**Applying Cell-Key Perturbation to 2021 Census Outputs**

Iain Dove, Stephanie Blanchard, and Keith Spicer

**Abstract**

In 2011, ‘pre-tabular’ protection in the form of record swapping was applied directly to the census microdata. After the standard releases, outputs could be requested by users, which then had to be individually created and assessed for disclosure risk by the SDC team. This process took considerable time, and after the redesign of tables for protection, some users were faced with outputs that did not meet their needs.

For Census 2021, ONS would like to allow much faster access to data and allow users to define/ create their own outputs. To remove the need for individual checking of tables, a post-tabular method of protection, ‘cell-key perturbation’, will be needed alongside record swapping. This perturbation will be required to protect against ‘differencing’ of tables but will consequently have an impact on outputs that will need to be explained to users.

These notes describe and demonstrate the proposed ‘cell-key perturbation’ protection method, the possible outputs system for 2021, how this differs from previous census methods and the resulting benefits and trade-offs for users.

**1. Disclosure control for the census**

**1.1 Why disclosure control is important**

To start with, it is prudent to ask whether disclosure protection is needed whatsoever. Predictably, there are several big reasons disclosure control is necessary for National Statistical Institutes, such as ONS.

The Statistics and Registration Services Act (SRSA, 2007) defines “personal information” as information that identifies a particular person if the identity of that person—

(a) is specified in the information,

(b) can be deduced from the information, or

(c) can be deduced from the information taken together with any other published information.

It is a criminal offence to release personal information.

To comply with the SRSA (2007), ONS must prevent the release of personal information, by use of disclosure control. ONS also makes its own assurances to respondents on each survey (including the census), that it will keep information confidential. ONS enjoys a good reputation when it comes to public trust which it aims to maintain and improve upon. Disclosure breaches or other loss of data would breach that trust and would likely jeopardise response rates for surveys or trust in our outputs.

The primary aim of disclosure control is to comply with relevant legislation, and fulfil ONS’ responsibilities to respondents by ensuring confidentiality of responses. As a balance, our secondary aim must be to ensure this protection does the least possible damage to the data/maintains as much data utility as possible (at the extreme, we could successfully protect confidentiality by preventing all statistical outputs).

**2. Census 2011**

**2.1 Record swapping**

For 2011 Census data, the majority of protection came from targeted record swapping, illustrated below. The process involves ‘swapping’ two households between different geographical areas. This is targeted to the households thought most at risk of disclosure, to maximise the protection per swap, and households are swapped within Local Authorities, to preserve all information at Local Authority level and above. Households to be swapped with each other are also matched on some information (e.g. size of household), so that basic information is preserved even at the lowest geographies.

**Figure 1**

**Swapping households between areas**



In the figure 1, a one-person household in output area A is deemed at risk. A match is found at nearby output area B, a household containing one person of the same age. This swap has not affected the population of output areas A or B, or their age distribution. It has affected the ethnic group distribution, as well as others that the households did not match on. However, looking at higher geographies, say the aggregation of output areas A, B, and C, nothing has been affected. By keeping matches as geographically close as possible, record swapping has introduced changes, and uncertainty, to counts at low geography where disclosure is most likely, but none at higher levels.

**Figure 2**

**Counts are affected by swapping**



Figure 2 is an illustration of how the relevant frequency tables would be affected. One ‘other ethnicity’ 35-44 year old is swapped with one ‘mixed ethnicity’ 35-44 year old. As well as preventing the ‘other ethnicity’ individual from being the only person of that ethnicity in output area 1, swapping also introduces uncertainty in small counts that may not have been affected. Although swapping is targeted to the most at risk records, all records still have a chance of being swapped, so there is some uncertainty in all small counts.

**2.2 Table checking**

After applying record swapping, outputs were checked by the disclosure control team to assess if they had ‘sufficient uncertainty’, whether outputs appeared ‘safe’, and how often apparent disclosures were incorrect, as a result of swapping. If tables were too disclosive, they were redesigned, removing detail where possible, for example a table that was too disclosive at output areas (OA) could be released at lower super output areas (LSOA).

Tables also had to be considered for disclosure by differencing, illustrated below. In cases where tables are very similar, but not identical, extra information can be gained by combining independent tables, considered ‘safe’ on their own.

**Figure 3**

**Similar outputs can be ‘differenced’ to find smaller counts** 

This often meant that the same category breakdowns were used across different tables, or that tables were available for only one set of categories. It also made table checking more difficult, as other sources of information, including other census tables had to be considered.

This checking was a labour intensive process that took considerable time. In some cases, after efforts to redesign disclosive tables, the non-disclosive outputs did not meet the original user needs.

