



Office for
National Statistics

**4th Meeting of the Methodological Assurance Review
Panel**

Agenda & Minutes

**17 & 18 October 2019, Office for National Statistics
London**

Office for National Statistics Methodological Assurance Review Panel Meeting

Agenda

17 & 18 October 2019

Drummond Gate London

Chair: Sir Bernard Silverman

Day 1

Time	Item	Presenter
1 10.30-11.00 45 mins	Introduction & Actions from previous meeting	Rachel Skentelbery
2 11.00-12.30 90 mins	Communal establishments	Cal Ghee
12.30-13.00 30 mins	Lunch	
3 13.00-14.30 90 mins	Census to Census matching and Overcoverage (EAP121)	Rachel Shipsey
14.30-14.45 15 mins	Break	
4 14.45-16.15 90 mins	Statistical Design for 2021 Census combined approach (EAP122)	Cal Ghee
16.15-16.45 30 mins	Summary & Actions	Rachel Skentelbery

Day 2

Time	Item	Presenter
09.00-09.15 15 mins	Introduction & Opening	Rachel Skentelbery
5 09.15-10.45 90 mins	Census Coverage Adjustment (EAP123)	Alison Whitworth
6 10.45-12.15 90 mins	Hard-to-reach populations (EAP124)	Jon Wroth-Smith
12.15-12.45 30 mins	Lunch	

7 12.45-14.15 90 mins	Statistical Disclosure control (SDC) (EAP125)	Keith Spicer
8 14.15-15.45 90 mins	Address list quality – Charlie/Mike	Charlie Dacke/Mike James
15.45-16.15 30 mins	Summaries & Actions	Rachel Skentelbery

Attendee List

External Panel Members

Sir Bernard Silverman (Chair)
Prof Natalie Shlomo (External Panel Member)
Dr Nik Lomax (External Panel Member)
Dr Oliver Duke-Williams (External Panel Member)

Office for National Statistics

Rachel Skentelbery (Vice-Chair, Chief Methodologist)
Cal Ghee (ONS Panel Member)
Owen Abbott (ONS Panel Member)
Sarah Henry (ONS Panel Member)
Jon Wroth-Smith (ONS Panel Member)
Jennet Woolford (ONS Panel Member)
Alison Whitworth (Presenter)
Charlie Dacke (Presenter)
Iain Dove (Presenter)
Jas Gakhal (Presenter)
Kirsten Piller (Presenter)
Rachel Shipsey (Presenter)
Viktor Racinskij (Presenter)
Gareth Powell (Secretariat)
Chris Lydiat (Secretariat)

Apologies

Prof David Martin (External Panel Member)

Actions

Agenda item¹	Action
[4,1]	A51 – The creation of IDs to identify items for the panel was agreed, with the inclusion of previous items if necessary.

¹ [M,N] – M denotes the panel number, N denotes its position on the agenda.

[4,2]	A52 – The panel would like a more detailed update on Communal Establishments in the future, including imputation and use of administrative data.
[4,3]	A53 – ONS to consider modelling approaches for the census to census matching to reduce the need for clerical reviews.
[4,4]	A54 – The panel would like to see the census scenarios framework, including the risk assessment, mitigations and the ONS Board's view on these.
[4,5]	A55 – Review consistency between household weights and individuals in benchmarking to make sure that totals are consistent, especially across household size.
[4,5]	A56 – Create the most suitable optimisation algorithm for Combinatorial Optimisation and ensure this method documented and reproducible.
[4,5]	A57 – Consider how to archive the census production systems and retention of methodology documentation for reproduction in future.
[4,6]	A58 – Review the wording of the census questionnaire with respect to “sofa-surfers” and ensure cognitive tests support the wording proposals.
[4,6]	A59 – Investigate ‘property guardians’ to account for individuals who usually live in non-residential property, as they may be missed because the address register does not classify them as residential.
[4,7]	A60 – The panel requested a more in-depth conversation into the perturbation methodology.
[4,8]	A61 – Present an evaluation report on the 2019 Census rehearsal to a future panel.

