

The census and future provision of population statistics in England and Wales

Public consultation

September 2013

Contents

Summary	2
1. Introduction	3
2. Possible approaches for the future	6
3. Advantages and disadvantages	8
4. Managing the risks	12
5. Consultation questions	15
Appendix	
Statistics available from different approaches	17

Other reports accompanying this document are:

C2 – [Summary of the uses of census information, September 2013](#)

C3 – [Summary of the benefits of census information, September 2013.](#)

For more information

Go to www.ons.gov.uk and search on Beyond 2011

Email us at beyond2011@ons.gov.uk

Summary

Every 10 years, for over 200 years, every household in England and Wales has been required to respond to the census.

The 2011 Census successfully provided population statistics that will be used for the next decade by planners, policy makers and researchers across the public and private sectors.

Our population is changing rapidly, and the need to understand these changes will continue. The Beyond 2011 Programme in the Office for National Statistics (ONS) is currently reviewing these needs, and how they might best be met in future.

Improvements in technology and in government data sources offer opportunities to either modernise the existing census process, or to develop an alternative census method that reuses existing data already held within government.

Our research has resulted in two approaches for census taking in future:

- once a decade, like that conducted in 2011, but primarily online.
- using existing government data and compulsory annual surveys.

Both approaches would provide statistics about the size of the population, nationally and for local authorities. A census using existing data and surveys would provide more statistics about the characteristics of the population every year, while an online census would provide more detailed statistics once a decade.

This consultation document describes these approaches, their strengths and weaknesses and the different types of information they could provide.

No decision has yet been made, and we welcome your views.

Please read this document fully and then respond using the [online questionnaire](#) by **13 December 2013**.

We will publish our findings in 2014.

For further information, please contact:

Beyond 2011 Programme
Office for National Statistics
Email: Beyond2011@ons.gov.uk

1. Introduction

1.1 The purpose of this consultation

After each census, the Office for National Statistics (ONS) reviews the future needs for information about the population and housing in England and Wales, and how these needs might be met.

In 2003 we published proposals for a future population statistics system making greater use of existing data, for example the list of people registered with a doctor, paying tax, receiving benefits, or registered to vote. Our conclusion at the time was that there was significant potential for the use of such data, described in this document as 'administrative data', but ONS did not have access to enough data sources to develop approaches further.

The Statistics and Registration Service Act 2007 enabled ONS to access further key data sources and, after the 2011 Census, the Beyond 2011 Programme was established to consider the needs and options for the next decade and beyond.

Our research used the 2011 Census information as a baseline in order to understand whether different sets of administrative data could be used to produce reliable statistics in future. The findings to date indicate that there is potential to conduct a census in future using administrative data and annual sample surveys, instead of asking every household to complete a questionnaire once a decade.

This kind of approach could enable the increasingly rapid change in society to be measured more frequently than our current system. However, it could not provide as much detail about the population and housing as the census has provided in the past, and we are also considering a census more like the one conducted in 2011 – although data would primarily be collected online rather than on paper.

Census approaches using administrative data are becoming more common around the world. Scandinavian countries, the Netherlands, Switzerland, Austria, Israel and Germany have moved to systems based primarily on administrative data. Without exception, these countries have well established population registers, and all residents are legally required to keep their information up to date. This is not the case in the UK, so the available information is less accurate, and is not consistent between sources.

Other countries, such as the USA, Canada, Australia and New Zealand, continue to use more traditional census approaches.

Moving to a census method using administrative data and surveys would be the most radical change in the production of population statistics in England and Wales since 1801. ONS is therefore consulting on the possible approaches, and this document sets out the advantages and disadvantages of each. Everyone who has an interest is invited to contribute to the discussion.

1.2 The value of census information

The census has been the bedrock of the UK statistical system for the past 200 years. It provides a wealth of consistent, comparable information from the national level down to very small areas and for small population groups.

Population and housing statistics underpin the allocation of billions of pounds of public spending each year. They are fundamental to decision making, policy formulation and outcome monitoring across government, the commercial sector and the wider community.

They also enable the UK to fulfil international obligations, determine voting rights for the European Union and are the basis for allocation of European regional aid.

Annual population estimates are produced every year, based on the previous census, with adjustment using information on births, deaths and migration. These form the basis for many other important statistics, such as mortality rates, fertility rates, unemployment rates and road traffic accident rates.

Annual household surveys use census information to adjust for bias in responses to the survey. Results from the Labour Force Survey, for example, which inform monetary and fiscal policy, would otherwise be less reliable. Such surveys provide information annually in between censuses, but can provide information only at national or regional levels.

