2011 Census Coverage Assessment and Adjustment Strategy

Owen Abbott
Office for National Statistics

Introduction

The central objective of the 2011 Census is to provide high quality population statistics as required by key users such as policy makers and service providers, on a consistent and comparable basis for small areas and small population groups.[1] The key mission critical aims include:

- Provision of high quality, value-for-money statistics that meet user needs
- Maximising overall response rates and minimising non-response in specific areas and among particular population subgroups
- Building user confidence in the final results

Every effort is made to ensure everyone is counted in a census. However, no census is perfect and some people are missed. This underenumeration does not usually occur uniformly across all geographical areas or across other sub-groups of the population such as age and sex groups. The measurement of small populations, one of the key reasons for carrying out a census, is becoming increasingly difficult. In terms of resource allocation, this is a big issue since the population that are missed can be those which attract higher levels of funding. Therefore, without any adjustment, the allocations based upon the census would result in monies being wrongly allocated. It is therefore traditional that census undercount is measured and the outcome disseminated to users. Hence in order to achieve the mission critical aims outlined above, a coverage assessment and adjustment strategy is required. This article outlines the proposed strategy for the 2011 UK Census. While the strategy is applicable to the UK, it is expected that there will be slight differences in the detailed methodology between countries to reflect local circumstances. This strategy covers the production of the census estimates and adjustments to the census database. It does not cover the Quality Assurance of those

Every effort is made to ensure everyone is counted in a census. However, no census is perfect and some people are missed. This underenumeration does not usually occur uniformly across all geographical areas or across other sub-groups of the population such as age and sex groups. Coverage levels in censuses across the world are declining leading to an increasing need to firstly slow the decline through improved enumeration, and secondly in improved methods and data for measuring coverage. In order to achieve the mission critical aims of the 2011 Census, a coverage assessment and adjustment strategy is required.

This article outlines the proposed strategy for the 2011 UK Census. The strategy is to significantly improve upon the 2001 One Number Census, and use it as a platform to develop an improved coverage assessment and adjustment methodology. The article also outlines the key areas of innovation for the 2011 strategy and the options that will be considered in order to develop the final strategy.

Stakeholder management is also an important part of the strategy to ensure that key users both buy into and understand the methodology. ONS will look to build on the consultation carried out prior to the 2001 Census and widen the user base with which it engages on this important topic.

estimates, since that is covered by the Census Quality Strategy,[2] although it is recognised that there are close links between the two.

'Background' below provides information on the strategies from previous UK censuses, and the lessons learnt from the most recent. It also summarises the changes in census design proposed for 2011 that are pertinent for coverage assessment and adjustment. The high level strategy is presented in 'The strategy for 2011', and then 'Key innovations and options' outlines the key areas of innovation for the 2011 strategy and the options that will be considered in order to develop the final strategy. 'Implementation' highlights some of the practical considerations and 'Communication strategy' discusses communications with stakeholders. After a brief discussion of future developments leading up to 2021, the conclusions are drawn out.

Background

Most census taking countries undertake some form of coverage assessment and adjustment, usually using some form of post-enumeration survey (PES). Measured undercount levels have on the whole been increasing over the past few decades. More importantly, the differential nature of the undercount has worsened with, for example, young males and ethnic minorities becoming increasingly difficult to enumerate. This has led to increasing priority and focus on the methods for measuring this differential undercount. The following sections outline the strategies adopted for measuring coverage in the 1991 and 2001 Censuses, and give a summary of the changes proposed for the 2011 Census design that are relevant for coverage measurement.

The 1991 Census

The 1991 Census post-enumeration survey was called the census validation survey (CVS), which sampled around 6000 households across England and Wales. The CVS had a dual role to estimate underenumeration and to assess the quality of the data. It has been described fully by Heady *et al*.[3] In the event, the CVS estimated the underenumeration to be around a quarter of a million. However, demographic analysis suggested that the underenumeration was more like 1.2 million. In addition, the CVS did not indicate any geographical variation in underenumeration, which was not plausible. Holt *et al*[4] summarises how the census estimates were revised in light of the demographic analysis. The final results, after a number of releases, were not internally consistent since estimates of underenumeration below local authority level could not be provided. This was deeply unsatisfying for the user community.[4]