Minutes

1. Actions from the previous meeting

A51 – The creation of IDs to identify items for the panel was agreed, with the inclusion of previous items if necessary.

1.1 – Item 4 on the agenda covered work on risk to census operations, closing action A45.

1.2 – The Census Coverage Adjustment session under item 5 covered points raised on optimisation in action A16, testing the strategy as raised by action A17. These actions were agreed to be closed.

1.3 – The addition of international practices and expanded data sources in hard to reach populations presented at item 6 closes A5, A6 and A22. This item also covered intentions for counting homeless individuals, closing A7 and A20.

1.4 – The inclusion of the SDC paper at item 7 closes actions A2, A3 and A4 that requested information on transparency, record swapping and perturbation of geographies.

1.5 – A conversation on K-means occurred in relation to Census Statistical Design, closing action A44.

1.6 – ONS updated the panel on VOA address lists, stating as VOA data is used to build the Census address frame, every current VOA address should be present. Further VOA/council tax data is used to create 'signs of life' to assist in determining if an address is residential. This update closes actions A48 and A49.

1.7 – The methodology used for VOA data has been changed from Cramer's V, closing A50.

2. Communal Establishments – Jas Gakhal, Jon Wroth-Smith, Gareth Powell

Background

The presentation detailed the collection design for communal establishments (CE) and the methodology of estimating small and large CEs. The design includes different collection methods for different types of CE. Not all CE types were being tested in the 2019 Census rehearsal. The methodology to account for non-response in small and large CEs was also discussed, with use of administrative data for large CEs highlighted.

Discussion & Suggestions

2.1 – The definition of a CE was queried by the panel to ensure they were satisfied that CEs are not counted as part of the housing stock.

2.2 – The hand delivery of questionnaires was queried, with the panel informed that this ensures greater ease of access for potential follow up visits. The panel recommended that ONS consider which types of CE need to be rehearsed or followed up with experts. The panel also recommended that ONS consider how to validate processes that are not tested in the rehearsal.

2.3 – The status of private owned university halls was queried and established as matching the status of public halls from an enumeration procedure perspective.

2.4 – How conflicts between ONS and respondents' classification of a CE are dealt with was discussed. It was explained that respondents could request a different form, and this would update the classification on ONS systems.

2.5 – Multiple matters relating to prisoners were discussed, including the definition of usual residence and the preservation of confidentiality within the prison system. The panel were told that following advice from MoJ, ONS defines the usual residence as the prison if a sentence is longer than 12 months, which is in line with the standard definition. ONS also acknowledged that information from prisoners has been less reliable historically. The panel recommended ONS consider further mechanisms to ensure prisoner confidentiality is assured.

2.6 – The panel queried the definition of large and small CE's and recommended ONS consider if the chosen 50 bed-spaces cut off for the Census Coverage Survey is the most

appropriate threshold. Investigations into the quality of administrative data for small CE's was also recommended.

2.7 – Small CE geographic distribution was discussed, leading to the panel recommending ONS consider how the non-uniform nature of this impacts outcomes when using the current CCS sampling methodology.

2.8 – The use of donors for imputation was also discussed, with the panel satisfied with the selection within the same CE process but noted known issues regarding small donor pools and repeated use of donors. It was recommended that the ONS continue to research potential overuse of donors.

Actions

A52 – The panel would like a more detailed update on Communal Establishments in the future, including imputation and use of administrative data.

3. Census to Census Overcount Matching – Rachel Shipsey, Viktor Racinskij

Background

Census to census matching aims to find individuals who have been included on the census multiple times at different addresses. Common occurrences include individuals with multiple houses, students living away from their parental homes during term-time and children of separated parents. A two-step process using inverse sampling followed by an algorithm to determine if a case is a real match, false match or requires clerical review forms the core of this method. The matching outcomes feed into the assessment of over-coverage.

