

CENSUS 2011: INDEPENDENT REVIEW OF COVERAGE ASSESSMENT, ADJUSTMENT AND QUALITY ASSURANCE

SUPPLEMENTARY REPORT, JUNE 2011

IAN PLEWIS⁽¹⁾, LUDI SIMPSON⁽²⁾ and PAUL WILLIAMSON⁽³⁾

⁽¹⁾ Ian Plewis is Professor of Social Statistics, University of Manchester.

⁽²⁾ Ludi Simpson is Honorary Professor of Population Studies, University of Manchester.

⁽³⁾ Paul Williamson is Senior Lecturer in Population Geography, University of Liverpool.

Introduction

On 21 February 2011, we produced a review for the Office for National Statistics (ONS) on the methodology for coverage assessment, adjustment and quality assurance in the 2011 Census in England and Wales. Our detailed comments were accompanied by a set of 23 recommendations which we believed would, if followed, improve the 2011 Census estimates. ONS responded positively to nearly all our recommendations on 16 March 2011. This supplementary report should be read in the light of our initial report and the ONS responses: both documents are available on the web from:

<http://www.ons.gov.uk/census/2011-census/2011-census-project/independent-assessments/independent-review-of-coverage-assessment--adjustment-and-quality-assurance/index.html>

Since the publication of our report and the ONS responses, ONS have published more material that relates to many of our recommendations. In this supplementary report, we consider these contributions from ONS in the light of our recommendations. We do so by listing the original recommendation, the ONS response followed by our comments on the most recent material.

First, however, we repeat the comment from our review about the success of the Census being dependent both on a high overall response rate and on good response rates in all areas:

“We are reasonably optimistic that, having taken account of our recommendations to develop, document and consult on specific aspects of methodology, the 2011 Census in England and Wales will provide population estimates that can guide resource allocation and social policy in the right direction for the next ten years. It must, however, be recognised that the target 95% confidence intervals set by ONS for the population counts – a maximum interval of $\pm 3\%$ for all LAs – are entirely contingent on achieving local as well as national targets for non-response.”

We are delighted to learn from ONS that early indicators of response to the Census suggest that the targets will be met. If confirmed, this will be a considerable achievement at a time of falling response rates to official enquiries.

In summary, we feel that the further procedures for Quality Assurance (QA) and adjustment significantly strengthen ONS' strategy for successful population estimation. We look forward to clear documentation and communication of the detailed methodology to all users. We urge a sceptical attitude to those comparator datasets proposed for QA which have not shown a close and consistent relationship with population, to free resources for the detailed but time-consuming QA necessary in areas where stronger checks indicate problems.

Numbered ONS documentation referred to in this report is available from the ONS website at:

<http://www.ons.gov.uk/census/2011-census/2011-census-project/independent-assessments/independent-review-of-coverage-assessment--adjustment-and-quality-assurance/reference-papers-for-independent-review.doc>

Recommendation 4.1

Provide a more timely breakdown than was achieved in 2001, at the same level of detail as in 2001, of response rates for the CCS. This will enable users to assess the robustness of the DSE adjustment.

ONS response:

Agree. This information will be provided as supporting information when the census population estimates are published in July 2012. An outline of the material to be published with the census population estimates will be published in July 2011.

Supplementary comment:

ONS have provided us with some information about the first release of Census estimates although they do not intend to publish this in detail until July 2011 (ONS32). We encourage them to include CCS response rates as

part of the first release and to clarify the distinction between information that will be provided to LAs as part of the first release from that to be provided at the second (and, if necessary, subsequent) releases. We feel that it will help the estimates to be accepted in an informed manner if as much as is practical is included in the first release, and if the standard numerical content for each LA is made available in spreadsheets with all LAs included as well as in individual LA packs.

Recommendation 4.2

Consider capturing data about household non-contact and refusal rates by CCS interviewer and analysing the data to provide additional intelligence for use in QA and ratio estimation.

ONS response:

Agree. Information about non-contact and refusal rates will be collected for every household that does not respond to the CCS (Census Coverage Survey) through the completion of CCS 'dummy forms'. Further consideration is being given to capturing information from the CCS interviewer record books about calling patterns and contact outcomes. However, this information would need to be keyed manually using additional, unplanned resource, so any exercise to capture this information would be prioritised in the areas where CCS response was lowest.

Supplementary comment:

We are pleased that ONS intend to capture this information and recommend that it is included in the first release in July 2012.