The 2001 Census

To overcome the risks to the credibility of the census which would be posed by a similar problem in 2001, the ONS established the One Number Census (ONC) project.
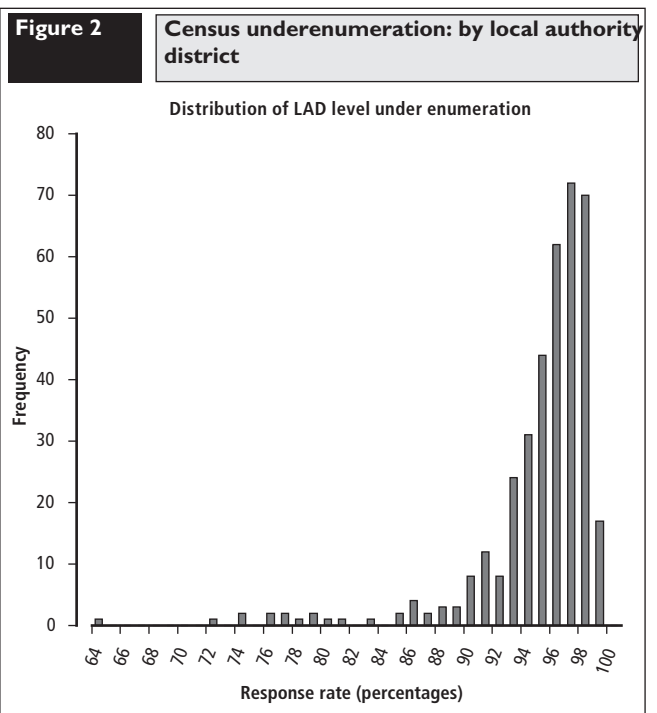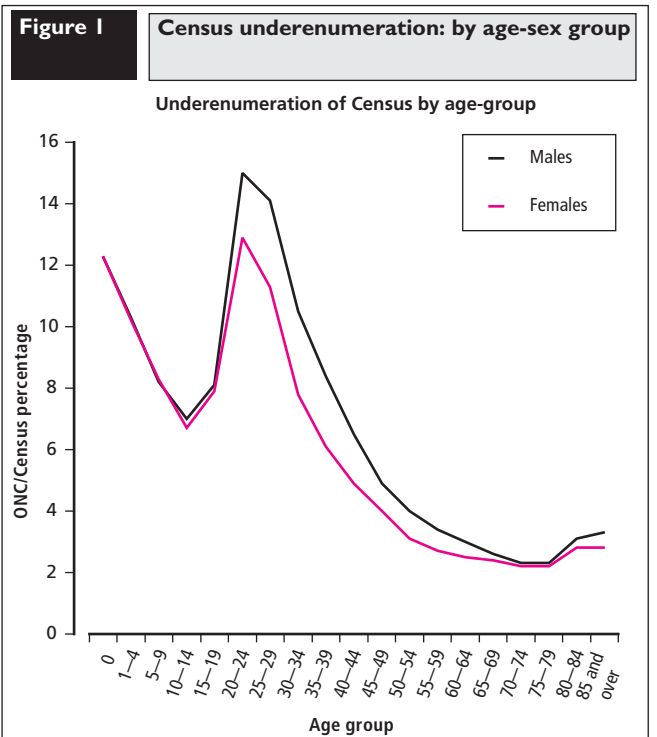
The One Number Census project
The One Number Census project had the goal of providing a methodology and processes to identify and adjust for the number of people and households not counted in the 2001 Census.[4, 5] The aim was to provide a population estimate that would be the basis for the 2001 mid-year estimate (with a minor time lag correction), and for which all census tabulations would add up to. The extent of the underenumeration was identified using a large survey covering approximately 320,000 households, the Census Coverage Survey (CCS), which provided an independent enumeration of a sample of areas. The results of the CCS were then matched, at individual level, to the corresponding 2001 Census data. The combined census and CCS information was used to produce an estimate of the number of people missed by the Census, using a combination of Dual System and ratio estimation. The people estimated

to have been missed were then added to the database by imputing using a donor method.[6] The resulting outputs were then subject to a rigorous quality assurance process. This type of strategy is used amongst other census taking countries, such as the US, Australia and New Zealand (although these countries do not undertake the full imputation process).

The One Number Census measured the underenumeration in the 2001 Census to be 6.1 per cent of the total population (See Box 1 below).

Box one

2001 Census Response Rates

Figure 1	Census underenumeration: by age-sex group

Underenumeration of Census by age-group

Age group

Figure 2	Census underenumeration: by local authority district

Distribution of LAD level under enumeration

Response rate (percentages)

Lessons learnt from the 2001 Census
The ONC was a big step forward. Both the Statistics Commission[7] and the Local Government Association[8] have published reviews that concluded that the methodology used in 2001 was the best available and no alternative approach would have produced more reliable results overall. However, there were some issues with the results which led to further studies and adjustments. These were summarised by Chappell and Dobbs[9] and ONS[10].