Discussion & Suggestions

3.1 – A discussion around the tolerances of the algorithm was initiated, with the panel concerned that rejection of a match if the exact date of birth was not the same was too strict. They were informed that this issue had been examined, and that the loosening of this condition caused a too large rise in false positives (due to the birthday problem).

3.2 - A second discussion into the application of the algorithm was initiated. The algorithm itself was well regarded but its application was questioned. The panel suggested that the algorithm could be modified to produce probabilistic outcomes, and these outcomes could then be applied to confirm a duplicate without the need for clerical review. This might reduce the need for clerical matching even further.

3.3 – The definition of a 'visitor' on census night was also briefly discussed, with focus on children of separated parents. The panel were assured that over-count of children was known and addressed through the current method.

Actions

A53 – ONS to consider modelling approaches for the census to census matching to reduce the need for clerical reviews.

4. Statistical Design for 2021 Census: combined approach – Cal Ghee

Background

The combined approach for 2021 census is the methodology for mitigating risks to quality when certain assumptions that underpin the statistical design are stretched. In particular, low response rates and mitigations for possible scenarios that may cause them were discussed.

Discussion & Suggestions

4.1 – An issue around how to communicate use of administrative data was raised. It was agreed that the messaging required care to mitigate against the risk of response rates falling due to perceived use of or linkage to administrative data.

4.2 – Mitigations around large-scale incidents followed this, where options to temporarily halt collections in the affected area were discussed, along with potential methods of collecting data after the incident. The panel requested ONS consider the feasibility of collecting data after the planned data collection exercise was completed as a mitigation strategy for such an event.

4.3 – A discussion of the risk of widespread low response followed this, with discussions focusing around messaging to the public to ensure responses are at high enough rate and quality for the subsequent statistical methods to be robust. Discussions also noted the potential use of administrative data as further mitigation.

4.4 – Following these discussions, the panel recommended the creation of a mitigation framework for each scenario. The panel also stated that they would only form a part of the review for this, as mitigation would require assurance from areas outside of methodology, such as communications and at board level.

4.5 – A final discussion was held where the possibility of enumerator redeployment was considered. It was suggested that the ONS research on the quality of administrative data may be used as a factor to decide which areas enumerators focus on raising response rates in.

Actions

A54 – The panel would like to see the census scenarios framework, including the risk assessment, mitigations and the ONS Board's view on these.

5. Census Coverage Adjustment – Kirsten Piller, Alison Whitworth

Background

An update on Census Coverage Adjustment was presented to the panel. The previous three stage system was compared with the new two stage system and research into both discussed.

Discussion & Suggestions

5.1 – The panel questioned the retention of all census responses and only adding to them. They were informed that over-coverage was assumed to be low, as this had been the case historically. In addition, there is not enough robust information to decide which records to remove. An additional suggestion was to use a larger area to ensure over-coverage was a lesser issue.

5.2 – In discussion of the two-stage research the panel recommended that in addition to pursuing this approach, the three-stage approach be retained for possible use. In addition, the panel recommended that a scenario of widespread over-coverage be explored.

5.3 – The committee also queried how bias adjustment was performed. The panel accepted the ideal method would be to run the systems multiple times and select the benchmark with the lowest error, as this would provide the best fit. ONS will now consider the ideal implementation of this.

5.4 – The committee also questioned the definition of household in houses of multiple occupancy. A potential mitigation of using single person households was suggested.

5.5 There was a technical discussion about the current optimisation algorithm used within the Combinatorial Optimisation methodology and the best algorithm for this case, and it was recommended more research was done to establish this.

Actions

A55 – Review consistency between household weights and individuals in benchmarking to make sure that totals are consistent, especially across household size.

A56 – Create the most suitable optimisation algorithm for Combinatorial Optimisation and ensure this method documented and reproducible.

