Scotland's Census
Shaping our future
A' dealbhadh ar n-àm ri teachd

Scotland’s Census 2022

External Methodology Assurance Panels

Summary Note PSR002: Panel 2

Tuesday 23rd June 2020

National
Records of
Scotland

Contents

PSR002: Summary Report of the findings of EMAP Session 2 – Tuesday 23 June 2020

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: -

PMP004: Developing a Hard-to-count Index

PMP005: Name Reordering Methodology

PMP006: Statistical Quality Assurance Strategy

Comments and queries welcome to:

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

Email: censussqa@nsrcotland.gov.uk

1. PMP004: Developing a Hard-to-count Index

Main points of discussion:

The purpose of this paper is to describe the methodology used to create the Hard to Count Index. The index partitions small area geographies across Scotland so the Data Collection operation can know where there may be difficulties with enumeration, and be able to estimate response rates and plan operations accordingly.

The approach taken was based on that used for the 2011 census and has been validated against rehearsal data from 2019. A multiple regression model was selected where variables include: age distribution, population aged 16-29, student population not in communal establishments, Scottish Index of Multiple Deprivation, proportion of dwellings privately rented, proportion of households that are flats, Urban/ Rural classification, proportion of people who have English as an additional home language.

1.1 The panel agreed that the methodology was sound and the paper was well written.

1.2 There was a wish that papers coming to the panel have consistent formatting for tables and graphs.

1.3 Comparisons on how other UK statistical authorities were tackling this problem were also suggested as additions to the paper. The reasons articulated in the meeting for differences in the approach adopted in Scotland seemed sound and it was agreed that the report would benefit from such reflections.

1.4 The explanatory variables were scrutinised.

1.4.1 Members asked to know whether the same variables had been used in 2011 and whether other variables such as long term illness were considered. Other social survey results were suggested for selecting variables.

1.4.2 The selection of variables for testing was partly driven by the availability of data at lower levels of geography.

1.5 It was requested that the rationale behind the segmentation into index groups be made clear in section 8.1.3 of the paper.

1.6 Caution in the use of stepwise selection of variables for the model in section 5.2 was raised for large datasets having biased R squared values.

1.7 More detail on the Census rehearsal, mentioned throughout many papers, was requested since no thorough explanation has been provided as yet to the panel.

Further justification in the paper on other potential methods was requested. The expert panel advised that a beta regression would be the standard approach rather than an OLS regression. The panel suggested that NRS may wish to actually undertake some analysis to confirm the panel expectation that the key practical results will not differ under the beta regression approach compared to an OLS regression. Alternatively, NRS might justify the OLS approach along the lines of a. previous practice and b. practice at other national statistical agencies whilst noting its value for future derivation of HTC indexes. Future work on the HTC index might explore the use of a Beta regression in more detail, particularly if later iterations enable additional explanatory variables to be included since this additional complexity may lead to differences in results across different modelling approaches.

1.8 Analysis of geographic spatial autocorrelation had not been performed on the explanatory variables and it was recommended that this was explored further.

Conclusion:

The panel agreed that the methodology is sound and was interested in comparing similar efforts of ONS and NISRA for this purpose in the paper. The panel recommended NRS consider notes where appropriate to note differences/similarities with approaches of other National Statistical agencies.

A paragraph on why the particular segmentation of scores is used for the index should be added to the paper.

Several suggestions for novel or alternative methods were made and NRS will explore the feasibility of carrying out some of these analyses in order to provide additional quality assurance for the results already obtained.

The panel recommended confirming a beta regression provides broadly the same HTC index as the OLS regression.

An explanation for variable selection was requested to be added to the paper. This might include a note on the data limitations that restrict implementation of some variables.

NRS to consider how the census rehearsal is referenced across methodology papers since it may have value in earlier papers considered by the panel and other later papers.

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: 20th July 2020

2. PMP005 - Name Reordering Methodology

Main points of discussion:

On the paper census forms there are three places where respondents need to write their names. The information for each person on one part of the form needs to match up with the information for the corresponding person on the other parts. In 2011, sometimes people appeared in different orders in the different parts of the form. This caused conflicts in processing where relationship inferences were wrong.

This paper outlines methods to look at all possible orderings for matches. A score is calculated for a reordering of names and scores are compared, by trying to minimise the score an ordering is then selected.

In 2021 the aim is for the majority to complete their census forms online thereby reducing the volume of paper questionnaires. Even so there is expected to be around 1,000 cases in 2021 from paper forms where the ordering of person information is not correct.. Even with this additional check proposed here there will still be a need for clerical review.

