It is crucial to have that 10-yearly benchmark to give local data against which you can calibrate all the other large surveys. Much of the rest of research, as well as spending decisions and so forth, is dependent on having that regular common approach done on the same day, the same basis and with the same definitions. There may be alternatives to that, to which we are certainly open, but that is the unique feature that even the largest surveys cannot easily provide. There are methods like imputation of data for local areas, but that can become unreliable. If you do not have that single benchmark, although you need other things as well between the 10-yearly intervals, you cannot calibrate back to that single objective point. The difficulty is that you will get increasing bias in all the surveys and will not know its extent unless you have that benchmark.

Q113 Pamela Nash: We know that, at best, at the moment census data are two years old, and therefore two years out of date, before they are in a usable form. If more up-to-date data become available to the Government, do you feel it is flexible enough to respond to that and change policy in time to make a significant difference?
Richard Bartholomew: Absolutely. A lot of the role of members of Government Social Research, statisticians and others is to collect other sorts of data in between the 10-yearly intervals. You cannot look at these in isolation; it is a combination. Many of the policy evaluations we do look at the more immediate impact of policies, changing course if they are not working properly, and exploring new options. It is not simply that we can rely on the census for these things; we need a combination of these different surveys. But, to make sure the surveys themselves use accurate samples, they need to relate back to a common basis that is provided by the census. It takes time to produce accurate statistics. The first results of the most recent census are coming out in July, but, again, you can use other data to supplement and compensate for that 18-month or two-year gap. That lag is a very important issue for local authorities, but you have to look at what the alternative would be to that. If there is one, we are open to it, but even large surveys—which may be an alternative—do take time to do.

Q114 Pamela Nash: Do the Government have access to the raw data before it is published?
Jenny Dibden: No, they do not; it is released to everybody at the same time.

Q115 Pamela Nash: So there is no pre-fixed data. Do you think that, apart from the timeliness and the amount of time it takes to publish, any other aspect could be improved that would allow the census data to be of more benefit to you and your colleagues?
Jenny Dibden: One of the benefits of the ONS work on Beyond 2011 is that it is taking a requirements approach. It will look at the requirement for data within Government in our case, and externally, and see whether the census as well as the other methods for delivering that data can be improved in any way. The programme of work more generally has value because it allows us to explore all data sources and see whether we can make better use of them.

Q116 Pamela Nash: Is there anything you would like to add?
Richard Bartholomew: It is a very good source at the moment. It is not suitable for everything. A choice may need to be made looking at the different options between the core purpose, which it has had for 210 years, of the head count—the numbers of people—and the other questions, some of which could be and are collected through other survey methods. If the need is to reduce expenditure—I know it is a costly exercise, although you have to look at the overall value for money of what it provides and who uses it—you might look at reducing the amount of data you collect from every household and supplementing a basic collection with a sample-based collection of more detailed information, which I think is the model used in the United States.

Q117 Stephen Mosley: I want to ask about the mechanics of how a
Department uses the data. If, say, the Department of Health wants some data from the census, how does it get them in a usable format? Does it approach ONS and get the raw data? Does it say to ONS what data it needs and they are supplied, or does it just get the data ONS publishes, which are the overall figures but not specific details about individuals? What is the information, and how does a Department get it?
Jenny Dibden: The individual records from the census are not released for 100 years; so nobody has access to those at all. The ONS works on the principle that data are made available to all. My working assumption is that Government have no extra access to census data. It is very important for us that the external community has equal access, because a lot of the work on which Departments depend in terms of analysis and use of the census is done by external academics.
Richard Bartholomew: We would never see the individual identity of the people who respond. Equally, ONS has a very good approach in safeguarding individual data so that there is no indirect identification. One of the two big innovations in the last 10 to 20 years that have enabled more analysis by both Government and academics of individual level data is the release of samples of anonymised records where there are individual-level data but they are disguised so that it is impossible to work out the identities even indirectly. It is quite a complex process, which statisticians could explain to you, for doing that.
The other one, which I think has tremendous benefits, is the linking of the 1% longitudinal sample of the people who take part in each census. It is the same individuals tracked through time, again under lots of safeguards. That is available both to us and academics under very secure conditions. You cannot take the data away; you have to go to an approved centre, which is ONS or now other approved centres, I believe, where you can analyse anonymised individual records. That is a tremendous resource for looking at causal impacts over time, such as the effects of unemployment on health and various other conditions, because it is linked with vital statistics on death registrations and so forth, and certain registered health conditions. That is a really positive development. We look for further work on that because it gives you that causality, and some interesting results are now coming through. We can link four or five censuses. It is more advanced in Scotland.
Jenny Dibden: An example of that in the Department for Work and Pensions is that the longitudinal study allows us to calculate mortality rates by social class, which is very important in forecasting state pension.

Q118 Stephen Mosley: As to those particular data, the Department for Work and Pensions has its methodology as to how it interprets them and comes out with results. Do you share that methodology with other Departments—for example, the Department of Health?
Jenny Dibden: It is very important in scientific work, including social science, that methods are shared. I do not know specifically what we do in terms of estimating state pension using the longitudinal work, but certainly we would routinely get external experts to look at our methodologies both before we actually use them and then when they are published and used. It is very important that we expose our work. Just as academics expose their work, we would expect to expose our work. If you think about potential policy changes, Government now routinely publishes on the internet equality and impact assessments, and you can see the way in which analysts have worked through the potential policy changes.

Q119 Stephen Mosley: You are talking about mortality rates and so on. DWP is doing the work and the Department of Health effectively does the same work. They can share it, but there does not seem to be a formal process to allow them to share those data as yet. Does that mean they are both doing it separately?
Jenny Dibden: The community of statisticians and economists and the rest of analysts in Government attempt to join up, and colleagues work across Departments. One of the benefits of this very forensic examination of whether we should have the census in the future and other sources is that all Departments—everybody—are being exposed to the real detail of

uses of the census. Certainly, within Government there are groups looking at whether we are maximising the use made of a whole range of data. We are also part of external groups—for example, those run by the Economic and Social Research Council—looking at the social science data requirement for the country that affects both external and internal users.

Richard Bartholomew: A key part of the role of the four analytical professions within Government is to network and share methods and approaches so that there is not duplication. I do not think you would find much duplication—certainly not in methodology—and people separately developing it. If we can see it happening, it is our role as heads of profession to make sure those people start working together; so I do not think that is a major concern for us.

Q120 Stephen Metcalfe: Departments across Government have seen their budgets squeezed over the last 18 to 20 months. Has that had an effect on the individual Departments' investment in research and development in this area in the way they gather, analyse and use data?

Jenny Dibden: As Heads of Analysis, we are always concerned about the capability and capacity of the analyst community in Government. Clearly, with the spending review settlement it was very important that we began to understand the implications of that for analytical capability and capacity both within Government and externally. Heads of Analysis undertook an exercise, which Sir John Beddington has spoken to in the House of Lords Science and Technology Committee, that looked at how Departments were dealing with the spending review settlement specifically in relation to analysts and what they spent on the research budget. The conclusion that we reached on the early returns was that analysis was not being disproportionately affected, which was important. There were indications that in some Departments spending would be preserved, or potentially increased, and in other Departments there would be reductions. We did that work at the very beginning of the spending review period, and it is quite difficult for Departments to anticipate exactly how they will spend over the coming years of the spending review. We have gone out again to see how that is panning out, what the current estimates of spend are and whether there are any issues.

On the back of that, the Heads of Analysis group also wrote to permanent secretaries saying, "We have done this exercise. We will continue to monitor, but it is very important that within your own Departments you have an understanding of your analytical capability." Another aspect is that, when Departments were proposing to eliminate or reduce the scope of particular data sources, we asked them to make sure there were not dependencies in other Departments on those data sources.

Richard Bartholomew: My experience is that it is some of the evaluations of the initiatives of the previous Government where the need has changed. My Department has brought to an end some evaluations, but we have also started other new research—for example, in the area of adoption breakdown, which is a priority for the Government. Therefore, there are opportunities. There has been some overall reduction, but it is not disproportionate compared with other reductions in expenditure. There are certainly definite areas of growth in demand for research.

Q121 Stephen Metcalfe: Presumably, there has been some reduction. Therefore, the Departments have to ensure that the research they undertake is of most value.

Richard Bartholomew: Yes.

Q122 Stephen Metcalfe: Is consideration given to the longevity of that research—for how long it will be useful?

Richard Bartholomew: Yes.

Q123 Stephen Metcalfe: What is the life expectancy of a piece of work, and what should someone be aiming for?

Richard Bartholomew: It is difficult not to give a very woolly answer because it depends on what it is and how precise it is. If you are monitoring a policy, you probably need much more regular sweeps of data to see its unfurling impact. If you are looking at a broader change in social attitudes, such as the British Social Attitude Survey, with an underlying change, clearly there is less justification, because from year to year things will change less rapidly. That is a debate we always have with our colleagues and ministers about the frequency we need. It is difficult not to give a woolly answer, but I suppose the spending review horizons are around three years; so there tends to be a regular survey at least every three years, and sometimes there is a good case for doing them annually. It is a question of affordability and the likely change you will get in that period. Some you would need to do much more regularly.

Q124 David Morris: Health and education are the most obvious parts of Government undergoing significant structural change. Do you feel these changes mean that the Government need for social data gathering is changing?

Richard Bartholomew: As a result of the structural changes?

David Morris: Yes.

Richard Bartholomew: My view, working in education, is perhaps different from DWP where the Department for Education is responsible but delivers through a number of different bodies such as schools and local authorities. It is not fundamentally a different picture in terms of the structural change. There is more emphasis now on making sure that the data we provide and collect is the right data for the local commissioners and providers of services, not just for Government. I think Government, certainly in the statistical and research areas, has a key role in making sure there is a common set of data available using common definitions—this is related to making the data more transparent and usable to commissioners of services and to the public—to ensure they are consistent data. In terms of statistics on social care for children and in our educational statistics it is very important that data are there for the people to judge

schools against a common benchmark of standards. All the evidence suggests that is the way to get improvement and consistency. There is autonomy in the way schools are run, but there is a common process of accountability using common standards, both nationally and, as far as you can, internationally, hence the emphasis on the new PISA study that has been running for a number of years. Those international comparisons have become increasingly important in education.

Q125 David Morris: Given the level of funding dependent on formulae, where do you think that the greatest effort is needed in firming up social data collection?

Richard Bartholomew: In a sense it comes back to the reliability of those local estimates of numbers, the point about timeliness and, if we continue with the 10-yearly census, finding ways of updating that. If we replace it with a different system, we should still get the regularity, because I know that for local authorities it is a key issue, particularly in large urban areas where there is a lot of change and very mobile populations. We need to be fleet of foot to address that and make sure we use other sorts of data where we know there is rapid turnover. There are possibilities in relation to the school census, which is quite sensitive in terms of the intake of five-year-olds, for example, to local change. By using that in combination with other data—and other departments will have data that are equally sensitive to very quick changes—that is where one would seek more development.

Q126 David Morris: It has been said that administrative data would be a more valuable resource if the data collected were cleaner and more easily collated. What would be the best first step to standardise administrative data collection?

Jenny Dibden: The Beyond 2011 programme will look at the question of standardisation. One of the difficulties here—or it is certainly something you need to consider in some depth—is why we have administrative data. Essentially, it is to administer things like benefits. If we move to a process of standardisation, we have to ensure that the purpose of the system remains. We still need to be able to pay benefits; that is the primary driver of that system. There are parameters within which we need to consider standardisation. I am sure there is scope for some standardisation, and we will need to make sure that we get the incentives right to do that and to ensure that there is a good early warning system so that the administrative system does not get changed and the first we know about it is when the statistics are impacted. There is a whole basket of issues being worked through, but I do not think we can get away from the fact that these are administrative systems for particular purposes and we need to see how we can make more use of them for statistical purposes. But it is quite difficult to see how you can turn them into statistical systems when their primary purpose is administrative.

Richard Bartholomew: To add to that, a specific thing that I think is important is the identity of individuals. There has been some discussion in other evidence about address registers. Clearly, there is a big problem about identifying that it is the same individual, if you are going to use this for census purposes, to avoid double counting. We know that sometimes people are not in administrative databases at all or their information is very out of date, and some may appear twice. We spend a lot of time trying to iron out those problems. That matching process to make sure you count each individual only once is a key challenge. It needs to be looked at very hard before we rely entirely on those administrative sources for population counts.

Q127 Chair: As an aside, Mr Bartholomew, clearly this Committee is interested in why young people end up studying science and engineering. Has the Department ever commissioned any longitudinal work in this area to try to track the exposure that young people get to good science education and the outcomes?

Richard Bartholomew: We have a longitudinal study of young people that has been running for some years. It has followed one cohort, and they were 13 to 14 years old when we first interviewed them and their parents. The purpose of it has been to track their decision making in some detail, because it is interviews about their intentions and ambitions to look at their subject choice before they finalise their GCSE decisions. We have then tracked them through to 19. Those data are now available. At least in the early stages of their career decisions we can see their educational choices, what influences they had and how that varies by social class and other background factors like parental influence.

Q128 Chair: But, interestingly, the Association of Science Education would argue that that is starting far too late; you need to start at school or pre-school.

Richard Bartholomew: That is right. It is probably a little more difficult with primary age children, but some research has been done in that area. We have certainly tracked them from 14. I agree that may be an area we need to look at. Maths, in particular, is the key to a lot of these things in terms of interest in science. It would be interesting to try to pick that up in younger age groups.

Q129 Chair: The reason I ask is that it is not just obvious to this Committee, but to a very wide part of the population that that would be a very interesting study to do. However, is it the case that it is not done because there is not a sufficiently strong drive from a Minister to get that work done because the Minister will not be around when the outcomes are published?

Richard Bartholomew: No, I do not think so; I think that underestimates Ministers' interest in future developments. That is why they are interested in longitudinal data.

Chair: I have known a lot of Ministers in my 20 years here.

Richard Bartholomew: Even if that was the case, I do not think it is true. It is our role as the people responsible looking to the longer term. It is a key role of Government Social Research to look to the longer term and what Ministers and their successors will need several years down the road, and to lay the basis

for that. You have to argue quite hard for that investment, but the announcement by David Willetts on the decision on the birth cohort study shows there is a real understanding of the value of that longitudinal data.

Q130 Pamela Nash: If a decision is made to discontinue the census and replace it, what do you envisage the Government will have to do to allow more data sharing between different Departments?
Jenny Dibden: The ONS work will identify the things that need to be done to facilitate data sharing. One of the important aspects of that work is the testing process through which ONS will go to identify how best to make use of alternative data. One of the things to which we need to pay close attention is the overlap between any new methods and the census.
As to issues that it needs to look at, there are legal issues to do with access to data, linking data and also matching data. There are also technical issues about definitions. Do you know whether an individual in one source is the same individual in another source? There is also a public aspect to it. All of the proposed alternatives that ONS will need to look at must be assessed against the criteria of whether the public will find them acceptable. One of the reasons we have a census in this country is that we do not have a population register. There is a perception that the public in this country would not be keen on a population register, although they exist in other countries. There is definitely a public aspect to being able to match data.

Q131 Pamela Nash: Is that perception because of the extended level of data sharing of personal information?
Jenny Dibden: Population registers in other countries have been running for many years. I guess there is a question about how long in this country we have not been comfortable with the concept of a population register. If we set one up, it would take many years to get to the point where it was robust. The public seem to have some difficulty with the concept of data matching and sharing, much as, on the face of it, they had with the census.

Q132 Chair: Who says that?
Jenny Dibden: Quite a lot of work has been done on trust in science, data and statisticians. To take the example of statisticians, if you look at surveys and work done on whether people trust statisticians, yes, they do. Do they trust statistics? No. There is a whole range of reasons for that.

