CENSUS ADVISORY GROUP

2011 Census Coverage Assessment and Adjustment: Progress Report
Owen Abbott, Methodology Directorate

The advisory groups are invited to note the progress made since the 2008 Advisory Group paper, and to provide any comments on any aspects of the work.

Introduction

Advisory Group paper 08/05 outlined the proposed methodology for measuring coverage in the 2011 Census, which was updated for the article in Population Trends 137 in Autumn 2009 (see Abbott, 2009). This paper outlines the work carried out since the 2008 advisory group paper.

The methodology is due for sign off in late summer 2010 following the completion of the remaining key methodological components

In the summer, once the research underpinning the final components has completed, the 2008 paper will be fully revised and issued to key stakeholders. Along with this paper will be a number of supporting documents, including:
- easy to understand guides
- full details of the CCS sample size
- the Estimation Areas proposed for 2011
- relevant papers that outline the methodological development.

Stakeholders will be invited to comment on the proposed methodology prior to the sign off. The sign off is of the high level components rather than the full details, and there will be a further period of detailed development that can take on board suggestions or comments made by stakeholders.

Methodological progress since 2008

Papers outlining the research described below are available on request.

a) Census Coverage Survey Design
Completed the work to define the Census Coverage Survey design and drawn some example samples. We are proposing the following changes from 2001:

- Use of a better 'design variable' which is used to allocate the sample - this uses the 2001 coverage patterns to drive the allocation, with some constraints to avoid over-optimisation for the 2001 patterns. This skews the sample into the hardest to count areas and as a result there will be more LAs that have enough sample to be estimated directly (i.e. by themselves);
- A consequence of this is that the costs of the survey have increased beyond the allocated budget. Therefore we have agreed a sample size cut of 10%, to live within the budget. This means a national sample size of around 15,000 postcodes (still around 300,000

households). Despite this cut, the revised allocation method still provides better precision at a national level (and more consistency at local level) than the allocation method in 2001. In addition, an amount of money has been made available that will enable the sample size to be boosted in a limited number of specific Local Authorities mid way through the census field period, where the census response rate is much lower than expected (and cannot be recovered fully within the census fieldwork) and there is insufficient sample to make robust estimates. This will allow the sample size to be reactive to these extreme circumstances;

- using Local Authorities directly at the design stage to improve efficiency and make the design simpler - therefore we allocate the sample directly to LAs, which is both efficient and easier to understand. This means we have done away with forming Estimation Areas at the design stage - but will form them specifically for the estimation stage, as we will not have enough sample to make good direct estimates for individual LAs;

- the choice of how much to cluster taking costs into account - we are proposing to sample output areas and then just under half the postcodes within each output area, which is a less clustered design than in 2001 (so is more statistically efficient);

- the form of the Hard to Count index will be a 5 level index with a 40%,40%, 10%, 8%, 2% national breakdown. This is a refinement of the 2001 index, and will mean that most LAs will have around 3 HtC levels. The methodology for deriving and initial index is outlined fully in Hopper (2010).

A paper on the CCS design has been accepted by the RSS for publication in its Series A journal.

b) Matching
We have completed an external review of the matching strategy which confirms that the strategy and the methodology which is mainly that used in 2001 is sound. However, we are exploring whether we can make improvements to the probabilistic methods used. The implementations of the automatic and clerical matching elements are both underway.

c) Estimation
We have confirmed the methodology as outlined in the previous advisory group paper is the preferred option through extensive simulation. We have built an element of dependence into the simulation studies to explore the impact of this on the estimates. In summary the methodological improvements include:

- How to apply Dual System Estimation - and minimise the risk of heterogeneity bias - we proposed to apply DSE at the 'cluster' level (that is the sampled postcodes within the Output Area) by age and sex. A lot of work has gone into examining the trade offs and assessment of heterogeneity bias. See Brown and Sexton (2009).

- How to generalise the DSEs to the non-sampled population - we have decided to use a simple ratio estimator. The information from 2001 has helped us to be able to make the estimator much simpler, and we expect its performance to be better. The simple ratio estimator works well with the DSE and fits well with the intentions for making adjustments for over-coverage and DSE bias. We have also examined the performance of this approach for estimating other domains such as ethnicity, and found that it works well. See Brown and Sexton (2009).