A57 – Consider how to archive the census production systems and retention of methodology documentation for reproduction in future.

6. Hard to Reach Populations – Jon Wroth-Smith

Background: Hard to Reach Populations focused on individuals that have been historically harder to count in the census. An overview of the Address Based Design was presented along with a number of methods aiming to reach all individuals in the census.

Discussion & Suggestions

6.1 – The panel raised concerns around the wording of the questionnaire. They questioned if people who were sofa-surfers would indicate they have no fixed address, as they may view their situation as temporary.

6.2 – The panel were told that no census night counts were planned to count rough sleepers. The reasons for this include the cost, health and safety considerations and the low data quality.

6.3 – The panel suggested there would be benefits of using local authority data to inform enumeration of rough sleepers, and that this could be part of the strategy.

6.4 – The panel also emphasised the need for careful messaging over the census count of rough sleeping. Care was needed as the census would only include those who had completed the census, and as this was likely to be low in homeless individuals, the messaging alongside the outputs needed to reflect this.

Actions

A58 – Review the wording of the census questionnaire with respect to “sofa-surfers” and ensure cognitive tests support the wording proposals.

A59 – Investigate ‘property guardians’ to account for individuals who usually live in non-residential property, as they may be missed because the address register does not classify them as residential.

7. Statistical Disclosure Control – Iain Dove

Background

Statistical Disclosure Control is the design of mechanisms to prevent census outputs being used to identify individuals, or characteristics of individuals. Business rules, the table builder and the use of record swapping were all presented to the panel.

Discussion & Suggestions

7.1 – How the table builder generates outputs was questioned. ONS clarified that given the same inputs, the table builder will always generate the same tables.

7.2 – The panel noted that there had been reconstruction attacks on other nation’s census data, with the main weaknesses being the number of dimensions available to each request. The main mitigation for this was highlighted as ensuring business rules for the table generator are robust.

7.3 – The level of transparency of the business rules was also discussed. The panel queried the level of benefit obscuring the specific rules would provide against a determined attacker. It was recommended to reconsider the level of detail disclosed about business rules.

7.4 – Record swapping was also discussed, with the panel noting the known issues that swapping records lead to lower quality of data at small geographic areas as the trade-off for adding uncertainty to provide disclosure protection.

7.5 – The tiers of output release were also discussed, with the panel establishing that three levels of detail would be available based on security and business requirement, with a small percent of the population available at higher levels. This led the panel to raise an issue on inference. It was recommended that ONS ensure that individuals with access to the small amounts of more detailed data cannot use this knowledge to infer details of individuals perturbed or suppressed in the public tables.

7.6 – A discussion into the technical details of perturbation was initiated, with the panel noting that the current method was close to achieving differential privacy².

² Differential privacy is a concept where information of a group can be disclosed without disclosing information on any individual within the group.

Actions

A60 – The panel requested a more in-depth conversation into the perturbation methodology.

8. Address Quality in the Census Frame - Charlie Dacke, Jon Wroth-Smith

Background

ONS aim to ensure the addresses used for the census are as close as possible to all residential locations. The method for achieving this involves bringing together multiple address datasets from multiple government sources.

Discussion & Suggestions

8.1 – Information on address base and council tax records was presented, with their matching rates discussed with the panel. Reasons for differences between them included the way individual local authorities coded properties. The panel expressed the importance of local authority intelligence in the address base.

8.2 – The near perfect match rate between the address base and the patient register was discussed, with the panel stating that the address base and patient register most likely share sources to achieve such a high matching rate.

8.3 – Research into the process on confirming addresses was explained, with administrative data and desk-based methods discussed. The panel raised that commercial real estate sites hold datasets that may be available for use, so an additional automated matching system could be used.

8.4 – Proximity of non-response areas was then discussed, with the panel recommending that clustering be assessed from the rehearsal and discussed at a future EAP.

Actions

A61 – Present an evaluation report on the 2019 Census rehearsal to a future panel.