Scotland's Census 2022

Statistical Disclosure Control Household Record Swapping Rehearsal Evaluation

July 2020

Contents

2019 Rehearsal Data – Household Record Swapping Methodology

1.   Summary of Record Swapping Rehearsal Evaluation

Household record swapping will be a key component of the NRS Statistical Disclosure Control (SDC) methodology. SDC methodology will be used to protect the confidentiality of households and individuals in published output tables from Scotland's Census 2022. This report outlines testing of the household record swapping method using data collected from the 2019 Census rehearsal in Scotland.

The Scotland's Census 2022 SDC & Outputs team applied the household record swapping methodology to 2019 Census rehearsal data. Diagnostic measures were then assessed to determine whether or not the methodology had functioned correctly and that swapping had been applied as expected.

The post-swapped rehearsal data was then assessed to determine whether or not the swapping had successfully applied an acceptable level of disclosure protection without excessively damaging the original data such that it would no longer be useful for data users. The methodology successfully provided protection against disclosure whilst maintaining acceptable levels of data utility for users.

Certain aspects of the evaluation were limited given that the volumes of data from the 2019 rehearsal are much lower than volumes during Scotland's Census 2022. Nevertheless, this record swapping rehearsal provided a useful functional test of the transfer of data across Census team, the application of the swapping method and the subsequent checks on disclosure protection and data utility.

Section 2 provides a brief overview of the household record process as part of the overall statistical methodology for Scotland's Census 2022. Section 3 provides detail on each of the steps carried out by the SDC & Outputs team during the record swapping rehearsal. Section 4 briefly summarises the outcomes and next steps based on findings from these rehearsal activities.

2. Household Record Swapping Background

This section provides a brief description of the household record swapping process which will be applied to data from Scotland's Census 2022 as part of the NRS Statistical Disclosure Control (SDC) methodology. As for Scotland's Census in 2011, NRS is developing this SDC approach to protect the confidentiality of individuals and households in published Census outputs. For more information on the swapping process as part of our SDC approach and in the context of the broader statistical methodology for Scotland's Census 2022, please see the slides from our Scotland's Census 2022 - Statistical Methods events held in February 2020.

For Scotland's Census 2022, SDC methods will be applied as the last statistical process before outputs are made available. A key component of the SDC methodology for the Scotland's Census 2022 will be the household record swapping method.

Household record swapping involves assessing every individual for uniqueness or rarity on a number of demographic characteristics. This then leads to every household being assigned a household "risk" score, a metric which indicates whether or not any of a household's residents possess a rare or unique combination of characteristics that might make them easy to identify in Census output tables, therefore posing a disclosure risk.

Once risk scores have been assigned, a sample of households is selected to be swapped with other demographically similar households in nearby geographic areas. Swapping demographically similar households in nearby areas helps to reduce damage to data. All households have a chance to be selected or "flagged" for swapping, however those with higher household "risk" scores are much more likely to be selected to ensure that individuals with rare or unique demographic characteristics are protected.

Details on the parameters used during the swapping process (risk variables, demographic matching variables and swapping sample size) are highly sensitive are therefore not given in this paper.

Section 3 outlines the specific steps undertaken by the SDC & Outputs team during the record swapping rehearsal using data from the 2019 Census rehearsal. These steps broadly mirror the process for applying the swapping methodology to data from Scotland's Census 2022.

3. Record Swapping Rehearsal Using 2019 Census Rehearsal Data

3.1 Receipt of Data from Data Processing Team – Clerical Checks

As detailed in our [Scotland's Census 2022 - Statistical Methods](#) slides, the record swapping is applied as part of the Scotland's Census 2022 SDC methodology, the final statistical process before Census outputs are made available.

Hence, the first step for the SDC & Outputs team during the record swapping rehearsal involved a clerical check of the 2019 rehearsal datasets received from the Scotland's Census Data Processing team. This was to ensure that the format of the datasets was suitable for use with the household record swapping methodology and to check that all variables used during the record swapping process were present. The SDC & Outputs team found no issues with the format of the datasets and all variables necessary for the swapping methodology were present.

3.2   Project to Create Swapping Input Data

The next step was to prepare the rehearsal datasets for the record swapping process by incorporating all relevant swapping information.

The main purpose of this first step is to combine the person and household records in the datasets with all necessary geographic information used during record swapping. A number of geographies are used during the swapping process and in subsequent checks on disclosure protection and data utility, including but not limited to Output Areas, Datazones, Electoral Wards and Council Areas.

Hence, during the swapping rehearsal, the 2019 rehearsal datasets were combined with the latest version of the [Scottish Postcode Directory](#) lookups created by the NRS geography team to match the postcodes of persons and households with all relevant higher level geographies used in the swapping process. Note that given 2022 Census output areas are still in development at this stage, 2011 Census Output Areas were used for the purposes of this rehearsal.

