

Scotland's Census 2022

External Methodology Assurance Panels

Summary Note PSR001: Panel 1

Wednesday 27 May 2020

Contents

1. PMP001: Estimation and Adjustment Methodology 4

2. PMP002: Census Coverage Survey (CCS) Sample Methodology..... 7

**3. PMP003: Census Coverage Survey (CCS) Sample Allocation and Reserve
Sample Methodology..... 10**

PSR001: Summary Report of the findings of EMAP Session 1 – Wednesday 27 May 2020

1. This paper summarises the main points of discussion during the external methodology assurance panel, including overall conclusion and advice given.
3. This paper will be published on the Scotland's Census website, following approval by the panel.
4. The methodology papers reviewed by the panel were: -

PMP001: Estimation and Adjustment Methodology**PMP002: Census Coverage Survey (CCS) Sample Methodology****PMP003: Census Coverage Survey (CCS) Sample Allocation and Reserve Sample Methodology**

Comments and queries welcome to:

Head of Statistical Quality Assurance team
Scotland's Census 2021
National Records of Scotland

Email: scotlandscensus@nrscotland.gov.uk

1. PMP001: Estimation and Adjustment Methodology

Main points of discussion:

The purpose of the paper is to provide an overview of the Estimation and Adjustment process, which aims to create a census database that is fully adjusted for any under or over enumeration, both at a household and individual level, at all geographies.

To create population estimates, Dual System Estimation (DSE) is applied to each stratum to estimate the level of undercount in sample areas. Ratio estimation is then used to extend these estimates to non-sampled areas at an Estimation Area level, with small area modelling being used to derive Local Authority (LA) level population estimates. Following the production of population estimates at all levels, the Census database is adjusted to account for any undercount and overcount. Synthetic households are created to reflect those missed by the Census, and synthetic people are imputed into real households as well as the synthetic households to reflect within-household undercount.

1.1 The panel agreed that the methodology was sound, with some minor suggestions made to aid clarity. These included a glossary of acronyms and descriptions of technical concepts (for example the Design Variable and bootstrapping), additional signposting across papers, flowcharts summarising processes, and a clear reference of definitions of geographical areas used.

1.2 In the absence of subsequent papers on the processes being considered, the panel highlighted that any final evaluation of the presented methodologies will be subject to the panel getting sight of subsequent papers and evaluating these elements.

1.3 The panel requested further details on Section 2.1, the limits on field staff that can work in both the Census and CCS. NRS subsequently clarified that to balance with practicality there is a maximum constraint of 25% overlap of the workforce. Further, anyone working on both operations are not permitted to work in the same area – offsetting independence violations.

1.4 Section 2.1.1 was discussed, the 10 visit maximum fieldwork constraint. NRS stated that after 10 visits the likelihood is that interviewers will either have received a response or a refusal.

1.5 Further detail was requested around the assumption of independence, although the panel was reassured that DSE assumptions are recognised to hold sufficiently for Census purposes based on precedent and the literature. Specifically administrative data and rolled forward population estimates were suggested as a form of assessment and bias correction.

1.6 Panel members discussed the feasibility of the assumption of homogeneity, and further information was requested.

1.7 The Census to CCS matching process was discussed, noting the assumption of a high level of accuracy. NRS subsequently reassured the panel by clarifying that simulations based on 2011 have provided very strong matching outcomes, ensuring that false positive and negative errors are balanced with a lack of systematic error.

1.8 The mode of collection was highlighted as a potential factor in data collection, for example telephone data capture may impact results.

1.9 The Hard to Count (HtC) index methodology is agreed by the panel to be sound, although more description of the origin of the information that created this index was requested. The makeup of the HtC index and more detailed methodology are scheduled for a future panel.

1.10 A conversation occurred around the need for contingency planning related to Covid-19, given the significant risk of this impacting on project delivery. NRS reassured the panel that this work is being undertaken currently.

1.11 Section 2.1.2 was highlighted, with panel members discussing the risk of the property listing phase resulting in an address list similar to the Census Address Register (CAR). Further detail was requested on how the risk will be mitigated.

1.12 The panel suggested including further detail on the probability matrix approach to estimate household size distribution – specifically the estimation of this, and the source of variables used.

1.13 The panel required more detail on whether the datasets including synthetic records were used for NRS purposes only or were shared externally. Flagging imputed records was suggested to distinguish them from legitimate un-imputed data, as in other longitudinal surveys.

Conclusion:

The panel agreed that the current methodology appears sound, subject to evaluation of subsequent papers and clarification on specific details.

