United Nations International Seminar
on Population and Housing Censuses:
Beyond the 2010 Round
27-29 November 2012
Seoul, Republic of Korea

SESSION 5: Future of Population and Housing Censuses (Roundtable discussion)

BEYOND 2011:
THE FUTURE OF POPULATON STATISTICS IN THE UK

Prepared by

Mr. Ian White
The Office for National Statistics (UK)

* This document is being reproduced without formal editing.

Beyond 2011: The future of Population Statistics in the UK?

The Office for National Statistics (UK)

Summary

The Office for National Statistics (ONS) is currently considering options for the future production of small area population and small area socio-demographic statistics in England and Wales. The *Beyond 2011* programme has been set-up to take a fresh look at approaches to meet future user needs for 'Census-type' data.

While innovative ways of carrying out a traditional Census in 2021[1] will be assessed most of the work is focusing on making better use of existing administrative registers and other data sources. These data sources are improving and maturing, and the *Beyond 2011* programme aims to assess whether, properly combined, such sources will be able to provide a viable and cost effective means of generating the required census statistics in future. A full understanding of user needs, quality, costs and public acceptability will all play a part in informing the recommendation on the way forward which will have to be made by 2014 in order to carry out the next census in 2021.

Any recommendations from the *Beyond 2011* work will have implications for all population based statistics in the United Kingdom (UK) and perhaps, in the longer term, for the statistical system as a whole. This paper describes progress to date in England and Wales and future plans. Parallel developments in Scotland and Northern Ireland are also described briefly.

Background

The 210 year-old Census is a cornerstone of the UK statistical system and provides a wealth of consistent, comparable data for small areas and for a range of topics from the local to the national level. It provides a key source upon which other statistics are built and a benchmark against which many are judged.

The most recent Census in England and Wales (in March 2011) was successful. A response rate of 94 per cent was achieved nationally with over 80 per cent in every local authority (LA) – a clear improvement on the response in 2001. Across London, the estimated response rate in most boroughs was between 5 and 15 percentage points better than in 2001. An option for online completion was introduced for the first time, and around 16 per cent of the 23 million census returns were completed this way.

Nevertheless, the traditional approach to census taking is becoming increasingly costly, and changes in society are making it more challenging to carry out. A more mobile population and the increasingly complex ways in which people live make the process of taking a Census more difficult – and the concept of a snapshot every 10 years is, arguably, becoming less relevant. At the same time improvements in technology and the growth of computerised registers and other records about people and services (referred to here as 'administrative sources') would seem to suggest an alternative approach – we should pull together what we already know about the population from these records and only supplement this where there are gaps.

In May 2010, Sir Michael Scholar, Chair of the UK Statistics Authority (UKSA) – which oversees statistical matters in the UK - 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."*

[1] A traditional Census here is defined as the self-completion of a Census form (online or on paper) by every household and person in the country at the same time.

ONS has established the ***Beyond 2011*** programme to assess user need and investigate options for producing the population and socio-demographic data required in England and Wales. Good progress has been made on establishing the programme and the list of options which are to be tested and assessed has been agreed.

The Statistical Options

In summary, ONS is, at the current stage of the programme, considering eight options, falling into three broad groups:

CENSUS TYPE OPTIONS

Option 1: Traditional Census

This option is a continuation of the current approach, that is be conducting a full field enumeration of the population at a single point in time (most recently, in March 2011). Everyone completes a 'long form' questionnaire and the responses alone 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 since 1801 (with the exception of 1941). Between censuses the results are supplemented by a series of mid-year population estimates which use the 'cohort component method' to take account of births, deaths and net migration. Social surveys on particular topics are used to fill in details on particular socio-demographic characteristics, and update estimates as required in the inter-censal period.

Option 2: Long/short form Census (long form *(to a sample of the population)* ***and short form*** *(to everyone else)*

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 was used in the USA until 2000.