Q133 Chair: If you asked a simple question of the public that was posed in the interests of children, "In child support cases, should the CSA have the right to access the DWP and HMRC records of both parents and so on?", you would get an overwhelming yes for that.
Jenny Dibden: Yes, but I think that is about making the case. It is like the census. The census had to work really hard this time to make the case for why people should fill in their forms. If you are a member of the public, you may well see the link between your census form and the allocation of resources for health or education, but you may not. ONS did a very good job of explaining to all of us just how important it was to fill in the census form and that that data were subsequently used. That is the case for the whole range of data sources. It depends on the issue. You have cited a specific issue, but generally on a source like the census there is a whole range of issues.
Richard Bartholomew: I think there is a case for that, and we would welcome a public debate about it. As analysts we can see the benefits of linking up those data, but clearly there are public concerns. Whether they are shared by everybody, or the majority, we do not know, and it depends on what you are actually talking about. As an example, the ContactPoint database for children that my Department established under the previous Government, which has now been discontinued, was one in which each child would have a record but with very minimal data. There were huge safeguards created around that to reassure people about the uses of those data. It was not going to be possible to use them for many research purposes because of those safeguards, but clearly there was a very strong view in some quarters that that was an infringement of civil liberties and people were concerned. Yet it was set up for the very laudable reason of trying to track children at risk to see whether they had had contact with social services and other agencies. Only about 5% of children or something of that order would have that kind of record; 95% would not. But there were huge concerns about that process and that contributed to the decision to discontinue it, as well as the overall cost of the exercise. There is public concern and you need a debate about it, because there are trade-offs here in terms of the cost of the census versus better use of administrative data. We can see the analytical value, but the public need to be convinced and reassured that they will be used properly.
Jenny Dibden: One implication if they are not reassured is that they will stop responding to surveys. By definition, surveys are voluntary; we do not make people fill them in, so we need them to buy in.

Q134 Pamela Nash: Over recent years how have the Government changed their way of using the data collected from the census and from other sources that are used between every 10 years, and has the collection changed as well?
Richard Bartholomew: From my perspective, we are making much more use of administrative databases. Certainly, in education there has been a huge transformation. We have had the schools census for about a decade. It took a number of years to develop and it has evolved over time. That provides data on about 93% of children. It does not include the independent sector so you have a gap there, but it includes basic information about children and their educational performance. It does not include a lot of information about their social background. It is a census and it gives us tremendous analytical power. We do not need to do some of the more ad hoc surveys we might have had to do 10 years ago to look at more basic characteristics. We can use those administrative databases in combination with surveys to look at

attitudes—obviously administrative databases do not include information on attitudes—to pinpoint, perhaps, primary children and their interest in science. That is a huge transformation. It has not happened overnight. There has been a lot of very hard work done by statisticians in setting it up, but it is a tremendous and internationally recognised data source. That is also true in other Departments. One of the major changes is greater use of statistics and greater power of statistics, but there are still opportunities clearly in terms of data linkage.

Jenny Dibden: Within my own Department, two decades ago we would have been able to look at the data in a single system. We now link for analytical purposes data from a whole range of systems. We also link data with HMRC, which now allows us to look at, for example, not just whether somebody has moved off benefit but whether they have gone into work.

Q135 Pamela Nash: I am very interested in how the information collected is shared with the Scottish Government and the Welsh and Northern Ireland Assemblies. How has that progressed since the devolution of power? How much power do they have in collecting the data in the first place? Are they still centrally collected? Personally, I had to complete two forms and I noticed they were different.

Jenny Dibden: The census we have been talking about today mainly is that which covers England and Wales. There are separate censuses in Scotland and Northern Ireland. Clearly, it is in everyone's interest for there to be co-ordination of what is in the census returns so that we are able to do comparative analysis across countries. As part of Beyond 2011, ONS will be looking not only at the census in England and Wales but also working with colleagues in Scotland who are looking at the future of the census in Scotland. There is a lot of join-up between the different censuses at a professional, statistical level.

Q136 Pamela Nash: The data collection is co-ordinated, but are the data collected analysed together?

Richard Bartholomew: You need to discuss that with the statisticians at the ONS, but they are combined for UK statistics. Some of the devolved Administrations may want slightly different questions. There are Welsh language questions and other issues like that. That also gives freedom. There are benefits in terms of slight differences, but from the UK perspective, those statistics are combined. Certainly, the professional statistics network, the Government Social Research network and the economists are cross-UK, and that is a benefit. We link up quite regularly with our colleagues in the other three countries to make sure we are joined up.

Q137 Chair: There is a seamless link between you and your colleagues?

Richard Bartholomew: There are different needs and they are devolved Administrations, so there are different emphases, but we make sure there is common ground on data collection.

Q138 Chair: You work on a collegiate basis.

Richard Bartholomew: Yes, absolutely.

Jenny Dibden: We have a head of profession for social research in Scotland, Wales and Northern Ireland, and we go through a self-assessment process for each head of profession within his or her own Department. I do one for the Department for Work and Pensions. That is peer-reviewed, as it happens, by Scotland, so that we can look at issues in my Department and the strength of analytical capability in social research in Scotland and learn from each other.

Richard Bartholomew: My own staff at a more junior level have annual meetings with their opposite numbers in the other Administrations to look at the data collected on children—children's care and so forth.

Q139 Pamela Nash: I suppose that is what I am trying to get to. There are different levels of power devolved to the different Administrations. Health is completely devolved to Scotland, whereas you cover the entire United Kingdom.

Jenny Dibden: We cover it; that is right.

Richard Bartholomew: Education is just England, so there are different challenges there.

Jenny Dibden: Going back to administrative data, DWP, for example, has information on the whole population.

Q140 Graham Stringer: You mentioned a couple of examples of longitudinal studies in the evidence you have given. The written evidence we have received says how valuable those studies are. Can you explain to the Committee whether the information in those studies and data sets has affected policy?

Jenny Dibden: The example I gave was to do with state pensions. The longitudinal survey allows us to look at mortality rates by social class, and that is used in forecasting state pensions. Clearly, that is available for policymaking and informs it. There have been changes in policy on state pensions which will have been informed by that. We have a whole modelling system for looking at pensions, both state and private.

Richard Bartholomew: I do not know whether you are talking of administrative data or longitudinal studies like the birth cohorts, but one of the big messages from those in recent years, certainly from the 1958 and 1970 cohorts, is the social mobility question. That is very much on the political agenda of all parties, is it not, in terms of the apparent reduction in social mobility between the 1958 and 1970 cohorts? That is a really strong influence.

Q141 Graham Stringer: But that has exposed an issue without leading to policy change so far. Are there any areas where you can say, "This is the information we have from these longitudinal studies, and this is the policy change we have enacted"?

Richard Bartholomew: I think that is one, in terms of Governments addressing it. For example, the pupil premium in my Department is certainly influenced by the concern about social mobility and improving the chances of children from disadvantaged backgrounds through a specific set of interventions. Clearly, you need a number of interventions with different age groups for the purposes of social mobility, but it is

crucial in terms of the arguments as to why one should invest both in early years and the pupil premium. Another one, which has justified the increased investment in early years, is the EPPE study—the Effective Provision of Pre-school Education study—which was set up about 12 or 13 years ago to look at three-year-olds and their experience of early years and the different qualities of early years. That had a seminal influence subsequently in all spending reviews in making the case for improving quality in early years' provision. Through that study we can relate the quality of the provision to the outcomes of the children at five and subsequent stages. We are now trying to put forward proposals to repeat that study. Obviously, 12 or 13 years is a long time and we need new evidence on that.

Q142 Graham Stringer: When there are policy changes are you consulted in terms of what impact they will have on your data sets? Generally, is consideration given to the coherence of your data sets when there are changes in policy?

Jenny Dibden: By "policy", you mean changes in the way it impacts on a computer system that has a bi-product of producing statistics. It is that sort of change.

Graham Stringer: Yes.

Jenny Dibden: Yes.

Graham Stringer: I had not thought of precisely that.

Jenny Dibden: All major change projects, even small ones, in Government are managed and go through a process of impacting. Within a specific Department we would be among a range of stakeholders that would have sight of that change and would be able to input as part of the impacting process the implications of that change. If a change was to be made to one of DWP's administrative systems as a result of a policy change, we would have sight of that and be able to say that that will impact on our ability to produce statistics or forecast. That does not mean to say the change will not happen, but it allows us to know what that change is so that we can make provision for explaining any discontinuity in the data series.

Q143 Graham Stringer: That is interesting. Can I go back to David Morris's series of questions where he was asking about the input of data sets into public expenditure? Is there any difference in the accuracy of your data sets on a regional basis?

Jenny Dibden: One of the strengths of the census is the ability to look at very small geographical units for the variables that the census collects. One of the reasons why at the moment the census is so valuable to Government is that it allows you to go down to those very small geographical areas.

Q144 Graham Stringer: I understand that. I am pursuing a slightly different point. It is not really about the focus but the accuracy of it. It is fairly well reported that fewer forms are returned from poorer inner city areas—sometimes from ethnic minority communities—and therefore there is less accuracy in those areas. Do you take account of that? How do you deal with that? When I talk to my health bodies, they say that our population is massively underestimated, and they always have done, not going back just to the last census but two or three censuses.

Jenny Dibden: That is right. ONS does some very important work alongside the census to look at the quality of it and also its coverage so that it can understand where there has been under-coverage. You are absolutely right. Looking back at the response rates for the 2001 census, although the overall response rate was 94%, for the main metropolitan areas it was 92%, and it went down to 78% in Inner London. That was a huge cause for concern, and certainly in some cases local authorities challenged the basis on which they were allocated resources. ONS has put in a huge amount of work this time to try to boost response rates, particularly in London, but also in other metropolitan areas, to push them back up to a certain level, but if you are using those data you need to take account of that.

Q145 Graham Stringer: Just forget the 2001 census, which had particular problems. If you were to go back to the 1991 census where there would still be lower rates, when you are doing the statistics and advising a spending Department on population, do you say, "These are the actual statistics we have got, but you really need to add on 3% or 4% to get to the real number", or do you just leave the statistics as they are?

Richard Bartholomew: That is what the ONS do in terms of adjustment to make overall population estimates from the census data they have collected to allow for that under-counting and poor response rates you will get from some particular areas or groups, like young men, who we know do not respond well. Therefore, an adjustment to the figures is made before they are used to make sure we are doing everything to compensate for that potential under-representation.

Chair: Thank you very much for your attendance this morning; it has been very helpful.

Written evidence

Written evidence submitted by HM Treasury (Census 00)

THE ANALYTICAL COMMUNITY

There is an analytical community within government to support those involved in strategy, policy and delivery to develop and use an evidence base.

Government analysts contribute to all stages of policy development and implementation though helping to:

- identify issues;
- understand and solve problems;
- present options;
- assess impacts (costs, benefits and risks);
- interpret existing data;
- specify and gather new evidence;
- test and assess ideas;
- translate research into practical options;
- inform decisions on difficult trade-offs; and
- inform decisions on whether to stop, continue or change a policy.

As well as using their own knowledge and skills, government analysts draw on appropriate external expertise, from the academic and broader research community both in the UK and overseas. They can advise on the quality of available external evidence and the confidence that should be given to it in decision-making.

Analysts are organised differently in different departments, but generally there are five different groups. Members of each discipline have a unique contribution to make as well as shared skills across the disciplines. At a very high level, the particular contribution can be characterised by:

- Economists (members of the Government Economic Service). Maximising welfare from scarce resources. Microeconomics by providing decision metrics for choosing one option or course of action over another; macroeconomics by fostering prosperity, high employment and stability economy-wide.
- Operational Researchers (members of the Government Operational Research Service). Helps people find solutions to complex problems through problem structuring and mathematical and statistical modelling to understand real-world systems, policy options and impact.
- Social Researchers (members of the Government Social Research Service). Understanding the potential and actual social impacts of policy decisions/practice, including understanding public perceptions and the opportunities for behaviour change. Advising across government on research methodologies and ethics.
- Statisticians (members of the Government Statistical Service). Ongoing measurement and monitoring of specific and general economic and social trends.
- Scientists and engineers (members of the Government Science and Engineering Community). Applying science and engineering knowledge to understand problems and develop policy solutions, including the application of knowledge from individual specialisms.

Each of the five analytical disciplines works together to provide the best possible evidence. Each have different, skills and backgrounds—further details provided later—and has their own leadership and management structures but all should be able to look at the problem and indicate which team of analysts would best be able to answer a problem, or signpost to the appropriate analyst for each particular question.

Each discipline ensures highly skilled and professional staff through rigorous recruitment procedures; adherence to competency frameworks specific to that discipline; continuous learning and development programmes. Many government analysts are also members of relevant professional bodies and learned societies. As a result, government analysts represent a professional and skilled resource that can add expert analysis and advice to inform choices across government and frontline delivery services.

Over the last five years or so the five analytical professions in government have been working increasingly closely together. The most robust and thorough evidence base for a problem will most often come from considering it from a number of perspectives/evidence streams. This can be achieved by ensuring that a multi-disciplinary group of analysts consider the problem. Many departments now ensure analysts work within multi-disciplinary teams to facilitate this approach.

Developing or implementing strategy and policy is, of course, possible without the input of the analytical community, but there are risks associated with this course of action. These risks include implementing a policy that does not work; that is more costly than expected or need be; that has unexpected consequences or poor public acceptance. Working collaboratively with government analysts can help mitigate these risks.

Signed:

Government Economic and Social Research team (GESR)

Government Operational Research Service (GORS)

Government Statistical Service (GSS)

Government Office for Science (GO-Science)/Government Science and Engineering (GSE)

HM Treasury

January 2012

Written evidence submitted by ESRC International Centre for Life Course Studies in Society and Health (Census 06)

I am responding on behalf of ESRC *International Centre for Life Course Studies in Society and Health*. Our answers to your questions are:

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

In our work, Census data are used in many ways.

(a) The Census serve as the definitive source for describing the health and social circumstances of the contemporary population, particularly for relatively small population groups. For example, the Census is one of the only data sources that allow us to examine the characteristics of specific ethnic minority groups, occupational groups, or residents of small geographic areas such as electoral wards. Census data, including the Samples of Anonymised Records, can serve a source of data for detailed research in their own right or can be used to substantiate or inform findings arising from sample surveys.

(b) Census allow us to examine and understand social change. With 100% data we can look at change over time at a population level, for example in access to basic amenities, or household structure. Census data linked to other administrative records in the ONS Longitudinal Study open up unrivalled opportunities for looking at change across individuals' lives, inter-generational change in social circumstances, and for understanding socio-demographic differentials in mortality and fertility. For example, the ONS Longitudinal Study is the key source in the UK for understanding differences in mortality by occupational class and provides a rich source for understanding the fertility patterns by ethnic group.

(c) Census data provide a reference point for ensuring that sample surveys are representative of their intended population, through the generation of population weights or grossing factors. This is a vital function of the Census for the social science community, as it allows us to understand the extent to which research findings from surveys are generalisable to the population at large.

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

Given full population coverage of Census, the long run of available Census data for monitoring social and demographic trends, and the relative detail of the UK census, the end of the UK Census is likely to have detrimental impact on the quality of social science research in the UK. The Census is part of the UK social science infrastructure; it is therefore difficult to assess the full ramifications of ending it. We assume that the Census would be replaced without some other method of enumerating the population. The absence of a firm proposal for this alternative, makes it difficult to understand the relative impact of ending the Census.

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

In our opinion, it is unlikely that alternatives to the Census would yield data of equivalent quality for social science research. Countries that monitor population trends using register data, have operated in a societal context where this accepted and supported by the general public.

To date, there appears to be little public appetite for a central population register in the UK. Any attempt to impose such a register without consensus runs the risk of yielding poor quality information and relatively high rates of data missingness.

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

Administrative data collected by Government departments could be improved with more attention to the nature of data collection and quality of completion. For example, Her Majesty Revenue and Customs could improve its recording of the occupation and industry of individuals, although this might place burdens on business. However, we are not aware of single data source, or universal point of interaction with state that would provide the opportunity for collecting high quality data about a range population variables. For example, it is not clear where opportunities would present themselves to equally detailed data on issues such as household structure.