While all of these uses and requirements are certain to continue, the increasing rate of change in society means that there is demand for statistics on a more frequent basis. A census using administrative data and large annual surveys, rather than asking every household to fill in a questionnaire once a decade, would provide more frequent statistics, but each approach has advantages and disadvantages, as described in section 3.

1.3 The census as a source for historic and genealogical research

Many users of census information make use of the individual records released after 100 years from historic censuses to support genealogical or social history research.

Clearly the world will be a very different place by the time that records from a 2021 census are released, and the information environment unimaginably different. At this point in time it is impossible to know how census information would be used, but it is clear that a set of records about the population would be of value.

ONS's role is to produce statistics rather than register individuals or households, but we are keenly aware of these uses of census information and have been discussing options with The National Archives, the UK Government's official archive.

It is not yet clear what would be possible if we were to move to a census method using administrative data, but we are interested in views on what opportunities there might be, as well as the risks from a change of approach.

1.4 The decision making process

Census legislation *allows* a census to take place, but does not *require* a census; each census requires Government to come to Parliament with proposals, which are set out in secondary legislation. The content and conduct of the census is therefore for the Government to propose to Parliament, taking account of the advice of the National Statistician and the UK Statistics Authority.

So ultimately it will be for the Government and Parliament to agree the arrangements for conducting a census in future. A census method based on administrative data and surveys would require legislation to enable easier data access; to make it a legal requirement for households to respond to any new surveys; and to determine how to create historic records for genealogical purposes.

This consultation plays a critical part in ensuring that the advantages and disadvantages of the different approaches are fully understood. The findings of this research, including the results of this consultation, will be published and presented to Government in 2014.

Note that the questions to be included in an online census once a decade, or in the surveys that would be required for a method using administrative data, will be decided separately. We will consult on these as we have done for previous censuses. For planning purposes we are assuming that questions similar to the 2011 Census would be used for both approaches.

Whatever the outcome of this review, the questions asked will be decided based on what adds most value for users overall, while maintaining quality, maximising continuity with existing statistics, and minimising cost and public burden.

1.5 The UK context

ONS is responsible for population statistics for England and Wales, and existing census legislation gives the National Assembly for Wales a formal role in the approval process for census.

The Welsh Government is very interested in the views of Welsh users about the different approaches and any consultation responses will be forwarded to the Welsh Government.

The Scottish and Northern Ireland Governments have responsibility for their own censuses and are undertaking their own research. Close collaboration is in place with all of the devolved administrations in order to manage the consistency of UK statistics.

2. Possible approaches for the future

A number of approaches have been considered by the Beyond 2011 programme, and [each has been assessed against a published set of evaluation criteria](#) including statistical quality, cost, technical and legal feasibility, public acceptability and public burden.

This has resulted in two potential approaches to taking the census in future:

- A census once a decade, like that conducted in 2011, but primarily online.
- A census based on administrative data and large annual compulsory surveys.

These approaches are summarised in Boxes A and B. The statistics each method would provide are described in more detail in the Appendix.

Box A: An online census once a decade

What it is

Collection of information from all households in England and Wales once a decade.

- The majority of responses would be online and it would be compulsory for households to respond.
- As in 2001 and 2011, a survey covering around 1% of the population would be used to adjust for those who did not respond.
- The census would also cover residents of 'communal establishments' such as university halls of residence and military bases.
- Administrative data would be used to help check the quality of the final statistics.
- Annual population estimates would be produced in much the same way as at present, using information about births and deaths, and information about migration using the International Passenger Survey and administrative data. The role of administrative data would continue to increase over time.

What it provides

- This method would provide a range of outputs every 10 years on a wide range of topics, similar to those produced from the 2011 Census.
- Outputs would be produced for a range of geographic area types, down to very small areas (groups of a few postcodes called 'Output Areas') and for small population groups.
- Detailed cross tabulations would be produced for a wide range of topics, with accuracy similar to the 2011 Census outputs.

Box B: A census using administrative data and surveys

What it is

Re-use of administrative data, with annual compulsory household surveys.

- Administrative data already held by government would be used by ONS to produce an annual estimate of the population in local areas.
- Sources currently being researched include those held by the Department of Health, the Department for Work and Pensions, HM Revenue and Customs, the Department for Education, the Higher Education Statistics Agency, NHS Wales and the Welsh Government.
- None of these sources provides a completely accurate count of the population so a compulsory annual survey of around 1% of households would be used to adjust for those who are not included in administrative data, or are recorded in the wrong location.
- Administrative data cannot provide sufficient information about the characteristics of the population that have previously been provided by the census such as ethnicity, or languages spoken, so a second compulsory survey of around 4% of households per year would be required.
- The survey questionnaire could be similar to that used for the 2011 Census, but with scope for more regular updating. The majority of responses would be collected online.
- A separate compulsory annual survey would be run to cover residents of 'communal establishments' such as university halls of residence and military bases.

