



Office for
National Statistics

**2nd Meeting of the Methodological Assurance
Review Panel**

Agenda & Minutes

**19 & 20 March 2019, Office for National Statistics
London**

Office for National Statistics Methodological Assurance Review Panel Meeting

Agenda

19 & 20 of March 2019

Drummond Gate London

Chair: Sir Bernard Silverman

Day 1

Time	Item	Presenter
1 10.30-11.15 45mins	Introduction & Actions from previous meeting	Rachel Skentelbery
2 11.15-11.30 15mins	Point from last meeting on whether the questionnaire reflects the primary need to count the number of people	Rachel Skentelbery
3 11.30-13.00 90mins	Matching strategy for Census, linkage and population spine (EAP107)	Rachel Shipsey, Pete Jones
13.00-13.45 45mins	Lunch	
4 13.45-14.15 30mins	Remove False Persons (EAP108)	Scott Redgwell
14.15-14.30 15mins	Break	
5 14.30-16.00 90mins	Process Flow (EAP109)	Steve Rogers, Cal Ghee
16.00-16.30 30mins	Summaries & Actions	Rachel Skentelbery

Day 2

Time	Item	Presenter
09.30-09.35 5mins	Introduction & Opening	Rachel Skentelbery
6 09.35-11.05 90mins	Use of alternative data for 2021 Census imputation and Imputation of different modes separately (EAP110)	Fern Leather
11.05-11.20 15mins	Break	

7 11.20-12.50 90mins	Estimation of stocks and flows and International migration flows with admin data “at core” (EAP111)	Becky Tinsley
12.50-13.30 40mins	Lunch	
8 13.30-14.00 30mins	Overcoverage (EAP112)	Viktor Racinskij, Ceejay Hammond
9 14.00-14.30 30mins	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)
Peter Jones (Presenter)
Rachel Shipsey (Presenter)
Scott Redgwell (Presenter)
Steve Rodgers (Presenter)
Fern Leather (Presenter)
Becky Tinsley (Presenter)
Viktor Račinskij (Presenter)
Ceejay Hammond (Presenter)
James Redmore (Secretariat)

Apologies

Prof David Martin (External Panel Member)

Actions

Agenda item¹	Action
[2,1]	A19 – The panel asked for more detail on why both types of perturbations are necessary
[2,2]	A20 – The panel discussed how homeless people will be counted and if there should be a separate survey to count them. Perhaps asking MHCLG to provide any data they have on homeless populations. The Panel requested a paper outlining work on this area.
[2,2]	A21 – The panel asked to see the online census questionnaire, and to be assured that ONS has been checking international practices in this area. This will be brought to the June panel meeting.
[2,2]	A22 – The panel asked if other potential datasets have been looked at to estimate the uptake of internet, for example DVLA and the electoral register data.
[2,2]	A23 – The panel requested if there is information regarding international practices of the CCS design and experiences of overcoverage in administrative sources, for example, in New Zealand.
[2,2]	A24 – The panel requested a dedicated future session on coverage adjustments, imputation of whole records and optimization methods.
[2,3]	A25 – The ONS are to investigate machine learning possibilities – speak with the Data Science Campus & the Alan Turing Institute.
[2,3]	A26 – The matching team need to re-evaluate the ‘Priority order’ used for the census to census matching – it is necessary to look at whether this order is sensible. With regard to the groups deemed most likely to be overcounted, the matching team need to confirm that the hierarchical order is sensible and articulate the impact of changing ‘people who responded online’ used in 2011 to ‘people who responded on paper’ for 2021.
[2,3]	A27 – Consider utilising households as a hierarchy within probabilistic matching. Explain further how associative blocking can be used within the probabilistic matching algorithm.
[2,3]	A28 – Deterministic linkage assumes that all match variables have the same weight. Consider ‘weighting’ variables within the RMR (resolving multiple responses) process.
[2,3]	A29 – Consider whether the manual checking of the linkage can be improved through better software to facilitate the process.
[2,4]	A30 – The RFPs team are to provide the distribution of removed date of births.

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

[2,4]	A31 – The RFP team to consider the proposal to use a probability based record linkage.
[2,4]	A32 – The ONS are to consider transparency, that is after the completion of the Census
[2,5]	A33 – The team to demonstrate what is being done to optimise overall response and digital response. The team are to provide some information on Matching quality vs DSE quality.
[2,6]	A34 – The E&I team are to update their paper changing figure 1 in the paper from a proportion to a percentage and clarify ‘preserving observed distribution’ statement.
[2,6]	A35 – The E&I team are to provide data on reused donors, specifically how often they are used, and geographical distribution of donors.
[2,7]	A36 – The team are to provide clearer communication regarding which ADC design will be used.
[2,7]	A37 – There is need for more discussion on how to incorporate administrative sources into the IPS.
[2,8]	A38 – Panel members will work through the paper and the model and will provide feedback at a later date.
[2,9]	A39 – ONS to arrange a meeting on Behavioural insights: completing online questionnaires and experiences in other countries.
[2,9]	A40 – Forward plan: ensure measuring uncertainty is included.
[2,9]	A41 – Arrange for Enhanced outputs to be on agenda for June Meeting.
[2,9]	A42 – Arrange for CCS Sample design and further work on estimation models to be on the agenda for June meeting.

