

Social Research Association: Response to the ONS Census consultation 2013

December 2013. www.the-sra.org.uk

From the ONS online questionnaire:

This questionnaire includes the following nine questions/opportunities to comment (you do not have to respond to all of the questions):

1. What are your views of the different census approaches described in the consultation document?
2. Please specify any significant uses of population and housing statistics that we have not already identified.
3. Please specify any significant additional benefits of population and housing statistics that we have not already identified.
4. What would the impact be if the most detailed statistics for very small geographic areas and small population groups were no longer available? High, medium, low or no impact?
 - 4.1. If medium or high impact, please give further information.
5. What would the additional benefit be if more frequent (i.e. annual) statistics about population characteristics were available for areas like local authorities and electoral wards? High, medium, low or no additional benefit?
 - 5.1. If medium or high, please give further information.
6. Please specify any significant uses of census information for historical research that we have not already identified.
7. What advantages or disadvantages for genealogical or historical research can you see from a move to a solution based on archiving administrative sources?
8. What are your views of the risks of each census approach and how they might be managed?
9. Are there any other issues that you believe we should be taking into account?

The SRA has responded to all these questions except for 3, and 6 and 7 which are outside our remit.

1. What are your views of the different census approaches described in the consultation document?

While doubts have been raised about the changes in general there is an acceptance that some development is inevitable. Our concern, therefore, is to minimise the adverse consequences of any changes for the social research community.

On the basis of what we have read and heard in ONS presentations, we strongly prefer option A, a census largely collected online, with face to face follow-up to non-responding households to ensure 100% coverage. We are assuming that if this option is adopted, all cases of online non-response will be followed up thoroughly using alternative data collection modes so that there are comparable response rates to those obtained in 2011. This is essential.

We have read the methodological report by Skinner, Hollis and Murphy (2013) and agree with their view that if this option is adopted, a substantial programme of methodological research should be undertaken to assess whether the increase in online responses relative to 2011 are associated with changes in measurement error and non-response bias.

Based on feedback from our members and from senior research colleagues, we have very strong concerns indeed about proposal B, the re-use of administrative data coupled with compulsory annual surveys. These concerns are both outlined here and then described in our responses to subsequent questions.

- We are concerned that, although ONS's work on deriving population estimates from administrative datasets shows promise, it has not demonstrated that the method will be accurate at all geographical levels.
- ONS's work to date has not yet provided sufficient evidence to convince us that the administrative data option will provide data of comparable quality to that produced by a more conventional Census methodology. We note that the independent review commissioned by ONS (Skinner et al, 2013) concludes that before one could safely conclude that the administrative data option would provide such data further evidence would need to be collected demonstrating that:
 1. The estimation methodology used realistically captures the effects of administrative data over coverage and under-coverage, and dataset matching error; and that it can be unequivocally be shown that these problems do not present serious threats to the quality of population estimates.
 2. The estimation methodology provides sufficiently accurate population estimates for sub-Local Authority geographies (MSOAs and LSOAs).
 3. Quality standards for national level population estimates will be met.
 4. A methodology has been established for checking the validity of population estimates and for producing alternative estimates if these prove to be unacceptable.

In addition there are real risks that a methodology based on administrative data will not be robust over time. Administrative datasets are designed primarily with policy implementation rather than statistical estimation in mind and can change. We can envisage situations where policy changes require major changes to their supporting datasets which run counter to the established data collection requirements. For example, the recent changes to Child Benefit mean that it is no longer a universal benefit for all children and this will have reduced its coverage of new babies and parents of young children. Clearly, had Child Benefit records provided an important source of statistics, these changes would have had profound effects on the trustworthiness of observed trends. Equally changes to benefit rules could have a major impact on statistics based on DWP administrative data.

There are likely to be future situations where policy changes require major changes to their supporting datasets which run counter to the established data collection requirements. Even if legislation is enacted to ensure that datasets are not changed without consultation with ONS, in practice this may be insufficient to mitigate the impact of changes to data (eg where / how they are collected) resulting from wholesale policy changes. Even if ONS were able to adapt its methodology to such changes, residual estimation biases may change in magnitude or direction, thereby muddying the interpretation of trend data. A key issue, therefore, would be the extent to which there is clarity on who owns statistically important data sets. Is it ONS or individual departments? Furthermore, if nobody has central oversight of **all** statistically important administrative datasets there is a real risk of one department altering or withdrawing a data set in a way which adversely affects other users [inside or outside government]. The citizenship survey's demise is a case in point.