David Blane
Professor of Medical Sociology,
Imperial College London

Deputy Director
ESRC International Centre for Life Course Studies in Society and Health

17 November 2011

Written evidence submitted by Professor Heather Joshi, Institute of Education (Census 10)

1. *Introduction*

My current affiliation is as a Professor Emerita of the Institute of Education, and the President of the Society for Lifecourse and Longitudinal Studies. The latter is an interdisciplinary international society of scholars interested in research on the lifecourse and the development of longitudinal data. Our members come from epidemiology, psychology, geography, sociology, demography, education and economics. I have formerly been president of both the British Society for Population Studies and of the European Society for Population Economics. My illustrations are taken only from my own work for economy of time, but I am sure I have the support of many colleagues in saying that social scientists working in and on the UK find the census a very valuable resource underpinning research as well as vital for informing central and local government.

2. *Statement on Conflict of Interest*

I have retired from my post at the Institute of Education, where from 2000 until earlier in 2011 I was director of the Millennium Cohort Study (MCS), making use of some census data as indicated in para 4 below. As also indicated in para 4, I was the head of the team providing academic access to the ONS Longitudinal Study, and I was involved in making the case for it to be linked to the 2001 census. In my retirement I am working part-time on research (education in social mobility, child residential mobility), which will make some use of existing census data. I, personally, would not expect to be a very extensive user of any data which might become available from any future census, but I do need to say that it would be very useful to colleagues in the social science community, as well as those responsible for making and administering policy and the public at large. I would hope you find my interest constructive rather than conflicted.

3. *How do social scientists use Census data?*

Turning to your questions, here are some examples, far from exhaustive, of how a social scientist uses the census, which are elaborated in paragraph 4:

- To compare change over time in the demographic and social structure;
- To make international comparisons;
- To compare locations at various levels of geography, from regions down to very small neighbourhoods;
- To track internal and external migration;
- To compare workplaces and places of residence, and the generated "travel -to-work areas";
- To examine detailed breakdowns of characteristics of individuals;
- To compare the combinations of individual characteristics found among members of the same family or household;
- To chart the living arrangements of individuals at different ages;
- To chart the demographic composition of detailed occupations;
- To classify areas as rural or urban;
- To inform the drawing up of sampling frames for sample surveys;
- To inform the assessment of bias in social survey data;
- To assist estimates and projections nationally and locally;
- To perform micro-level analyses on Samples of Anonymised Records;
- To perform—micro-level analyses using the ONS Longitudinal Study.

4. Here, I briefly amplify these points and give examples of such use from my own (and co-authors') work. References are appended at the end of this document. I do not pretend to be offering a comprehensive account of the uses made by the social science community of the census. I am offering this eclectic exercise in self citation as the most cost-effective way I have, in the time available, of illustrating a wide range of uses on the basis of a sample of one user. I am not suggesting that the Select Committee needs to consult the publications cited, just to noting a "tip-of-the iceberg" selection of social science research which has used the census.

4.1 *Change over time in the demographic and social structure:*

Censuses offer the building blocks for assessing long-term change over many decades in demographic, social and economic indicators. An example from my own work is the construction of secular time series of female economic participation rates.

4.2 *International comparisons:*

Just as census offers the possibility to gauge continuity and change over time, they also enable comparisons of conditions cross-nationally. An example from my work includes comparing Britain and France on childbearing by women's occupation. Ongoing work on residential instability as experienced by young children draws on comparison of US and British censuses.

4.3 *Comparing locations at various levels of geography:*

The finer the level of geographical resolution—down to electoral wards, or smaller approximations of neighbourhood such as Output Areas, the more is social science dependent of census indicators on the social profile of inhabitants. Sample surveys cannot cover all areas in any detail, administrative data (even where reasonably accurate) do not record all relevant indicators, eg ethnicity, housing quality, educational attainment of adults, household composition etc. I have made various uses of census-based small area statistics, such as analyses of child outcomes in the Millennium Cohort and National Child Development Study second generation by census (and other) indicatorss of deprivation at ward and Lower Super Output Area level.

4.4 *Tracking internal and external migration:*

Though I have not used them myself the special census migration tables seem to me to be invaluable for identifying the characteristics of people moving between localities within the UK. Duration of residence data can provide one half of the demographic profile of international migration.

I have used less detailed census migration tables, by age, to inform patterns of non-response to Millennium cohort study.

4.5 *Comparing workplaces and places of residence, and the generated "travel to work areas":*

Have supported other researchers using commuting data in the ONS Longitudinal StudyS (see below).

4.6 *Detailed breakdowns of characteristics of individuals:*

Examples include: economic activity or carer status, longstanding illness, educational qualifications, country of birth, ethnic group, type of household or housing into gender, details of age by single years and/or detailed geographies. Most valuable is the cross classifications among such characteristics.

For instance a comparison of the educational, occupational and health status of ethnic groups controlling for country of birth and duration of residence, and in different regions of the UK. I have not conducted such work myself. Although sample surveys may be adequate to provide details of these cross-classification at a national level, the results become statistically unreliable at more local levels, where they may be needed for planning services.

4.7 *Combinations of individual characteristics among co-residents: eg "workless households", homogamous or mixed couples with respect to age, ethnic group, occupation or educational attainment.*

Besides looking at the distribution of paid work across dual, single and zero earner couples, this feature of the census will be used in ongoing work on gender and social mobility, where it will be necessary to compare the education and occupations of spouses.

4.8 *Charting living arrangements of individuals at different ages—eg one-person household, one parent family, extended family, unrelated adults living together, and the number of dependent children. These can also control for marital status, housing conditions etc.*

Census evidence on the living arrangements and location of ethnic groups will be of academic as well as policy interest. I have not been involved in work on ethnic minorities or international migrants, but we have used longitudinal census data to look at different groups' experience of lone parenthood over time.

4.9 *Segregation of occupations, industry etc by gender, ethnicity and age to record the extent and evolution of occupational segregation.*

I have used such results to identify, gender typed occupations in coding occupational data collected in sample surveys in the analysis of gender pay gaps.

4.10 *Classification of areas as rural or urban:*

The Rural Evidence Research Centre at Birkbeck College has devised a classification of areas for DEFRA based upon the settlement pattern of the whole population (which could probably not be implemented from survey data). I have applied this Census-based indicator of rurality to the analysis of the Millennium Cohort and the ONS Longitudinal Study.

4.11 *Drawing up asampling frames for sample surveys:*

Although there may be alterative lists of addresses from which sample surveys could be drawn (eg postal or electoral) they do not contain much information on which to stratify samples. The over sampling of areas of high ethnic minority settlement for the Millennium Cohort was based upon ward-level census data for 1991, with adjustments for boundary changes provided by Professor Ludi Simpson.

4.12 *Assessing bias in social survey data:*

One example is the study of attrition through mobility in the MCS.

4.13 *Assisting estimates and projections nationally and locally:*

Census data forms the benchmark for estimates and projections of population, households and pension population by marital status.

4.14 *Micro-level analyses of Samples of Anonymised Records:*

No personal use of SARS.

4.15 *Micro-level analyses of ONS Longitudinal Study:*

I was head of LS User Support for seven years, during which time many academics made use of the 1% sample of longitudinally linked individual census records, with their household at each census also linked to vital registration records. I presume there will be fuller evidence about this remarkable resource and rich resource from the CeLsius team currently responsible for access and from the user community. Its classic value, which should remain a raison d'être, is to provide valid denominators for estimating social inequalities in mortality. My own research on it includes a study of the timing of childbearing by the woman's occupational attainment comparing Britain and France, social profile of movers between urban and rural England, analyses of the geographical and gender dimensions of health inequalities, and the dynamics of children's living arrangements.

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

Just at the time when the technology and governance arrangements permit far better academic use of census data through micro data sets, the existence of this increasingly valuable source of evidence is going to be withdrawn. Given the range of uses to which is it currently put, this will be a blow to empirical academic research and the policy uses which flow from it, removing the bedrock on which much else is built.

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

The routine recording of residential moves, including those out of the country, seems to me to be a top priority. This might be achieved through a better use of the NHS central register and better recording of movements across borders. Clearly the current system is inadequate.

The social and economic details collected at birth, and marriage registration could be improved. For births I would like to see true birth order, educational attainment and duration of residence at current address included in on the birth certificate. But without supplementary data on the size of the population at risk to give birth at each age and number of previous births, data on birth registration will be hard to interpret.

The most important priority for mortality analysis is to have adequate denominators of the population at risk.

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

The sample size of national surveys should be increased to allow for adequate coverage of small categories and cross classifications.

If the census goes, there should be better ways of sample surveys accounting for visitors and transient residents in households, and allowing for people (like children whose care is shared between separated parents, or weekly commuters) being covered.

The linkage of vital events from birth and death registration to the existing Longitudinal Studies' databases could usefully be continued. Consideration might be given to collecting other information from their sample members, if a way could be found to safeguard confidentiality. This would need to be very carefully thought through but there may be an instructive example in France, which has just recruited a large fraction of its national birth cohort (ELFE) from the Echantillon Demographique Permanente (EDP)—the French equivalent of the LS.

References by Paragraph

4.1
Joshi, H E, Layard, P R G and Owen, S J (1985). Why are more women working in Britain? *Journal of Labor Economics*, Vol 3 January 1985, S147–S176.
Joshi, H E and Owen, S J (1987). How Long is a Piece of Elastic? The measurement of female activity rates in British censuses. *Cambridge Journal of Economics*, Vol 11, 1, 55–74.
Joshi, H E (1985). "Motherhood and Employment: Change and Continuity in Post-war Britain" in *Measuring Socio- Demographic Change*, Occasional Paper No 34, London: OPCS, 70–87.

4.2
Rendall, M S, Ekert-Jaffé, O, Joshi, H, Lynch, K and Mougin, R (2009). Universal versus economically polarized change in age at first birth: A French-British comparison, *Population and Development Review,* Vol 35 Number 1, 89–115 and work currently in progress

4.3
McCulloch, A and Joshi, H (2001). Neighbourhood and family influences on the cognitive ability of children in the British National Child Development Study. *Social Science and Medicine*, 53, 5, 579–591.

4.4
Plewis, I, Ketende, S C, Joshi, H and Hughes, G (2008). The contribution of residential mobility to sample loss in a birth cohort study: evidence from the first two waves of the UK Millennium Cohort Study, *Journal of Official Statistics*, Vol.24, No.3, 2008. pp. 365–385.

4.5 and 4.6 nothing to report here at present

4.7
Joshi, H (1998). The opportunity costs of childbearing: more than mothers' business. *Journal of Population Economics*, 11, 161–183.
Davies, H, Joshi, H, Peronaci, R (1998). Dual and Zero-earner Couples in Britain: Longitudinal Evidence on Polarization and Persistence. Discussion Paper in Economics 8/98, Birkbeck College, London.

4.8
Clarke, L and Joshi, H (2001). Accessions and Abdications: Men and women as family markers in linked data for England and Wales. In B. Zaba & J. Blacker (eds.) Brass Tacks, Essays in Medical Demography. Athlone Press, London, 207–230.
Rendall, M, Joshi H, Oh, J and Verropoulou, G (2001). Comparing the Childrearing Lifetimes of Britain's Divorce-Revolution Men and Women. *European Journal of Population*, 17, 365–388.

4.9
Sullivan, A, Joshi, H, Leonard, D (2010 in press). Single-sex schooling and labour market outcomes' *Oxford Review of Education*, Vol. 37, Number 3.
Paci, P, Joshi, H and Makepeace, G (1995). Pay gaps facing men and women born in 1958: differences within the labour market in J. Humphries and J. Rubery (eds) *The Economics of Equal Opportunity*, Manchester: Equal Opportunities Commission, 87–112.

4.10
Joshi, H, Dodgeon, B and Hughes, G (2008). A profile of population change in Rural England. CLS Working Paper 2008/4, London: Centre for Longitudinal Studies. http://www.cls.ioe.ac.uk/publications.asp?section=0001000100060009

4.11
Plewis (ed) *et al* (2007) Millennium Cohort Study: Technical Report on Sampling. CLS. IOE.

4.12 see 4.4

4.13
Diamond, I, Testaghiorghis, H and Joshi, H E (1990). The Uses and Users of Population Projections in Australia. *Journal of the Australian Population Association*, Vol 7, 2, 151–170.
Joshi, H E and Davies, H B (1992). Pensions, Divorce, and Wives' Double Burden, *International Journal of Law and the Family*, 6, 289–320.

4.14 No personal use of SARS

4.15
Ekert, O, Joshi, H, Mougin, R and Rendall, M (2002). Fécondité, calendrier des naissances et milieu social en France et en Grande Bretagne: des évolutions contrastées. *Population* 2002 (3) 485–518 English edition: Fertility, Timing of Births and Socio-economic Status in France and Britain: Social Policies and Occupational Polarization 2002(3) 475–507.
Wiggins, R D, Joshi, H, Bartley, M, Gleave, S and Lynch, K, Cullis A (2002). Place and personal circumstances in a multilevel account of women's long-term illness. *Social Science and Medicine*, 54, 5, 827–838.

Wiggins, R D, Bartley, M, Gleave, H, Joshi, H, Lynch, K (1998). Limiting long-term illness: a question of where you live or who you are? A multilevel analysis of the 1971–1991 ONS longitudinal study. *Risk Decision and Policy*, 3,3, 181–198.
Gleave, S, Wiggins, R D, Joshi, H, Lynch, K (2004). Identifying Area Effects on Health: A Comparison of Single- and Multi-level Models in P Boyle, S Curtis, E Graham, E Moore, (eds.) *The Geography of Health Inequalities in the Developed World*, pp. 245–266, Ashgate ISBN 0 7546 1398 4.
Joshi, H, Wiggins, R D, Bartley, M, Mitchell, R, Gleave, S, Lynch, K (2000). Putting health inequalities on the map; does where you live matter? in H Graham (ed) *Understanding Health Inequalities,* Open University Press, 143–155.
Sloggett, A and Joshi, H (1998). Indicators of Deprivation in Longitudinal Perspective *Environment and Planning A,* 30, 1055–1076.
Sloggett, A and Joshi, H (1998). Indicators of Deprivation as Predictors of Life Events *Journal of Epidemiology and Community Health* 54, 4, 228–233.
Sloggett, A and Joshi, H (1994). Higher mortality in deprived areas: community or personal disadvantage? *British Medical Journal*, December, Vol 309, 1470–1474.
Weatherall, R, Joshi, H E and Macran, S (1994). Double Burden or Double Blessing? Employment, Motherhood and Mortality in the Longitudinal Study of England and Wales. *Social Science and Medicine*, 38, 2, 285–297.
Clarke, L and Joshi, H (2003). Children's Changing Families and Family Resources. In Jensen and McKee (eds) *Children and the Changing Family: between transformation and negotiation.* London: Routledge Falmer, 15–26. ISBN 0 415 27773–6

Professor Heather Joshi
Institute of Education

November 2011

Written evidence submitted by the Economic and Social Research Council (ESRC) (Census 26)

PREAMBLE

1. For over 40 years the Economic and Social Research Council (ESRC) has been at the forefront of developing the UK's social science data infrastructure, supporting the creation of world leading data resources and major, pioneering data services which have played a leading and longstanding role in shaping the international data landscape for social and economic research. The ESRC works closely in collaboration with a wide range of stakeholders across government and elsewhere who both produce and fund access to social and economic data resources to develop the UK's data infrastructure. The National Data Strategy, led by the ESRC, provides a critical framework for identifying future data needs for the academic, policy and practitioner communities.

2. The ESRC Delivery Plan 2011–15 (http://www.esrc.ac.uk/news-and-events/news/13751/esrc-delivery-plan-2011–2015.aspx) reiterates ESRC commitment to support strategic investment in core data infrastructure, including the ESRC Census Programme. This Programme provides academic access, expert user support and the representation of academic census user requirements to the census organisations through a network of data support units and a co-ordinator. The ESRC has invested extensively in its Census Programme to support data access services, expert support and training, research and development to underpin social science research using UK census data. This investment indicates the widespread and enduring importance of the census datasets as a key source for understanding British population and society. At present, over 20,000 registered users in UK Higher and Further Education take advantage of these services.