As a result, there were a number of key lessons from the ONC project that are pertinent to the strategy in 2011. These were explored by Abbott and Brown.[11] In summary, these lessons were:

- the ONC was not able to make adjustments in all situations, particularly when there were pockets of poor census response. While all efforts will be made to reduce this risk in 2011, there must be recognition that both national and sub-national adjustments are likely to be needed
- engagement with stakeholders is critical, particularly Local Authorities, and efforts to provide simple messages are essential
- for 2011, strategies for measuring and adjusting for the lack of independence between the census and CCS must be developed (this was the failure of one of the assumptions underpinning dual system estimation, leading to underestimates of the population). This assessment of dependence should form an integral part of any future dual system estimation methodology, as the ONC has indicated that the assumption of independence is not realistic
- two of the weaknesses of the ONC were its reliance on the CCS, and the perception that it would solve all 'missing data' problems. The potential for census, survey and administrative data to supplement the CCS (such as use of a triple system estimator) requires a full programme of research (co-ordinated with work on administrative sources across ONS), and the issue of perception should be addressed through more robust communication
- the measurement of overcount requires greater attention
- the balance of 'measurement' resource between easier and harder areas needs careful consideration – for example how should the sample be allocated given a fixed overall size

The 2011 Census design

Planned changes to the 2011 Census design are likely to influence the methodology used to estimate census coverage. The proposed changes to the census design where there may be an impact on coverage or coverage measurement methodology are identified below. Most of these are likely to be similar across the UK, but there are differences between countries. Each section notes where this is the case.

Population base
The population base for 2011 is to enumerate persons at their usual residence and collect information from visitors. There is a possible impact on coverage if the extra burden placed on the public through collecting visitor information results in visitors being missed at their usual residence. However, it may also improve coverage of persons who would be missed under a usual residence collection only, such as those who consider themselves not to have a usual residence. There is a risk that deliberately collecting information from persons twice will result in increased overcount. Further information on the impact of the population base on coverage will be obtained from question testing and the 2007 Census Test.

Topics
In the 2001 Census, there were very few controversial questions. For 2011, the possibility of including a question on income has been debated

and it is being tested in the 2007 Census Test. Inclusion of such a question will undeniably reduce response levels (the 1997 Test showed an overall difference in response of 2.8 per cent), and may also affect the response patterns. If income or another sensitive question is included, further research would be needed to examine the effect on the coverage measurement methodology.

Delivery method
The 2011 Census in England and Wales is likely to involve post-out of census forms in the majority of areas. This strategy is dependent upon an up-to-date and comprehensive national household frame. The census household frame will be constructed from the best available sources, and will be extensively checked through field visits prior to the census to maximise its coverage. However, inadequacies in the household frame would have a huge influence on census coverage, hence it will be essential to assess coverage of the register as well as coverage of persons within households identified on it. However, the household frame itself may provide a valuable information source that can be used within the coverage assessment strategy.

Furthermore, post-out reduces face-to-face contact with enumerators. While extended nation-wide publicity and call centres will be implemented in an attempt to counter this, the reduced contact with the public may increase underenumeration, perhaps for particular subgroups of the population such as the elderly.

Scotland and Northern Ireland are likely to adopt a more traditional delivery and postback methodology, similar to that employed in 2001.

Targeted enumeration resources
A key component of the 2011 Census design in England and Wales is a targeted allocation of enumerator resources, allowing resources to be focused on areas where response is expected to be lowest. This requires a national categorisation that identifies areas (or population subgroups) to support this targeting of resource. This has implications for the assessment of coverage, as the achieved response levels will be influenced by the allocation, and hence may be an important explanatory variable. In addition, the categorisation is similar to the 'hard to count' index used in the 2001 CCS design. Research will need to establish whether this enumeration categorisation is suitable for CCS stratification, or whether an alternative is more appropriate. Since Scotland and Northern Ireland are using a 2001 Census enumeration model, this targeting is not going to be a significant factor in those countries.

Response management
Each census questionnaire will be tracked centrally using the household frame to facilitate follow-up of non-returns. The process will identify multiple returns from the same address, and hence will aid the detection of duplicate records resulting from the use of multiple response routes. In addition, central tracking will provide early indications of response rates to the lowest geographical level – there is therefore the potential for this information to be used within the coverage assessment strategy. As Scotland and Northern Ireland are using a 2001 Census enumeration model, the questionnaires are likely to be controlled locally and therefore there may not be the same opportunity in those countries.

Response methods
In the 2011 Census, respondents may be able to provide information through a number of modes, such as the internet. Having more than one mode has an impact on coverage assessment due to mode effects. If response patterns are different between paper and internet returns, it may be sensible to estimate coverage separately. Most importantly, multiple response routes raise the potential for significant overcount in the 2011 Census if not carefully controlled.

