

Scotland's Census 2022

External Methodology Assurance Panels

Summary Note: Panel 8

Thursday 25 February 2021

Contents

PMP020: Census Coverage Survey (CCS) - Sample Balance methodology 4

PMP021: Resolve Multiple Responses Prioritisation and Resolution methodology..... 6

PSR008: Summary Report of the findings of EMAP Session 8 – Thursday 25 February 2021

1. This paper summarises the main points of discussion during the external methodology assurance panel, including overall conclusion and advisory recommendations.
2. Where appropriate, the panel's reasons for any advice that proposed methodology is not fit for purpose will be stated.
3. This paper will be published on the Scotland's Census website, following approval by the panel.
4. The methodology papers reviewed by this panel were: -

PMP020: Census Coverage Survey (CCS) - Sample Balance methodology**PMP021: Resolve Multiple Responses Prioritisation and Resolution methodology**

Head of Statistical Quality Assurance team
Scotland's Census 2022
National Records of Scotland

Email: censussqa@nrscotland.gov.uk

PMP020: Census Coverage Survey (CCS) - Sample Balance methodology

Main points of discussion

This paper sets out the methodology for assessing and adjusting sample balance for the Census Coverage Survey (CCS). The CCS is a survey which takes place after the census, with a sampling methodology aiming to provide an accurate representation of the overall population in sampled areas. However, every sampling process has an associated risk that the sample drawn may be an outlier amongst all possible samples. An unbalanced sample, that is, one not representative of the population, could result in inaccurate population estimates, so the balance of a sample must be assessed and corrected for if required.

1.1 The paper was an interesting and enjoyable read, and there was general support for the overall methodology. There was specific praise for the discussion of alternative approaches and their respective pros and cons.

1.2 The subject matter is quite complex and some sections of the paper are not straightforward to understand on first read. Some panel members found it easier to follow on a second read, and gave some suggestions for improvement, such as signposting within the paper, a greater use of diagrams and flowcharts, and the inclusion of a worked example (e.g. with figures from 2011).

1.3 Some terminology needs a clearer explanation, such as “placeholders” (examples of information taken from these forms?), “cluster-based census non-response rates estimated from DSE”, “unbalanced” (with respect to what?), “Sample bias” vs “Sample balance”

1.4 Notation needs to be checked and standardised, e.g. to make it clear whether a variable is an estimate or a known quantity; standard deviation/variance; confidence intervals.

1.5 It would be helpful to start the paper with some background involving what was done in 2011 and why. In addition, the methodology of NISRA (Northern Ireland) was mentioned but not ONS (England & Wales). This would help put the NRS (Scotland) methodology in context.

1.6 Likewise, an indication of the scale of the issue - what would be the impact of these adjustments - would also help provide context.

1.7 To assess placeholders as an appropriate proxy for non-response, DSE-based non-response rates are used. Placeholders are then used to get adjustment factors for CCS to produce better estimates for DSE. Is there a potential element of compounding bias? More justification is needed to demonstrate whether this is the case.

1.8 Similarly, DSE relies on the independence between the census and CCS. While placeholders and census respondents are two distinct groups, they may not be independent. If they are not independent, then it may have implications to use this information to adjust the CCS data, which in turn is used for DSE.

1.9 There was some discussion about alternative sources (i.e. in addition to, or instead of, placeholder info) to assess balance.

1.10 More discussion requested on choice of correlation coefficients & confidence interval threshold (Why 95% and not 99% confidence interval, given the large numbers? Why correlation of 0.5?) Could different thresholds be considered in live running and the outcomes assessed? Can the correlation in the strengths & limitations section be quantified?

Conclusion

The methodology was broadly approved, albeit with reservations about the use of placeholders, for which more justification was requested.

The panel was interested in comparisons with the methodology to be used in other UK censuses as well as the methodology used in 2011, and an indication of the scale of the issue, to provide the necessary background.

The panel also gave suggestions for improving the readability of the paper, such as the use of diagrams, and clear explanations of terminology.