Where names do not match Names are matched against corruptions and misspellings. Known nicknames are also used in relevant cases.

1.1 The panel found the paper to be well written and the methodology suitable for purpose. They requested more information around some score terminologies. They also suggested moving the flow diagram to main body of text and some other drafting suggestions. They suggested the addition of definitions for 'weak and strong link' to the glossary. The panel also asked for an explanation of the remove false persons process to be mentioned in the paper.

1.2 The cost function for matching and the resulting scores were queried. Panel members wanted to know why the cost function was chosen over others and for this to be reflected in the paper. It was noted that the absence of a middle name would have an impact on matching scores but NRS explained that the relative difference in scores is used so this should be ok.

1.3 Requests for the code to be added to the paper for transparency or pseudo code to be written if the actual code is not suitable for sharing.

1.4 Details around the wider process were discussed. Including the cut off point for clerical review and when in the process that will occur. The panel also asked whether the imputed data will be flagged as such for later in the data journey and how continuation forms would impact the applicability of the algorithm. NRS explained that changes made to data is flagged and checked at multiple stages of processing and where changes can't be made with a degree of certainty, records are automatically sent for review. Continuation forms are dealt with separately and evidence from 2011 and rehearsal indicates that there was no need for corrections of this type in the small number of submitted forms.

1.5 The scale of the issue the method was trying to resolve was discussed. It was noted that this process needs only to be carried out on paper returns. This discussion prompted panel members to recommend checks on automatic changes so that errors aren't being introduced. NRS explained that there will be a quality assurance process to check the various transformations and edits to the data on its journey.

1.6 The panel sought and were provided with an explanation of the nickname database used for fuzzy matching.

Conclusion:

The panel was broadly happy and approved of the methodology.

Suggestions were made to improve the clarity of the paper. These suggestions included, adding in pseudocode if possible, definitions and descriptions of other processes mentioned, reordering of figures and small continuity improvements.

The scale of the issue the method is aiming to solve is expected to be a relatively small subset of the paper returns. The panel recommended checks to changes introduced by this process to ensure errors aren't being introduced.

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: 20th July 2020

3. PMP006 - Statistical Quality Assurance Strategy

Main points of discussion:

This paper is a high level overview of the proposals for Statistical Quality Assurance for Scotland's Census 2021. The work for phase 1 of the SQA accreditation has been completed and the programme is moving into phase 2.

The paper summarises the end to end journey of census SQA, starting at development of questions and ending with final outputs. In the secondary level of SQA checks, after every time data is changed, analysis is carried out to ensure changes haven't negatively impacted quality.

The panel is invited to comment on whether the processes laid out in this paper are going to ensure statistical quality in the intended way.

1.1 The panel liked how this paper pulls together all the high level aspects of Quality Assurance work taking place with Census.

1.2 The panel would like to see this paper serve as a high level summary and also link to further documentation on each piece of work.

1.3 The panel felt that the paper could be improved by inclusion of

- Timescales and information on when certain outputs would expect to be published
- A flow chart for each stage of QA work
- A comparison of the online and paper questionnaire
- Measures that suppliers are taking to ensure quality
- Mention of field force training
- The risk that COVID-19 poses to quality results

1.4 The Census Quality Survey was discussed. Respondents are asked to answer as if it were Census day. This was requested to be made clearer in the text in section 4.8. It was also noted that CQS responses could be biased towards people who are good and accurate responders.

1.5 Stakeholders should be provided with QA metadata such as proportion of records which have been imputed.

1.6 Signposting and guidance should also be provided to stakeholders who are wanting to use Census outputs for COVID analyses.

1.7 The tense in which the Census rehearsal is written about needs updating to a post rehearsal sense. Building upon this by adding in information from rehearsal results with a short explanation on the scope and how data collected from rehearsal is being used to improve quality

1.8 Use of the Census Address Register and its impact on quality was raised. There were concerns around households not in the list being missed out. The procedure for addresses found in the field was not mentioned in the paper but would maybe be more appropriate in a more detailed paper in the context of field force.

Conclusion:

The panel found the paper overall to be a well written and effective high level explanation of quality assurance measures. They expect to see further detail on the QA work in other papers dedicated to each part mentioned in this paper.

Members expect this paper to work as a directory for understanding the process at a high level and then linking to further detail.

NRS will be adding some QA aspects that were not included in the paper.

The way that the rehearsal has been written about needs altering from its pre-rehearsal sense, to reflect the fact that it has now happened.

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: 20th July 2020