What it provides

- This method would provide statistics every year on key aspects of the population and housing stock.
- The majority of statistics about the characteristics of the population in each Local Authority would be produced using one year's survey data.
- Once the survey had been running for a few years, statistics for smaller areas within local authorities, such as electoral wards, would be produced annually, by combining three years' or five years' data from the survey.
- A limited set of basic estimates by age and sex would be available for small groups of postcodes, but the most detailed statistics for Output Areas that the census has provided in the past would not be available using this method.

3. Advantages and disadvantages

The two census approaches would produce different sets of statistics, and would have different advantages and disadvantages. Our assessment of these is described below.

An online census once a decade

Strengths

The census delivers a **rich set of statistics for a range of geographic areas** – nationally, for Local Authorities, for wards and groups of a few postcodes (typically around 125 households). It provides detailed cross- tabulations for a wide range of topics.

Because all households are required to complete a questionnaire, the census has a **proven ability to deliver detailed statistics for small geographic areas**.

The census is a familiar and tested approach both in the UK and internationally – 2011 was the 22nd census in England and Wales and was run very successfully.

The census provides a **high degree of continuity** with the statistics provided by previous censuses and so supports detailed analysis of change over time

The single reference date means that the census delivers a **single, high-quality ‘snapshot’** of the nation.

Weaknesses

The census produces statistics only every 10 years. ONS uses births, deaths and migration data to provide adjusted estimates of the size of population every year, but the majority of statistics are published only every 10 years. For some topics, and in areas where change is rapid or complex, this reduces the usefulness of the statistics. For example, the population growth resulting from expansion of the European Union in 2004 was not fully understood until 2012.

The online census costs more than using administrative data and surveys– the long term cost is estimated at around £625 million (in 2013 prices) over each 10-year period. This is equivalent to £1.10 per person per year, and is comparable to other countries.

The census is a major operation and requires large scale recruitment of staff. The 2011 Census employed around 40,000 staff at its peak – the majority employed for only six weeks. **The scale and cyclical nature of the census makes the build-up and run-down of the operation challenging** both in management terms and in its effect on funding. A more stable profile where skills were retained might be more efficient.

The census imposes a burden on all households because every household is required to respond. Only a limited number of questions can be asked, and it does not include questions such as household income.

An online census once a decade – *continued*

Opportunities

Online collection of census data gives the opportunity for **faster and more efficient data processing**, compared with the time and cost of processing over 20 million paper questionnaires. The online responses to the 2011 census were found to provide higher quality data than the responses from the paper questionnaires.

Risks

International experience, also reflected in the UK, is that **it is becoming increasingly difficult to achieve high levels of response to a census**. The 2011 census was successful and achieved similar response to the 2001 census, but at significantly increased cost – about 35% more in real terms. This trend looks set to continue.

Not all households have access to the internet, and **an online approach risks excluding some people and households**. Only 16% of households completed the 2011 Census questionnaire online, and a step change would be required. Other methods of responding would still be made available.

Some people consider the census to be **an invasion of privacy**, and that some of the questions asked are too intrusive. As with all previous censuses, ONS would take extensive measures to protect privacy (as discussed in section 4).

A census using administrative data and surveys

Strengths

Statistics would be based on continually updated government sources, so it would be possible to generate population estimates (counts by age and sex) for England and Wales, for each Local Authority and for areas the size of electoral wards **on an annual basis**.

Similarly, because statistics about the *characteristics* of the population would be produced using a compulsory survey, these could also be produced annually – although not in as much detail as an online census once a decade could provide.

Annual outputs **would allow changes and trends** such as growth in population or migration **to be identified much more quickly** than is possible with a ten-yearly census.

This method would cost less than an online census once a decade, but there would still be costs of carrying out annual surveys. The cost of a new system over a ten-year period is estimated to be around at £460 million (in 2013 prices), which is equivalent to 80 pence per person per year. This estimate will be revised as research continues.

This solution would depend partly on the re-use of data already collected by government. The new surveys would cover around 50 percent of the population over a decade, so **this approach would reduce the burden on households**.

Weaknesses

An administrative data solution would never produce the most detailed statistics available from an online census once a decade. An annual survey of around 4% of the population would not provide statistics for small groups of postcodes, and statistics for wards would be less detailed.