Many social researchers working in or for local government or local organisations are very concerned indeed at the loss of information for very small geographies such as Output Areas (OAs). Small area Census data are often used as building blocks to construct bespoke geographies or to identify areas for policy or practice pilots where local level data are needed to inform pilot area selection and supplement the data collected in the course of the pilot. Small area data have been especially important in the study of ethnic minorities both directly and indirectly. Simpson and Jivraj's analysis of the 2011 Census showed that every single ethnic minority group within England and Wales had become more dispersed geographically and so they were able to challenge the UK press' loudly proclaimed view that increased immigration meant that there was increased ethnic polarisation within Britain, and show that they were wrong. This would have been impossible without access to very local data which only the Census provides.

Our members are concerned that decisions on the future of the census should give sufficient weight to likely long term data needs and should avoid making decisions based on the data that is needed at this particular point in history. We acknowledge that this poses challenges, but there is specific concern about the assumptions being made, albeit implicitly, about the (lack of) need for small area data in national level policy making.

It seems highly likely that a future national government will at some point become concerned about (widening) inequality between different neighbourhoods and the nature of disadvantage suffered. The kinds of questions likely to be posed include, for example: how wide is the gap between rich and poor areas? is there a clustering or dispersion of deprived areas? is there a tipping point at which areas go into a spiral of decline? Answering these and other questions will require an accurate picture of multiple deprivation at the small area level for the whole country, using a common measurement yardstick. This is also particularly important given the medium / long term effects of the Localism Act and the potential growth of Neighbourhood Forums with key inputs into local planning issues.

A critical challenge in constructing a reliable picture is the need for accurate population denominators for small areas. We do not see how this could be done in a sufficiently robust way without the census. Even though denominators would need to be updated between censuses, the census provides a critical 10 year check to ensure that population estimates are not becoming inaccurate. We recognize that it may well be that some local areas are capable of constructing their own denominators using local knowledge, but a key point is that it would not be possible to construct a national or even regional picture of the situation across different small areas using locally generated knowledge.

We are aware of on-going work on options for updating the indices of multiple deprivation but we feel that ONS need to give sufficient weight to the overall point made above that such data is very likely to become far more important in underpinning policy decisions, including decisions on the targeting of public expenditure.

The Census is also crucial in the design of samples for surveys of ethnic minority groups.

We can, however, see the potential benefits in extending the use of admin data to supplement census data in-between 10 year collections. Therefore we support the further testing of administrative data with a view to using it to *enhance* information collected through a (largely) online Census but not to replace it. We would like to press for significant efforts be devoted to **improving the quality of admin data** which we know to be often incomplete and full of inaccuracies. Staff, who collect admin data, will need to understand the importance of accuracy and have adequate time to do careful data entry. ONS will need to be able to check for the integrity and completeness of admin data used for Census purposes and take steps to ensure it is fit for purpose.

2. Please specify any significant uses of population and housing statistics that we have not already identified.

Small area statistics are used extensively in policy evaluation and especially policy pilots, where local level data are needed to inform pilot area selection and supplement the data collected in

the course of the pilot. Many areas of policy focus have a local dimension – poverty, ethnicity, unemployment and related to these health, education, crime, and housing, to name a few.

We are also aware that small area census data is used as a matter of course by survey research organisations in designing efficient sampling processes for large-scale government surveys. This is particularly an issue when trying to get representative samples of ethnic minorities: without small area information on the distribution of ethnic minority groups the issued sample for these surveys would need to be considerably larger and more expensive, to ensure that sufficient numbers of ethnic groups were located. The PSI Fourth National Survey of Ethnic Minorities could not have been implemented as effectively as it was if detailed census data were not available for sampling purposes. Similarly, for the Muslim boost for the Citizenship Survey, by using OAs as PSUs it was possible to achieve higher screen eligibility rates for a given level of population coverage than would have been possible using larger geographic units (eg. wards) thereby making the survey more affordable. It is not clear if even ward / MSOA data would be available for ethnic groups under the proposal B? If not, sampling ethnic minority groups cost-effectively would become difficult or even impossible.

Following consultation with our members, we are aware that there is an 'ecosystem' of small-scale local voluntary groups who use small area census data on population and housing characteristics to identify specific areas of inequality and hardship, as a focus for their charitable and other activities. These groups depend on census information but may not be aware of the ONS consultation. Many of their funders require socio/economic and demographic details of the locality for which funding for local projects is sought. Only the Census provides this and it will not be possible for these sorts of groups to collect this themselves. The Coalitions' Localism Act which makes possible the setting up of Neighbourhood Forums is leading to / requiring these Forums to prepare detailed socio- economic/demographic descriptions of the geographic area of the forum- much of the key information is Census derived.

