Northern Ireland
Statistics and Research Agency
Gníomhaireacht Thuaisceart Éireann
um Staitisticí agus Taighde

Statistical Disclosure Control Methodology for 2021 Census

GUIDANCE NOTE

CENSUS OFFICE FOR NORTHERN IRELAND

Introduction

Statistical Disclosure Control (SDC) refers to a range of methods that aim to protect individuals, households, businesses, and their attributes from being identified in published tables.

For the 2021 Census, NISRA is considering two strategies to ensure individuals are protected from identification while minimising the impact on the quality of results, these strategies are:

1. Targeted record swapping
2. Cell key perturbation

1. Targeted record swapping

Targeted record swapping was applied to the 2011 Census data and NISRA propose to use this same method in 2021. This method involves detecting records that are most at risk of identification and swapping them with similar records from other geographical areas. All households and individuals are allocated a risk score based on a small number of characteristics that are considered to be unique or rare. A sample is then selected for swapping. The selected households are swapped with similar households – these similarities are grounded on basic characteristics so data quality can be preserved (i.e. household size so the number of people in households in each area will be unspoiled). This method can achieve a high degree of confidentiality while minimising damage to results.

The targeted record swapping method is illustrated in figure 1. A one-person household in output area A is deemed at risk of identification. A similar household is found at nearby output area B, containing one person of the same age: these records are swapped. This swap has not affected the total population of output areas A or B, or their age distributions but has affected the ethnic group distribution, as well as others that the households did not match on. However, by looking at higher levels of geography, say the aggregation of output areas A, B, and C, you will see that nothing has been changed. By selecting records that are as geographically close as possible, record swapping has introduced changes, and uncertainty, to totals at low levels of geography where disclosure is most likely, but not at higher levels.

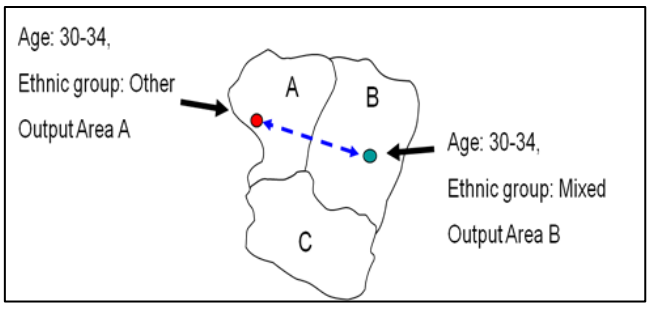

Figure 1: Targeted record swapping method

2. Cell key perturbation

For the 2021 Census, NISRA is considering the development of a flexible dissemination system to specifically meet the needs of users. It is crucial that SDC methodology is incorporated into this product in order to prevent individuals, households and businesses being identified. To achieve this, NISRA propose to continue the use of targeted record swapping and is considering the additional use of a method known as "cell key perturbation".

Cell key perturbation involves making small changes to some cells in a table with low counts. This adds more uncertainty to the table with the goal of protecting against disclosure, especially disclosure which occurs when one table is compared to others with similar information (differencing). In order to make these small changes, the method adds 'noise' to some cells. Although adding noise disrupts the differencing process, it may lead to inconsistent table totals. However, in these instances the proportions would remain the same (see figures 2 and 3). The key benefits of using this method are that customers will be able to obtain 2021 Census data much faster than in 2011.

TABLE C12: SELECTED OCCUPATION BY AGE

	355 Conservation and Environmental associate professionals	Proportions (%)
All usual residents	**215**	**100%**
Aged 0 to 24	2	1%
Aged 25 to 44	119	55%
Aged 45 and over	94	44%

TABLE C13: SELECTED OCCUPATION BY AGE BY SEX

	355 Conservation and Environmental associate professionals	Proportions (%)
All usual residents	**215**	**100%**
Aged 0 to 24	2	1%
Aged 25 to 44	119	55%
Aged 45 and over	94	44%
Males	**190**	**88%**
Aged 0 to 24	2	1%
Aged 25 to 44	98	45%
Aged 45 and over	90	42%
Females	**25**	**12%**
Aged 0 to 24	0	0%
Aged 25 to 44	21	10%
Aged 45 and over	4	2%

Figure 2: Tables C12 and C13 before cell key perturbation

TABLE C12: SELECTED OCCUPATION BY AGE

	355 Conservation and Environmental associate professionals	Proportions (%)
All usual residents	**216**	**100%**
Aged 0 to 24	3	1%
Aged 25 to 44	119	55%
Aged 45 and over	94	44%

TABLE C13: SELECTED OCCUPATION BY AGE BY SEX

	355 Conservation and Environmental associate professionals	Proportions (%)
All usual residents	**217**	**100%**
Aged 0 to 24	3	1%
Aged 25 to 44	119	55%
Aged 45 and over	95	44%
Males	**190**	**88%**
Aged 0 to 24	2	1%
Aged 25 to 44	98	45%
Aged 45 and over	90	42%
Females	**27**	**12%**
Aged 0 to 24	1	0%
Aged 25 to 44	21	10%
Aged 45 and over	5	2%

Figure 3: Tables C12 and C13 after cell key perturbation