Minutes

Firstly, a congratulations to Dave Martin on his OBE, by the panel members. The panel would like to thank Dave for his comments on the papers and accept apologies for his absence.

1&2. Introductions and Actions from the previous meeting

The actions from the previous meeting were discussed. The following points were raised:

1.1 – The panel followed up the actions from the SDC paper, asking why are both types of perturbations based on pre-tabular and post-tabular are necessary. The panel also wanted confirmation on whether national level tables would be perturbed. This is to be picked up with SDC.

1.2 – Action A8 was closed, the panel were happy with the additional work carried out and provided to the panel, and the conclusion that the strategy will remain unchanged.

1.3 – Further discussion on the transparency Action A2, panel happy with progress made and for this to be carried forwards.

1.4 – The panel discussed the electronic census questionnaire, the research that had gone into the questions, positioning, and language style. They asked whether the behavioural insights unit had been consulted or involved.

1.5 – The Overcoverage item on the agenda was requested by Action A15, which has subsequently been closed.

1.6 – Research into the AZTool found the main cost pattern to be unchanged, leading to unchanging results. This has led to the closure of action A9.

1.7 – Item 7 on the agenda contains discussion of the population spine and administrative data sources, closing actions A12 and A13.

Actions

A19 – The panel asked for more detail on why both types of perturbations are necessary.

A20 – The panel discussed how homeless people will be counted and if there should be a separate survey to count them. Perhaps asking MHCLG to provide any data they have on homeless populations. The Panel requested a paper outlining work on this area. This supercedes Action A7.

A21 – The panel asked to see the online census questionnaire, and to be assured that ONS has been checking international practices in this area. This will be brought to the June panel meeting.

A22 – The panel asked if other potential datasets have been looked at to estimate the uptake of internet, for example DVLA and the electoral register data.

A23 – The panel requested if there is information regarding international practices of the CCS design and experiences of overcoverage in administrative sources, for example, in New Zealand.

A24 – The panel requested a dedicated future session on coverage adjustment, imputation of whole records and optimization methods.

Panel were happy with the progress made with all other actions.

3. Matching strategy for Census, linkage and population spine – Rachel Shipsey

Background

This paper outlines the matching process for the 2021 Census for the purposes of measuring coverage. That is determining whether two or more records belong to the same person, or household. The paper includes outlines of the three census matching processes, a review of the lessons learned from 2011, the forward intentions and lastly how matching quality will be determined.

In addition, the paper outlines the ONS's progress on developing a population spine to support the transformation of outputs based on integrated census, surveys and administration records. The purpose of the spine is to enable the production of official statistics, and support research.

Discussion

3.1 – Panel raised some concerns about Soundex and its inability to match some types of names and agreed that alternatives such as Jaro-Winkler and Levenshtein were better.

3.2 – Discussion about the potential for using machine learning techniques and if they can improve the efficiency of the process.

- 3.3 – Was Fellegi-Sunter the best method? Panel members believe this is an acceptable method but agreed that some additional exploration of machine learning would provide assurance.
- 3.4 – Some on the panel raised how this process could be risky from a public perception perspective, although noted that ONS have been undertaking testing and engagement in this area.
- 3.5 – The panel suggested that census records should be used in the next iteration of the spine.
- 3.6 – The panel requested that Machine learning techniques be considered.
- 3.7 – The team were asked to identify if it's feasible to utilise the ONS longitudinal study to check the accuracy of spine.
- 3.8 – The panel recommended pursuing a UK spine rather than just England & Wales.
- 3.9 – The panel asked the team to consider how to integrate Census records into the spine as a data source.

Actions

- A25 – The ONS are to investigate machine learning possibilities – speak with the ONS Data Science Campus & the Alan Turing Institute.
- A26 – The matching team need to re-evaluate the 'Priority order' used for the census to census matching – it is necessary to look at whether this order is sensible. With regard to the groups deemed most likely to be overcounted, the matching team need to confirm that the hierarchical order is sensible and articulate the impact of changing 'people who responded online' used in 2011 to 'people who responded on paper' for 2021.
- A27 – Consider utilising households as a hierarchy within probabilistic matching. Explain further how associative blocking can be used within the probabilistic matching algorithm.
- A28 – Deterministic linkage assumes that all match variables have the same weight. Consider 'weighting' variables within the RMR (resolving multiple responses) process.
- A29 – Consider whether the manual checking of the linkage can be improved through better software to facilitate the process.

4. Removing False Persons (RFP) process for the 2019 Census Rehearsal and 2021 England and Wales Census – Scott Redgwell

Background

This paper analyses the Remove False Persons (RFP) process from the 2011 Census and 2017 Census test. This includes an evaluation of its requirements, analysis of its performance, and the potential options and recommended approach for the 2019 Census rehearsal.