The strategy for 2011

Objectives

The primary objective of the coverage assessment and adjustment strategy in 2011 is to identify and adjust for the number of people and households not counted in the 2011 Census. A secondary objective is to identify and adjust for the number of people and households counted more than once in the 2011 Census.

There are a number of other objectives:

- simple methods should be developed where possible, to aid communication of the methodology
- there are a number of ways in which undercount can occur (such as missing a whole household or missing a person from a counted household), and an objective is to be able to measure the extent of each of these, permitting more transparent adjustments
- gaining acceptance of the methodology from users is a key objective. Users will not accept their census population estimates if they are not confident about the methodology used to derive them
- since all census outputs will be influenced by the methodology, we will communicate with all users through appropriate channels and with tailored materials

The strategy will aim to address the lessons outlined in 'Lessons learnt from the 2001 Census', looking for improvements and taking into account the changes to the census design outlined in 'The 2011 Census design'. Target precision levels (for sampling errors only) are relative 95 per cent confidence intervals of 0.1 per cent around the national population estimate and 1 per cent for a population of half a million. Local Authority District (in England and Wales) and age-sex level population estimates should aim for minimal variation of precision, therefore ideally being the same precision across all estimates. This requirement will have an important influence on the methodology, particularly the sample design. The intention is to try to deliver results that are better than the 2001 results, and in particular ensure that there are no areas with a worse precision than the worst that was achieved in 2001 (i.e. there is no relative confidence interval for a Local Authority total population that is wider than 6.1 per cent, although 5 per cent is perhaps achievable).

Strategic review

During 2005, ONS undertook a strategic review of the high level options for coverage assessment and adjustment in the 2011 Census. This looked at international practice and the different sources of information used to assess coverage. What was clear from this review was that some form of Census Coverage Survey is required. The risk to the census results of relying on alternative sources being suitable to replace a CCS is too high. However, one of the key lessons learnt from the 2001 Census is that reliance on a CCS alone is also not desirable, given the likely dependence between the census and CCS. Therefore the use of additional sources is a natural extension of the 2001 strategy, and this will be explored.

The review therefore concluded that the coverage assessment and adjustment strategy in 2011 should be to build on the ONC framework, addressing the lessons outlined in 'Lessons learnt from the 2001 Census' looking for improvements and taking into account the changes to the census design outlined in 'The 2011 Census design'. The coverage assessment and adjustment workstream is taking forward the development of this strategy, in conjunction with colleagues from administrations in Scotland, Northern Ireland and Wales. To address some of the lessons learnt from 2001, there are a number of key innovations and options that require exploration. These are discussed in 'Key innovations and options'. However, the following represents the current thinking on the overall strategy for 2011.

Survey design

The overall strategy will be similar to that used in 2001, with a CCS of a similar size and field methodology. It is clear that for robust coverage assessment the sample would still need to be area based as there would not be a household listing of very high quality that was independent of the census process (which will use a household listing as described in 'Delivery method'). Therefore, the basic CCS design strategy is likely to follow the model adopted in the 2001 Census. It will be a stratified multi-stage sample selection of areal units that will be independently re-enumerated. 'Key innovations and options' discusses some of the improvements to the CCS design that are being explored, and Appendix A outlines the work required to define the survey design.

Matching

Matching the Census and CCS data will again be required. However, there may be additional sources of unit level data that are also needed to be matched if they are suitable sources for use in the strategy. In addition, there may be a requirement to carry out searching across the census database to measure overcount, another form of matching. Sources of data and measuring overcount are discussed in 'Key innovations and options', and the work required to develop the matching methods and systems is outlined in Appendix A.

Estimation

It is highly unlikely that Triple System Estimation (which uses the Census, CCS and a third independent high quality list to estimate the population in the sample) will be feasible in 2011, due to uncertainty around access to high quality third lists. Therefore, in practice, Dual System Estimation (DSE) will again be employed, although the research will be focused on evaluating and improving the methodology. In particular, we will explore the opportunities for using external data sources to either improve the overall precision or to reduce the bias (such as dependence) in the Dual System Estimator. This is explored in 'Key innovations and options' and the work required is outlined in Appendix A.

A ratio type estimator to derive population totals will be developed in a similar fashion to the approach in 2001, and small area estimation techniques are likely to again be needed to derive reliable estimates for smaller geographies and subgroups. Appendix A outlines the work required to develop and implement these methods.

Adjustment

Following the production of the estimates the census database will again be fully adjusted to take account of the undercount, although an adjustment for overcount will also be developed. Wholly missed households will be imputed, located using the census household frame, and persons within counted households will also be imputed. A similar methodology to that used in 2001 is likely to be developed, although additional sources of data will be used to explore whether improvements can be made in the imputation accuracy. Appendix A outlines the work required to deliver the methodology and implemented system.