3. In the preparation of this submission, the ESRC has received input from members of its Policy Committees, key investments, and a from its Census Programme investments via the Census Coordinator, Professor David Martin.

How do social scientists use Census data?

4. Social scientists use census data for a diverse range of purposes and it is impossible to articulate one single position in response to this question. Many social scientists are specifically concerned with major thematic research questions in fields such as migration, ageing, health and social circumstances. These researchers use a range of research data sources, of which the census is one very important source.

5. The Census serves as the definitive source for describing the health and social circumstances of the contemporary population, particularly for relatively small population groups. For example, the Census is the only data source that allows us to examine the characteristics of specific ethnic minority groups, occupational groups, or residents of small geographic areas smaller than electoral wards. Census data, including the Samples of Anonymised Records (SARs), can serve a source of data for detailed research in their own right or can be used to substantiate or inform findings arising from sample surveys.

6. Social scientists are often concerned with understanding long-term societal change, which is only possible by the comparison of census data over multiple decades, and historians in particular have developed unique insights from individual census responses which are released after 100 years, allowing analysis of life in the

late 19th and early 20th centuries. With the end of the census such comparative study will no longer be available to future generations. Census data allows social scientists to study change over time at a population level, for example in access to basic amenities or household structures. A particularly unique strength of the current census is that its near-comprehensive population coverage in a single data collection exercise facilitates a unique range of integrated data outputs and cross-tabulations:

Example 1 from Academic A

In order to obtain rates of births, deaths, marriages etc., counts of events (such as births) obtained through the vital registration system are combined with denominators of the population at risk of experiencing this event—for example, women of childbearing age. Of particular importance is the need for the denominators to be broken down by detailed attributes (social class, ethnic group, country of birth or by small geographical area). The census is currently the most reliable source of such data. For example, the ONS, only publishes age-specific fertility rates by country of birth for census years, where they can be more confident of knowing the correct population by sex, age and country of birth. In my own social research, I rely on these census-based population estimates: (i) to investigate fertility rates of migrants by country of birth; (ii) to map and explain spatial trends in teenage pregnancy in Welsh local authorities over time; (iii) to compare local geographies of family size in Britain and hence understand and predict demand for school places.

7. Whereas other social surveys may ask for fuller details of housing (for example), no other source provides the ability to tabulate housing and other data for small geographical areas. Similarly, there is currently no other source which permits so many different variables to be cross-tabulated and integrated for small areas. The census currently permits creation of multiple output data products, essentially aggregated data for small areas, geographical boundaries for those areas, flow data representing travel to work and migration flows between areas, samples of anonymised records and a link to the ONS Longitudinal Study (LS):

Example 2 from Academic B

The census covers the entire population, both those living in private households and those in institutions. It therefore allows us to look at the size and characteristics of sub-groups of the population who are not usually represented sufficiently in sample surveys eg those in institutions and those who are too small in number to be represented in a sample survey. In looking at family formation among minority ethnic groups I have relied on micro-data samples from the 1991 and 2001 censuses—the SARS and CAMS. Whilst summary statistics can be gained from large social surveys such as the Labour Force Survey, it is only by using samples from the census that we have enough sample size to look, for example, at the factors associated with inter-ethnic marriages and cohabitations, and to compare trends for first and second generation migrants.

8. This richness is only possible because of the existence of a single census database derived from near-comprehensive coverage of the population across all census questions, geographically referenced at the address level within the Office for National Statistics. Currently there is no other source of such data:

Example 3 from Academic C

Social research is often interested in the role of local area characteristics in affecting individual behaviour. Currently, census data, alongside other information, are used to construct local area classifications, eg to assign each small area with indices of deprivation. These local area classifications and indices can then be linked with individual-level survey data to identify possible contextual effects. In my own research we have linked local area classifications and information on local housing markets to individual level data from the British Household Panel Study in order to assess the impact of local area unemployment rates and housing markets on the ability of young adults to move out of the parental home and make the transition to residential independence.

9. It should be noted that census data also provide a key input to social science teaching at undergraduate and masters levels, with the aggregate outputs in particular forming a focus for many courses, both in data analysis and in substantive fields such as demography, sociology, public health and population geography:

Example 4 from Academic D

Longitudinal data allow research to move beyond cross-sectional associations to examine potential causal mechanisms. We have in the UK some excellent longitudinal data sources for social research, such as the Birth Cohort Studies and the British Household Panel Study. However, these do not contain sufficient numbers of particular sub-groups (ethnic minorities; recent migrants) to be able to say a great deal about the wellbeing and life course experiences of minority populations. This may change in the future eg if the ethnic boost within Understanding Society is successful and these minority groups are retained in the study. Linked census data that make up the ONS LS provide a unique opportunity to study eg the longitudinal determinants of health and wellbeing of ethnic minorities and the dynamics of family formation among recent migrants. Recent research that colleagues and I have completed uses the ONS LS to examine the childbearing patterns of recent migrants and hence address part of the question as to extent to which recent increases in the birth rate in Britain has been caused by recent increased migration.

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

10. The implications are heavily dependent on the nature of the replacement data sources. Implicit in the question is also an assumption that census data can continue to be collected and published at comparable levels of quality and detail as is presently the case, and there are good grounds for believing that this may not be the case. While methodological understanding and development are very important themes, the majority of social science researchers are fundamentally concerned with the nature and quality of the data available to answer key substantive research questions over and above the mechanism which has been employed to collect and publish these data. However there are concerns within the social science community that the end of the UK census without a firm proposal for a viable alternative is likely to have a negative impact on the quality of social science research in the UK.

11. The most important potential implications of ending census data without an adequate replacement relate to the loss of explanatory power which comes from the integrated nature of the census. Census data allow, for example, the cross-tabulation of qualifications and employment status at small area level, facilitating important insights into the relationships between education and life chances. Non-census replacements which may offer independently-derived counts of qualifications and classifications of jobs cannot be employed to adequately answer these questions unless there is integration of data at the individual level. The ESRC is proactively working with partners to drive forward use of administrative data through the establishment of an Administrative Data Taskforce (due to report to Ministers in the summer 2012). It is hoped that barriers and issues, such as data quality, can be addressed through a coordinated plan of action.

12. A further concern is that there are some census questions (eg ethnicity, self-reported health) for which there are no comprehensive alternative sources, hence it would not be possible to directly continue with some important existing analysis. In some cases it is possible to conceive of alternative sources for some of these data, but they would take time to develop and are likely to result in significant changes of definition (eg the difference between asking households how many cars they have available and counting cars by registered address from DVLA records), with the consequence that many analyses of change would be considerably disrupted. If existing large scale surveys are used as an alternative source, there is a question whether they can continue to provide reliable information, given many social surveys are suffering from declining overall response rates. Whilst non-response is an issue for the census as well, non response rates are far higher on sample surveys were response is voluntary.

13. Another danger is that whereas the definitions and frequency of census data collection are under the control of the national statistical organizations, many potential alternative sources are determined by other agencies and policies which are liable to change at short notice resulting in loss of data, for example, the imminent discontinuation of universal child benefit, which might have been a potential complementary source of data on counts of children and families.

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

14. Importantly, any alternatives to the Census, whether it is a rolling census, a population register from administrative sources, a survey based approach, will have to be able to provide similar information in a timely and accurate manner and provide similar level of coverage and quality.

15. The first results of recent censuses have taken around 18 months to become available and detailed statistical products of interest to social science researchers take much longer. In some cases these may not be available until two to three years after the census, hence census data at their most valuable are often two to three years out of date. The census is also challenged by complex household formations, coverage of short-term visitors to the UK, persons with multiple residences, and a range of difficult-to-enumerate populations, including young men in large metropolitan areas and students. In 2011 it appears that many of the response difficulties experienced in 2001 have been addressed and coverage rates are still high, but at significantly increased cost and it is currently too soon to examine what biases may have been introduced as a result of the different enumeration strategies adopted. The 2001 experience suggests that once census coverage falls even moderately, it becomes extremely difficult to adequately correct the population counts by means of statistical adjustment and imputation.

16. In this context, potential alternatives to the census do offer some significant advantages, in theory. To provide an adequate population data source for social science research, it would be essential to develop some form of statistical population register—although this might be created temporarily in a secure setting by matching existing lists and does not necessitate the creation of an identity register, against which individual data are continuously matched. The UK is undoubtedly hampered by the absence of a population register, or even common personal identifiers across many of the key datasets. A population register has played an important role in other countries (for example Nordic population registers which represent some of the best sources of data for demographic research, especially when linked to administrative records) which have moved to administratively-based alternatives to the census.

17. The matching of individuals and addresses is essential in order to be able to recover the multivariate cross-tabulation of data routinely produced by the census and to have adequate counts for small geographical areas. ONS is currently conducting a programme of research on alternative sources, among which the NHS

register and DWP/HMRC Customer Information System appear to be the most promising reference frames. Also very welcome is the recent establishment of the GeoPlace national address gazetteer. A substantial amount of additional cross-matching would be necessary in order to overcome some of the current known weaknesses of these lists in terms of under- and over-counting.

18. Against these lists, a wide range of additional registers could be matched to derive additional characteristics for example annual school census data, HESA records, benefits data, DVLA information etc. No system of matched registers would provide the levels of detailed social characteristics available from the census but many of these could be included within an integrated and extended social survey system, which drew on the same address gazetteer as a sampling frame and was appropriately stratified. This could in theory address all the current census themes and more, providing greater responsiveness for the inclusion of additional questions, but with data inevitably not all available for such detailed geographical areas. The quality of many of these sources remains unknown in detail, which is why the current ONS Beyond 2011 programme of work is to be strongly supported, and is being support by the ESRC.

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

19. The creation of an integrated non-census population data system would require deficiencies in existing data systems to be addressed and a significant expansion in the administrative systems available to the statistical agencies. Matching at the individual record level is essential, potentially based on the NHS register, although there are concerns about selective over- and under-counting.

20. The current system of monitoring international migration using the International Passenger Survey would need to be massively augmented by a proper flow of data from the e-borders system, which over time would provide the most accurate information on numbers of UK residents travelling abroad temporarily and long-term and would also provide the most accurate figures for overseas citizens within the UK.

21. The Schools Census (Pupil Level Annual School Census pre-2007) records a variety of information about pupils in state education in England, including the student's home addresses. Access to data from the Schools Census is currently limited, especially fields such as home address postcode which act as personally identifying. The Schools Census 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. Differences in questionnaire wording between Scotland and other parts of the UK have prevented production of these data.

22. 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 higher education. These have the potential to produce migration matrices showing the movement of students to universities and colleges, and there are no obvious alternative sources of such data. 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.

23. The Annual Survey of Hours and Earnings (ASHE) includes information about employees' workplace and home postcodes through a questionnaire sent to employers. It does not include any socio-demographic information about the individual, nor does it include information about the usual mode of transport to work. The current collection model does not allow any additional personal data, such as mode of transport to be collected, and with the small sample size (1%) no detailed spatial disaggregation could be made even if this data were collected.

24. The success of the current census is evaluated using an extensive census coverage survey and it would be essential that a similar survey instrument were used to evaluate the coverage quality of any alternative system: this would enable users to gain confidence in the counts and also to understand the patterns of bias and omission which may exist in the new sources.

Economic and Social Research Council (ESRC)

30 November 2011

Supplementary written evidence submitted by the Economic and Social Research Council (ESRC) (Census 26a)

14 DECEMBER 2011 EVIDENCE SESSION

You asked if I could send the list of the top 10 most heavily exploited datasets used by social scientists, I referred to yesterday in my oral evidence. This is data collected by the Economic and Social Data Service which is the national service ESRC funds to provide access to social and economic data. Those with an asterisk are datasets collected by ONS or other government departments.

1. Census data*.
2. Labour Force Survey*.
3. Health Survey*.
4. British Household Panel Survey.

5. British Social Attitudes Survey.
6. 1970 British Cohort Study.
7. General Household Survey*.
8. British Crime Survey*.
9. National Child Development Study.
10. Workplace Employee Relations Survey.

Jeremy Neathey
Deputy Director of Policy
Economic and Social Research Council (ESRC)

December 2011

Written evidence submitted by the Royal Statistical Society (Census 30)

1. The Royal Statistical Society (RSS) is the UK's only professional and learned society devoted to the interests of statistics and statisticians. Founded in 1834, it is one of the most influential and prestigious statistical societies in the world drawing its membership from over 50 countries. It aims to promote public understanding of statistics and provide professional support to users of statistics and to statisticians.

2. The RSS Census Study Group provides a cross-sectoral forum for debate and education among census users, focusing on the broader methodological issues which should concern all users of census data rather than on specific sectoral or topic interests. The group convenes meetings and contributes to the work of other special interest groups, bringing together census users from across academia, business, central and local government and the voluntary sector.

3. The Academy of Social Sciences, which exists to be the voice of the social sciences in the UK for the public benefit and whose membership encompasses 750 individual Academicians and 43 learned societies (with a total reach of 85,000 social scientists), drawn from right across the social science disciplines at the academic, policy and practice levels, is glad to add its support for and endorsement of this submission. The issues that it raises are shared by the wider social science community.

General Points

4. The RSS would stress the importance of the Census in providing more than just national figures; its value and use depends on a key feature—the ability to provide highly disaggregated figures (often for small geographies) that also permit the joint analysis of a range of variables. The Census is also resource that is in constant use as a key information source throughout the ten year inter-censal period. Census outputs take two key forms—highly detailed cross-tabulations for separate locations of interest and anonymised micro-data. Both forms of output must be available, not just to government, but to a much wider user constituency. (In addition there is a third form of output in the longitudinal study.) The current outputs are available to a huge constituency of users outside government and provide enormous economic benefits. If the current Census is to be replaced by something that combines administrative data at the personal level then the issues of access for this wider constituency need to be addressed as an integral aspect of the evaluation.

5. The RSS supports the "Beyond 2011" project and agrees that it is right to investigate whether an alternative to the traditional Census would prove advantageous. Nevertheless the traditional census provides a huge range of information. As well as providing the most vital facts about society at local and national level, this is a crucial underpinning of large parts of the evidence base for government, both central and local, for the NHS, and for business and researchers. It is a major foundation block of the UK's information system both in respect of the uniquely detailed information it provides and also as it is the benchmark on which numerous sample surveys are founded and against which they are calibrated. We believe it provides substantial value for money.

6. Thus it is crucial that alternatives are fully and properly explored. While it would clearly be desirable to find a lower cost method to replace the census, it would be a false economy to reduce costs if this resulted in overall much lower information content. The current ONS "Beyond 2011" programme of research is to be welcomed as a means of establishing as clear as possible a picture of the genuine potential for assembling a viable census replacement system. It is anticipated that any such system will need to comprise a combination of administrative and survey sources, the latter required to address important variables which are not collected in any comprehensive administrative systems.

7. The work of the ONS covers England and Wales. Separate studies are under way in respect of Scotland and Northern Ireland. It would be a major concern if there were substantial differences in the systems adopted for the different parts of the UK.

Question 1: *How do Social Scientists use Census data?*

8. Social scientists use census data for an extremely wide range of purposes, reflecting the breadth of the census as a data collection exercise which is designed to meet the needs of many sectors and organizations. It

should also be remembered that social scientists are located across all sectors, with significant census use in central and local government, business and academia in addition to the large body of social science researchers and students within the academic sector. Many social scientists are specifically concerned with major thematic research questions in fields such as migration or ageing and will use a range of research data sources, of which the census will be one.