Panel Advice

Tick (✓) where appropriate

The Panel's advice is that the proposed methodology is fit for purpose.



The Panel's advice is that the proposed methodology is not fit for purpose (reasons must be stated below).

Reasons for advice (if to not proceed with proposed methodology):

Chair: Katherine Keenan

Date: 21st March 2021

PMP021: Resolve Multiple Responses Prioritisation and Resolution methodology

Main points of discussion

This paper explains the details of how multiple responses from a single individual or household are resolved into a single record. There are various reasons we may receive multiple responses for a household or an individual - such as a person filling in an individual response separate to the rest of their household, partial response via the online questionnaire followed by a paper return, or two members of a household independently filling in a census return for the household. This paper looks at how we decide from which return(s) we take responses for the census dataset, and in what order we prioritise these returns.

2.1 The panel agreed that the paper was well-structured, written very clearly, and there was broad support for the methodology, with some suggestions for improvement as outlined below.

2.2 In particular, there was agreement that it is important that individual returns are prioritised, as they give people the opportunity to answer honestly, in privacy, and from their personal perspective about themselves on questions about identity.

2.3 As with other papers, the diagram showing where this process sits inside data processing was appreciated.

2.4 The use of date-of-birth was a good example of how merging responses to multi-part questions can create invalid data. However - how do we check for introduced inconsistencies when merging information from multiple responses?

2.5 The panel expressed concern that the prioritisation decision for voluntary questions (that a missing response on a higher priority return should not be overwritten) would result in a loss of information, and there were suggestions to combine the responses somehow to encode the prioritised version and the lower priority information. It was however acknowledged that this decision was meant to respect a respondent's wishes, in that a lack of response may be a deliberate decision to not provide a response to a voluntary question.

2.6 The panel asked if there would/could be any consideration into the timing of responses received in terms of prioritisation (e.g. the last response is the one which counts?)

2.7 Another question about prioritisation was about the prioritisation of online responses over paper responses, when it requires more effort to request a paper response and return it. Does this mean that relevant information in the paper form might be missed?

- 2.8 There was disagreement over the amount of information to be included in the paper. Some thought that the 2011 methodology description was too long and could be reduced to an overview of the parts of the methodology covered in this paper, and the rest replaced with links to other published papers. Others appreciated the level of detail.
- 2.9 The extension of the search for duplicates to checking within the postcode was commended.
- 2.10 The panel asked if there could be information added to show the scale of the application of the Resolve Multiple Responses methodology. Roughly how many records would this affect? How much does it reduce overcount? Would 2011 figures be comparable, considering that the extension of the search to postcodes would find more duplicates?
- 2.11 As with other papers - how does this compare to methodology being used in the England & Wales and Northern Ireland censuses?
- 2.12 What are the alternatives to this methodology?
- 2.13 It would be good to explain “duplicate person records within a postcode” with a list of possible scenarios where apparent duplicates are different people, or people living at more than one address (e.g. child with parents who live apart, people with same name on same street).
- 2.14 Where an address is split into two households, how does the second household receive information about the need to get a separate internet access code?

Conclusion

The methodology was deemed to be statistically sound, with broad support from the panel. The topic was considered to be an interesting one, with suggestions and discussions around potential alternatives for prioritisation such as the timing of a response.

The panel appreciated the importance of prioritising individual returns in order to support respondents who wish to answer separately to other members of their household, allowing them to answer sensitive questions in private.

There were some reservations about the treatment of voluntary questions, with concerns about discarding information, however it was acknowledged that including non-response for a voluntary question on a priority form was considered to be respecting the respondent's wishes.

Some additional information was requested, such as a comparison with other UK censuses, some examples of duplicate names within a postcode, and possible alternative methods.

Panel AdviceTick('✓') where
appropriate**The Panel's advice is that the proposed methodology is fit for purpose.****The Panel's advice is that the proposed methodology is not fit for purpose (reasons must be stated below).****Reasons for advice(if to not proceed with proposed methodology):****Chair: Katherine Keenan****Date: 21st March 2021**