Key innovations and options

This outlines the key innovations for the coverage assessment and adjustment strategy in 2011.

Census Coverage Survey design

The CCS will be the primary source of information that will feed into the measurement of the undercount in the 2011 Census. It is clear that for coverage assessment the sample will still need to be area based as there will not be a household listing of very high quality independent of any list created for the 2011 Census. Therefore, the basic CCS design strategy is likely to follow the model adopted in the 2001 Census. It will be a stratified multi-stage sample selection of areal units that will be independently re-enumerated. What we need to decide is the most appropriate variables for stratification and clustering.

From previous experiences, we know that undercount in a census varies by geography and demography. Therefore, to ensure our sample design is efficient, it is prudent to consider stratifying our sample to make use of this knowledge. 'The 2011 Census design' noted some additional operational factors that could have an influence on undercount (such as post-out in England and Wales) and we may be able to increase the precision of our estimates by considering further stratification in the sample design.

In 2001, the main geographic stratification came from forming Estimation Areas (EAs) by grouping contiguous Local Authorities to create populations of around 500,000 persons. This contiguous grouping did cause some issues at the estimation stage in areas that were very heterogeneous. The geographical layout of England and Wales means that in many cases, urban areas had to be grouped with rural areas. However, it is important to form some kind of Local Authority grouping since they will be one of the key subgroups for which we will want to derive estimates of population.

For 2011, EAs will again be formed from groups of whole Local Authorities and samples will be drawn from each of these strata. We are investigating potential improvements in design efficiency by grouping Local Authorities by some form of area type indicator, rather than restricting the groups by geographical constraints. This might have implications for processing time, and so a cost-benefit analysis is being undertaken as well as exploration of post-stratification as an alternative

Other areas where improvements are being sought are:

- to look for better and more up-to-date proxies for 2011 patterns of 'Hard to Count' to help form a more robust stratification
- to explore stratifying the sample by a 'size' measure to ensure that the sample contains a good representation of the key population sub-groups for which we expect the undercount to be high. The main potential improvement would be access to updated population counts
- to examine the potential for improving the sample allocation. The intelligence obtained through the 2001 One Number Census can be utilized to provide a better idea of where the sample needs to be concentrated

Appendix A provides an outline of the work required to define the survey design. It is likely that the sample design will be similar in Scotland and Northern Ireland, although there will be some differences due to geography and socio-demographic characteristics.

Sources of data

One of the key innovations for coverage assessment in 2011 is to build in more sources of data, particularly those that are up to date. There are

three places where the additional data can be used – in the design of the CCS (discussed in 'Census Coverage Survey design'), in the estimation methodology or in the adjustment process. The greatest gain from using additional sources is in the estimation process, as it should help to either improve the overall precision or to reduce the bias in the Dual System Estimator. Here we discuss three sources that have the potential to be used in the estimation process:

- visitor data from the Census
- the Census household frame
- government (ONS in England and Wales) survey data

Appendix B provides a list of the other sources that may be suitable for use in the methodology.

Use of these sources would reduce the reliance on the CCS as the primary source, a key lesson learnt from 2001. The work required to decide whether these sources are used and how they should be used is outlined in Appendix A.

Visitor data
One source of information that has potential for use in coverage assessment is that provided by visitors about their usual residence and basic demographic data. The idea behind the collection of visitor information is to match them back to their usual residence and thus discover whether they contributed to the undercount. Therefore, the visitor information could supplement the CCS by helping to identify residents who were not in the Census or in the CCS by adding the visitor information into the Census/CCS matching process. The benefits of this information are that it will be a national source of information, covering every area in the UK; it provides data from population subgroups that are likely to be fairly mobile (as they are likely to be visiting), a group which we know are difficult to count; and attempts to collect information from population groups who do not consider themselves to be usually resident anywhere, even though they are. Again, these are known to be hard to count and any information would help measure coverage.

The Census household frame
The household frame described in 'Delivery Method' that the 2011 Census will use to control its field operation is a source of data that is likely to be an integral part of the estimation strategy. It will be a source of data that was not available in 2001, and could be used as the basis for estimation of household spaces and types (vacant, second home etc), probably using some form of dual system estimator. However, this will depend on the quality of the register, which we will be able to measure through the CCS. This then provides additional controls and auxiliary information for the adjustment process, which in turn would use the register to place imputed households and control the types of imputation.

Government survey data
ONS is currently developing an Integrated Household Survey (IHS), which is planned to be available at the time of the 2011 Census, with a national annual sample of around 500,000 persons, with a quarterly survey period.