Option 3: Short form Census plus Continuous Survey

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, that is a questionnaire covering some basic topics only. Population estimates would use the existing cohort component method in the inter-censal period. Socio-demographic statistics would then be collected through a large scale continuous survey. A similar approach is now in use in the USA, with the 'American Community Survey' replacing the long form census element.

Option 4: Rolling Census

In this case, a census is conducted in different areas throughout the country on a rolling basis. For example, 10 per cent 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 been commissioned into the suitability of a rolling census approach in England and Wales.

SURVEY OPTIONS

Option 5: Address register plus survey

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 OPTIONS

Option 6: Administrative data – aggregate data only

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.

Option 7: Administrative data – unit record level

In this option unit 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 (for example, dual system estimation, ratio estimation) would then 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.

Option 8: Administrative data – intermediate approach

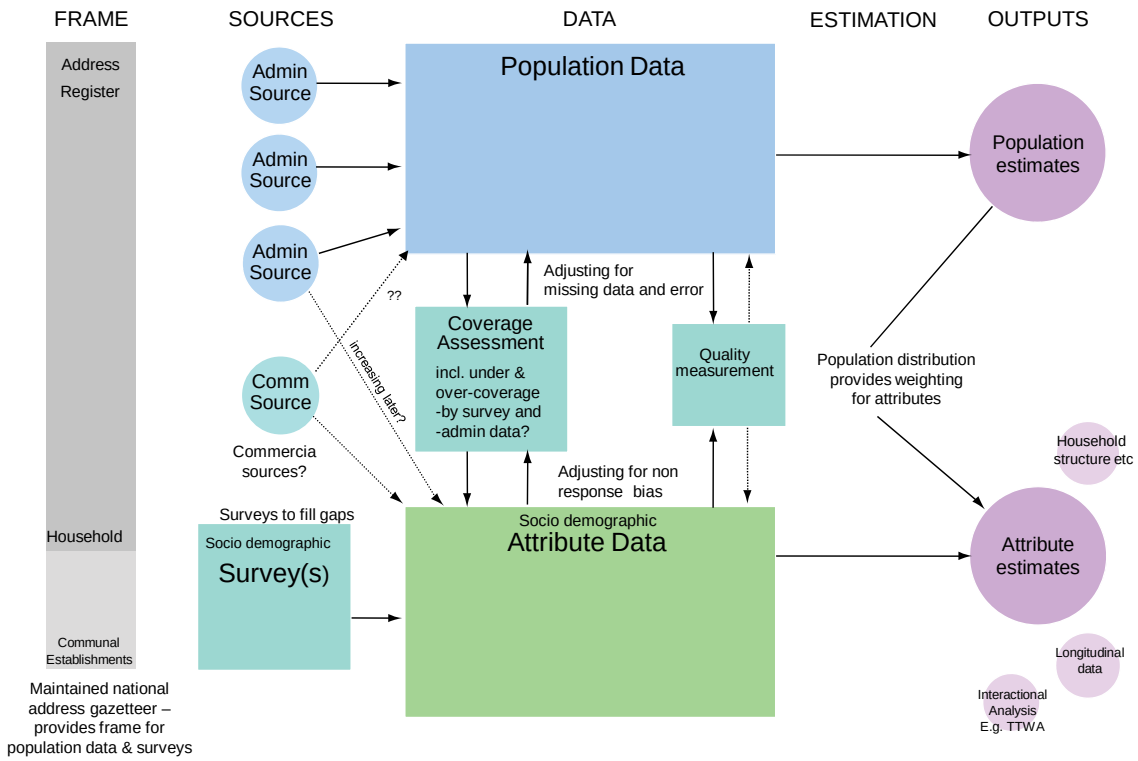
This option would make use of both aggregate and unit record level administrative data approaches. The aggregate administrative data approach would be used to produce initial population counts. Data from a coverage check survey, measuring both under-coverage and over-coverage, would be linked to record level administrative data sources and the address register, in the coverage check areas only. An estimation process (dual system estimation, ratio estimation) would, as in option 7, 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.