For many topics several years of survey data would need to be combined, particularly to produce statistics for areas as small as electoral wards. **This means that the date to which the statistics refer would be less clear.** Experience from a similar survey used in the USA suggests that this adds complexity to the interpretation of statistics.

This method would not result in the detailed historical record of people and households used by family historians and other historical researchers. Archiving administrative sources may provide an alternative approach, but a method has not yet been developed.

New legislation would be required for approaches using administrative data and surveys, to give access to the required data and to make surveys compulsory.

A census using administrative data and surveys – continued

Opportunities

As with the census, surveys could only ask a limited number of questions, and would not include questions such as household income. However, over time **it should be possible to extend the use of administrative data to cover topics not covered by the survey**. Anonymised tax records, for example, might provide local statistics on household income – a frequently identified user need that cannot be met by an online census.

An annual national survey has the **potential to be more flexible than an online census in the questions that are asked**. It would provide the opportunity to adjust questions to respond to changing needs.

Archiving administrative datasets might result in **new opportunities for historical research in 100 years' time**. For example it might be possible to archive sources more frequently than once a decade.

Risks

This approach **would depend upon a number of new and partially untested methods**. Whilst early research has given promising results, a change in approach on this scale requires extensive development of new statistical methods and computer systems. Experience from countries that have made such a move, for example Denmark, Sweden and Finland, is that it can take a number of decades.

Any change would lead to some discontinuities from statistics produced previously. New methods will inevitably result in some differences in results, and it would be difficult to know whether apparent changes in the population were real or not.

The approach would rely on access to the required administrative data, and would also be susceptible to changes in the scope of the population included and the data items collected in the administrative systems. Recent changes to child benefit entitlement for example will result in some children not being included in data from the Department for Work and Pensions in future.

As with the online census method, there is a risk that **it would be difficult to achieve the required response rate to the new surveys**. Again, international experience is that response rates to surveys are falling over time, and some voluntary surveys run by ONS currently have response rates of around 60%, compared with around 80% in the 1980s.

This method **would require public acceptance of the use of administrative data for statistical purposes**. Bringing together data from multiple sources raises concerns about privacy for some people. ONS would take extensive measures to protect privacy (as discussed in section 4).

4. Managing the risks

4.1 Managing the risks of an online census once a decade

An online census once a decade, whilst different in the method of data collection, would be similar to the 2011 Census, which is a tried and tested approach.

However, the risk of falling response rates may require increasing investment each decade, and could ultimately result in a fall in response to the extent that it damages the quality of the statistics that it produces. One option, if this were to occur, would be to use administrative data in place of completed questionnaires for the individual addresses that failed to respond. Such an approach is being considered in other countries.

The risk of excluding people who do not have access to the internet could be managed by providing paper questionnaires, enabling completion by telephone, or collecting responses through doorstep interviews.

The main reason that only 16% of households completed the 2011 Census online was that all households were provided with a paper questionnaire. This was similar to the 2006 Census in Canada, where 18% of households responded online.

Having had the opportunity to introduce use of the internet in 2006, the Canadian census in 2011 was conducted primarily online (censuses are taken every five years in Canada). 54% of households responded online, with an overall response rate of 98%. This suggests that the risks of moving to a primarily online approach in 2021 would be manageable.

4.2 Managing the risks of a census using administrative data and surveys

New approaches and discontinuity in results

The risks relating to new methods are less well understood, and new approaches may not work exactly as planned at the outset.

Similarly, we may not be able to fully understand any discontinuities in the statistics compared with previous years' population estimates.

A transition period of ‘dual running’, moving towards a census using administrative data and surveys, but also with an online census covering all households in 2021, would enable us to compare the results from a new system to the results from previous approaches. However, this would increase costs during the transition period, and would increase the burden on the households who would be required to respond to both the online census and the new surveys. A reduced set of questions could perhaps be included to reduce the public burden.

As a minimum, in 2021 we would need to collect questionnaires from at least 10% of households to enable a comparison of new and old methods to be made, but even then some risks would remain. This would also impose an extra burden on these households.

Our research has shown that producing statistics using administrative data is most difficult for large cities, because the population changes most rapidly in these areas and administrative data tends to be out of date. It might be that an online census could be taken just in these areas to check whether the new methods were working.

Data access and survey response rates

To manage the risks of data access, new legislation would be needed to give ONS permission to receive the required datasets from other government departments, and to give ONS a formal role in decisions about changes to key data sources. Ireland, Holland and Scandinavian countries have legislation of this kind.