This source could be used to assess coverage of persons within households (as in theory it should contain all the persons in each sampled household), either independently or in conjunction with the CCS (for instance, to reduce bias due to dependence). The advantage of this source is that it is independent of the census although since the survey will have been conducted around the same time as the census there may be some correlated non-response. However, some early work using 2001 survey data has indicated that the ONS continuous surveys do capture persons that the Census missed. A similar strategy could be used in Scotland and Northern Ireland using their equivalent surveys.

Dual System Estimation

ONS are anticipating that correcting for underestimation of the population, due to bias in the DSE (which includes dependence), will be part of the methodology developed to address the lesson learnt from 2001. One strategy is to develop the framework used in 2001, set-out in Brown *et al*.[12] This utilized external data on the aggregate number of households in an area to correct for underestimation in the DSEs.

There are some potential improvements that we could make to the 2001 approach to reduce bias:

- more sophisticated use of the household frame – the census household frame should be much better in 2011 and this accuracy should provide a direct Estimation Area level correction (removing a weakness of the 2001 method which was only able to make regional level estimates)
- develop a more realistic model for translating a household level correction to person level, perhaps by using a non-uniform household size distribution
- explore the potential for using sex ratio information from another source (such as Demographic analysis) to provide a better distribution of the correction across age-sex groups

Overcount

The 2001 ONC primarily focused on measuring undercount. Overcount has not historically been a problem within the censuses of England and Wales, and therefore measurement of it was given a low priority. The CCS collected information about potentially overcounted individuals by asking whether there was anywhere else they might have been counted in the census. A matching study was undertaken based on the responses collected, resulting in an estimate of less than 0.1 per cent overcount. Further studies have indicated that this might have been an under-estimate. Based on its matching process, the Longitudinal Study estimated that 0.38 per cent of the population responded twice. A study of duplicates within the census database backed up this finding, estimating that there was potentially around 0.4 per cent duplicate persons. However, no adjustments were made to the 2001 Census estimates for overcount.

'The 2011 Census design' noted that changes to the census design are likely to raise the potential for overcount in the 2011 Census. Therefore, methods for estimating this will need to be explored. We will need to consider two alternatives:

- An approach that estimates overcount separately from undercount (the path adopted in 2001). Hence separate adjustments for under and overcount are made; or
- An approach that adjusts the individual postcode counts (and thus is integrated within the DSE and the resulting estimates and adjustment processes).

It is expected that a targeted matching exercise, where the whole database is searched for duplicates, is likely to be used as part of this strategy in addition to using data collected from the CCS and national analyses.

Implementation

The implementation of the coverage assessment and adjustment methodology will be a key component of downstream processing – that is the part of the census processing that occurs after data capture. There is likely to be three components to the implementation, as there was in 2001:

- a matching system that provides the ability to match the Census responses to the CCS responses, and also (dependent on the methodology) matches these to other data sources. This system may also be used for searching (e.g. for duplicate records as part of the overcount measurement strategy). It will include both an automatic and clerical interface. A team of matchers will be required to operate the system
- an estimation system that uses the outputs from the matching system to produce the best estimates of the population for a number of domains (e.g. Estimation Area and Local Authority District level, age-sex groups, households etc). The outputs from this are fed into the Census Quality Assurance process, and also into the adjustment system. A team of statistical staff will be required to operate and analyse the outputs from the system
- an adjustment system that imputes households and persons into the census database so that the outputs will be consistent with the outputs from the estimation system. There is a possibility that this system may also mark records that are overcounted and therefore should not appear in the outputs (i.e. they are effectively deleted). There is a possibility that CANCEIS, the standard edit and imputation tool used by ONS, could be utilised. A team of statistical staff will be required to operate and analyse the outputs from the system

These systems are likely to be built specifically for the purpose of processing the Census, albeit using existing software where feasible. Interfaces with other parts of downstream processing will also need to be developed. Where practical, we will explore the potential for re-use of these systems across ONS – particularly the matching system which might suit many survey and administrative source applications.

Communication strategy

As the research progresses and the methodology evolves, it is intended to keep stakeholders informed of progress and allow input through many of the established consultation routes, this article being a part of that process. The key England and Wales stakeholders for this strategy are:

Stakeholder
Local Authorities and their strategic partners
Dept for Communities & Local Government (DCLG)
Department of Health (DoH)
Other central government departments
Scottish Executive
Welsh Assembly
Business user groups
Health user groups
Census advisory groups
Special population groups
Academics
Statistics Commission
Royal Statistical Society
ONS Centre for Demography
International groups – UNECE, Australia/New Zealand/Canada/United States

ONS noted that we will seek to describe, clearly and unambiguously, statistical methodologies in terms that can be understood by non-expert audiences, and avoid black-box processes.[1] Therefore, communication methods will include both technical documentation and papers for stakeholders such as academics, and also more accessible and easy to understand papers on the methodology and principles aimed at users such as Local Authority Chief Executives. There is likely to be an ongoing series of such documentation, all made available through the National Statistics website.