Suggestions were made by the panel to aid the clarity of the paper, including a glossary of acronyms and more detailed explanations of technical terms; the inclusion of diagrams to summarise key processes and more signposting within all three papers.

More operational information was provided by NRS regarding constraints on fieldwork between the Census and the CCS, and the protocols surrounding the 10 visit decision.

The plausibility of assumptions underlying DSE were discussed, and more detail is required on perfect matching and the assumption of independence.

More detail was sought by the panel on the Hard to Count Index, which will be presented to the panel at a later stage.

The need for contingency planning related to Covid-19 was highlighted by the panel, particularly given the interview based approach of the CCS. Work is being undertaken by NRS on this subject.

More detail is required on the source of variables used to feed into the probability matrix approach used in the estimation of household size distribution.

The panel highlighted the need for more detail on the imputation process, specifically how synthetic records are used out-with NRS, and if they are flagged in final datasets.

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: Alan Marshall

Date: 27th May 2020

2. PMP002: Census Coverage Survey (CCS) Sample Methodology

Main points of discussion:

The CCS sample must be representative and distributed across the population of Scotland, in order to minimise the variation and ensure precise and unbiased estimates. The purpose of the paper is to investigate the sample selection to be used in the 2021 CCS, including sample clustering and size. The Relative Standard Error (RSE) is used to compare the precision associated with different sampling strategies, and is derived from both the variability of the Primary and Secondary Sampling Units (PSU, SSU) within the sample frame, and simulated runs of the estimation process. It was highlighted that some additional simulation results from further sampling fractions had been omitted – which were sent to panel members separately. These results contained the final sampling proportion chosen of 9% PSU and 17.5% SSU, with an estimated RSE of 0.1778%, and have since been included in the document.

2.1 Similarly to paper 1, overall agreement was reached by the panel that the methodology described was sound. Feedback was given on formatting and graphical representations of tables 1 and 2 were suggested.

2.2 The panel discussed presentational points of the paper, including early clarification of technical concepts, mention of stratification variables in the 2021 approach discussed in paper 3, and inclusion of 2011 RSE values for comparison as well as explanation of the benefit of lower RSE values – in terms of narrower CIs resulting in increased precision of estimates.

2.3 The panel discussed the change from Datazones to Planning Areas, and requested more information on the potential impact of this on the homogeneity of socio-demographic composition of Primary Sampling Units.

2.4 The panel requested more further detail around the creation of the synthetic CCS. The NRS explained that response rates are simulated for both the CCS and Census independently, so no independence violation is built in.

2.5 The panel suggested that altering the divisions of the Hard to Count Index could mitigate the risk of collapsing strata if the sample size is insufficient. Literature relating to this can be sent to panel members as a follow up on this point.

2.6 The panel discussed table 3 in Section 2.2.2, and based on the consistent sample sizes for each scenario, further information was requested around whether to assume random non-response patterns.

2.7 Panel members suggested including a test of significance in Section 2.2.3 to analyse the impact of sample clustering on travel times.

2.8 The panel suggested extending Table 5 to include PSUs of 10-12, to clearly demonstrate that increased PSUs decrease the RSE. NRS reassured panel members that this analysis had been conducted, and would be included within the paper.

2.9 There were discussions around the possibility of increasing the SSU proportion or the sample size, and the benefit of investigating the impact of even lower Census response rates. NRS reassured the panel that this is covered in paper 3.

2.10 The panel discussed several technical points including the possibility of implementing a non-adaptive approach (as well as adaptive approaches), and the use of uncertainty measures of estimates, that were slightly out of scope of the discussion. These points were subsequently discussed outwith the panel and have been resolved separately.

2.11 The panel discussed if different clustering had differential impacts on the KPI of Local Authority population estimates. Additional information is required.

Conclusion:

The panel agreed that the current methodology appears sound, subject to clarification on specific details. The panel suggested including flowcharts of the simulation approach, and graphical figures to illustrate these findings.

The panel discussed implications of changing from Datazones to Planning Areas on the homogeneity of socio-demographic compositions.

The panel considered altering the divisions of the Hard to Count Index to reduce collapsing of strata of insufficient sample size.

More information was requested on the modelling around field force travel times.

More detail is required on rationale for maintaining current SSU proportions and the sample size, and the panel suggested more consideration of low Census response rates (for example, contingency planning for Covid-19).

Technical comments were made from panel members that have been discussed outwith the panel, and have been resolved separately.

Extra detail is sought to demonstrate the impact of clustering on estimate precision at a Local Authority level.

