

# **Scotland's Census 2022**

## **External Methodology Assurance Panels**

### **Summary Note PSR006: Panel 6**

**Thursday 29 October 2020**

---

## Contents

<b>1. PMP015: Census to census linking and Overcount Correction .....</b>	<b>4</b>
<b>2. PMP016: Household Record Swapping methodology .....</b>	<b>7</b>
<b>3. PMP017: Cell Key Perturbation .....</b>	<b>9</b>

**PSR006: Summary Report of the findings of EMAP Session 6 – Thursday 29 October 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: -

**PMP015: Census to census linking and Overcount Correction****PMP016: Household Record Swapping methodology****PMP017: Cell Key Perturbation**

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

Email: [censussqa@nrscotland.gov.uk](mailto:censussqa@nrscotland.gov.uk)

## 1. PMP015: Census to census linking and Overcount Correction

### **Main points of discussion:**

This paper looks at proposed methods of resolving over-counting in Scotland's Census 2022. The method proposed in this paper particularly considers how overcount caused by people being recorded in multiple locations, or in the wrong location can be resolved.

The proposed method links the census dataset to itself. The linked records are then linked to administrative data. Using these links, a probability that the record represents a distinct genuine individual can be found. The total of these probabilities will then be the estimate for the total number of individuals represented in the census dataset. The difference between this and the number of records in the dataset can be used to account for the duplicates and misplaced records in the census when producing population estimates.

1.1 The panel were impressed by the amount of work that had gone into this paper.

1.2 There was agreement from the panel that the method was sound, well-thought through and justified. However, it was thought that the paper would benefit from being re-structured and re-written in parts to help guide the reader through the paper because of how technical it is.

1.3 There was considerable discussion around how the quality of the administrative data would affect this method. In particular, the impact of someone being missing from the administrative data and, therefore, impossible to link to from the census. NRS acknowledged that this will occur as administrative data will have gaps, however, this should not be an issue as this does not affect the probability calculations. It was agreed that an explanation of this in the paper would be beneficial.

1.4 Similarly, there were thoughts on whether differences in coverage of different subsets of the population in the administrative data would affect how effective the methodology is. NRS will consider how differences in coverage will affect the methodology and add information to the paper about this.

1.5 Several members of the panel requested information about how the thresholds used were decided upon. NRS explained that the thresholds had been chosen through a clerical review of linkage using 2011 Census data and choosing a sensible value based on the findings from this. NRS will look at adding some information about this to the paper.

1.6 The possibility of using additional administrative datasets, such as birth registrations, was mentioned. NRS explained that they have not pursued legal permissions to use other administrative datasets for this purpose.

1.7 The panel suggested that some information on the uncertainty around the probabilities calculated as part of this would be useful. NRS agreed that it would be worth looking into whether standard errors could be calculated, and also explained that while the uncertainty of these probabilities is not calculated, uncertainty of the overall estimation process is calculated.

1.8 The panel thought the section looking at what is done when records link to two or three other records was incomplete and the paper was tentative about how to deal with these records. NRS confirmed that the process for these records still needs to be thought through in more detail before a final decision is made.

1.9 The panel suggested that the paper could be improved by:

- making the scale of this issue clear from the outset and how it compares to other types of error in the census
- moving the details of some of the calculations to an annex
- making it clear throughout the paper which types of overcount this method is addressing
- including information on how this new method would have affected estimates in the 2011 Census.
- expanding the section on students to include an introduction of why they are an issue and a conclusion on the implications of this section.

**Conclusion:**

The panel approved the methodology and praised the amount of work that had gone into it. The panel recommended for the paper could be re-structured so it is easier to read as it is very detailed and technical.

NRS to consider the comments on the structure and amend the paper as appropriate.

<u>Panel Advice</u>	Tick('✓') where appropriate
<b>The Panel's advice is that the proposed methodology is fit for purpose.</b>	✓
<b>The Panel's advice is that the proposed methodology is not fit for purpose (reasons must be stated below).</b>	

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

**Chair: Alan Marshall**

**Date: 19<sup>th</sup> November 2020**

## 2. PMP016: Household Record Swapping methodology

### **Main points of discussion:**

In order to protect the confidentiality of Scotland's people and households in outputs of Scotland's Census 2022, National Records of Scotland (NRS) will employ a number of Statistical Disclosure Control (SDC) techniques.

This paper looks at one of the proposed SDC techniques: household record swapping. This technique was used in the 2011 Census, however, an amendment is proposed for 2022 to provide NRS greater flexibility in specifying swapping parameters that achieve the appropriate balance between disclosure protection and the preservation of data quality. This is particularly useful as outputs from Scotland's Census 2022 will be available via a flexible table builder tool.