Recommendation 5.1

Set out in detail, ideally with some examples, the way in which DSE is applied to data from the Census and the CCS to produce estimates both for each age-sex group and also for each ethnic group, making it clear when the assumption of independence between the Census and CCS has been relaxed, explaining which levels of aggregation have been used and how different forms of post-stratification (for example, tenure and ethnic group) have been used to strengthen the results. The methods used for small CEs should be included.

ONS response:

Agree. We recognise the user need for this information to support their understanding and confidence of their census population estimates. We will publish a large amount of detailed information, including the points specifically made in the recommendation, about the coverage assessment and adjustment and quality assurance process. This will be done in two tranches, one with the first release and one with more detail shortly after, to ensure that preparation for the second tranche of information does not delay the publication of the census population estimates. ONS will publish the content and timetable for each tranche in July 2011.

Supplementary comment:

We note that ONS intend to provide the information we suggested and stress the importance of including some examples.

Recommendation 5.2

Publish a detailed assessment of Census-CCS matching success rates for automatic and manual matching (and for each EA).

ONS response:

Agree. This will be published as part of the second tranche of supporting information described in the response to recommendation 5.1.

Supplementary comment:

We note that ONS intend to provide the information we suggested.

Recommendation 5.3

Clarify whether the IHS is capable of being used for bias adjustment for persons in counted households.

ONS response:

Agree. Information from the IHS and other ONS Social Surveys will be used as part of the within household bias adjustment during the coverage adjustment methodology. An outline of the methodology for this component will be published in May 2011.

Supplementary comment:

We think the planned investigations set out in *ONS23* – linking Census responses to survey responses from the same households around the same time – will provide a useful independent source of within household dependence. We do not recommend relying on aggregate data on household size. The quality of the matching and the characteristics of the matched samples will, however, be important issues.

Recommendation 5.4

Given their use in subsequent outputs, give serious consideration to extending the quinary age groups to 90+, if necessary collapsing geography rather than age groups to secure sufficiently robust Census-CCS estimates amongst these higher age groups.

ONS response:

ONS will publish in May 2011 the age groups that will be used within the estimation process. The collapsing of geography and age groups during the application of the coverage adjustment methods depends on the numbers found within each of these strata during the estimation phase. As the numbers within a hard to count group, within an estimation area, within the CCS sample aged 90+ are likely to be very small it would take considerable geographic aggregation to reach acceptable levels of numbers to produce reliable estimates. Therefore ONS does not think it is prudent to commit to this now as this relies heavily on the size of the numbers found within the individual DSEs.

Supplementary comment:

We accept ONS' point about reliable estimates for small groups and agree with the rationale for selected age-sex estimation groups (*ONS27*) subject to:

- (i) Using 9 rather than 0 as the terminal digit for the upper bound where 0 is currently proposed.
- (ii) Add an explanation to the paper about how the aggregation of age groups will be handled, where necessary, including confirmation that ONS will consider spatial aggregation to ensure robust estimates of the eldest age groups.

Recommendation 6.1

Clarify precisely how ONS will integrate the separate overcount propensities identified from its Census self-match and Census-CCS matching exercises as inputs to the DSE process.

ONS response:

Agree. This work is currently being finalised and will be published in May 2011.

Supplementary comment:

ONS24 provides the requested clarification. We are satisfied with the approach outlined. However, to provide additional confidence, ONS should consider simulating a more extreme scenario where a higher overcount rate coincides with an atypically low response rate.

Recommendation 6.2

If faced with timetable or resource pressures for estimating overcount, prioritise nationwide Census self-matching of those with a stated second address, or different address one year ago (not a currently listed ONS strata) above self-matching other strata less at risk of being a duplicate within the same GOR.

ONS response:

Agree. The timetable for processing and publishing the data relies on a number of assumptions about volumes and timings for each processing step. Should these assumptions be significantly wrong, thus incurring delays, then this would be considered as a way of improving processing times. However if overcount is a significant problem, reducing processes which assess overcount may reduce the quality of the estimates.

Supplementary comment:

ONS response noted and agreed.

Recommendation 6.3

Publish estimates of the components of overcount associated with each EA and, where possible, LA, including both removal of duplicate returns from within same postcode and the net overcount adjustment arising from Census-CCS matching.

ONS response:

Agree. This information will be provided as supporting information when the census population estimates are published in July 2012. An outline of the material to be published with the census population estimates will be published in July 2011.