The situation is even more acute for Local Authorities in identifying areas of need and deprivation notably for funding applications for both EU and national UK funding. Furthermore, as increasing powers are passed down from central to local government, their need for relevant local data will increase just as their ability to collect or pay for it will reduce because of cuts in analytical budgets and personnel.

4. What would the impact be if the most detailed statistics for very small geographic areas and small population groups were no longer available? High, medium, low or no impact?

High.

4.1. If medium or high impact, please give further information.

The reasons are given in our response to Q2.

5. What would the additional benefit be if more frequent (i.e. annual) statistics about population characteristics were available for areas like local authorities and electoral wards? High, medium, low or no additional benefit?

Medium.

5.1. If medium or high, please give further information.

We can certainly see the benefit of more frequent statistics at local authority and electoral ward level. However, in the light of our concerns about basing population statistics on admin data, we are firmly of the view that this benefit would in no way be sufficient to justify the wholesale *replacement* of Census data with administrative / survey data.

8. What are your views of the risks of each census approach and how they might be managed?

The risks of proposal A, a ten-year census mostly collected online, seem minimal, as (i) the approach (including online) was partly used in 2011 and (ii) the transition from paper-based to primarily online data collection has been successfully made in other countries.

The risks of proposal B, using admin data plus compulsory surveys, are currently very much greater, and our concerns are based on the fundamental issue that government admin data is not collected with national statistics in mind. For this and other reasons there is a strong risk that admin data is too unstable and unreliable to become the main basis of population statistics.

Government admin databases are prone to change, mainly in response to shifts in policy. For example, the Child Benefit database will have changed fundamentally after it became no longer universally available. Other changes to state benefit rules will produce similar changes to data held, as will the complete removal and introduction of benefits.

If national statistics are based on admin data, these shifts in the data collected would call into doubt the accuracy of changes in statistics over time, as to whether these are real effects or a result of changes in the data available.

Over time, again due to policy changes, existing databases may no longer be available. For example the patients' database may not continue to exist under health service privatisation.

We note that the National Statistician must be consulted on the data implications of fundamental changes such as privatisation. But this commitment to consult does not guarantee that such changes won't be made, whatever the concerns raised.

The proposed use of admin data heavily depends on the assumption that such datasets continue to be available and can be pooled and cross-referenced by ONS. This cannot be guaranteed, however, and the increasing level of concern over the need to protect individual data may lead to unforeseen restrictions on their use. For example, the EU is currently

reviewing potential legislation that could dramatically restrict the amount of data sharing by governments.

Government departments in the past have often regarded admin data as belonging to the department and have been reluctant to pass it on. And legally it may not be clear who 'owns' the data, in which case there is a risk that individual departments may withdraw consent for ONS to use their data, or may amend it in ways that make it less useful.

Our other main concern is over the validity of admin data. Social researchers have a unique perspective on this, because in conducting research for government departments they frequently use admin data to draw a sample of respondents. It is a commonplace among researchers that the admin data they are given contains many inaccuracies. Address information may be years out of date or simply wrong; names, ages and phone numbers are incorrect, household members are missing, and so on. There are many possible reasons – for example if benefits are paid into the recipient's bank rather than by cheque, there is no incentive to keep address records up to date – but no doubt some inaccurate information is provided by citizens, as well as recorded by officials. But this doesn't change the nature of the problem – one that many researchers are very familiar with, and habitually allow for by over-sampling. Therefore we strongly recommend that before ONS goes ahead with any serious use of admin data, they first carry out rigorous checks on the integrity and completeness of large samples of the data they plan to use.

The ONS consultation document notes that 'without exception', countries that use admin data to supplement national statistics, rather than a census, also have a compulsory population register. We can see that such a register would be an essential record for checking admin data and building accurate statistics. In effect, it is like a continuous census of basic-level data. A population register appears to have been the basic building block of all administrative data based Census systems developed to-date. However, it is widely believed that there is no appetite for introducing such a register in England and Wales. Without it, there is therefore a considerable risk that admin data – even supplemented by annual compulsory surveys – will not be sufficient for the vital purpose of providing valid and reliable, local and national population statistics.

9. Are there any other issues that you believe we should be taking into account?

We are doubtful that Parliament would agree to making compulsory the public's participation in a very large (4%) annual survey. If, as noted above, there is no political appetite for a compulsory population register, it seems unlikely that a large compulsory survey would be any more acceptable. But without compulsion, the survey would unquestionably suffer from non-response bias and there would be a considerable risk that for some measures this would be substantial.