

# **Scotland's Census 2022**

## **External Methodology Assurance Panels**

### **Summary Note: Panel 7**

**Tuesday 26 January 2021**

**Contents**

**PMP018: National Adjustment methodology ..... 4**

**PMP019: Administrative Data-based Population Estimates methodology ..... 8**

**PSR007: Summary Report of the findings of EMAP Session 7 – Tuesday 26 January 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: -

**PMP018: National Adjustment methodology****PMP019: Administrative Data-based Population Estimates methodology**

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

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

## PMP018: National Adjustment methodology

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

The purpose of the National Adjustment methodology paper was to give an overview of the National Adjustment strategy for Scotland's 2022 Census, should there be a need to implement one, and details on how this would fit within the Estimation process.

Estimation is a statistical process that is used to produce population estimates, run at an "estimation area" level (a type of geographic grouping). The aggregate estimation areas will form a national estimate. Once produced, these estimates are also checked against other comparator sources, to ensure they are in line. If not, a National Adjustment may be applied.

In 2011, both ONS (the Office for National Statistics, who administer the England and Wales Census) and NISRA (Northern Ireland Statistics and Research Agency) made a national adjustment in their Census data to account for bias that may arise in calculating the estimates - bringing ratios of males to females within particular age groups in line with these comparator sources. It was deemed unnecessary for Scotland to do so in 2011, but there remains a possibility a similar method will be required for the Census in 2022.

The paper outlines an approach to how Scotland's Census will determine if national adjustment is necessary in 2022, and what can be done to implement one.

1.1 The panel was pleased with the quality of the paper, and comments were made on how enjoyable they found the presentation. They appreciated that the history of the methodology and comparison with other Census offices were explained, as it gave context and depth to the issue at hand.

1.2 The panel expressed concerns around the use of Scottish Longitudinal Study (SLS) as the main comparator source mentioned in the paper. There were questions on whether the use of other comparator sources would be better suited, such as commercial sources. NRS explained they have so far found the SLS to be the best fit, as it is "real" or unadjusted data, and that ONS and NISRA are using their equivalent to the SLS so there is a harmonised approach. However, NRS also stated that although they haven't yet been specified, a range of other sources will be considered, and that work was ongoing in this regard. Additionally, the Statistical Quality Assurance team (SQA) will have input on the appropriateness of the comparator sources used. It was suggested that the paper would benefit from having such explanations included, including more about the data that comes from SLS and its potential bias.

1.3 Part of the concern expressed by the panel was that the SLS uses a 5% population sample while a Census was 100%, so other comparator sources closer to

the mark were suggested, such as data from the Department of Work and Pensions (DWP), or Mid-Year Estimates. NRS responded that although they are considering other sources of data, they did not have currently have access to DWP data. Additionally, ONS also did some work on mid-year estimates, and it was found that they were not appropriate for a comparator source in national adjustment as the data is adjusted based on the previous Census. A discussion was had about using a comparator source with potential bias, as it risked propagating or compounding the bias. As with point 1.2 above, such comparisons should be included in subsequent drafts of the paper.

1.4 Another consideration that the panel had about the SLS was the timeliness of up to date SLS data. Some members expressed personal experience with the lag in time that it takes to obtain up to date data. NRS acknowledged that the data refresh has, in the past, been slower to update than would be preferred. However, they have been in contact with the NRS SLS Project Manager about their potential use of the data for the Census, and he seemed confident that the delay will be minimised by Census day.

1.5 The panel felt that it was good that the approach would be naturally reticent about making a national adjustment, and that it showed in Scotland's decision in 2011 not to make the adjustment while ONS did. However, the paper did not compare the scale of the issue between Scotland and England and Wales, so it was difficult to get a feel for differences or similarities of the problem - for example, what were the respective targets? NRS indicated that as part of 2011 quality assurance there is published documentation which explained the evaluation of national adjustment when ONS chose to implement it and NRS did not. NRS will add pointers to this information in the paper.