3.3   Household Record Swapping Process

Once the 2019 datasets had been prepared for swapping, the household record swapping methodology was applied. During and after the record swapping process, a large number of diagnostic metrics are used to indicate whether the process has been completed as expected. These include key measures on the number of individuals and households identified as being risky, the overall size of the swapping sample & proportion of high risk households within the swapping sample and the "quality" of the matches for households that are being swapped (i.e. are households flagged for swapping being successfully matched with demographically similar households in nearby areas?).

During the swapping rehearsal, diagnostic measures indicated that the record swapping process had successfully identified "risky" households, had selected a sample containing an appropriately high proportion of high disclosure risk

Scotland's Census
Shaping our future
A' dealbhadh ar n-àm ri teachd

households and had successfully found demographically similar matches in nearby areas without issues.

One anomaly was observed here relating to the volume of data within the rehearsal datasets rather than an issue with the swapping process itself. One of the swapping diagnostic metrics provides a warning for Output Areas (OAs) where a high proportion of households have been swapped. 2022 Census OAs are currently in development, however a key requirement for these, as in 2011, will be that they contain no fewer than 50 persons and no fewer than 20 households.

Given the lower response rate in the 2019 census rehearsal relative to our anticipated return rate for Scotland's Census 2022, nearly all of the OAs within the 2019 rehearsal datasets fell below the 50 person and 20 household threshold. As such, the swapping diagnostics correctly indicated that a large number of OAs had a proportionally high number of households being swapped relative to their overall household counts.

During live 2022 Census operations, this would be a cause for concern indicating that the swapping had caused significant damage to a large number of OAs. This would lead to a revision of swapping parameters by the SDC & Outputs team. For the purposes of rehearsal, this warning was ignored and evaluation proceeded to the next step.

3.4   Post-swapped Data Utility Checks

A key tenet of the Census SDC methodology is to balance the protection of respondent confidentiality with maintaining an acceptable level of data utility for data users. Hence, during live 2022 Census, once the household record swapping methodology has been applied, the SDC & Outputs team will assess the post-swapped data to ensure that the swapping process has not caused an excessive level of damage to the data such that it is no longer useful for data users.

Hence, once the household record swapping methodology had been applied to the 2019 rehearsal data, a separate analysis was carried out to compare the pre and post swapped data. This analysis helped to determine the extent to which the original data had been damaged by the swapping process (and data utility preserved).

As with the record swapping parameters, details on data utility metrics is restricted as information on these could be used to infer information on how the swapping process is applied. However, these data utility metrics have been developed in coordination with colleagues at the other UK Census Offices, the Office for National Statistics and the Northern Ireland Statistics and Research Agency.

Evaluation of the post-swapped rehearsal data found that an acceptable level of data utility had been maintained within the post-swapped rehearsal data. Data utility in the

post-swapped was found to be within acceptable thresholds for all of the metrics used during the Scotland's Census 2011 SDC methodology and in development for Scotland's Census 2022.

3.5   Post-swapped data disclosure control checks

As discussed, the during the swapping process, a number of diagnostic measures are used to provide an overview of whether or not the swapping has achieved the intended aim of applying protection to persons and households at high risk of disclosure. In addition to this, during Scotland's Census 2022, the SDC & Outputs team will carry out additional manual checks on the post-swapped data to determine whether the swapping process has applied an appropriate level of disclosure protection.

Again, details on disclosure control metrics used during the swapping process are highly sensitive as these could be used to make inferences on swapping parameters. Nevertheless, as with the Data utility metrics, doubt thresholds have been developed in coordination with SDC & Outputs experts at the other UK Census Offices, ONS and NISRA.

Assessment of the post-swapped rehearsal data indicated that an acceptable level of disclosure protection had been introduced by the record swapping process, such that an appropriate level of uncertainty was present in output tables around whether or not cell values represented real counts or had been altered during the swapping process.

4.   Swapping Rehearsal Evaluation and Conclusion

As noted in earlier sections of this report, certain attributes of data received during the 2019 Census rehearsal in Scotland (chiefly the small volume of data relative to what will we receive during Scotland's Census 2022) restricted some aspects of the record swapping evaluation.

Nevertheless, this rehearsal provided assurance that the household record swapping method can be applied to Census datasets to introduce an acceptable level of disclosure protection whilst preserving data utility for users of Scotland's Census 2022 data.

These rehearsal activities also provided a useful functional test of the transferal of data during the statistical processes that will be carried out during Scotland's Census 2022.

The NRS SDC & Outputs team are continuing to investigate the parameters that will be used during the household record swapping for Scotland's Census 2022. Work is also ongoing to finalise the metrics on disclosure protection and data utility that will be used to evaluate whether the swapping process has introduced an appropriate

level of protection whilst preserving data utility for users of Census data. This investigation is being carried out in conjunction with UK Census colleagues at the Office for National Statistics and the Northern Ireland Statistics and Research Agency.