Almost all of these options will use an **address register** either as a source or as a frame for surveys. Furthermore, these descriptions have been simplified in the interest of brevity. It is very possible that the solution adopted may use elements of more than one of these.

In practice, as the census and survey options are well understood already, research is focusing primarily on the administrative data options. Figure 1 provides a summary view of how administrative sources and surveys might be used in combination to create population and attribute estimates.

The programme is being organised around six-monthly cycles of research, with the options re-assessed and the business case updated at the culmination of each cycle. This will involve the production of biannual research reports setting out what has been learned about the options, data sources and modelling methods. There will be five such 'research cycles', with a final recommendation being made in September 2014.

Figure 1 : Beyond 2011 – Administrative sources options - Overview



Potential data sources

A number of administrative data sources have already been identified as possible alternative sources for estimating the size of the population by age and sex. These include:

- the National Health Service (NHS) Patient Register;
- the National Insurance system;
- the Electoral Registers
- the Census of Schoolchildren (an annual canvas of children attending state maintained schools); and
- information on students attending universities and other higher education establishments.

ONS has legal access to all this data.

But the challenges associated with using such administrative data will mean that a coverage survey (similar to that conducted and used alongside the 2011 Census) will be needed to adjust for coverage error. In a full census, coverage adjustment is primarily about controlling for under-coverage, while over-coverage is a relatively minor issue. But with data from such sources as health and national insurance systems the reverse often the case, and the information needed to adjust for coverage problems is correspondingly different. This is for a variety of reasons, but one key issue is that when emigrating there is no requirements to de-register from such systems. (For further information see the ONS paper *Beyond 2011: Exploring the Challenges of Using Administrative Data*, ONS Beyond 2011 Methodology Paper M2, 2012- at http://www.ons.gov.uk/ons/about-ons/what-we-do/programmes---projects/beyond-2011/news/reports-and-publications/index.html)

Criteria for assessing the statistical options

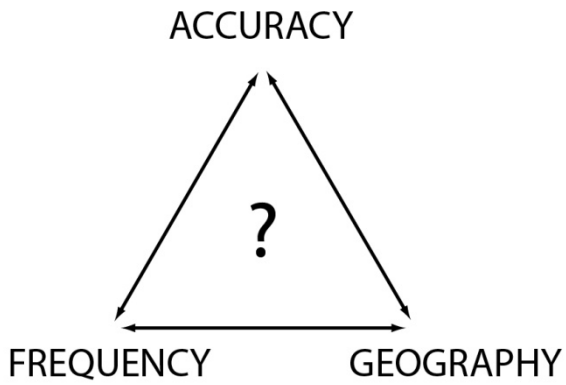
The options will be iteratively assessed against a set of criteria including:

- Fitness for purpose – that is, the ability to meet user requirements including:
  - accuracy of the statistics produced
  - frequency of the outputs, for example whether updated annually, or every 5 or 10 years
  - geographic level at which outputs can be produced
  - consistency and comparability of the outputs across geographic areas
- Cost (both total cost and cost profile over time)
- Technical and legal feasibility
- Risk
- Public burden and acceptability.

Scoring the options against these criteria will help to determine our final recommendation.

The key trade-off: accuracy versus frequency versus geography

Our final recommendation will need to balance user needs against the other criteria above, but it is inevitable that there are going to be some compromises in the solution adopted. For any given topic and approach a more frequent supply of outputs may require us to compromise on a higher level of geography or lower quality – or vice versa. The requirement will vary by topic. Outputs for small areas every five years might be sufficient for some topics, while high level outputs every year might be a better solution for others.



The different statistical options that we are investigating have different strengths and weaknesses, and we are consulting widely to help us understand users' priorities on this difficult three-way trade-off.

Costs

The cost of the research phase of the programme is £25m over four years (2011/12 to 2014/15).

The cost of implementation of any solution recommended is currently unknown but there is clearly potential for the overall costs, and the cost profiles, of the different options to be substantially different. One possible relationship between the traditional census (Option 1) and the administrative data options (Options 6,7,8) is illustrated in Figure 2.