**3. Census 2021**

**3.1 Proposed option for 2021**

In preparation for 2021, the disclosure control team is investigating several methods of protection, including use of targeted record swapping plus cell key perturbation, illustrated below.

This would specifically protect against disclosure by differencing and allow user-defined outputs to be distributed through an online table builder. The protection and checking will have been applied before the tables are made, so anything available to be built will not need to be checked, and tables will not need to be re-designed. Most protection would still come from targeted record swapping as before, with cell perturbation also protecting against differencing.

**3.2 Cell key perturbation**

**Figure 4**

**The cell key perturbation method**

****

The cell key method adds a small amount of noise to some cell counts, based on the random number ‘cell-key’ for each cell, and the ptable. Because the cell key is created using the records in each cell, the random number can be replicated, and the same noise can be added in a repeatable way. This also has the benefit that the same records appearing together in a cell in unrelated tables still having the same cell-key, and thereby receiving the same noise (if any).

We can control how much noise to add using the ptable. As the majority of disclosure protection will come from applying the targeted record swapping, the cell key perturbation will be a ‘light touch’ and most cells will be unchanged. This is done by ensuring that most entries in the ptable are +0, so no noise is added in most cases.

Adding noise to cells in the outputs disrupts the differencing process. If noise is added to these tables, the resulting differenced table will also contain noise (highlighted in figure 5). This may lead to illogical counts being obtained by differencing, or counts may remain sensible, but importantly, would-be intruders will know perturbation has been applied and that small counts obtained by differencing will be unreliable.

**Figure 5**

**Observed differences may be real or may have been introduced by the perturbation**



**3.3 Trade offs**

With risky households protected by record swapping and differencing attacks protected against from cell perturbation the need for checking by the disclosure branch could be eliminated. This would allow much earlier access to data for users. It would also make access easier, with users able to define their own tables through the online table builder, rather than accepting re-designed tables from the disclosure branch.

Although adding noise to cells provides protection against differencing attacks, it also means that cell counts are changed from their original values. These changes happen in a consistent way at the cell level, but taking aggregations can lead to inconsistencies between different tables, particularly:

1) High geography table vs aggregate of low geography tables

2) The same marginal totals appearing in different tables

Examples are given below in figures 6 and 7.

**Figure 6**

**Inconsistencies between geographies**



When frequency tables are built at different geographies, the perturbations added will be different as well. Because records are further split at lower geographies, the cell keys will be different. E.g. a cell with 2 records with cell key 161 at a high geography could be 1 cell with 1 record and cell-key 100 plus another cell with 1 record with cell-key 61. This may alter the noise they receive from the ptable. If aggregates are taken from output area to higher geographies (e.g. Local Authority level), there will be different perturbations applied to the same cells.

In the example shown here in figure 6 of a small local authority table built at output area and local authority level, many small differences (1 & 2) are observed with few larger differences (5+), mostly in the totals. Totals are calculated by the sum of internal cells, so are much more likely to contain larger changes than internal cells, though it is also possible that several additions of noise cancel each other out (+1 +1 -2 in the case of occupation=6). Some values remain unchanged (non highlighted cells), though this is uncommon. These changes are a very small proportion of the estimates themselves (±2 out of several thousand), but could nonetheless be big concern for users in how they approach totals that do not agree, or how to explain these differences to stakeholders. It’s important for us to get feedback from users on how seriously these issues would affect them, and whether accepting the inconsistencies would be worth the increased flexibility.

**Figure 7**

**Minor inconsistencies between marginal counts of age**



Also shown are inconsistencies between different marginal (univariate) totals in figure 7. None of the internal cells are shared between the two tables. There are small differences between margial totals, of up to ±2.

To reduce the impact of these inconsistencies we are aiming to release population counts unperturbed and frequency tables at higher geographies also unperturbed. Tables at higher geographies are much less prone to differencing partly because they already contain a greater level of detail for which overlaps are harder to extract, and also because intruders are much less likely to find an individual or recognise an individual within an entire region, rather than a local area (within several nearby postcodes). These additions mean that population counts would not be disputed in any area and that higher tables would not contain inconsistencies (at national, regional level etc.)

**4. Summary**

Use of the cell key method could allow users better and earlier access to data from an online table builder in which users can specify their own outputs. It also causes inconsistencies between some tables, especially between aggregations of low geographical areas. The question for us to answer is whether this is a worthwhile trade-off for users. As next steps we have a trial implementation of the method being developed. It’s vital that we gather as much user feedback as possible on the proposed trade-offs to evidence whether this approach should go ahead in 2021.