A broad (draft) timetable outlining the key communication steps is given below:

- Autumn 2006 – Strategy Statement circulated
- Autumn 2007 – Methodology paper circulated
- Spring 2008 – Updated methodology paper and technical documentation presented at stakeholder Workshops/Roadshow (TBC)
- Autumn 2008 – Updated methodology paper circulated
- Autumn 2009 – Update methodology paper circulated and presented at stakeholder output roadshows (TBC)
- Spring 2010 – Final Methodology paper(s) circulated and presented at a variety of forums
- Autumn 2010 – Further communication in conjunction with other census consultations
- Autumn 2011 – Further communication in conjunction with other census consultations
- Autumn 2012 – Census results released, with associated coverage assessment and adjustment metadata

Future developments and looking forward to 2021

The 2011 Census provides the only opportunity to evaluate new methods that might be applicable in the 2021 Census, whatever form that may take (including the proposed Integrated Population Statistics System[13]), in a realistic census environment. The coverage assessment and adjustment strategy will therefore seek opportunities to evaluate new methods, possibly in partnership with other areas of ONS. This may include:

- matching multiple sources of administrative data with survey data
- evaluation of a triple system estimator (TSE)
- asking administrative source data holders to keep a snapshot of their data as at Census day (in anticipation that the data might be made available in the future)

Conclusion

In the context of the outcomes from the 2001 Census and the increasing pace of change in society, it is clear that for the 2011 Census, a coverage assessment and adjustment strategy will be required. As noted above, coverage levels in censuses across the world are declining leading to an increasing need to firstly slow the decline through improved enumeration, and secondly in improved methods and data for measuring coverage.

The 2011 Census project has a number of initiatives to improve the enumeration process and deliver a high quality census. This article outlines the proposed coverage assessment and adjustment strategy for the 2011 UK Census. The strategy is to significantly improve upon the success of the 2001 One Number Census, and use it as a platform to develop an improved coverage assessment and adjustment methodology. The strategy is similar across all the countries within the UK, although the final methodology may be slightly different and use different data sources.

This article describes the key components of the methodology, and then focuses on four key areas of innovation and options, namely:

- Census Coverage Survey design
- sources of data
- Dual System Estimation
- overcount

The development of this methodology for 2011 is underway and early research has focused on the design of the Coverage Survey, the estimation process and the suitability of the visitor data arising from the census questionnaire. Appendix A gives a fairly comprehensive breakdown of the work required to deliver the methodology.

Stakeholder management is also an important part of the strategy to ensure that key users both buy into and understand the methodology. ONS will look to build on the consultation carried out prior to the 2001 Census and widen the user base with which it engages on this important topic.

Key points

- Coverage assessment and adjustment will be critical to the success of the 2011 Census

- The strategy is to build on the success of the One Number Census.

- Ongoing consultation and engagement with users is an important part of the strategy

---

References

1. ONS (2004) *2011 Census: Strategic aims and key research in England and Wales*. Information Paper. Available at www.statistics.gov.uk/downloads/theme_population/Strategic_aims.pdf
2. ONS (2006) *2011 Census Quality Strategy*. Advisory Group paper AG(06)06. Available at www.statistics.gov.uk/census/pdfs/ag0606.pdf
3. Heady P, Smith S and Avery V (1994) *1991 Census Validation Survey: coverage report*, HMSO: London.
4. Holt T, Diamond I D and Cruddas M (2001) Risk in official statistics: a case-study of the 2001 one-number census project. *J. R. Statist. Soc*. D, **50** 441–456.
5. Brown J J, Diamond I D, Chambers R L, Buckner L J and Teague A D (1999) A methodological strategy for a one-number census in the UK. *J. R. Statist. Soc. A* **162**, 247–267.
6. Steele F, Brown J and Chambers R (2002) A controlled donor imputation system for a one-number census. *J. R. Statist. Soc. A* **165**, 495–522.
7. Statistics Commission (2003) *The 2001 Census in Westminster: Final Report*. Available at www.statscom.org.uk/media_html/reports/report_022/contents.asp
8. Local Government Association (2003) *The 2001 One Number Census and its quality assurance: a review*, Research Briefing 6.03. Available at www.lga.gov.uk/Documents/Publication/onenumbercensus.pdf
9. Chappell R and Dobbs J (2005) *Are densely populated inner city areas easy to measure and estimate? Lessons learned from the 2001 Census in England and Wales*. Paper presented at IAOS satellite meeting, Wellington, New Zealand, April 2005. Available at www.stats.govt.nz/about-us/events/satellite-meeting/default.htm
10. ONS (2005) *One Number Census Evaluation Report*. Available at www.statistics.gov.uk/census2001/pdfs/onc_evr_rep.pdf
11. Abbott O and Brown J (2006) *A review of the 2001 One Number Census methodology and lessons learnt*. Paper presented at GSS Methodology Conference, London, June 2006. Available at www.statistics.gov.uk/events/gss2006/downloads/D1Abbott.doc
12. Brown J, Abbott O and Diamond I (2006) Dependence in the 2001 one-number census project. *J. R. Statist. Soc. A*, **169**, 883–902.
13. ONS (2003) *Proposals for an Integrated Population Statistics System*. Information Paper. Available at www.statistics.gov.uk/downloads/theme_population/ipss.pdf