9. A particularly unique strength of the census is its near-comprehensive population coverage which facilitates a unique range of integrated data outputs and cross-tabulations which are available down to a very small area level. No other single source provides detailed tables for small geographical areas. The small area aggregated data provided by the census allow a range of analyses to be undertaken investigating neighbourhood influences on aspects of life such as educational achievement, health and crime. Business users in particular make extensive use of these multivariate small area data for area classification and to inform decisions about target markets and site location. Government users are more likely to be concerned with items such as the assembly of deprivation indicators and denominator datasets which permit the characterisation of areas to inform the delivery of services. These applications demonstrate the enormous importance of the census as a baseline dataset, with which unique analyses are possible due to its unique combination of breadth and depth. The census also provides more specialised research datasets such as flow data permitting the investigation of travel to work and migration patterns, made possible by asking place of residence, place of work and place of residence one year before the census and providing important insights into travel patterns, housing demand and patterns of economic development.

10. A further family of unique analyses is made possible by samples of anonymised records, which permit users, for example, to create bespoke cross tabulations of census variables and to analyse household structures. Although these represent only a small fraction of the census dataset, the near-comprehensive census coverage ensures "sample" sizes far greater than even the largest social surveys. A further most important research linkage is provided by the ONS Longitudinal Study, which now contains a sample of linked records from 1971 to 2011 and permits unique insights into individual life histories, for example understanding the role of past employment and educational history on health and mortality later in life.

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

11. The answer to this question depends entirely on the nature of the replacement data sources. Public acceptability will be a major determinant in the success of a potential alternative system although that issue is not addressed in any detail here. There is no certainty that census data can continue to be collected and published while maintaining current quality levels. In 2001 substantial adjustments were necessary to the census estimates for some local authorities due to failings in the census process. Initial feedback suggests that 2011 enumeration has been successful in addressing these challenges, but at increased cost and also by using a variety of different data collection methods (online, post-back, post-back following enumerator visits) which may introduce further biases. It will not be possible to fully assess the implications of these changes until quality assurance of the 2011 census outputs is complete.

12. Perhaps the greatest risks of ending the census would result from the lack of integrated data which permit the cross-classification of different variables, with a resulting loss of explanatory power. Administrative data bases do not usually provide detailed cross-tabulations of outcome variables against socio-economic and other characteristics. It is also not certain whether alternative sources will be able to provide all the richer datasets such as flow data, microdata and links to longitudinal studies. The extent to which alternative sources will be able to overcome the weaknesses of the census model, while replacing its combination of geographical and social detail, will be key to evaluating the impact of ending the census. There are some census questions for which there are no alternative sources, or for which alternative sources may take many years to develop for example questions on health and languages.

13. While the use of administrative sources may offer the prospect of more timely data, the overall data infrastructure would become more dependent on the policies of a range of government departments and organizations whose primary objective is not data collection. From a social science perspective this creates a risk that data series may change in unplanned ways and that comparable datasets through time may be difficult to achieve. Government departments supplying information to the statistical system replacing the census should be required to consult with the National Statistician before implementing major changes to their data collections.

14. Despite this, there is no reason why a non-census system could not be developed to continue to create microdata samples and links to longitudinal studies—indeed, the longitudinal studies already demonstrate much of the power of linked administrative systems. However, the lack of a decennial census validation would present real barriers to long-term analyses due to the unknown effect of such a major break in the methodologies employed. International experience suggests that it takes a long time to develop alternative data sources to the point where they can provide replacement data series of equivalent quality to that which has been achieved by recent UK censuses.

15. Even if an alternative to the Census could be developed which would satisfactorily meet the needs of social scientists, this will take time. A main current criticism of the Census is that it is decennial and that there is a long wait for new data—we will continue to use 2001 data until well into 2013 for example. A concern is

that if the 2011 census proves to be the last, it may be an unacceptably long time before a suitable alternative can be developed.

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

16. The answer to this question is contingent on the levels of coverage and accuracy which may be achieved by a notional future census. Whereas census data typically appear 18 months to three years (in the case of some of the richest research datasets) after the enumeration, the long time period between each data refresh can be a challenge to many users. The census continues to be challenged in terms of overall response rates, difficulties of identifying and gaining access to properties and a range of hard to reach population groups, including young men, visitors to the UK and those with multiple residences and complex household structures. Initial signs are that the 2011 census has achieved a good result in the face of these challenges, but with a considerable increase in cost. International experience suggests that the difficulties of achieving near-comprehensive coverage will continue to increase. In the 2001 census coverage in some areas fell to levels which were not amenable to satisfactory adjustment using statistical methods.

17. There are thus attractions to the use of a system of alternative sources which might address these difficulties more effectively, but it is unlikely that this can be achieved by use of any single approach. Rather, a combination of methods and alternative sources needs to be evaluated. It seems likely that any alternative would need to include some form of statistical population register or database which would allow the matching of other lists at the individual record level, otherwise it will not be possible to generate multivariate datasets comparable to those obtained from a census.

18. ONS is currently conducting a welcome programme of research on alternative sources, among which the NHS register and DWP/HMRC Customer Information System appear to be the most promising reference frames. Also very welcome is the recent establishment of the GeoPlace national address gazetteer. A substantial amount of additional cross-matching would be necessary in order to overcome some of the current known weaknesses of these lists in terms of under- and over-counting. Against these lists, a wide range of additional registers could be matched to derive additional characteristics for example annual school census data, HESA records, benefits data, DVLA information etc.

19. No system of matched registers would provide the levels of detailed social characteristics available from the census but many of these could be addressed by an integrated and extended social survey system, which drew on the same address gazetteer as a sampling frame and was appropriately stratified. This could in theory address all the current census themes and more, providing greater responsiveness for the inclusion of additional questions, but with data inevitably not available for such detailed geographical areas. The quality of many of these sources remains unknown in detail, which is why the current ONS Beyond 2011 programme of work is to be strongly supported.

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

20. The creation of an integrated non-census population data system will require a range of different, coordinated actions. Firstly, it will be necessary to identify ways of improving the quality of existing sources. For example, reconciliation of addresses from a definitive address register and other sources such as the NHS register will result in improvements to both. The NHS register provides the most promising data source to contribute to population counts and internal migration but it is subject to recognised weaknesses, including both over- and under-counting of specific population groups. Moving to a post-census statistics system would require issues such as this to be seriously addressed and supported to an extent which has not previously been done—and also recognition that complete resolution is most likely unachievable. Secondly, some existing sources would need to be augmented by considering the need for additional data collection, for example whether more systematic ethnicity information could be included in one of the major data series such as the NHS register. Developments in other areas may yield specific new variables of interest: for example the current census question on central heating might in due course be more usefully replaced with information on energy use consumption from utility records.

21. Despite recent improvements, the measurement of international migration still remains a major weakness and the availability of well-structured data from the e-borders system would be an essential element of a workable census replacement system. A third element would be close integration of the current social survey system with the census replacement, almost certainly involving both the integration and expanding of existing surveys so as to provide greater coverage of socioeconomic detail and the continuation of a substantial coverage survey which would allow both user and producers to better understand the characteristics of the post-census population data. Declining response rates to the major government surveys is already a concern and it should not be assumed that using larger or more integrated surveys as part of a census replacement will overcome

these challenges without further detailed attention to design and implementation. Increased use of internet-based data collection may offer some advantages in this respect but again its potential role in a greatly changed data environment requires substantial research effort.

Royal Statistical Society

30 November 2011

Written evidence submitted by the Royal Geographical Society (with The Institute of British Geographers) (Census 33)

1. The Royal Geographical Society (with The Institute of British Geographers) welcomes this opportunity to comment on the inquiry into the census and social science.

2. Given the spatial nature of data collected within the Census, social scientists within geography are amongst those who make greatest use of its data within their research, along with 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. The Society—as the learned Society representing the discipline—focuses on use by the geographical research community.

3. Formed in 1830, the Society's Royal Charter is for "the advancement of geographical science". We are a charity that seeks to develop, promote and support the discipline of geography and its practitioners in the areas of research and higher education, teaching and fieldwork, policy and wider public engagement. The Society has more than 15,000 Fellows and members, of whom a substantial number are academics and other researchers whose work we support through a range of activities. These include holding the largest geographical research conference in Europe, publishing three of the leading international peer-reviewed geography journals in the world (including Transactions of the Institute of British Geographers which is often ranked first), co-ordinating 27 specialist research groups, and providing small grants for researchers at all career stages. We work very closely with all Higher Education (HE) geography departments in the UK.

4. In the preparation of this response, the Society consulted with and elicited responses from two of the Society's research groups (the Quantitative Methods Research Group; and the Population Geography Research Group). In addition we received comments through a number of the Society's Fellows.

5. The key points of the response can be summarised as follows:

- Geographers, as social scientists, make wide use of the Census in their research, through aggregation of population data to the areas and regions; flows of population between areas; and examining how these change over time. This includes the understanding of international migration; internal patterns of migration; population estimates and services; public health planning; and the creation of classifications of areas; and transport, including mode of travel to work.
- The specific impact of ending the Census is heavily dependent on the exact nature of the replacement population data system, but the main risk would mean deterioration of the information social science research has at its disposal which would compromise Central and Local Government's ability to make accurate decisions about funding services.
- To achieve an alternative to the Census with data of equivalent or higher quality would need a full investigation of the geographical coverage of the alternatives as it is not clear that any proposed so far provide the same combination of small area detail and accuracy.
- Other existing data sources that could be improved upon include the NHS Patient Register; the Schools Census; the Higher Education Statistics Agency (HESA) record information for students at Higher Education Institutes; the Annual Survey of Hours and Earnings (ASHE); and migration data (based on the NHS register and International Passenger Survey / e-borders system data).

How do social scientists use Census data?

6. The census provides a common geographical framework for the integration of data from many sources and its small area counts are essential in allowing:

(i) Aggregation of population data to the areas and regions—including mapping, modelling, classification and analysis of small areas such as neighbourhoods. This work ranges from the construction of deprivation indicators and area types to the investigation of geographical patterns of inequality and the relationships (for example) between local social conditions and ill health or life chances, or in providing denominators for studies of disease prevalence or crime rates. This also allows for analysis of sub groups of the population eg ethnic or migrant groups; lone parent families; people in institutional accommodation.

(ii) Flows of population between areas—for which the census provides data both on travel to work patterns and residential migration. Understanding of such flows is essential to our understanding of daily travel patterns, urbanization, regional development, demand for transportation and housing.

(iii) Understanding long-term societal change by comparing census data over multiple decades, and even with the individual responses which are released after 100 years.

7. Specific examples of research undertaken by geographers using Census data include:

(i) Studies of international migration: measuring neighbourhood level population change of ethnic groups. There is considerable evidence to indicate processes of spatial de-concentration are taking place, particularly in London, where the majority of ethnic minority populations are located.

(ii) Understanding internal patterns of migration: millions of people in Britain change their place of usual residence each year. The Office of National Statistics (ONS) Longitudinal Study enables Census records at the individual level to be linked from one Census to the next, allows these internal movements to be understood.

(iii) Population estimates and services: 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.

(iv) Public health planning: health geographers make extensive use of census data on socio-economic conditions, demographic characteristics and health related topics and census data are routinely used in public health planning and assessments of service activity and outcomes.

(v) Creation of classifications of areas: The Census has been an important source of denominators in the creation of indicators of local disadvantage using more regularly up-dated information sources. These 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.

(vi) Transport, including mode of travel to work: this is particularly well used by researchers as it is by transport planners, including use to define the official Travel-to-Work Areas (TTWAs) used by various Government Departments. In addition Census "flow" data (eg journey-to-work) has been used to delineate functional urban regions and city regions, which identify the hierarchy of cities and towns.

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

8. The Census is an important tool for shaping and evaluating policy in the UK, and is unique in its universal (100%) coverage of the population and in the breadth of subject areas for which information is recorded. It is the only current study which can provide the geographical detail needed to accurately show variations in British society and it is hard to see how attempting to reproduce this coverage by other means would be less resource intensive or expensive.

9. The specific impact of ending the Census is heavily dependent on the exact nature of the replacement population data system. If the alternative provides the same detail of data, but at greater frequency of update, then the data environment for social scientists will have been improved by abandoning the Census. However if, as is more likely, the replacement will be a series of partial products—such as population estimates enhanced by administrative records, combined with local data on housing, economic activity and health based on government surveys, then there is a real risk it will not allow aggregation to a wide enough range of areas needed for research and policy purposes.

10. The design of any census replacement system would need to include a considered strategy for reviewing and updating the geographical basis on which all socioeconomic data are to be published. Further research is required to assess the geographical scales at which data might be produced from an enhanced system of social surveys.

11. 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: further research is also needed on the ability of alternative data systems to capture data such as this—including travel to work and migration flows—for which we currently rely on the Census.

12. The main risk identified is a compromising of Central and Local Government's ability to make accurate decisions about funding services (including schools and hospitals). For example, unless an alternative data source, based on civil registration of the resident population, is developed it will become much more difficult to make local needs assessments in for health service provision.The census is also important in sparsely populated areas, which are likely to be poorly represented in sample surveys. A very considerable effort to enhance the power of national sample surveys would be necessary to substitute for the census in this respect.

13. There would be further loss in terms of the sources of information compiled in the Longitudinal Study, which is an internationally admired source of data originally based on a sample from the 1971 census. This provides a unique source of information on the currently aging population and other cohort studies have nothing like the same power to analyse local variation in important dynamic processes of aging, migration and health outcomes.

14. Assuming no satisfactory replacement then this will mean deterioration of the information social science research can use to investigate the well-being of the population, with the real difficulties starting to emerge from around 2022–23.

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

15. The ONS, NRS, NISRA and WAG are all engaged in a project, overseen by the UK Statistics Authority, called Beyond 2011 (2011 to 2014) which aims to develop options for replacing the census. Eight options are being investigated: (1) traditional census (eg UK Census 2011), (2) short and long form census (eg US Census 2000), (3) short form census and continuous survey (eg US Census 2010) and (4) rolling census (eg France's current methodology) which uses a very large rolling survey and interpolation and extrapolation methods, (5) address register plus survey, (6) administrative data—aggregate (cf. Index of Multiple Deprivation), (7) administrative data—record level (eg Finland) and (8) administrative data—intermediate.

16. For further details of alternatives being considered, the committee is recommended to refer to the National Statistics Beyond 2011 project website for further details. For example, Option 7 might be based on the reconciliation of existing administrative registers—including the new National Address Gazetteer and population lists such as the NHS and DWP/HMRC lists as the basis for an initial population register. Another person identifier is the National Insurance and the DWP maintains the "Lifetime Labour Market Database" which can trace individuals over time and across space.

17. An important consideration in the evaluation of alternatives would be to thoroughly investigate the evenness of their geographical coverage and would be likely to require monitoring via some form of coverage survey, as is used in the 2001 and 2011 censuses, to provide confidence in the coverage of the alternative system across different local authorities and area types.

18. Although the matching of individual records in administrative databases my provide substitutes for some Census variables there are others (eg hours of work or employment status) where it is likely that research will be dependent on existing government surveys (such as the Labour Force Survey/Annual Population Survey) for information. Their existing sample sizes mean that it is not possible to use data from these sources for some smaller local authorities, let alone more detailed local areas. For them to be effective substitutes for the Census their sample sizes would need to be significantly increased—which would involve significant costs.

19. If the UK had a proper population register, as other EU countries do, with a legal requirement on all residents to report their addresses and changes of address and with an universal person identification, then the population spine would be as accurate as it could ever be. This would, as in other countries such as Finland, provide a sensible alternative to the Census. Parliament rejected the project to create a National Identity Card and existing cards have been cancelled. Historically, a population register was first proposed (and rejected) in the House of Commons in 1753. The National Health Service Register covers about 98% of the population at present but could be extended to cover all the population by including all individuals not entitled to NHS health care or receiving health care from other bodies.

20. Monitoring of travel and migration could exploit new data sources from traffic and public transport monitoring, together with more determined development of existing systems such as the NHS register for internal migration. Some census questions would inevitably need to be replaced by enhanced social surveys in order to ensure continuity of information on topics not readily captured from any administrative source.

21. Alternative sources of commuting data are very limited and 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 because the small sample size does not permit the publication of detailed disaggregated flows.

22. None of the proposed alternatives in any of these areas will provide the combination of small area detail and accuracy that make Census data the one essential data resource. 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".

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

23. 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 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.