1.6 The panel had a question about the breakdown from the national estimates to lower geographies and how it would calibrate to the undercoverage correction that the Census Coverage Survey (CCS) is used for, and if you could check the differential response between men and women through CCS. NRS indicated that CCS will still be required to account for most undercoverage, but that national adjustment is looking for residual bias. One option, for example, would be to look for areas where female response rates were lowest and check if there was a mismatch between sex response rates. If it is not as much as expected, it might be because there are more males missing in those areas. While these checks aren't set, they are the types of things that NRS would look for when considering where to place people in lower geographies. Further information will be included in the paper.

1.7 There was some confusion expressed about the graph included in the paper, and some suggestions were made for improvement:

- Why was the 20 - 49 age range used as an example? The paragraph prior seems to be comparable at an age range up to 60, so some clarification would be appreciated.

- Why are the estimates in this range weighed? This could be clarified as well.
- Had other ratios been considered?

NRS indicated that this information was used to illustrate why/where ONS used national adjustment in 2011 (and hence why NRS may need to in 2022). All age groups will be looked at, not just those in the paper, and clarification will be made.

1.8 A suggestion was made by the panel about the confidence interval used on page 10, and that clarification was needed to understand that it was used in ONS’ methodology (as not a ‘true’ confidence interval). Clarification will be made in the paper.

1.9 It was suggested by the panel that some local authorities may challenge the addition of people to certain areas. NRS indicated that part of the quality assurance process will be focused consulting local authorities as part of the quality assurance process.

**Conclusion:**

The panel was largely happy with and approved the methodology. However, suggestions were made to improve the clarity and understanding of the paper, primarily around the comparison between ONS and NRS, and including more detail about the primary source proposed (SLS).

Suggestions were also made in regard to other issues that may need to be considered, such as the use of other comparator sources, and potentially practicing the method using synthetic data.

**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: 17<sup>th</sup> February 2021**

## PMP019: Administrative Data-based Population Estimates methodology

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

The Administrative Data-based Population Estimates (ABPE) paper details the methodology that the National Records of Scotland (NRS) has used thus far in testing the feasibility of creating admin data based population estimates without using a census as a starting point. The process, so far, has produced an overall overcount compared with mid-year estimates). NRS is looking to continue to iteratively develop the methodology and consider alongside other evidence when making recommendations for future censuses.

The purpose of bringing the paper to the panel was for slightly different reasons than other Census papers. Although the team that worked on this methodology is part of the Census programme, it is not directly Census focussed, and the ensuing discussion was around input on future improvements to the methodology rather than approval, as the paper is already published. The team welcomes all comments and suggestions for improvements that can be made to the methodology.

1.10 In the paper, the panel found only mention on public-sector administrative data sources. They wondered if other data sources, such as commercial sources, had been considered, although they acknowledged the potential need to pay for them. NRS mentioned that they had initially started considering commercial sources, such as credit scores (Experian), but that speaking with the public panel, the preference was for the use of public-sector data sources. There were concerns with the linking of commercial data to public sector data.

1.11 Another potential source that the panel considered was (the previous) Census data, at least as a first step. NRS acknowledged that they deliberately did not use Census data as the original goal was to “see what happens if we don’t have a Census”. Additionally, mid-year estimates were not used because they depend on the Census, and the goal was to compare with the next Census

1.12 A suggestion that the panel made was in regard to holding accessory or other types of coverage surveys, to update/check linkages from administrative data sources (in cases where these sources may be less dependable). NRS was considering the use of surveys.

1.13 The panel felt that the linking methodology was strong, but wondered how it compared to ONS and NISRA and if Scotland was likely to adopt a similar approach. NRS indicated that ONS have considered ‘stocks’ approaches (where each year is considered in isolation) and ‘flows’ approaches (where changes are made to a base year depending on known movements of individuals). Once NRS have multiple years available the relative benefits of these approaches will be explored.