Appendix A	Work required to deliver objectives

Work area	Deliverables required	Notes
Census Coverage Survey Sample Design	a) Develop sampling strategy, including: • Sampling units and Estimation Areas • Sample size • Sample allocation • Define stratification • HtC index b) Select samples: • Develop sample selection system • Select sample for rehearsal • Select sample for 2011 Census	The sample design is expected to be similar to that employed in 2001. However, we are looking to make gains in efficiency by using the information we obtained on coverage patterns from 2001, and also make the sample more representative (or balanced), ideally using more up to date data. We will use both empirical and simulation techniques to assess the efficiency of different options.
Matching	a) Develop matching strategy, including: • Matching methodology b) Implement matching system: • System Requirements • System testing c) Run matching system: • Develop matching protocols • Recruit and train matching staff • Run rehearsal matching • Run 2011 Census matching	The matching work will build on the methodology used in 2001. However, since there might be additional sources of data to match (such as Visitor data or ONS survey data), further work will be required to work out the best method for matching those to Census and CCS. Also, the system may be required to carry out extensive searching for measuring overcount which will require additional research to define the best way of achieving.
Estimation	a) Develop estimation methodology: • Estimation Strategy • Use of Dual System Estimation • Bias adjustment in DSEs methodology • Estimation of population totals • Small area estimation methodology • Variance estimation methodology • Contingency methods b) Explore how to handle overcount, particularly whether it can be integrated or separate from the assessment of undercount. Also assess likely characteristics of overcounted population. c) Evaluate potential for using other sources beyond the CCS: • Cost-benefit analysis of use of visitor information • Cost-benefit analysis of use of ONS survey data • How to use household frame in estimation d) Implement and run estimation system	The estimation methodology will again build on the research carried out in 2001. A (more advanced than 2001) simulation environment will be used to evaluate different techniques and options. This can be used to ensure that lessons learnt from 2001 are addressed, and define the circumstances under which the methods are not robust. Overcount assessment will require a lot of work as this will be the first time that ONS attempts to measure it accurately. There are a number of options that need to be explored. Furthermore, there are many sources of data that could be added to the estimation process – and in order to decide whether it is worthwhile including them (as they will cost a lot to match) a cost-benefit analysis is required for each.
Coverage Adjustment	a) Develop coverage adjustment strategy: • Models for characteristics • Calibration methodology • Imputation methodology • Potential use of other sources such as the household frame • How to handle overcount b) Implement and run coverage adjustment system	Coverage adjustment work will commence later than the sample design and estimation work, so that the basic framework is laid down. The key pieces of work here are to re-examine the models used in 2001 and explore improvements, particularly around the calibration to control totals. The imputation methodology is unlikely to change significantly, although consideration of overcount needs to be factored in.
Support work	Work required to support the delivery of the methodology including: • Stakeholder management and communications (see 'Communication strategy') • Project management • Integration with downstream processing plans and technical environment (see 'Implementation') • Prototyping • Quality Assurance of methodological proposals • Consideration of CCS field methodology	

Appendix B	List of data sources that might be suitable for use in coverage assessment and adjustment methodology

This list is indicative of the sorts of data that could be used.

- Child Benefit records (individual level)
- Pension records (individual level)
- Birth registrations (individual level or aggregate level)
- Second residence information (from the 2011 Census)
- 2001 Census data (individual level and/or aggregate level)
- Electoral Roll data (individual level and/or aggregate level)
- Longitudinal Study (individual level, as a reverse record check like that used in Canada)
- Community Health Index (individual level and/or aggregate level (Scotland specific))