24. Researchers highlight that 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.

25. Suggested areas where improvements to existing data sets could be achieved, include:

— Migration data: both internal (based on the NHS register) and international (based on combination of existing International Passenger Survey data with robust counts from the e-borders system). Until such time as the latter are available it is difficult to conceive how adequate counts can be produced for international flows. There may be an opportunity to utilise additional data sources from business, particularly relating to energy use, car ownership, communications (including mobile telephony traffic, a potential new alternative for tracking population mobility), financial status and consumption patterns, but these sources are unlikely to be individually comprehensive.

— 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, but the lack of socio-demographic data in the Patient Register limits its scope for use in social science research.

— 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. 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 and 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.

— 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.

— 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. 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.

References

[i] David Martin, Professor of Geography, University of Southampton and Coordinator of the ESRC Census Programme; Phil Rees, Emeritus Professor, School of Geography, University of Leeds; Professor John Stilwell; Danny Dorling, Professor of Human Geography, Sheffield University; Dr Martin Frost; Dr Darren Smith, Chair PGRG; Christopher Brunsdon, Chair QMRG and Tony Fielding, Research Professor, Department of Geography and, Sussex Centre for Migration Research, University of Sussex.

[ii] Riva, M Curtis, S, Norman, P (2011). Residential mobility within England and urban–rural inequalities in mortality; Social Science and Medicine, doi:10.1016/j.socscimed.2011.09.030).

[iii] ONS (Office for National Statistics), NRS (National Records Scotland), NISRA (Northern Ireland Statistics and Research Agency_) and the WAG (Welsh Assembly Government).

[iv] ONS (2011a) Beyond 2011. Online at: http://www.ons.gov.uk/ons/about-ons/what-we-do/programmes—projects/beyond-2011/index.html

[v] http://www.ons.gov.uk/ons/about-ons/what-we-do/programmes—projects/beyond-2011/index.html

[vi] Home Office (2011) Identity cards. The UK National Identity Card and the Identification Card for EEA (European economic area) nationals ceased to be valid legal documents on 21 January 2011. Online at: http://www.homeoffice.gov.uk/agencies-public-bodies/ips/about-us/suppliers/identity-cards/

[vii] Coleman, D (2010) Memorandum from David Coleman, Professor of Geography at Oxford University. Ev 184, House of Commons Treasury Select Committee.

Royal Geographical Society (with The Institute of British Geographers)

30 November 2011

Supplementary written evidence submitted by the Royal Geographical Society (with The Institute of British Geographers) (Census 33a)

EMAIL TO THE COMMITTEE FROM PROFESSOR PHIL REES, ROYAL GEOGRAPHICAL SOCIETY, 15 DECEMBER 2011

ORAL EVIDENCE 14 DECEMBER 2011

I make a couple of observations on a remark made and a question asked by The Right Honourable Graham Stringer, Labour MP for Blackley and Broughton.

The role of an information infrastructure in building the welfare state

Mr. Stringer made the point that the reforms of the Labour Government of 1945–51 which established the foundations of Britain's welfare state were uniformed by a census because in 1941 the exigencies of the Second World War meant that such an enumeration was not possible. However, I would point out that crucial to the war effort and the foundation of the National Health Service was the second National Register, 1939 to 1952. Conscription, labour direction, rationing and patient lists were founded on the National Register. I would argue that if you do not have a census after 2011, then such a register is absolutely necessary, where set up, as I argued for, as a full Population Register or constructed virtually by matching records across existing national administrative registers, including the NHS Register.

Do poor people report less or more illness than richer people?

Mr. Stringer stated that he was puzzled by higher rates of illness (Limiting Long Term Illness) revealed in the Census for Buckinghamshire than inner London or Manchester and suggested that the question on LLTI might achieve different responses from richer than poorer people. I don't think this observation is correct. The maps of LLTI from the 2001 Census (eg Daniel Dorling and Bethan Thomas (2004. *People and Places: A 2001 Census Atlas of the UK*. The Policy Press, Bristol) show clearly that Buckinghamshire has lower LLTI rates than inner London or Manchester. A map from their atlas showing LLTI for persons age 16+ (adults) by local authority is available online at:

http://www.sasi.group.shef.ac.uk/publications/pandpexamples.htm#illness

However, I would observe that this map underestimates the inequalities because before you can compare illness rates in different areas, you must standardize for the effects of age structure. When you do that the gaps between poorer inner city and affluent suburb are even greater. Studies of the variation of standardized LLTI across local areas in the UK suggest that about 70% of the variation is associated with deprivation and a part of the 30% with urban living (linked to pollution, for example). There is a concern that the question about self-reported health or illness may elicit perceptual rather than real differences. However, a large number of prospective studies that link the answers to a health or illness question to subsequent presented illness or death confirm that people are telling the truth. That is why we have a health service that enables people to report their symptoms to their doctor.

Professor Phil Rees
Royal Geographical Society

December 2011

Written evidence submitted by the Office for National Statistics (Census 36)

1. The Office for National Statistics (ONS) has recently, at the request of the UK Statistics Authority, established a new Programme—"Beyond 2011"—to look at the most appropriate way of providing population and small area socio-demographic data beyond the 2011 Census.

2. ONS welcome the inquiry into how social scientists use census data and how changes to the approach might impact upon social science. Our recommendations, which will be made in 2014, will be strongly influenced by an understanding of what matters to users and how statistics are used—and the Committee's findings will contribute to thinking on the best way forward.

3. This note provides a high-level summary of the work being undertaken under the Beyond 2011 programme and is provided to the Committee for information. Further details on any aspect of the programme and the options being considered can, of course, be provided on request.

THE BEYOND 2011 PROGRAMME—BACKGROUND

4. The system for providing population and socio-demographic statistics for the UK, like that in many other countries, has traditionally been built around a census of the whole population. The traditional censuses in England and Wales, Scotland and Northern Ireland provide a population count and a snapshot of the nation at a single point in time. Supplemented by its coverage survey and quality survey, it currently provides the most reliable national source for a number of socio-demographic topics, and can produce accurate, multi-variate outputs for small areas. Many statistical series are re-based every ten years to take account of census results.

5. However, the Census has a number of disadvantages not least the fact that its detailed outputs are updated infrequently (every ten years). This can be a significant issue in areas experiencing rapid population change, or when the importance of a particular socio-demographic topic suddenly changes in response to new or emerging Government policies and priorities.

6. In addition, in recent times it has become increasingly challenging and expensive to conduct censuses and household surveys partly because of the complexities associated with an increasingly mobile population but also because of the reliance on the general public's willingness to take part. At the same time, users want a greater range of outputs to be available and updated more frequently in order to have a better understanding of the population and how it is changing. Similar pressures are reported in other countries, indeed some countries have recently moved away from a traditional census approach while others are currently considering doing so.

7. Prior to the decision to hold the 2011 Census, a review clearly established that, in the short term, a traditional census was the best way to meet user requirements in the UK—and all of the signs are that the 2011 Census operation has been highly successful. Nonetheless there are growing concerns, including from the statistical and user communities, that census-taking in its current form may not be the right vehicle looking ahead. While a traditional census can produce high quality statistics for small areas, it only delivers these data every 10 years and there are significant costs and risks associated with the traditional approach to census taking.

8. In May 2008 the Treasury Select Committee report "Counting the Population" recommended that:

> "The Statistics Authority set strategic objectives to ensure that the data gathered throughout the UK can be used to produce annual population statistics that are of a quality that will enable the 2011 Census to be the last census in the UK where the population is counted through the collection of census forms".

9. Furthermore, in May 2010 Sir Michael Scholar, Chair of the UK Statistics Authority, wrote to the Minister for the Cabinet Office to say that:

> "As a Board we have been concerned about the increasing costs and difficulties of traditional census-taking. We have therefore already instructed the ONS to work urgently on the alternatives with the intention that the 2011 Census will be the last of its kind".

10. The Beyond 2011 Programme was formally established by ONS in April 2011 to look at the alternatives to running a Census in England and Wales in 2021. An iterative programme of research is being taken forward to:

- check and clarify user requirements for population and socio-demographic statistics including the frequency, quality and geographical level required; and
- assess the feasibility of alternative statistical options against a range of agreed criteria which will include consideration of issues such as cost, risk and public acceptability.

11. Close collaboration is planned with the devolved administrations in Scotland and Northern Ireland, who are considering parallel work, to ensure that the obligation to produce UK statistics is met.

12. Following a period of extensive statistical research, the Beyond 2011 Programme will report its findings and make a recommendation in 2014. The recommendation will be evidence based and supported by a full business case. The outcomes will have implications for all population-based statistics in England and Wales and potentially, in the longer term, for the statistical system as a whole.

Options for Meeting Users' Requirements

13. A long list of alternative options for meeting users' needs for small area population and socio-demographic statistics will be tested, evaluated and prototyped before making a recommendation in 2014.

14. At this early stage we are considering a range of alternatives including:

- census-type solutions;
- survey based solutions; and
- administrative data solutions.

Annex A provides a brief summary of the main options being considered. *(Note that this Annex is already published as part of the current "Beyond 2011 Consultation on User Requirements"—but is attached here for information).* Note that although a number of census based options are included, ONS has considerable knowledge and experience of census and survey operations and so the research programme will initially focus more upon assessing options for making better use of existing administrative or commercial datasets. Research will include close consideration of the value that might be added by the re-use of key administrative sources such as the electoral roll or school census data.

15. All statistical options will be assessed against an agreed set of criteria. The specification of these criteria will take account of users' requirements and assess each option in terms of quality, cost, risk, public burden and acceptability (amongst others). When considering possible solutions, we recognise there may be trade-offs to be made between some of the factors important to users—for example, accuracy and frequency—and we are aware of some concerns in this area, for example that a change in model may result in a loss of scope and

quality but may bring cost savings. The specification of the criteria to be used will be crucial to ensure that the right balance is struck when considering the options.

Plans for User Engagement

16. A full understanding of user needs and priorities is essential in order to ensure that maximum benefit is gained from the opportunity to refresh our approach.

17. Beyond 2011 will be undertaking a comprehensive programme of stakeholder and user engagement. This will build on the work undertaken for the 2011 Census. In particular, we shall be making use of our extensive list of contacts to ensure that key users are aware of the work we are planning and have an opportunity to comment on our proposals.

18. The first of two public consultation exercises has already started and will run until 20 January 2012. This is designed to give both established users and other interested parties an opportunity to:

- contribute to our understanding of users needs and priorities; and
- provide information on methods and data sources which could be used to produce population and other socio-demographic statistics.

19. The consultation also provides an opportunity to properly assess the importance to social scientists and the wider community of research data such as the sample of anonomised records (SARs) and the ONS Longitudinal Study (LS) which are currently entirely dependent upon the census. We are very aware of the central place played by these datasets in much research and more generally of the role of the census as a historical record. Ability to support research work will be one of the considerations that feed into our final recommendations.

20. The results of this consultation will be published in spring 2012 and will contribute to the development of evaluation criteria for the assessment of the alternative statistical options.

21. A second public consultation will take place in 2013 to seek comments on the leading options and their relative merits in order to inform the final recommendation. These public consultations will be supplemented by meetings and events with all key stakeholders—including central and local government, commercial users and suppliers, academics, community groups, utilities, privacy groups, genealogists and other interested parties. The views of social scientists will form an important input here—hence our strong interest in the results of this inquiry.

22. Regular feedback will be provided on emerging research findings, the refinement of the options, evaluation criteria and option scores, in order to ensure that our decision making processes and the rationale behind the final recommendations are understood.

Conclusion

23. The Beyond 2011 Programme provides an opportunity to refresh our approach to producing population and small area socio-demographic data in England and Wales. In order to ensure we identify the right approach for the nation, the views of all users, including social scientists, are essential. The outcome of the current inquiry will provide a valuable input to help inform the criteria used in selecting the final options—and so the recommendation on the best way forward.

Annex A

STATISTICAL OPTIONS

(Note that this Annex is already published as part of the current "Beyond 2011 consultation on User Requirements"—but is included here for information)

A1. This annex provides a high level summary of the main options that will be considered by the Beyond 2011 programme. Each of these options (and sometimes variations upon them) will be tested and assessed against an agreed set of criteria.

A2. ONS has considerable knowledge and experience of census and survey operations and so the research programme will initially focus more upon the administrative data-based options outlined below.

A3. The descriptions here are simplified in the interest of brevity. Equally it is very possible that the solution adopted may use elements of more than one of these options—and we may recommend different solutions for different types of data. Much more information on the lead options and the methodologies that lie behind them, will be released over the course of the Programme, including a more detailed paper in 2012.

Traditional Census

A4. This option is a continuation of the current approach ie conducting a full census of the population at a single point in time (for example, most recently, in March 2011). Everyone completes a "long" form questionnaire and the responses provide the basis for producing a count of the population and details on key

characteristics. Historically a census has been taken every ten years in England and Wales, but some countries use a five year cycle instead. In-between censuses the results are supplemented by a series of mid-year population estimates which use the "cohort component method" to add and subtract to take account of births, deaths and net migration. Social surveys on particular topics are used to fill in details, and update estimates as required in the inter-censal period.

Short and Long Form Census

A5. This option is similar to the current census approach in that information would be collected from the full population at a single point in time. A subset of the population would be asked to complete a traditional "long form" census questionnaire. A "short form" questionnaire collecting basic demographic, household and family information would be sent to the remainder of the population. Estimates would be created from a combination of the two sources. Population estimates would be updated using the existing cohort component method, and socio-demographic statistics by social surveys. This approach has been used historically in Canada.

Short Form Census + Continuous Survey

A6. In this option, basic demographic, household and family information would be collected from the full population at a point in time through a short form census ie a questionnaire covering these basic topics only. Population estimates would use the existing cohort component method in the inter-censal period. Socio-demographic statistics would be collected through a large scale continuous survey. A similar approach has been used in the USA, with the "American Community Survey" replacing the long form census element.

Rolling Census

A7. In this case, a census is conducted in different areas on a rolling basis. For example, 10% of the country might be selected each year so that over a decade the whole population would be counted. Population estimates would be supplemented by the existing cohort component method, and socio-demographic statistics by social surveys. An approach of this kind has been implemented in France. Research has already been commissioned into the suitability of a rolling census approach in England and Wales.

Address Register + Survey

A8. Data from an address register is combined with a survey to estimate the population and its characteristics. In its simplest form, the average number of people living at each address included in the survey is multiplied by the number of addresses in an area to give an estimate of the population in that area. Estimates could be updated on an annual basis, but a longer time period may be required to produce estimates for smaller geographic areas. Careful stratification can be used to increase the quality of estimates but this approach is very sensitive to the quality and consistency of the address register—and local variations in household size. Administrative sources or a specific coverage survey would be used for quality assurance, or to supplement survey data to produce small area estimates.

Administrative Data—Aggregate

A9. In this option, aggregate data from a variety of administrative sources would be used to produce initial population counts. This would be achieved by a weighted average of the various sources, stratified by variables including age, gender and geography. Extra information, such as a coverage check survey would be used to refine weights applied to the stratified initial counts. These weights would be regularly updated, producing annual population estimates. Data to produce socio-demographic statistics would be collected through survey(s), with potential for including information from public and private-sector administrative sources in the future.

Administrative Data—Record Level

A10. In this option record level administrative sources (including the address register) would be linked together to produce initial population counts. A regular coverage check survey would be used to assess the accuracy of the initial count, measuring both under-coverage and over-coverage. An estimation process (eg dual system estimation, ratio estimation) would be used to derive weights to be applied to the initial population counts. Further statistical adjustment, for example through imputation, may also be possible. Data to produce socio-demographic statistics would be collected through survey(s), with potential for including information from administrative sources in the future.

Administrative Data—Intermediate

A11. This option would make use of both aggregate and record level administrative data approaches. The aggregate administrative data approach would be used to produce initial population counts. Data from a coverage check survey, measuring both under-coverage and over-coverage, would be linked to record level administrative data sources and the address register, in the coverage check areas only. An estimation process, eg dual system estimation or ratio estimation, would be used to derive weights to be applied to the initial

population counts. Data to produce socio-demographic statistics would be collected through survey(s), with potential for including information from administrative sources in the future.

