Population And Migration Statistics (PAMS) Committee (Scotland)

Scotland's Census 2021 : Outputs and Statistical Disclosure Control

Outputs for 2021

High level planning work has begun on the planned outputs for the 2021 Census. We are planning to deliver a flexible output system which will allow users to design their own tables. The level of flexibility available will depend on the design of the output system and the statistical disclosure control method that is chosen. We also plan to deliver the first set of outputs a year after the census date. This is an extremely tight timescale and so this publication will consist of a very limited set of outputs and will not use the outputs delivery system mentioned above.

We will be consulting with stakeholders later on this year to identify key outputs from the 2021 Census and to seek views on the content and design of the flexible output delivery system.

Statistical Disclosure Control 2011

In 2011, we applied the following statistical disclosure control methods to tabular outputs:

- Targeted record swapping: All households have a chance of being swapped with similar households in a neighbouring area. However, households which are assessed as being unique or unusual have a higher chance of being swapped.

- Table redesign: Tables are designed to limit the number of cells with small values (1's or 2's) and to avoid attribute disclosures. To ensure that this redesign is sufficient we also apply checks for differencing and doubt. Differencing involves checking whether disclosive information could be derived by comparing new tables to tables that have already been published. Doubt checking involves looking at whether the proportion of small cells that are real rather than imputed or swapped are at or above the agreed level.

This has the advantage that some small cells could be released and that all tables were consistent and added up to the same totals. However it is very time consuming and is a fairly manual exercise. It leads to long delays for customers and can result in them getting far less detailed data than they wanted. This is not ideal, so we are hoping to do something different in 2021.

Statistical Disclosure Control 2021

We are currently working with colleagues in Office for National Statistics (ONS) and Northern Ireland Statistics and Research Agency (NISRA) to investigate a range of statistical disclosure control methods for 2021. ONS's current preferred methodology is a combination of targeted record swapping - as used in 2011 - and the cell key method which is adapted from the Australian Bureau of Statistics (ABS) method. The cell key method

adds a small amount of uncertainty to tabular output as it randomly adds a small number (mostly 0's but also +1,-1,+2) to each cell in a table.

Targeted record swapping worked well in 2011 and would be the main method of protecting the confidentiality of the data. The cell key method would be a secondary and 'light touch' method that would add a small amount of uncertainty to each cell and would protect against differencing.

The advantages of the cell key method are as follows:

- Level of uncertainty: Adding another statistical disclosure control method will add increased uncertainty.

- Timeliness: As this method protects against differencing tables could be produced much quicker and manual checking would not be required.

- Flexibility: This method would enable you to have tables at a greater level of detail than previously available under table redesign due to increased uncertainty.

The disadvantages of the cell key method are as follows:

- No consistency between tables: Although each individual table is additive within itself, a table with different grouping of those variables or of a different geographical level would have different numbers. In general, the larger the table the more uncertainty would be added by the cell key method. This means if you calculate Scotland totals from two tables with different geographic breakdowns you will get different figures.

An example using 2011 data illustrating the effect of the cell key method is provided in Annex 1 at the end of this document.

Other options for 2021

The other options under consideration for 2021 are:

1. Doing the same as in 2011, which will result in many of the problems experienced in 2011. These were: delays in outputs, large amounts of manual intervention, table redesign resulting in customers not getting what they need, no flexible table generation (however, it should be possible to make a range of published tables available in a user friendly way).

2. Similar to 2011 but with a higher rate of targeted record swapping. There is likely to be less table redesign required with this option but tables would need to be checked for differencing so it is likely that less flexible table generation options would be possible.

3. A mixed approach with targeted record swapping used as in 2011 but with cell key method only used for tables at a more detailed level of geography or for those with more than a pre-agreed number of cells. Some tables, probably Scotland and local authority level tables could be made available without applying the cell key method.

Next steps

We are continuing to investigate statistical disclosure control methods for 2021 and no decisions have been made yet. This is not the only opportunity you will have to share your views, but rather the first of many as we will continue to consult and involve stakeholders throughout our development process.