Discussion

- 4.1 – Some discussion on assigning a probability to a false person rather than removing them entirely.
- 4.2 – The panel discussed why this stage could not be included in the RMR stage.

Actions

- A30 – The RFPs team are to provide the distribution of removed date of births.
- A31 – The RFP team to consider the proposal to use a probability based record linkage.
- A32 – The ONS are to consider transparency, that is after the completion of the Census publish the response rate and clearly and explicitly define what a response consists of.

5. 2021 Census Process Flow – Steve Rogers & Cal Ghee

Background

This presentation shows the proposed flow and structure of the processing, including current estimates of the time it will take to process each stage. The panel were invited to discuss the balance between achieving the census quality goals and timeliness, as the two are related due to the ambitious timetable for producing outputs.

Discussion

5.1 – Whether the publication timetable is affected by elections.

5.2 – A discussion around timeliness and quality, finding the right balance between the two.

5.3 – The panel recommended that the quality of outputs, particularly the key local authority level ones, should not be compromised just to meet the schedule timings.

5.4 – The biggest risk was considered by the panel to be the amount of paper received, and the overall response rate.

5.5 – The panel believed If LA level estimates would be at risk of revision after rushing to meet a publication date, it would be better not to risk that – instead consider publishing just a national figure which could be revised (adhering to CoP guidelines) with less impact.

5.6 – The team were asked to demonstrate the contingency time needed for each process (where risks and delays are to each element), to demonstrate the risks across the whole period.

5.7 – Specify contingencies, thresholds and decision making process in advance – not robust to make decisions on the fly.

5.8 – The team were asked to demonstrate what is being done to optimise overall response and digital response. The team are to provide some information on Matching quality vs DSE quality.

Actions

A33 – The team to demonstrate what is being done to optimise overall response and digital response. The team are to provide some information on Matching quality vs DSE quality.

6. 2021 Census Editing and Imputation Strategy – Fern Leather

Background

This paper outlines the Edit and Imputation strategy for the 2021 Census and sets out the principles of the 2021 strategy of which the outcome is to produce a single fully populated, consistent census database.

Discussion

6.1 – Questions raised regarding hard edits – the panel believe these should be applied in the online questionnaire to minimise cleaning.

6.2 – Panel were curious regarding the process of when a donor is not found.

6.3 – Discussion around rare cases, and what information is publicly known such as there is no imputation of voluntary questions.

6.4 – Panel agreed with modularisation approach and that it's acceptable to use different donors for imputation of a single record.

6.5 – Panel agrees that, where valid, deterministic edits should be carried out first.

6.6 – Panel supported continuing the methods from 2011.

6.7 – Discussion around how uncertainty is measured.

6.8 – The panel requested the QA team to consider whether comparisons with other data sources for travel patterns (such as Strava) would be useful.

6.9 – The panel also recommended the Census team should consider how to add imputation variance into uncertainty measures for Census.

6.10 – The panel mentioned the problem that record swapping changes place of residences and hence impacts on the work-place statistics. The ONS is to consider how to handle this problem.

Actions

A34 – The E&I team are to update their paper changing figure 1 in the paper from a proportion to a percentage and clarify ‘preserving observed distribution’ statement.

A35 – The E&I team are to provide data on reused donors, specifically how often they are used, and geographical distribution of donors.

7. Transforming population and migration statistics – Becky Tinsley

Background

This paper outlines the ONS’s transformation of population and migration statistics. ONS is working across the Government Statistical Service to put administrative data at the core of evidence on international migration and population by 2020. This is a part of the work being delivered through the ONS’s Administrative Data Census (ADC) project.

Discussion

7.1 – Some discussion as to whether there should be an external advisory panel for the migration statistics, as the panel may not be the correct people to provide assurance.

7.2 – The panel asked for consideration of the use of exit checks in the population spine, and what other information could be added.

7.3 – Discussion around terminology in the paper such as circular migrants, the panel recommended that the definitions should be improved and clearly stated.

Actions

A36 – The team are to provide clearer communication regarding which ADC design will be used.

A37 – There is need for more discussion on how to incorporate administrative sources into the IPS.

8. Over-coverage estimation strategy for the 2021 Census of England & Wales – Viktor Račinskij and Ceejay Hammond

Background

This paper outlines the strategy for dealing with over-coverage in the 2021 Census. Over-coverage occurs when individuals are enumerated multiple times or in the wrong location, and this is likely to occur at greater rates in the 2021 Census due to greater number of responses online.

Discussion

8.1 – Panel discussed the presentation of the paper and concluded that the paper needed clearer and more consistent notation.

Actions

A38 – The panel will work through the paper and the model and will provide feedback at a later date.

Other Actions

A39 – ONS to arrange a meeting on Behavioural insights: completing online questionnaires and experiences in other countries.

A40 – Forward plan: ensure measuring uncertainty is included.

A41 – Arrange for Enhanced outputs to be on agenda for June Meeting.

A42 – Arrange for CCS Sample design and further work on estimation models to be on the agenda for June meeting.