- We have also explored the methodology for making estimates of household size, and found that the 2001 approach works well – alternatives did not perform well.

- We have carried out substantial work to examine the options for making LA level estimates, to see if the model used in 2001 can be improved. The proposal is to use a simplified and easier to explain methodology, which also gives slightly better precision. See Baffour *et al* (2010) for full details.

d) Biases in the DSE

Overcount
- We have developed our ideas for how we will estimate the level of overcount in the census. We are proposing to measure duplication through a sample rather than trying to examine the whole population, as this will provide some protection from making false positive matches (John Smith with the same date of birth). We discussed our proposals with the GSSMAC (see Abbott and Large, 2009) who were broadly in agreement with our proposed approach. We have also developed our proposals for estimating the total overcount using data from the CCS. Again, we have discussed our proposals with the GSSMAC (see Brown *et al*, 2009) who were broadly in agreement with our proposed approach. However, we have yet to test the proposed methodology through simulations.

Dependence
- We have considered and made improvements to the 2001 methodology for adjusting for measurable bias in the DSE (i.e. 'dependence'). We discussed our proposals with the GSSMAC (see Brown and Abbott, 2009) who were broadly in agreement with our proposed improvements. We have completed the development work which shows improvements in removing bias from the estimates when we have an external robust estimate of the number of households. See Sexton and Brown (2010) for full details.

Outstanding methodological work

a) Estimation Areas
We have carried out some work to explore whether contiguous or non-contiguous Estimation Areas should be used. Previously, we have thought that non-contiguous groupings would provide substantial gains for this. However, our simulation studies have shown that there is no evidence of these gains. However, we have been challenged by the UKCDMAC subgroup to provide additional evidence to show we have fully addressed this. As a result we are currently simulating all of England and Wales to provide additional analysis.

b) Coverage adjustment
Work to develop the Imputation methodology has been lower priority as it is relatively self contained, and therefore progress has been much slower than expected. However, this is now one of the main priorities over the next 3 months and we are expecting to be able to determine whether significant improvements are possible within that timeframe.

References
Abbott, O. (2009) 2011 UK Census Coverage assessment and adjustment methodology. *Population Trends*, **137**, 25-32. Available at www.statistics.gov.uk/downloads/theme_population/PopTrends137web.pdf

Abbott, O. and Large, A. L. (2009) Measuring the level of duplicates in the 2011 Census. Paper presented to 17th Meeting of the GSS Methodology Advisory Committee. Available at www.statistics.gov.uk/methods_quality/nsmac.asp

Baffour, B., Silva, D., Veiga, A. and Sexton, C. (2010) Small Area Estimation Strategy for the 2011 Census. Available on request.

Brown, J. and Abbott, O. (2008) Response to Discussion at NSMAC on Estimation of Over-Count. Note for 14th Meeting of the National Statistics Methodology Advisory Committee. Available at www.statistics.gov.uk/methods_quality/downloads/Estimation_overcount_followup.doc

Brown, J. and Abbott, O. (2009) Tackling Biases in the Dual-System Estimator. Paper presented to 16th Meeting of the GSS Methodology Advisory Committee. Available at www.statistics.gov.uk/methods_quality/downloads/GSSMAC16_booklet.pdf

Brown, J. and Sexton, C. (2009). Estimates from the Census and the Census Coverage Survey, Paper presented at the 14th GSS Methodology Conference. Available at www.ons.gov.uk/about/newsroom/events/14th-gss-methodology-conference--30-06-09/programme/population-estimates-from-the-census-and-census-coverage-survey-paper.pdf.

Sexton, C. and Brown, J. (2010). Adjusting for Measurable Dependence and Correlation Bias in Dual-System Estimation. Available on request.

Brown, J., Taylor A. and Abbott O. (2009) Over-count in the 2011 Census: Estimation Issues. Paper presented to 17th Meeting of the GSS Methodology Advisory Committee. Available at www.statistics.gov.uk/methods_quality/nsmac.asp

Hopper (2010)

ONS (2008a) 2011 UK Coverage Assessment and Adjustment Methodology. Census Advisory Group paper AG (08)05. Available at www.ons.gov.uk/census/2011-census/consultations/user-adv-groups/census-adv-groups/statistical-dev/2011-uk-cen-cov-ass.pdf