1.14 The panel felt that assigning penalties for certain postcodes felt a bit arbitrary, and wondered if the method would benefit by doing some statistics/probabilities based on clerical review. NRS acknowledged this is an area to work on, but found that it was a bit difficult to assign on this basis. Penalties need to be scored based on comparability — so for example, what would the scores look like for two people in different postcodes with the same name, versus two people in different postcodes with slightly different names? For this reason, scores were made informally rather than rigorously, however, NRS recognises that this is an area that may need further review.

1.15 One idea the panel had for the type of data used was for SMR01 or extended CHI. NRS explained that the health activity data used was essentially a combination of activity on SMR00 (outpatients), SMR01 (inpatients), SMR06 (cancer) and other data from the last 3 years put together — so the data used in the paper essentially used SMR01. However, NRS did also acknowledge that they would like other sources of data to use, such as DWP (pensions), as it would provide good coverage on groups they would like to focus on.

1.16 The panel commented that it would be useful to see the analysis of the threshold of 25, to see how helpful it was in choosing the threshold. NRS indicated they had looked at the threshold, but it was not very sensitive if you go up or down a bit — but that it is an area of further work.

1.17 The panel had a question about how it is determined that estimates are “good” or not — for example, the Census has Key Performance Indicators (KPIs). The ABPE does not currently have KPIs. By publishing the accuracy (as indicated through comparisons with the mid-year estimates) NRS can facilitate discussions with users around whether the level of accuracy achieved meets their needs. NRS are planning to hold stakeholder engagements in the spring.

1.18 Related to point 1.9 above, the panel felt that using mid-year estimates as a measure has difficulties as there tends to be a drift in mid-year estimates over the ten year interval. NRS acknowledged that some work needed to be done around measuring the drift of the estimates (as they are based on Census as well).

1.19 A question that was raised by the panel was on the rule of insisting that people were on NHSCR; what happens if you could find them only on different datasets? NRS felt that the problem was continually adding people from datasets could produce inflation, but also felt that the business rules need to be looked at in closer detail.

1.20 The panel also brought a question about the “last 3 years” cut-off that was used in the data. They felt that this was a long time (as it provides people with more opportunity to move around), and wondered if a sensitivity test could be one on, for example, the last 6 months. NRS indicated that one of the issues here was in the way that the data was linked — it was necessary to include the last 3 years due to an

unexpected data issue before publication . Because of this, it will be an area for future development aiming to be updated in their next methodology review.

1.21 The panel wondered about how other countries (aside from ONS) are approaching such issues, and a brief explanation was given on countries who belong to the International Census Forum (ICF), who have Census procedures similar to the UK's. It was acknowledged that this has become an international trend, as more countries look to benefit from already-held sources of data to overcome issues that a traditional Census has (for example, pandemic, or natural disaster situations). However, no countries had yet to fully implement an administrative Census, although more countries were incorporating uses.

1.22 There was a question raised by the panel on the sustainability of data sources, and if sources such as Google, Facebook or Twitter could be used — and if so, if biases could be compensated for. NRS has considered these methods (such as webscraping), but has yet to take a view.

1.23 The panel had a comment about the heaping of links across the distance scores and the robustness of this method. NRS indicated that the issue was where some datasets did not have complete information for matching. For example, the electoral register does not have date of birth, so the scores of links involving it had to be manually adjusted to ensure that they were not discarded by the threshold.

### Conclusion:

The panel felt that the paper detailed some complex methodology and was overall enjoyable and explained clearly.

Much of the discussion centred around the data sources that could potentially be explored, as well as the question of sustainability of such sources, as ideally a persistent database would be used. Throughout the session, the larger implications of discussion circled back to how administrative population estimates could be used, and how it could support the evidence base for future Censuses. Its main problem is that the characteristics that the Census normally captured would be lost as there is currently not an administrative dataset that captures all this information. However, it might serve as a useful and complementary dataset that may become a new type of population estimate, as some things can also be captured that a Census cannot, such as more-frequent updates.

### 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: 17<sup>th</sup> February 2021**