1.1 The panel felt the paper was well-structured, easy to follow and an enjoyable read.

1.2 The panel agreed that the method was appropriate and justified based on the information presented in the paper. The amendment from 2011 to introduce an additional stage of swapping sounded fair when it provides census data users with greater flexibility to access the data. However, it was noted that the panel was unable to assess the application of the household record swapping fully, as information about the parameters and specifics of the swapping process could not be included in the paper. This information was not included as knowledge of this information would compromise the protection afforded to the Census data.

1.3 The panel asked if it would be possible to include any information on how the uncertainty introduced to the data by swapping records affects the quality of the data. NRS do have measures to check the data utility, however, as with some of the other information that was not included in the paper this information is restricted.

1.4 The panel asked for clarification on what made the additional stage of swapping different from what was already done in 2011. NRS explained that the additional stage of swapping directly targets rare or unique records at the Output Area level (the lowest geography for outputs) while the existing method considered a range of geography levels.

1.5 The panel suggested that the paper could be improved by including:

- a flow-chart to show how the various SDC techniques fit together
- more details about the flexible table builder to provide further explanation about why NRS is proposing an additional stage of swapping compared to the 2011 Census
- more information on the doubt metrics, possibly explicitly stating that they will be updated to consider new questions in the Census

- an explanation of what makes a household 'high risk' if this would be possible without giving away sensitive information about the process.

1.6 The panel thought the section describing the testing of this technique using 2019 Rehearsal data seemed quite light compared to other papers they had seen. NRS explained that due to size of the rehearsal this testing was restricted to checking that the process worked rather than testing how effective the technique was.

### Conclusion:

The panel approved the methodology with the caveat that they could only judge the methodology at a high-level as detailed information could not be included in the paper due to restrictions on who this information can be shared with.

NRS will take the suggestions for improving the paper on board and make amendments accordingly. NRS will also consider whether there is any information that can be provided to indicate how the quality of the data is affected by Household Record Swapping.

### 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: 19<sup>th</sup> November 2020**



### 3. PMP017: Cell Key Perturbation

**Main points of discussion:**

In order to protect the confidentiality of Scotland's people and households in outputs of Scotland's Census 2022, National Records of Scotland (NRS) will employ a number of Statistical Disclosure Control (SDC) techniques.

This paper looks at one of the proposed SDC techniques: Cell Key Perturbation. This technique was not used for the 2011 Census. However, the introduction of a flexible table builder for the 2022 Census means additional SDC techniques are required. Cell key perturbation adds a small amount of noise to some cells in a table, meaning that users cannot be sure whether differences between tables represent a real person, or are caused by the perturbation.

The paper also considers whether to use perturbation of zeros and explains why NRS are not proposing to this method.

1.1 The panel found the paper interesting and it was clear that a lot of research went into it.

1.2 The panel agreed that the proposed method was sound and the justifications for using Cell Key Perturbation and not perturbation of zeros made sense.

1.3 Panel members thought that the paper would benefit from having strengths and limitations sections for both Cell Key Perturbation and perturbation of zeros.

1.4 The panel asked for more detail about the testing of the method that was mentioned, particularly who the 'users' involved in the testing were. It was suggested that the paper could also include more detail about this to give the testing more credit. NRS explained that this was internal testing and involved staff from different areas and with varying degrees of statistical knowledge and knowledge about the census.

1.5 The panel asked whether tables for the same variables at different geographies would be inconsistent if this method is used. NRS confirmed that this would be the case but that this will be made clear to users, and that the flexible table builder should make it easier for users to obtain data at the desired geography levels.

1.6 The panel asked if Cell Key Perturbation could undo some of the record swapping that is done as part of other SDC methods. NRS confirmed that cells affected by record swapping could also be perturbed, but that it would not undo the swapping as the record would not return to where it was originally.

1.7 Panel members asked why the proposed methods for SDC were preferred to rounding figures. NRS explained that rounding had been used by other censuses in the past and this had been unpopular with users.

1.8 There was some discussion around whether census data users would be able to work out the cell key table by creating a large number of tables of census data via the flexible table builder. It was agreed that this would be difficult to do, and NRS are considering whether users have to agree not to do this before using the flexible table builder and whether to impose restrictions on the number of tables that can be generated over a specific time period.

1.9 The panel also suggested that the paper could be improved by including:

- a Plain English abstract as this has been beneficial in other papers
- less detail in the abstract around the limitations of perturbation of zeros
- more detail on the background of what SDC is, why it is needed and what is different about the 2022 Census that means different techniques are needed
- more information on how the perturbation table is created.

### Conclusion:

The panel agreed that the proposal to use Cell Key Perturbation and not to use perturbation of zeros was sound.

NRS to take comments on the structure and formatting of the paper on board and amend accordingly.

### 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: 19<sup>th</sup> November 2020**