Panel Advice

Tickwhere 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: Alan Marshall

Date: 27th May 2020

3. PMP003: Census Coverage Survey (CCS) Sample Allocation and Reserve Sample Methodology

Main points of discussion:

The purpose of this paper is to examine stratification options as well as how best to allocate the sample to these strata. Allocation options consist of allocating proportionally based on stratum size, optimal allocation using 2011 response patterns to minimise variance, and optimally allocating a portion of the sample according to 2021 Census return rates.

The paper also investigates the possibility of implementing an additional 20% reserve CCS sample to be used as a contingency if Census response rates appear to be critically low. This would involve extending the duration of the CCS fieldwork by around 6 days to support the additional sample.

3.1 Overall agreement was reached by the panel that the methodology was sound, subject to minor clarifications. The panel requested more emphasis on the reserve sample being implemented by extending the duration of live CCS operations to increase the overall sample size.

3.2 Feedback was provided by the panel regarding formatting, including diagrams to illustrate Section 1.2 and 2.1, and the accessibility of graphical figures.

3.3 In Section 2.1, further detail was requested around the decision that the RSE should be $\leq 0.19\%$ by the panel.

3.4 The panel suggested that increasing field force efforts to increase response rates could be considered, as well as investigating an increased sample size.

3.5 The panel required more detail around basing decisions on the flexible and reserve sample on partial Census return rates. NRS subsequently clarified that in 2011, by 10 days post census 74% of people had completed their Census, so there should be a fairly accurate impression of overall responses.

3.6 The panel requested more information around operational issues in specific areas that could impact assumptions around overall response patterns.

3.7 Further information was requested around the reason for not waiting until the end of the Census period to gather response data, which could mitigate some risk.

3.8 The panel discussed methods of ensuring consistent Census response rates. NRS subsequently clarified that the two main methods of boosting responses are using Communications campaigns, and implementing non-response follow up.

3.9 Further detail was required around potential examination of response patterns across household surveys; projecting trends from administrative data sources; or allocating based on a more contemporary source like the Scottish Household Survey.

3.10 The panel suggested that regressing to the mean could reduce the risk of over-allocating to 2011 Census response rates.

3.11 Further detail was requested by the panel around the rationale for using DSE when the ONS have changed their approach. NRS subsequently clarified that DSE has been proven to be efficient in 2011, and the ONS model may not be suited to Scotland given the smaller population.

3.12 The panel requested further information on the origin of the Digital Exclusion Index (DEI) and how recent it is, as well as how modifiers were calculated in Table 5.

3.13 Further information was requested around the random modified Census response rate scenario, including rationale for evaluation and technical details.

3.14 The panel discussed the number of randomisations conducted, and NRS subsequently clarified that 500 runs of the simulation were conducted with one version of response rate randomisation.

3.15 The panel discussed Figure 1 in Section 3.1.2, and NRS subsequently clarified that Figure 1 illustrates situations that could occur in 2021, by modelling response rate data based on various scenarios using simulations. Further work refining the simulation methodology should help further clarify the precise option, however the panel were satisfied with the simulation method used in the paper.

3.16 The panel requested further information on Monte Carlo estimates of variability in the results discussed in Section 3.2.

3.17 The panel requested further detail on the scenario where more than one area is affected in a critical way such as Glasgow, such as localised Covid-19 outbreaks.

3.18 The panel suggested that the size of the reserve sample could be a function of the scenario presenting itself 4 weeks after Census, as opposed to a static 20%.

Conclusion:

The panel agreed that the current methodology appears sound, subject to evaluation of subsequent papers and clarification on specific details. Additional emphasis was suggested to highlight the fact that the reserve sample is implemented by extending the duration of CCS live operations, as this was noted as being previously unclear.

Further detail was requested around several minor points, including operational challenges of decision making around the flexible and reserve sample in a short timeframe.

The panel requested further information regarding alternative sources that could be used as the basis of allocation, including household surveys or administrative data.

More detail was required around why NRS are using DSE when the ONS have changed their approach, however NRS provided clarification on this.

Further information was required around the source and recentness of the DEI, particularly in the context of Covid-19.

The panel requested further technical detail, relating to Monte Carlo estimates of variability and the possibility of the reserve sample size being a function of the scenario at the time.

The outcome of multiple areas of critically low response rates was queried, with Covid-19 being used as a specific example. Additional simulations would be beneficial.

Panel Advice

Tick where appropriate

The Panel's advice is to 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: Alan Marshall

Date: 27th May 2020