Supplementary comment:

Again, we recommend that this information is included in the first release.

Recommendation 8.1

Clarify the circumstances in which an LA fixed effect will be favoured, including the presence of atypical features within the EA such as substantial HE student residents.

ONS response:

Agree. An outline of the methodology for this component will be published in May 2011.

Supplementary comment:

ONS28 provides the requested clarification. We are satisfied with the approach outlined.

Recommendation 8.2

Publish the asymmetric confidence intervals derived from variance estimation in preference to symmetric confidence intervals, as those expert enough to use confidence intervals are likely to be expert enough to make use of this useful additional information.

ONS response:

ONS will publish asymmetric confidence intervals derived from the variance estimation process. The methods for estimating variance and deriving confidence intervals will be published in May 2011. This publication will also set out how confidence intervals will be reported, either alongside the census population estimates or as part of a methodological evaluation of coverage estimation and adjustment.

Supplementary comment:

ONS29 sets out the theory and application of the bootstrap confidence intervals that will, in principle, be asymmetric. We note that the simulation has been carried out at the EA level with supporting material at the LA level to follow.

Recommendation 9.1

Finalise and publish details of the planned large CE imputation process in sufficient time to receive user feedback before it is implemented and describe in the final documentation how imputation deals with small CEs.

ONS response:

Agree. The methods for imputing into small and large CEs will be included within the methods on coverage imputation to be published in May 2011.

Supplementary comment:

ONS25 provides the requested details. We are satisfied with the approach outlined, although we invite ONS to consider further the ordering of donor files prior to the start of the imputation process.

Recommendation 9.2

Identify areas and population sub-groups with significant overcount and assess the extent of any bias that might be introduced by retention of duplicates, through a comparison of identified duplicate and missed records.

ONS response:

Agree. This analysis is part of the plans to evaluate the overcount component of the coverage assessment and adjustment methodology. This evaluation will be done after the census population estimates are produced and aims to evaluate how well the methods worked and identify improvements for future use of the methods. Any significant biases found during this evaluation will be reported to ensure that users have a clear understanding of the quality and any limitations of the census data.

Supplementary comment:

We agree with ONS' response to this recommendation.

Recommendation 9.3

As in 2001, publish imputation rates for each LA, age, sex, ethnicity and intention to stay category, so that expert users can take account of these when undertaking their own analyses of Census data.

ONS response:

Agree. An outline of the material to be published with the census population estimates will be published in July 2011.

Supplementary comment:

Again, we recommend that ONS clarify whether this information is included in the first or subsequent releases.

Recommendation 10.1

In the light of our comments, identify those QA checks that are so strong that they are able to be used to improve the Census where necessary.

ONS response:

Agree. This will be included in the detailed QA methodology to be published in May 2011.

Supplementary comment:

We see some advantage in placing the comparators to be used in QA and set out in *ONS30* into three ordered categories: 'improvement' (the comparator source is strong enough to be used in an improvement to the Census estimates); 'explanatory' (the comparator source is strong enough to require an explanation for the discrepancy, even if no improvement is made); 'weak' (the comparator source is not, by itself, strong enough to question the Census). We would expect the majority of the comparators to fall into the 'explanatory' category. We would like to see an explanation of how our proposed categorisation, if adopted, is linked to the 'red/amber/green' categorisation in *ONS30* that will trigger supplementary QA. As part of this categorisation process, we recommend that more thought be given to the precise weight allocated to each age-sex group. We also recommend that *ONS31* is used to fully enumerate and categorise the proposed QA checks.

Recommendation 10.2

Provide a unified overview and detailed documentation of proposed QA methodology as soon as possible to allow users to understand its coherence.

ONS response:

Agree. ONS will publish a unified overview of the QA methodology in May 2011. Additional detail will then be provided in July 2011.

Supplementary comment:

ONS30 covers the QA issues we raised in overview form. However, for users to assess the adequacy of the QA methodology, the summary must indicate where more detailed information is available for each section. We are also pleased to learn about the research into age-sex ratios (*ONS33*) to be published later this year that should strengthen the QA process. We concur with the mechanism proposed (in *ONS34*) for distributing regional and national adjustments, should they be necessary, to LAs, based on the numbers of people or households estimated to have been already missed by both the Census and the CCS.

Recommendation 10.3

Undertake discussion in Spring 2011 to gain users' confidence in the QA and in the post-QA improvement procedures.