The risk of falling survey response rates would require surveys to be compulsory – again this would require new legislation. This would not fully address the risk, but data from administrative sources could potentially be used to understand the key characteristics of those who did not respond and make adjustments.

4.3 Public acceptability and privacy

Protecting confidential personal data is a fundamental principle for all statistical activities. It is particularly important for the census, whether it is an online census once a decade or for a census using administrative data and surveys, because data is collected from the whole population.

ONS has a strong track record of protecting the personal data from censuses and other surveys, and does not release personal census data for 100 years. This will continue to be the case, whatever the outcome of this review.

Data held by ONS is protected by law under the Statistics and Registration Service Act 2007. Unauthorised disclosure of personal data is punishable by up to two years in prison.

If a census were to be conducted using administrative data and surveys in future, before linking data from multiple sources together ONS would replace the names, addresses and dates of birth in the administrative data with unique but unidentifiable codes. We would also take extensive measures to control data access. This is consistent with [guidance from the Information Commissioner's Office](#). A [separate paper](#) provides further details of our approach.

Regular discussions have been held with a range of security and privacy experts and a Privacy Advisory Group has been established.

The methods used to protect the 2011 census data were independently reviewed and the findings were [published](#). This review found that ONS complied fully with data protection and government security requirements, and "*the public can be assured that the information they have provided has been well protected*".

We will conduct a similar review for any future census activities.

5. Consultation questions

This consultation asks for your views about the uses and benefits of census information, and the advantages and disadvantages of the two census approaches described in this document – both for statistical and historical purposes.

It builds on our research over the past two years to understand how people use census information, and the benefits those uses bring to society and the economy. The findings of this research are summarised in our [summary of the uses of census information](#), which includes users for genealogical and historical research, and our [summary of the benefits of census information](#). You do not have to respond to all questions.

5.1 General comments

Q1: What are your views of the different census approaches described in this document?

5.2 Uses and benefits of population and housing statistics

A good understanding of uses and benefits is critical to understanding the relative merits of each method. These questions ask you to tell us about any uses or benefits of census statistics that we have not yet fully understood.

In your answers, please provide as much evidence as possible. Uses that fulfil a regulatory or legal requirement are also of particular interest.

Q2: Please specify any significant uses of population and housing statistics that we have not already identified.

Q3: Please specify any significant additional benefits of population and housing statistics that we have not already identified.

5.3 Impact of different census approaches on statistical uses

These questions ask for your views on the real impact the different census approaches would have on operational or business decisions.

Q4: 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?

Q4.1 If medium or high, please give further information.

Q5: What would the additional benefit be if more frequent (ie annual) statistics about population characteristics were available for areas like Local Authorities and Electoral Wards? High, medium, low, or no impact?

Q5.1 If medium or high, please give further information.

5.4 Impact of different census approaches on historical research

ONS has worked with The National Archives and genealogists to understand how census information is used in historical research. These questions ask you to tell us about any uses or benefits of census information that we have not yet fully understood and to share your views on the potential impact of the different census approaches.

Q6: Please specify any significant uses of census information for historical research that we have not already identified.

Q7: What advantages or disadvantages for genealogical or historical research can you see from a move to a solution based on archiving administrative data sources?

5.5 Managing risks

As described in section 3, there are risks and opportunities with both census approaches. These questions give an opportunity to comment on these and to raise any other issues.

Q8: What are your views of the risks of each census method and how they might be managed?

Q9: Are there any other issues that you believe we should be taking into account?

5.6 How to respond

Please submit your views via the [online survey](#) by 13 December 2013.

All responses relating to uses of census information in Wales will be passed to the Welsh Government and all responses will be published, unless you indicate otherwise.

This consultation is part of an ongoing dialogue, which will continue until we publish our findings in 2014.

Further information, including numerous research reports, is available on the [Beyond 2011 pages of the ONS website](#).

If you wish to discuss any element of this work please contact us at Beyond2011@ons.gov.uk

Appendix: Statistics available from different approaches

This appendix provides further detail about the types of statistics that would be available using the different census approaches.

In order to do this, it is helpful to first describe the main geographic areas that are used for publishing population statistics. Please see Box E.

BOX E: Summary of main geographic areas used to publish census statistics

OA – Output Area – Groups of postcodes with a total population between 100 and 625 people, average 300. OAs are used across the UK as the base unit of census output.

LSOA – Lower Layer Super Output Area – Total population between 1,000 and 3,000 people, average 1,600. Each LSOA is made up of a number of OAs.