A12. Almost all of these options will use an address register either as a source or as a frame for surveys. At the same time, all options are likely to be supported by some form of coverage survey or independent check on quality to help correct for over and under-coverage, in addition to a quality survey to monitor the quality of results.

Office for National Statistics

30 November 2011

Supplementary written evidence submitted by the Office for National Statistics (Census 36a)

As requested at Q76, please find attached a short paper on international comparisons.

INTERNATIONAL COST COMPARISONS

In preparing the business case for the 2011 Census in England and Wales, ONS presented equivalent costs that had been supplied by some other countries who undertake similar field enumeration censuses. These are shown below. Note that the costs of the England and Wales Census had been converted to 2005–06 prices for better comparison.

	England and Wales 2011	*USA 2000*	*Canada 2006*	*Australia 2006*	*New Zealand 2006*
Cost per head sterling equivalent at 2005–06 prices	£7.53	£17.20	£7.40	£6.10	£7.50
Cost per head per year	£0.75	£1.72	£1.48	£1.22	£1.50

Note also that Canada, Australia and New Zealand carry out a census every five years, England and Wales and the USA, every 10 years. Countries with five-yearly censuses are more able to re-use infrastructure (such as processing and output systems) and thus spread the costs between censuses. They also make the census investment twice in the decade, giving significantly higher costs per year than in England and Wales.

These figures demonstrate that the England and Wales Census costs are comparable with those in Canada, Australia and New Zealand, who all do five-yearly censuses and have scope for re-using infrastructure, and are significantly lower than the US—where the use of census benchmarks to determine seats in Congress has justified higher levels of spend.

Independently, the United Nations Economic Commission for Europe (UN-ECE) issued a report in 2008 (Measuring population and housing*) in which it compared the practices of countries within the ECE region in the 2000–01 round of censuses. In examining the per capita costs it reported that the UK was ranked fourth lowest out of the 15 of those countries that conducted a traditional field enumeration with self-completion questionnaires.

* http://www.unece.org/stats/publications/Publication_on_2000_censuses.pdf)

I hope that this has been helpful.

Glen Watson
Census Director
Office for National Statistics

December 2011

Written evidence submitted by the Joseph Rowntree Foundation (Census 38)

DECLARATION OF INTERESTS: NONE

The Joseph Rowntree Foundation is an endowed charity that funds a large, UK-wide research and development programme. We seek to understand the root causes of social problems, to identify ways of overcoming them, and to show how social needs can be met in practice.

Our purpose is to influence policy and practice by searching for evidence and demonstrating solutions to improve:

- the circumstances of people experiencing poverty and disadvantage;
- the quality of their homes and communities; and
- the nature of the services and support that foster their well-being and citizenship.

We have no political affiliations and work in partnership with all sectors—private, public and voluntary. We aim to present evidence in a balanced, unbiased way and to stimulate debate on current and emerging issues.

In all our work, we look to reflect the diversity of the UK population, learn from others and operate in a sustainable way—socially, environmentally and economically—finding practical and realistic solutions and focusing on the needs of disadvantaged people. Through our housing, community and care services, we aim to achieve the highest professional standards.

Across our research programmes we have used information collected in the UK Census to support various projects and these incidences will be highlighted in this submission. It is our intention to answer the four questions posed as requested and to make the case that the Census is an invaluable data source in the field of social science.

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

1.1 As the census attempts to get information from every person in the UK it is the only survey which provides a detailed picture of the entire population. It is unique because it covers everyone at the same time and asks the same core questions everywhere. This makes it possible to compare different parts of the UK, Local Authorities and even smaller area statistics on many of the range of variables collected.

Although this response will focus on the social research value of the Census, it is important to note the key role it plays in providing and developing public services. The information the census provides allows central and local government, health authorities and many other organisations to target their resources more effectively and to plan housing, education, health and transport services for years to come.

1.2 Census data is used by social scientists to form the backbone and in many cases the core of many research projects. Aside from the obvious national representation (despite undercounts for certain types of people such as homeless people and the travelling community) the census data enriches many other datasets by linking with them and therefore really enhancing the analysis possibilities. For example, the ONS Longitudinal Study (LS) is a data set comprising linked census and event records for 1% of the population of England and Wales (about 500,000 people at any one census). It was set up in 1974 to address problems with the adequacy of occupational mortality data, and the lack of longitudinal fertility data, but since then it has been used to address a wide range of other research questions. The LS is invaluable for longitudinal and cross-sectional studies of census and events data as it contains all information from census returns since 1971 (including data on occupation, economic activity, housing, ethnicity, age, sex, marital status and education), administrative events data (on births & deaths, fertility, mortality, migration and cancer registrations) and also geographic data. The linking of the data to the LS allows researchers to look in detail at so many things which would not have been possible otherwise such as; survival rates of cancer in relation to socio-demographic characteristics of the population; and birth outcomes of second generation migrants.

1.3 Secondary data analysis from census data saves time that would otherwise be spent collecting data and, particularly in the case of quantitative data, provides larger and higher-quality databases than any individual researcher or research team could collect on their own. This coverage and robustness would not be possible to carry out or maintain by a small research team. In addition, analysts of social and economic change consider secondary data essential, since it is impossible to conduct a new survey that can adequately capture past change and/or developments.

1.4 Here at JRF we have used information from the Census to support a wide range of projects and programmes, some of which are summarised below:

- JRF recently commissioned two projects as part of its "A Better Life" programme, to stimulate and inform thinking on alternative approaches to a better life for older people with high support needs. We supported the projects as part of our programme of research and development projects which we hope will be of value to policy-makers, practitioners and service users. Key information from the Census used in this project includes the question asking respondents whether they have "any long-term illness, health problem or disability which limits activities or work". Media coverage from this project includes a round up about how to empower older people on Publicnet.co.uk (October 2010) and an article in The Guardian (October 2010)—"A better life for older people" discussing this research printed on Older People's Day.
- The other project as part of the Better Life Programme resulted in a research paper entitled, "Demographic issues, projections and trends: Older people with high support needs in the UK" (October 2010). The population aged 85 and over is the fastest growing age group in the entire UK population. More older people than ever before live alone and this has implications for residential arrangements and care among this group. This paper provides an overview of demographic issues (from the Census), projections and trends in relation to older people with high support needs in the UK and recommends continued investment in data sources to further understanding of health, disability, economic and social well-being in old age.

— The Census Programme run by the Place team at JRF produced a number of influential reports based on information collected in the Census. A study on "Population turnover and area deprivation "(Bailey and Livingston, 2007) draws on detailed analysis of flows of population for neighbourhoods in England and Scotland, based on 2001 Census data. It examines whether deprived neighbourhoods have less stable populations, whether they are poorly connected to the wider housing market through movements of people in and out, and whether they are losing better qualified individuals through net movement out of the area. Using data from the 2001 Census, this study looks at inward and outward flows of population for small areas or neighbourhoods in England and Scotland. The Census provides a unique opportunity to study these population dynamics, as it captures information on all people who moved in the year before the Census. It therefore provides data on inward and outward movement of population and hence net flows of people for every neighbourhood in the country.

— Changes in communal provision for adult social care: 1991–2001(Banks, Haynes, Balloch and Hill, 2006) examined shifting trends in care home provision for adults between 1991 and 2001. The report analyses detailed census data from 1991 and 2001 to reveal general trends in adult social care provision. A number of variables are used to understand these trends, including levels of local authority funded places, health and poverty indicators, property prices and provision of other types of social care. This project explored changes in residential and nursing care provision between 1991 and 2001. Using Census data, the researchers were able to look at general trends and also at relative changes in the age, gender and ethnicity of care home residents.

— A particularly influential project that attracted a lot of media attention, both nationally and internationally was on "Migration and social mobility: the life chances of Britain's minority ethnic communities" (Platt, 2005). This project investigated the impact of class background and ethnicity on class position. Drawing on data from the ONS Longitudinal Study (LS), the report traces patterns of intergenerational social mobility for children from different ethnic groups growing up in England and Wales. For the first time, this report measured their progress and class position compared with those of their parents. It provided a unique insight into the differences between the class positions of parents who had migrated to Britain and their children raised in the country. Taking advantage of the new question on religion in the 2001 Census, the report also asks whether patterns of intergenerational mobility vary by religious affiliation and whether religion can add to our understanding of ethnic group differences.

— The relationship between poverty, affluence and area was also analysed using 2001 Census data by Wheeler, Shaw, Richard Mitchell and Dorling. The report used Census data to map the nature and extent of geographical and social inequality in the UK at the beginning of the twenty-first century. Findings included higher numbers of practising, qualified medical practitioners living and working in areas where rates of illness are lower; an "inverse education law" in which areas with the highest proportions of young people with no qualifications tend to have the fewest teachers; well-qualified people living outside London and the South East who accept lower status jobs simply because of where they live; households which might be said to have more cars than they need, while others don't have any at all. The two groups of households tend to live in very different places.

— JRF also publishes annual publications that use information from the Census most notably, Monitoring Poverty and Social Exclusion (JRF's flagship publication) Monitoring Poverty and Social Exclusion is an essential resource for policy-makers and others wanting to take stock of what is happening and understand the challenges ahead. It highlights a number of issues that need to be addressed through government policy: in-work poverty, the number of children/young adults with few/no qualifications, young adult unemployment, health inequalities and low-income households' lack of access to essential services.

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

2.1 Aside from the rich and robust research that is possible from the census a fundamental concern about ending the census is its impact on overall social science data quality. Data quality in general will be greatly reduced. For many of the most valued and robust social surveys in the UK the census is used to select postcode sectors for the main sample and (in the case of surveys such as Understanding Society) the ethnic boost sample. Census information is therefore integral to the design of these surveys. Virtually all sampling in the UK is based to some extent on census information, as it enables researchers to pick a representative set of areas. Consequently, if the census was cut, it would reduce the quality of all social (and commercial) survey designs.

2.2 As well as the value of linked events data to the Census in the ONS Longitudinal Study (LS), it is also used to carry out invaluable longitudinal research on themes such as mortality, fertility, cancer survival, migration, infant mortality, deprivation, social mobility, ethnicity and ageing. It is the only study that has a complete set of census records for individuals, linked between successive censuses, together with data for various events. Thus, the LS represents a continuous sample of the population of England and Wales, rather than a sample taken at one time point only. It now includes records for over 950,000 study members. Fresh LS members enter the study through birth and immigration and existing members leave through death and emigration. In addition to the census records, the individual LS records contain data for events such as deaths, births to sample mothers, emigrations and cancer registrations. This rich data source based on the census has allowed us to undertake prospective analyses of census and event data. For example: studies of associations

between employment status and mortality, between economic status and cancer registrations, and between socio-economic factors and fertility; survival analysis of mortality by area deprivation: prospective analysis of successive census data. For example: analysis of patterns of retirement migration, studies of the effects of divorce and remarriage on housing tenure, comparison of changes in education, employment and migration between the 1970s and 1980s.

2.3 The 2011 Census will be linked into the LS study. The main purpose of this is to link a fifth tranche of census data, laying the foundation for the linkage of a fifth decade of life events data. Information gathered while linking the data will feed into the 2011 Census quality assurance process, identifying areas of possible undercount or overcount. If the 2011 Census is the last of its kind, this clearly has severe implications for the LS as the rich social research that has been carried out using this data source will no longer be possible after 2011.

2.4 Another impact will be the loss of the small area statistics and indeed microdata that enables analyses to be undertaken at the Local Authority level. This level of geography enables mapping, local level tabulations, and even multivariate analyses using local area level variables. These data differ from aggregate outputs such as the Census Area Statistics in that users can look at individual level characteristics within areas and define area level measures of their own choice. Users can define and use subsets, and can create new classifications by grouping existing classifications or combining information from more than one socio-economic characteristic. Microdata from the UK census are a unique source for studying residential mobility, and particularly for providing an insight into ethnic differences in internal migration and the processes of ethnic residential integration. The size of the Sample of Anonymised Records (SARs) and the Controlled Access Microdata Sample (CAMs) enable the migration patterns for relatively small population subgroups, such as minority ethnic groups, to be examined; and to be examined alongside other characteristics of individuals and neighbourhoods.

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

3.1 There are various potential alternatives to the census in providing population and socio-demographic data although each source has its strengths and weaknesses. The new UK Household Longitudinal Study, Understanding Society (USoc), offers the possibility of a much more fine-grained analysis of UK society, with a target sample size of 40,000 households and 100,000 individuals, bigger than any other comparable longitudinal study. The study design includes a significant sample boost for key ethnic minority groups and aims to collect biomedical measures and samples to enable new research on the social determinants and impacts of health in a household context. This will complement data from the English Longitudinal Study of Ageing (ELSA) and from national birth cohort studies. The main problem however with USoc as an alternative to the census, is that it's longitudinal, so over time because of wave on wave attrition (where the number of respondents decreases over time due to loss to follow up, emigrating or death) it becomes less representative of the population. Although it is large and robust, when you get down to small area statistics which census produces, then USoc is not appropriate.

3.2 There are a number of surveys in the UK that allow detailed analyses at a national level to be carried out due to advanced techniques in weighting the data so that population projections can be made. These surveys include:

- The Labour Force Survey (LFS).
- The Longitudinal Study of Young People in England (LSYPE).
- Households Below Average Income (HBAI).
- The Annual Population Survey (APS).
- The Family Expenditure Survey (FES).
- Health Survey for England (HSE).
- Families and Children Study (FCS).
- Growing Up in Scotland (GUS).
- Millennium Cohort Study (MCS).

Although incredibly rich and invaluable in many instances it is not possible to produce small area statistics from these datasets by many indicators due to the size of the samples and confidentiality issues. Another problem for national level analysis is that of missing data where most often the information that is missing is from the key population that we are most interested in. It is of course noted that the detail and information collected in these surveys are invaluable in their own right and that they are all hugely rich data sources.

3.3 The LFS Annual Local Area Data Series, for example, and the Unitary Authority/Local Authority (UA/LA series) have been withdrawn on instruction from ONS, due to confidentiality issues. Although LFS databases released for analysis to social science researchers by outside bodies have always been anonymised to ensure that users could not identify any respondent with the information given, advances in technology and software has made it easier to link survey records to either other survey files or other administrative or commercial databases. "Although the risk for most respondents is very small, there remains a risk of identification for people with unusual combinations of personal circumstances. Thus the release outside the

central government statistical services of social survey databases with small area identifiers, alongside a national database with detailed coding, has now been ceased" (ONS).

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

4.1 A key survey that it may be possible to build on would be The Annual Population Survey (APS). APS represents a major survey which comprises key variables from the Labour Force Survey (LFS), all the LFS boosts and the APS boost. The APS is now able to provide survey data that can produce reliable estimates at local authority level. Key topics in the survey include education, employment, health and ethnicity. There must be potential to build on this survey improving the variety of information captured. The APS combines results from five different sources: the Labour Force Survey (LFS); the English Local Labour Force Survey (LLFS); the Welsh Labour Force Survey (WLFS) and the Scottish Labour Force Survey (SLFS). The APS aims to provide enhanced annual data for England, covering a target sample of at least 510 economically active persons for each Unitary Authority (UA)/Local Authority District (LAD) and at least 450 economically active persons in each Greater London Borough. In combination with local LFS boost samples from Wales (WLFS) and Scotland (SLFS) the survey provides estimates for a range of indicators down to Local Education Authority (LEA) level across the UK.

The position of the Joseph Rowntree Foundation is that the Census should certainly stay due to the evidence collected and summarised above. It will always be the gold standard of data collection and social science research would not be as robust without it in terms of data quality, coverage and analysis. Practical and effective policy interventions based on evidence which is not as robust may be less effective. The analysis that is possible using the actual Census datasets and those combined and linked to others opens up realms of additional research opportunities that would no longer be possible if the Census was no longer carried out.