Key questions remain as to the height and shape of the peak in set-up costs of the administrative data options and the ongoing running costs. A key output of our research will be a better understanding of the total costs and cost profiles for the different options.

Figure 2 Potential (real term) cost profiles of traditional census and administrative data options

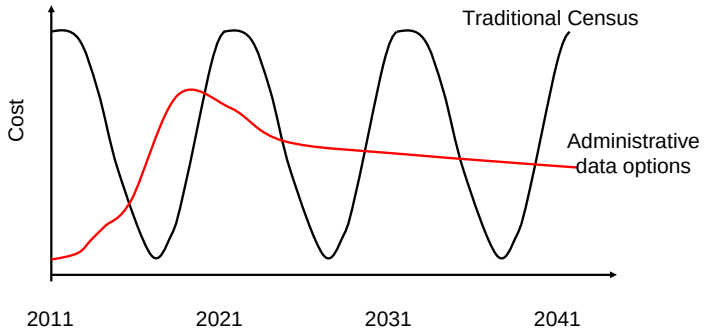
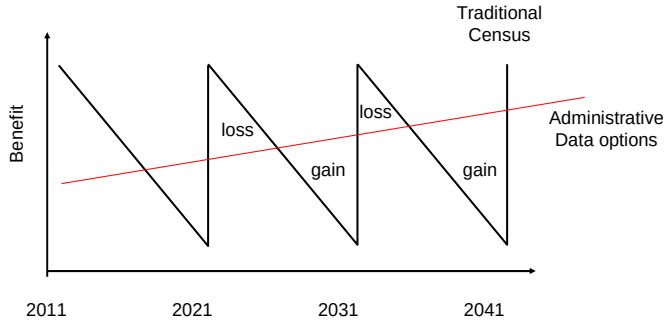


Figure 3: Potential benefit profiles of census and administrative data options



Benefits: long-term statistical benefit profiles

The Census provides a benchmark for the current UK system of population and socio-demographic statistics[2] and many statistical series are re-based every ten years to take account of census results.

However, the Census has a number of disadvantages - not least the fact that its detailed outputs are updated so infrequently (every ten years). This means that the accuracy and relevance of the data declines over the 'inter-censal' period. Moreover, due to the scale and complexity of census operations and quality assurance processes, there is a delay of approximately 18 months before detailed outputs are published. This can be a significant issue in areas experiencing rapid population change, or when the importance of a particular socio-demographic topic changes in response to new or emerging Government policies and priorities.

The benefits from a traditional census (Option 1) are realised in a cycle over a decade. At the point when the statistics are first published, they are highly accurate but as the population changes through births, deaths and migration for example, they provide a less relevant picture of the population – and so less benefit.

An administrative data model based on registers might never provide such an accurate or detailed snapshot of the population at a single point in time, but it does have the potential to provide more frequent updates, reducing the decline in accuracy over the decade – though the maintenance and updating of the registers themselves presents separate quality and relevancy issues.

[2] Separate censuses are run in England and Wales, Scotland and Northern Ireland. Together these provide a complete picture of the UK.

It may be, therefore, that the 'average' benefit derived over a decade is comparable to that of a traditional census, as shown in Figure 3. Clearly, as with Figure 2, this is only illustrative; the precise location of the profile is not known at present. Understanding the benefit profile for each option over the full 10-year cycle is a key part of the work being undertaken.

Stakeholder engagement

Effective and extensive engagement with stakeholders sits right at the core of the *Beyond 2011* programme.

The programme has a wide and complex set of stakeholders across all sectors of government, and society – not least established users of census statistics. A number of highly influential stakeholders and stakeholder groups have strong, potentially divergent, views on how the work of the programme should be taken forward. These views need to be fully understood, balanced against each other and fed into our approach to the assessment of options.

Ensuring that the needs and priorities of users are fully reflected in the solutions developed and in the final recommendation is also essential. A solution that is based on a poor understanding of what users need, both now and in coming years, is at risk of failing to deliver an effective approach – but also of missing a huge opportunity to re-think the way statistics are produced.