MSOA – Middle Layer Super Output Area – Total population between 5,000 and 15,000 people, average 7,800. Each MSOA is made up of a number of LSOAs. These are similar in size to an average electoral ward – although there is much more variability in the size of electoral wards.

LA – Local Authority – Total population between 2,000 and one million people, average 160,000. Each LA is made up of a number of MSOAs.

A.1 Statistics available from an online census once a decade

Population estimates

An online census once a decade would provide high-level estimates of the size of the population by age and sex (from national to OA level) plus [a wealth of detailed statistics](#) about population and housing characteristics. In total nearly five billion individual statistics will be published from the 2011 Census, giving very detailed cross-tabulations for small geographic areas and small population subgroups, and this would continue if we have an online census in future.

Each year in between censuses, ONS produces estimates for local authorities by:

- adding the number of births
- subtracting the number of deaths
- adjusting for the migration of people within England and Wales
- adjusting for immigration and emigration.

The statistics are not as accurate at the end of the decade as at the start, and this would continue to be the case.

We would also continue producing population projections, based on trends in births, deaths and migration seen in previous years, and making assumptions about the possible trends in future.

Population characteristics

Between 10-yearly censuses, ONS's annual surveys provide some statistics about the characteristics of the population. The largest of these is the Annual Population Survey, which receives responses from around 140,000 households each year. These surveys are large enough to produce statistics at the national and regional level, but generally cannot produce detailed statistics for local authorities or smaller areas.

These surveys are voluntary, and census statistics are used to make adjustments to the results to allow for the fact that the types of people who respond to voluntary surveys tend to be different to those who do not.

Timeliness

ONS published national and LA-level population statistics by age and sex from the census about 15 months after the reference date. For example, statistics from the 2011 Census relating to the reference date of April 2011 were published in July 2012.

Annual population estimates between censuses are currently published within 12 months of the reference date.

Detailed statistics about population and housing characteristics are currently published from the census over a period ranging from 18 months to 3 years of the reference date.

An online census would enable faster processing of census questionnaires, and the first census statistics would probably be available within about a year of census day. Similarly, other statistics might be published more quickly, but it would still be a complex process and it is not yet clear what would be feasible.

A.2 Statistics available from a census using administrative data and surveys

Population estimates

A census using administrative data and surveys would provide accurate statistics every year about the size of the population by age and sex – nationally, for local authorities and for small areas. It is likely that these would be available down to OA level but these are not yet fully tested.

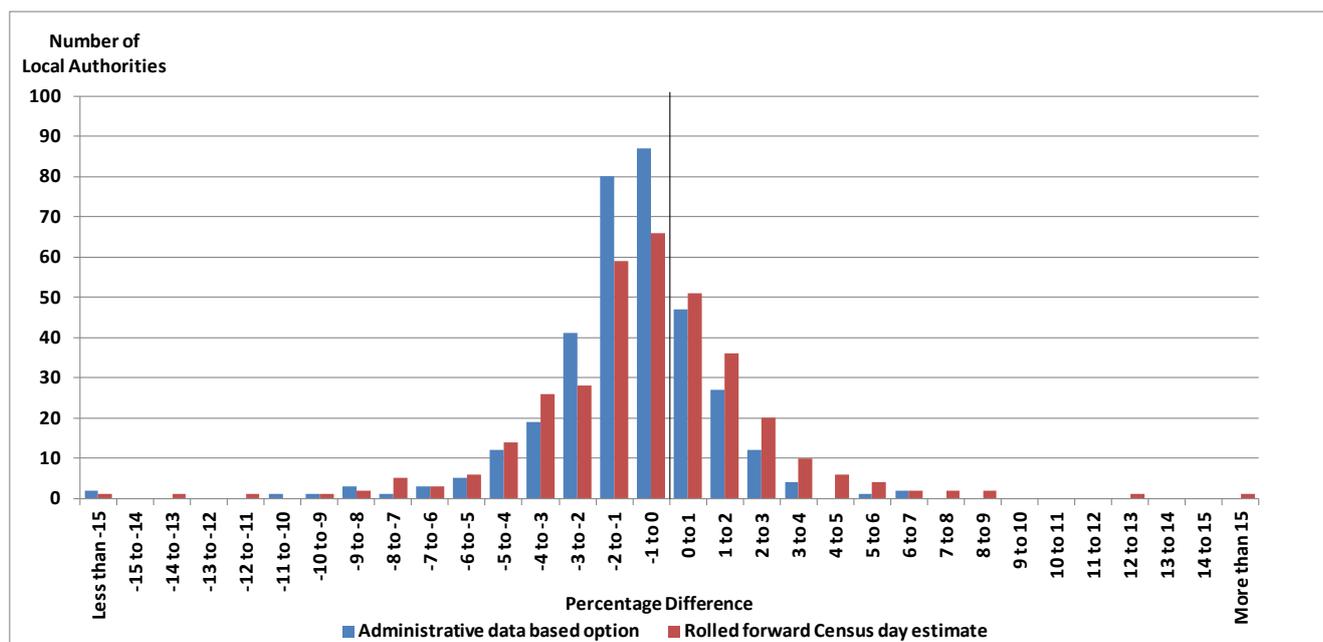
These statistics might not be as accurate as the estimates from the census, but could be more accurate than the annual population estimates for LAs produced between censuses in the current system. They would certainly be more accurate than the estimates for LAs produced towards the end of the decade between censuses.