Joseph Rowntree Foundation

30 November 2011

Written evidence submitted by Professor Leslie Mayhew (Census 39)

THE CENSUS AND SOCIAL SCIENCE

This memorandum is intended as a contribution to the inquiry into the impact of ending the Census on social science research. I consider the issues to be extremely important to the future of social science and set out in this note my views on certain key aspects, particularly the use of alternatives. I would also be very happy to provide oral evidence if invited to attend the hearing.

I am Professor of Statistics at Cass Business School and Director of Mayhew Harper Associates Ltd. I have been working with local administrative data in an academic and private capacity for a number of years both as a user and also as an advisor to local authorities and healthcare providers.

As an academic I have published extensively on demography, health and social care, social security, pensions etc. My use of data has spanned both Census and administrative data and so I am able to compare many different sources and also application areas. I am also a current member of the Office for National Statistics (ONS) National Population Projections Expert Advisory Panel.

I spent much of my early career in the civil service where I was a regular user of different sources of data (Census, surveys, administrative systems) both as professional analyst and then as a senior civil servant in the DHSS and then DSS. I also spent five years as a director in the Central Statistical Office (then part of HM Treasury) and subsequently in the Office for National Statistics itself.

Since leaving the civil service in 1998 to become an academic, I have accumulated 13 years of experience working with administrative data in local authority and PCT environment. My breadth of experience therefore covers national and local administrative data collection systems and surveys of different kinds.

I am a past user but also a strong critic of the Census and of the methodology used. I therefore consider myself to be uniquely positioned to express views particularly on questions 3 and 4 in the inquiry:

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

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

Two papers published in 2011 by myself and with my colleague Gill Harper are highly relevant to this debate and are recommended reading for the Committee:

1. "Applications of Population Counts Based on Administrative Data at Local Level".

   G Harper and L Mayhew (2011). Jnl. Appl. Spatial Analysis DOI 10.1007/s12061–011–9062-z.[1]

[1] http://www.cass.city.ac.uk/__data/assets/pdf_file/0009/82692/Applications_of_population_counts_harper_mayhew_full_text_JASAP_0411.pdf

2. "Using Administrative Data to Count Local Populations" .G. Harper and L. Mayhew (2011). Jnl. Appl. Spatial Analysis DOI 10.1007/s12061–011–9063-y.[2]

A key feature of our approach is that we create comprehensive population data bases by combining different local data sets. These data sets are used by local providers of health, local authority and other services. Applications of the data base are often cross-cutting in policy or research terms and enable useful linkages between sectors of the local economy eg education, health and social care.

The data base produced may be contrasted with information available from the Census:

— The results are timelier than the Census which can take several years to process before being released and be more than 12 years out of date before it is refreshed. Based on the 2011 Census basic population aggregates will not be released before July 2012 and more detailed data will not follow in some cases until much later.

— On Census day this year we obtained and processed snapshots of local administrative data covering the six Olympic Boroughs in London who commissioned the work. Fully geo-referenced databases down to household level by age and sex were completed and handed over to local authorities inside six months.

— The databases contain much greater granularity than is possible using Census data. Each database is able to produce statistics for any size or shape of geographical area or administrative unit and unlike the Census is not constrained by any pre-determined geographical boundaries.

— The data can also be flexed by age so that any group of any age width can be considered. This is important for applications in education such as forecasting school places and in health and social welfare for which eligibility or qualifying ages often do not fit easily into standard age groups adopted by the Census.

— Census data cannot be linked to information on users of different services but our data can. That makes it more valuable to service providers but also to academic users for example to measure inequalities, unmet need, inefficient provision or coverage of services.

Users believe that our data are more accurate than ONS Mid-Year Estimates (MYEs) especially in areas of London with high population influxes and turnover. For example, in the six borough project, we found that our figures were on average 8.3% higher than the ONS 2010 MYEs but these gaps may be smaller outside London. There are no comprehensive independent comparators but our data is very close to the Child Benefit caseload whereas ONS's for the six boroughs were not (12.5% under count).

The core variables we produce are population by age and sex, household type and size, benefit and tenure status, all by geographical location. Ethnicity data are available for children attending state schools but the same information does cover other age groups in most other administrative sources. To fill this gap, we impute ethnicity of individuals and households using probabilistic methods from our extensive data base of names.

By taking different snapshots of administrative data, it is possible to analyse churn and turnover at local levels which is currently a hot topic for research. For specific applications, for example in health, social care, and education, other administrative data sets or variables of interest may be linked together eg attendance at hospital, educational attainment.

The scope of the database is therefore potentially extensive and much wider than the Census can offer. Projects can span healthcare, public health education, social care, youth crime, housing and so on. On the other hand, variables covered by the Census but not adequately covered by locally available administrative sources include: Information about occupation and place of work, self reported health, caring responsibilities and religion.

In our own work, we have been able to fill these gaps to some extent using data from local surveys and linking them to administrative data. A similar approach combining survey information with administrative data could be considered at national level, possibly by using or adapting existing national surveys.

Under our method data are processed in a secure environment by a third party (in this case ourselves). No data are shared between data owners and the final data base is anonymised for statistical purposes only ie no individuals are identifiable. Our approach is cleared by local Caldicott Guardians and by the Information Commissioner.

The cost and timeliness of our approach has much to commend to the Committee. Expanding the system to cover the country would cost roughly one tenth or less of the present cost of the Census, results could be available with six months and the exercise repeated at more regular intervals such as annually. This would benefit both local users and academics whose research would instantly become more accurate, relevant and therefore useful.

There is scope to reduce costs further by introducing standard legally binding data sharing protocols and generally reducing bureaucratic impediments in order to make the process more automated. Better addressing

[2] http://www.cass.city.ac.uk/__data/assets/pdf_file/0019/82711/Using_administrative_data_to_count_populations_harper_mayhew_full_text_JASP_0411.pdf

and general conformance to national database BS standards etc would also help linking efficiency and reduce the need for any manual processing.

I hope this brief analysis is helpful to the Committee. In the time that we have been working on this approach, we have observed steady and substantial improvements to the range and quality of administrative data. In conclusion, I would stress that our approach is practical and there is continual scope for improvement as I hope this memorandum has demonstrated.

The main risks to the system would be if any if the local data sets used were scrapped or there were legislative changes that prevented their use. The government should consider therefore either securing their future access or replacing them in due course with an alternative that does the same or better job.

Professor Leslie Mayhe
Cass Business School, Mayhew Harper Associates Ltd.

November 2011

Written evidence submitted by the Local Government Association (Census 41)

1. Introduction

1.1 The Local Government Association is a voluntary membership body with 422 member authorities England and Wales. Together they represent over 50 million people and spend around £113 billion a year on local services. They include county councils, metropolitan district councils, English unitary authorities, London boroughs and shire district councils, along with fire authorities, police authorities, national park authorities and passenger transport authorities.

1.2 The Local Government Association welcomes this opportunity to offer written evidence to the Science and Technology Committee's inquiry into the Census and Social Science.

2. Key Points

2.1 At a time of public expenditure constraint, the Local Government Association welcomes an examination of alternatives to the Census as timely given both the cost and the weaknesses of a ten yearly approach. However, this examination needs to recognize that reliable, timely and appropriate statistics are key tools that underpin wise, cost effective local spending decisions.

2.2 Accurate local demographic and related statistics are a fundamental enabler for local public services and, to date, local authorities have relied on analysis of the Census as one of the best available sources of such evidence.

2.3 A key priority should be cost effective, better quality and more up to date local statistics that are fit for 21st century approaches to the use of data. Although there are differences of context and requirement, there are lessons to be learned from the way that successful private sector companies do this.

2.4 Abandonment of planning for a traditional census should therefore go hand in hand with the implementation of fit for purpose alternative arrangements. Alternatives will need significant work to offer consistent and reliable data.

2.5 Local government employs a wide range of people under the broad heading of social scientists, many of whom are demographers and researchers providing local evidence to underpin service planning, investment and the engagement of citizens in these processes. It is also, both directly and indirectly, an important customer for wider social research that relies on Census data.

2.6 In welcoming this inquiry and the Office for National Statistics consultation, the Local Government Association believes that they need to trigger technical dialogue to work through options, with local government recognized as one key customer of social science based on data currently supplied through the Census.

3. Detail

How do social scientists use Census data?

3.1 Local authorities currently use analysis of Census data to formulate local policy, to plan and deliver local services, to allocate local resources, and to monitor performance. The Census has also provided the most consistent data for both comparison between areas and for informed and customized analysis of the needs of localities. Census currently:

- Underpins local government financial settlements: crudely, some 75% of local authority funding is centrally funded with allocation based to a significant degree on population.
- Enables socio economic analysis to identify the distribution and geography of specific groups, for example, for school place planning, meeting the needs of the elderly, understanding the extent of single parent households, or meeting local housing needs.

— Provides consistent trend time series data, for example (and despite some weaknesses), it is a key source of travel to work and migration flow data for long term transport and housing strategies.

— Informs local authority work required by statute. Population data is important, for example in Waste, Minerals and Transport Planning, Joint Strategic Needs Assessment, Local Development Frameworks and the Survey of the Area (under existing legislation but proposed to continue in The Town and Country Planning (Local Planning) (England) Regulations [2012]).

3.2 Also:

— Census data has been a key component in commercial products that are often purchased by local authorities to provide local insight for policy and planning; and

— The wide use of census data by the academic research community is also relevant to local authorities, for example where used in longitudinal studies that inform aspects of health policy and social mobility.

— Census results are available for small geographies yet retain consistency nationally and locally, thus enabling examination of the area in question and its context, for example, to identify "hotspots" or spatial patterns. Small area data is a building block that can be configured to illustrate parishes, towns, and district council wards. This is important in supporting localism objectives such as helping neighbourhood forums to draft neighbourhood plans that local authorities will need to adopt.

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

3.3 This depends on what arrangements are put in place and ONS are yet to make firm proposals. However:

— Census has been important in providing information on which the applied social science, that is particularly important to local public services, relied.

— Whilst some Office for National Statistics alternative proposals involve something like 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: a number of our member local authorities are concerned about the potential negative impact if replacement arrangements provide less relevant data and / or data of inferior quality.

— As described in the previous section, local authority use of Census is not statutory, but significant plans and activities that use Census data are statutory requirements. Without alternative data of equal or better quality, these requirements would be less well evidenced and less authoritative.

— Census data is only published every 10 years with data processing typically taking a further one to three years so the results are rapidly out of date. More frequent surveys, or improvement and better use of administrative data on a regular basis would provide more up to date results.

— In addition, such an approach would have potential to make greater and more effective use of public investment in other data sources that fall outside Census, including data on income, taxation, and second jobs.

3.4 If Census were abolished without alternative arrangements and without "read across" or consistency between census data and the new arrangements, there are a number of questions raised that are illustrated by:

— The potential termination of Samples of Anonymised Records (SARs) and longitudinal studies, both of which enable in depth analysis such as that by Office for National Statistics on the "Golden Generation", health inequalities and trends in family and household formation, but available much more widely to the research community and used extensively, for example on work into social mobility; and.

— How to sustain improvements in population measures: this remains a core requirement for public policy and services. We note that Office for National Statistics report a "successful 2011 census" with good response rates.

— Household composition: we are unaware of other datasets that can provide such detailed information on households and their composition. Since family structures are becoming more complex and diverse, retaining a good source of information to understand this issue will be important.

— Data for small areas: a key capability of the Census is to provide detailed information for small areas which is consistent nationally and locally, and with close to 100% coverage and therefore malleable, for example, to create data for neighbourhoods or to respond to administrative boundaries change.

— Commuting and internal migration: Census currently provides information on commuting and migration, and is 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.

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

3.5 This is an area where further work and engagement with local government (and more widely across the public, private and voluntary sectors) is necessary before detailed conclusions are reached. The answer lies in technical assessment of the feasibility, the affordability of options available and any "trade offs" such as whether greater frequency of statistics militates against the ability to cross tabulate the current range of census characteristics against each other.

3.6 This consideration should address factors such as:

- The role of administrative datasets. These are generally designed to meet specific purposes. As a result, there are understandable inconsistencies between such data sources in terms of definitions, coverage and completeness. Some under-record populations (eg the electoral register) and others over-record (eg General Practitioner patient lists). This approach is a potential option, but further work is needed to establish and compare the effort and cost needed to make such data fit for wider purposes; however.
- Administrative systems such as Electoral Registers may have potential for development to improve records completion provided it is affordable. For example, General Practitioner Lists were first floated as an aid to an Improving Migration and Population and are now incorporated into the mid-year population estimates methodology.
- There are a number of experiments and activities that should be evaluated as part of the process. For example, modeling approaches in relation to population projections and small area geographies, or work commissioned by some local authorities that includes Newham, Barking and Dagenham, Luton and Birmingham where they constructed local population estimates by using administrative sources.
- The role of sample surveys, which have their place, but also have limitations and challenges in achieving sufficient and representative response rates.
- 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 only be reproduced by a survey, which if it is to produce reliable results for small areas, will need to have high percentage coverage or be repeated over a number of years to boost coverage. It may be possible to use modeling 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.

3.7 It will be important to learn from international experience, some of which is well documented: Scandinavian countries use a population register. France has a system of rolling censuses and the US combines a short-form census with surveys. None of these seem entirely successful and some, such as population registers, may raise questions about public acceptability.

3.8 We are also aware of the debate about whether and how private sector sources might offer part of the solution. This is a sensible avenue to explore, although at present, the use of central government data looks more promising (eg school census, DWP/HMRC information system, GP patient register). However, these offer a potential means of population count and some basic demographic and employment characteristics rather than the wider range of census variables.

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

3.9 Although not comprehensive, we suggest that the following data could be improved or made more widely accessible:

- Health data on the ageing population and disability would be useful.
- The availability of data on income and taxation.
- Second jobs.
- Broadband take-up.
- International migrants: sample sizes for international migration from the International Passenger Survey are small so it would be helpful to further explore whether e-borders data might be developed. Also, European economic migrants who stay for less than a year are not counted as residents, but because of "churn" this group is often being replaced by other short-term residents, and so this group is never counted.

Local Government Association

December 2011

Correspondence submitted by the Minister for the Cabinet Office, 25 July 2012 (Census 42)

You wrote to the Minister for Science on 20 June in relation to the Census. As the Minister with responsibility for the Census, I am best placed to reply to you.

Population and other socio-demographic statistics are currently drawn from a combination of the ten yearly census supported by regular updates from surveys and administrative systems. In recent years it has become more challenging and expensive to conduct censuses and household surveys, due to the fact that people are more mobile and because of the increasingly complex ways in which people live. Relying on the ten yearly census also means that information quickly becomes out of date, and we are keen to explore alternatives that provide more up to date data. Many other European countries, under similar pressures, have moved away from the use of the traditional census to make greater use of data already held by public bodies.

The Beyond 2011 Programme was established by the UK Statistics Authority in April 2011 in response to concerns expressed by the current and previous Governments about the accuracy and costs of current methods. The programme is considering eight options (including a traditional census approach) and will assess users' requirements and consider the best way of meeting these needs. The outcome will be an evidence-based recommendation which will consider the costs and benefits of each of the options being evaluated.

Any changes to the census will have far reaching implications for central and local government and for other users. All options will be assessed against an agreed set of criteria in order to ensure that they are capable of meeting users' requirements, providing population and socio-demographic statistics of the required quality that are publicly acceptable.

Ultimately, any recommendation from the Beyond 2011 Programme must offer value for money. The costs associated with each option will be considered longside other factors including the ability to meet user requirements and the quality of the expected outputs. While cost is a driver, the real issue is ensuring that the best possible approach is taken, and best use is made of existing public data and developments in technology. In light of the scale and national importance of the census, a full review of the approach is entirely appropriate.

The Government is committed to finding an alternative to the traditional census, and we are looking forward to receiving the Committee's report. We will make sure that the most appropriate Minister responds.

25 July 2012

Printed in the United Kingdom by The Stationery Office Limited
09/2012 017215 19585

ISBN 978-0-215-04848-6