ONS response:

Agree. A range of engagement activities have already taken place over the past two years to build users' understanding and a further meeting is planned with the QA advisory group in late spring 2011 to discuss the outstanding elements of the QA methodology. In particular, among the key items remaining to discuss is the guidance from moving from core to supplementary QA. This work will be covered in one of the documents published in July 2011.

Supplementary comment:

We are satisfied with ONS' plans to engage with interested parties over QA.

Recommendation 10.4

Confirm that knowledgeable, impartial support would be sought from the LA concerned during QA, when encountering a difficult and unusual pattern of discrepancies between the Census population estimates and detailed QA checks.

ONS response:

Agree. ONS recognises that where LA estimates are significantly different from other sources and contain unusual patterns that cannot be explained through the supplementary analysis, then support and input from the relevant LA would be valuable and will be sought. ONS would expect this to be the exception rather than the norm.

Supplementary comment:

We welcome this response and the plans set out in *ONS30* to engage with LAs.

Recommendation 10.5

Use the strength of each QA check to prioritise QA work, giving low priority to work-intensive checks that are unlikely to be used in improving Census estimates. This priority should apply both in the current preparation for QA before summer 2011, and during the subsequent QA itself. In particular, the proposed procedures and datasets for post-QA improvement should each be fully specified and fully prepared prior to their use in QA. This does not preclude judgement and methodological refinements during the QA.

ONS response:

Agree. ONS has already acted upon this advice in refining the methodology for QA. Prioritisation on specific checks will ensure that resources are efficiently used allowing enough resource and time for a robust quality assurance of all LAs but also to ensure sufficient resource for those LAs that require more detailed supplementary analysis. The QA methodology paper to be published in May 2011 outlines the core checks for each LA. This list has been prioritised to reflect the recommendations of the review team. The procedures and preparations for the post-QA improvement options are currently being developed and are on track for implementation when the QA begins in July 2011.

Supplementary comment:

As noted above, it will be important to be precise about the value of the comparator sources to be used at each stage of QA.

Recommendation 10.6

Prepare for and expect to undertake supplementary QA in LAs with the highest indications of concern from the HtC index and field reports.

ONS response:

Agree. The planning and preparation being undertaken for the supplementary QA is already focussed on the LAs that are expected to be the most challenging, and these were the focus of the QA studies conducted in 2010.

Supplementary comment:

ONS31 is expected to cover these issues when published.

Recommendation 10.7

Ensure that the timetable for coverage assessment and QA is sufficient for the likely requirements of in-depth investigation for a minority of LAs as well as the national investigation that can only be finalised when all LAs have been processed.

ONS response:

Agree. The timetables for estimation and QA have been integrated (along with earlier processing stages) to be able to have sufficient time to assess the quality of the census results and improve them if necessary during the processing of the data. This timetable includes time for supplementary analysis on a small proportion of LAs which require more detailed investigations. Within the timetable there is an explicit period of time for considering the national estimates and any further improvements that may be identified as a result of the national level QA. If additional time is required as QA reaches its final stages then ONS will consider delaying publication of the results.

Supplementary comment:

We welcome this response.

Recommendation 11.1

The QA indicators for ethnic group should not be considered for adjusting the Census.

ONS response:

Agree. The QA sources will not be used directly as the basis for an adjustment. ONS is also considering a range of other related variables which can further support and enhance the ethnicity check, including country of birth and language spoken. Where these all point towards the possible need for an adjustment, a final view will be taken by considering an alternative stratification of the Dual System Estimate, using the ethnic group classification.

Supplementary comment:

We welcome this response which links to our comment about the categorisation of the value of comparator sources for QA.

Recommendation 11.2

The plans to target large households and HMOs in early collection should be as comprehensive as possible. Where ethnically diverse LAs with substantial numbers of recent immigrants have not provided lists of HMOs for targeting by early collectors, these lists should be sought with urgency, targeted in those LAs for which published numbers of HMOs are high.

ONS response:

Agree. A final request for HMO lists was sent to LAs with published high numbers of HMOs. One of the tasks of early collectors, who start on 21 March, is to visit addresses which are known to be HMOs and areas where there is evidence of large households. The early collectors will offer householders continuation questionnaires, additional questionnaires for new households not on the Address Register and help in completing the questionnaire. Every Census Coordinator Area will have at least one early collector, and there will be more early collectors in areas where there are large numbers of HMOs or large households.

Supplementary comment:

We are content with this response.