Figure 1 shows a comparison between early results from methods using administrative data and the mid-2010 indicative population estimates, rolled forward to the 2011 Census day (ie the population estimates produced using current methods, 10 years after the census). The height of the bar indicates the number of local authorities within a given percentage difference from the 2011 Census estimate.

The chart shows that the administrative data-based method is generally more accurate than the current method at the end of a decade – it is roughly equivalent to year five of the current system. There is a bias towards underestimating the population overall and a coverage survey would be used to adjust for this.

A consistent level of accuracy would be achieved every year, because administrative datasets are continually updated, and further improvements are likely, as the new methods are in the early stages of development.

Figure 1: Comparison of administrative data-based methods and mid-2010 indicative population estimates (rolled forward to 2011 Census day) with 2011 Census estimates



Population characteristics

The compulsory annual survey of 4% of households would provide statistics about population and housing characteristics every year. These would be much more detailed than the annual statistics available from the current system – but not as detailed or accurate as an online census would provide once a decade.

The size of the population being measured and the percentage of the population included in the survey are the key factors that affect the accuracy of the statistics.

Statistics available using one year's data

Assuming a random selection of households is taken across England and Wales, a 4% sample of the population would produce a reliable¹ estimate of a population subgroup of about 800 people – whatever the geographic area concerned.

This might be for example 800 males who are unemployed or 800 Welsh speakers in a Local Authority.

Broadly speaking, a cross-tabulation using one year's survey data could be produced if the **average value** of the 'cells' within the table was 800 or more (some cell values would be much larger and some would be much smaller).

Similarly, for MSOAs, which have 7,800 residents on average, one year's survey data would provide meaningful statistics for a table that had around 10 cells. In practice, this means that most MSOA statistics using one year's data would be based on single variables .

Statistics using more than one year's data

Combining three years' data from the survey (a 12% sample of the population) would allow a reliable estimate for a subgroup of 230 people or more. For example, this might be the number of carers in an LA aged under 25 who provide less than 20 hours' care per week or the number of people in an MSOA with a limiting long term illness.

More generally, a cross-tabulation using three year's survey data could be produced if the **average value** of the 'cells' within the cross-tabulation was 230 or more. For MSOAs, this means that a cross-tabulation with about 30 individual cells could be produced.

Combining five years' data would enable a reliable estimate for a subgroup of 130 people or more.

¹ See later discussion of the precision that could be achieved

Box F summarises the types of cross-tabulations that might be reliable for geographic areas of different sizes, using different numbers of years' data from the survey.

Box F: Statistics possible using survey data				
Area type	Average number of residents	1 year's data (800 threshold)	3 years' data (230 threshold)	5 years' data (130 threshold)
LA	160,000	Detailed cross-tabulations (c 200 cells)	Detailed cross-tabulations (c 500 cells)	Very detailed cross-tabulations (c 1000 cells)
MSOA	7,800	Some single variable statistics (c 10 cells)	Very simple cross-tabulations (c 30 cells)	Simple cross-tabulations (c 50 cells)
LSOA	1,600	Not available	Some single variable statistics (c 5 cells)	Some single variable statistics (c 10 cells)
OA	300	Not available	Not available	Not available

Note that the values of 800 (one year's sample), 230 (three years) and 130 (five years) included here are subject to further research but are indicative of the population subgroups for which reliable statistics could be provided.

Box G summarises the statistics that might be available for a number of topics that were included in the 2011 Census.

Box G: Statistics available for high priority topics				
	Level of geography, and years of survey data needed			
Topic	LA	MSOA	LSOA	OA
Ethnicity	1 for large groups 3 for most others 5 for very small groups	3 and 5	5	Not available
Economic activity	1	3	5	Not available
Marital status	1	1 for large categories 3 for others	3 for large categories 5 for others	Not available
Tenure	1	3	3 for large categories 5 for others	Not available
Household composition	1	1 for large household types 3 for others	5	Not available
Accommodation type	1	3	5	Not available

Levels of precision

The more years of survey data that are used, and the larger the size of the population group being measured, the more precise the estimate would be. Conversely the precision of estimates for very small population groups would be lower.