Group views are invited, at this early stage, on the statistical disclosure control methods currently under consideration for 2021 and more specifically the cell key method. In particular, would the increased flexibility and timeliness, that would be possible with the cell key method, be acceptable to users as compensation for the loss of consistency between tables?

Input from the group, as key stakeholders, on developing a list of key outputs for 2021 would also be welcomed. We are also seeking volunteers who would like to be involved in the design and/or testing of the new output system.

NRS: Statistical Futures
29 April 2016

Annex 1: Example of the Cell Key method using data on age and proficiency in English.

Below is an illustration, using the 2011 table of Proficiency in English by age (LC2105SC) of how the cell key method will alter tabular outputs in practice.

1. Real 2011 output at Scotland level and for one data zone without using the cell key method:

		All people aged 3 and over	Speaks English very well	Speaks English well	Does not speak English well	Does not speak English at all
Scotland	All people aged 3 and over	5118223	4555104	489579	62128	11412
Scotland	3 to 15	739151	575567	145570	14516	3498
Scotland	16 to 24	632488	578315	47084	6193	896
Scotland	25 to 49	1813010	1657693	126990	24659	3668
Scotland	50 to 64	1043240	962300	70081	9062	1797
Scotland	65 and over	890334	781229	99854	7698	1553
S01006506	All people aged 3 and over	834	745	77	12	-
S01006506	3 to 15	85	61	22	2	-
S01006506	16 to 24	79	77	2	-	-
S01006506	25 to 49	365	335	24	6	-
S01006506	50 to 64	181	168	10	3	-
S01006506	65 and over	124	104	19	1	-

2. Cell Key method applied to 2011 table at Data Zone level. Resulting Scotland level totals and one Data Zone shown:

		All people aged 3 and over	Speaks English very well	Speaks English well	Does not speak English well	Does not speak English at all
Scotland	All people aged 3 and over	5118254	4555109	489640	62106	11399
Scotland	3 to 15	739126	575538	145569	14543	3476
Scotland	16 to 24	632440	578284	47076	6188	892
Scotland	25 to 49	1813086	1657731	127045	24649	3661
Scotland	50 to 64	1043260	962315	70080	9060	1805
Scotland	65 and over	890342	781241	99870	7666	1565
S01006506	All people aged 3 and over	835	745	77	13	-
S01006506	3 to 15	86	61	22	3	-
S01006506	16 to 24	79	77	2	-	-
S01006506	25 to 49	365	335	24	6	-
S01006506	50 to 64	181	168	10	3	-
S01006506	65 and over	124	104	19	1	-

3. Cell Key method applied to 2011 table at council area level. Resulting Scotland level totals shown:

		All people aged 3 and over	Speaks English very well	Speaks English well	Does not speak English well	Does not speak English at all
Scotland	All people aged 3 and over	5118219	4555105	489578	62125	11411
Scotland	3 to 15	739152	575570	145568	14517	3497
Scotland	16 to 24	632483	578312	47085	6191	895
Scotland	25 to 49	1813009	1657693	126990	24657	3669
Scotland	50 to 64	1043242	962301	70079	9063	1799
Scotland	65 and over	890333	781229	99856	7697	1551

So you will notice from this example that when applying the cell key method for two different geographies that;

1. The numbers in the rows and columns of all of each individual table above adds up to its totals.

2. The resulting Scotland totals from the cell key method applied at data zone level and at council area level are different from each other (totals highlighted in yellow).

3. The two cell key totals are both slightly different from the real total (real total highlighted in green, cell key totals highlighted in yellow). 31 more people in the case of the cell key method applied at data zone level and four less in the case of the cell key method applied at council area level.

4. More uncertainty is added to tables with a higher number of cells, so the data zone level table will have more numbers that will have been changed by the cell key method (The cells that were changed at data zone level by the cell key method are highlighted in turquoise). Likewise, if this table had also been produced by gender as well as age group (or by single year of age), the totals would be different again as even more uncertainty would be added by the cell key method.