It is essential that key stakeholders understand that there are a range of views that need to be balanced and are assured that we are taking the right approach.

The opportunity to review the existing approach for the production of population and small area socio-demographic statistics also brings with it risks relating to management of expectations. Whilst the programme will be fully open to new ideas, and will be looking for ways to extend outputs beyond those currently available, we again need to communicate clearly that any new approach will take time to implement, and will, at least initially, have limitations.

Public consultation

There are two major public consultation exercises. The first was run from October 2011 to January 2012 giving users (both established and new) an opportunity to tell us about changes to their information requirements and priorities in order to inform our scoring and weighting of the evaluation criteria. The consultation highlighted the complex nature of users' requirements for population and socio-demographic statistics. Despite this complexity, it was possible to draw some broad conclusions from the analysis of the responses:

- Users welcomed the opportunity to discuss their needs and supported the principle of a review to look at new requirements.

- In the absence of a proven alternative, there remains strong support for a traditional census. In addition, there is concern about losing consistent, accurate, multivariate data for small areas because users see this as the key benefit of a census. Also, genealogists and socio-historians are concerned about the potential loss of the census as a historical record.

- Overall, data requirements have increased. In particular, there is a requirement for high quality data to underpin local decision making

- There is a requirement for more responsive statistics about new and emerging topics.

- Users would like data which is more timely and available on a more regular basis than is currently available from the traditional census-based system. However, although the importance of this does vary by topic, most users are not prepared to trade accuracy or small area geography for frequency. There is a clear need for the *Beyond 2011* programme to focus closely on the balance between frequency and accuracy, and to understand how and where high accuracy adds financial benefit.

- For many users and topics there appears to be a dual requirement: frequently updated estimates at LA level, combined with a less frequent update (for example, every 5 years) at a lower level of geography.

This initial consultation exercise is being followed up by an ongoing series of more detailed and focused discussions with users, data suppliers, partners in the devolved administrations (Wales, Scotland and Northern Ireland) and other key stakeholders.

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 recommendations.

Public opinion and the acceptability (or otherwise) of alternative solutions will play a key part in our assessment and we are running surveys and focus groups to test public opinion as well as engaging with groups with a special interest in privacy and confidentiality.

Developments in the Devolved Administrations (Wales, Scotland and Northern Ireland)

The *Beyond 2011* work described above covers England and Wales and the Government in **Wales** is working closely with ONS to review the options being considered in order to ensure that Welsh interests and requirements are properly reflected in any proposed way forward.

In **Scotland**, alternatives to the census are being considered as part of a much broader programme of work across the Scottish Government aimed at enhancing the whole national statistical system. Plans include combining several sources of data to trial a population spine and modelling for small area estimates. The quality of the data will be assessed against the 2011 Census. An investigation is also taking place into the potential use of new technologies to enhance survey and census based models.

A detailed review of the 2011 Census in **Northern Ireland** will begin after the initial census results are published. This review will include the examination of options for the future provision of socio-demographic statistics in Northern Ireland. Any proposals stemming from this work will be subject to consultation.

Timetable

The first consultation, and supporting workshops, to capture user needs and opinion, ran from November 2011 to January 2012. A report on the outcomes of this consultation is available on our website (http://www.ons.gov.uk/ons/about-ons/what-we-do/programmes---projects/beyond-2011/index.html)

ONS is testing and prototype statistical options during 2012 and 2013 – and this will be supported by an ongoing dialogue with users. The second public consultation currently planned for Spring 2013 will provide more detail on the lead options and their relative benefits.

The *Beyond 2011* programme will provide a final costed recommendation in 2014.

Further inforamtion

Further information on the Options and the *Beyond 2011* Programme as a whole is available on our website - http://www.ons.gov.uk/ons/about-ons/what-we-do/programmes---projects/beyond-2011/index.html .

ONS very keen to share experience internationally and build on best practice developed elsewhere. To discuss further or for more details on any aspect of this work please contact us at: **beyond2011@ons.gov.uk**