Statistics for a characteristic that applies to a population of 800 people in any geographic area could be produced based on a single year's data with a 'confidence interval' of plus or minus 40% (480 to 1120 people); that is, we could be 95% confident that the true estimate lies in this range.

While at face value this seems quite a large range, it is worth considering the relative size of such a population. In an average-sized LA of 160,000 people a population subgroup of 800 represents 0.5% of the population. Expressing it in these terms would mean we could be 95% confident that the true estimate of the proportionate population would be between 0.3% and 0.7%.

Similarly, for a characteristic relating to a larger subgroup in an average sized LA, for example of 20,000 people (representing 12.5%), the confidence interval would be much narrower at plus or minus 6.4% (18,720 to 21,280 people). In relative terms, the true estimate would therefore lie between 11.7% and 13.3% of the total population of the LA.

It would be possible to produce statistics of similar precision for characteristics that apply to fewer people in the population group by combining survey samples across a number of years.

If we used three years' data, we would be able to produce statistics on characteristics that apply to a group of 230 people with a confidence interval of plus or minus 40% (140 to 320 people).

Combining five years' data it would be possible to produce estimates that apply to 130 people with a confidence interval of plus or minus 40% (80 to 180 people).

Implications for more detailed statistics

ONS would still include statistics within published tables where the numbers were smaller than the thresholds in Box F, but these would be of a lower precision, and indicated as such. Although it would not be possible to know with any certainty what the precise values were in these areas, it would nonetheless be clear that the numbers were small.

However, some of the very detailed cross-tabulations for small geographic areas that would be provided by an online census once a decade would not be available from a census using administrative data and surveys.

For example, it would not be possible to publish reliable statistics about the health of the population (from ‘very good’ through to ‘very poor’), for each ethnic group, in 10-year age groups, for each LSOA. This would result in 850 categories per LSOA (5 health categories, for 17 ethnic groups, for 10 age groups). Because each LSOA has on average only 1,600 residents in total, the majority of statistics in this table (if not all) would have a value far below the 130 threshold.

Over time, as more data becomes available from administrative data sources, it might be possible to produce more detailed statistics for small areas – household income for example, using data from the tax and benefit systems – but much more work is required to understand the possible data sources and develop the required methods.

Timeliness

A census using administrative data and compulsory surveys would enable national and LA-level statistics to be published within 12 months of the reference date.

Statistics about population characteristics based on three years’ survey data could be published within two years of the reference date. For example statistics relating to mid 2020 could be published in mid 2022 using data from 2019, 2020 and 2021. Statistics based on five years’ data could be published within three years of the reference date.

The Output Areas as a building block

Although many users make use of output area (OA) statistics, our evidence suggests that few actually make decisions or focus policy at the OA level itself. Instead they use OAs as a ‘building block’ to construct statistics for larger areas of their own. These might be the catchment of a school, hospital or supermarket for example, or they may wish to maintain areas such as old parish boundaries. By grouping together OAs to best fit their area of interest, users are able to produce statistics for these customised geographic areas.

It may be possible for ONS to offer more detailed cross-tabulated statistics for customised groups of OAs, without necessarily delivering the detailed OA-level information itself. More research is required in order to understand what could be achieved.

Box H: Example statistics possible using multiple years' survey data

With one year's survey data, reliable statistics could be produced on the number of:

- People with good, fair or poor health in each LA
- People with a limiting long term illness in each LA
- Welsh speakers in each LA in Wales
- Households in which no-one has English as their main language, for most LAs
- Carers (aged 24 and under) providing less than 20 hours' unpaid care, for over a third of all LAs.

With three years' survey data, reliable statistics could be produced on the number of:

- Carers (aged 24 and under) providing less than 20 hours unpaid care, in each LA
- Carers with good, fair or poor health in each LA
- People in good health, providing less than 20 hours' unpaid care, in each MSOA
- People with a limiting long term illness in each MSOA
- Welsh speakers in each MSOA in Wales
- Households in which no-one has English as their main language, for a third of all MSOAs
- People with good, fair or poor health in most MSOAs.

With five years of survey data reliable statistics could be produced on the number of:

- People in good or fair health in each LSOA
- Welsh speakers for over three quarters of all LSOAs in Wales
- People with a limiting long term illness in each LSOA.