



## **2001 CENSUS: DISCLOSURE CONTROL EVALUATION**

### **Project Objective**

To have a series of measures in place that will uphold the 2001 Census confidentiality commitments that published tabulations and abstracts of statistical data do not reveal any information about identifiable individuals or households.

### **Background**

The confidentiality of personal information is a legal obligation, reinforced in the National Statistics Code of Practice. The Northern Ireland Statistics and Research Agency (NISRA) has made commitments publicly to ensure the protection of information collected from the 2001 Census.

These commitments were given on the Census form:

*The information you provide is protected by law and treated in strict confidence. The information is only used for statistical purposes, and anyone using or disclosing Census information improperly will be held liable to prosecution. Census forms will be held securely under the terms of the Public Records Act (Northern Ireland) 1923.*

and in the White Paper "The 2001 Census of Population"(March 1999):

*Precautions will be taken so that published tabulations and abstracts of statistical data do not reveal any information about identifiable individuals or households. Special precautions may apply particularly to statistical output for small areas. Measures to ensure disclosure control will include some, or all, of the following procedures:*

- Restricting the number of output categories into which a variable may be classified, such as aggregated age groups;

- where the number of people or households in an area falls below a minimum threshold, the statistical output - except for basic headcounts - will be amalgamated with that for a sufficiently large enough neighbouring area; and/or
- randomly modifying some data before the statistics are released.

The White Paper can be found [HERE](#).

In order to honour these commitments the three UK Census Offices; Office of National Statistics (ONS), General Registry Office Scotland (GROS) and NISRA developed and implemented a robust disclosure control strategy which helped secure public confidence in the three Office's resolve to safeguard the personal information provided through Census forms and ensure full co-operation with the Census thus enabling them to provide authoritative statistics to the highest quality possible. The following sections describe the procedures that were used in 1991 and outline the work that was undertaken to establish the procedures implemented in 2001. The development work was mostly conducted by methodological specialists in ONS on behalf of the three Census Offices.

## **Disclosure Protection in 1991**

In 1991, the disclosure of information in Census output was protected by:

- the setting of minimum population thresholds for tables (e.g. 16 households and 50 persons for Census Area Statistics, and 320 households and 1,000 persons for Standard Tables);
- collapsing available categories for certain variables into appropriate bands; and
- the technique of Barnadisation. This method introduced uncertainty into all published Census output by modifying cell counts by up to +2 and -2 in published tables.

## **Methodology**

The development of the disclosure control strategy for the 2001 Census consisted of two main elements; namely, (i) a review of the disclosure methods deployed in 1991 and increased risk of disclosure since 1991 and (ii) a programme that researched possible disclosure control options

***A review of disclosure control methods used in the 1991 Census and the increased risk since 1991***

The review of the 1991 methodology concluded that it effectively protected information about identifiable persons and households and that at least the same level of protection should be provided in 2001. The table design and population thresholds worked well in 1991. However, an alternative tabulation method to Barnadisation was required due to inconsistencies that appeared within, and between, tables, through a loss of additivity as a consequence of Barnadisation.

The Census Offices also needed to consider options of protection that would address increased risk of disclosure as a result of improvements in technology and the impact this has on the availability of data and the ease with which an intruder may identify individual information. The review resulted in a number of important observations that would give rise to increased risk which are set out below.

- The 2001 Census results would be very widely disseminated via the Internet meaning that users and the general public could acquire Census data more readily and easily than ever before. The increased accessibility also increases the risk of misuse of Census data.
- The 2001 Census would provide outputs through a variety of media and provide information that would be more detailed than published for previous Censuses. Data for output areas would be provided that are considerably smaller in geographical size than the lowest geographical level provided in 1991. For these small geographical areas that contain about 125 households, there would be a need to manage the risk of revealing information about any persons or households with unique characteristics in such small areas.
- Census data users can obtain large volumes of Census statistics freely and any increased risk from attempts to break the confidentiality protection provided would need to be managed.

The findings of the review concluded that there was a substantially increased risk of disclosure in the 2001 Census than previously and the Census Offices would have to develop methods to protect against it. It was believed that if counts of 1 and other small values were simply left in the tables, then there would be a perception that the Census Offices were not doing all that they could to (i) fulfil its legal obligations in respect of preserving confidentiality and (ii) ensure that all possible steps were taken to prevent inadvertent disclosure. There was a requirement for disclosure measures that made persons and households with unique characteristics within an area not visible in tabular output.

### ***Research of possible disclosure control options***

A research programme was developed that explored possible options for addressing the increased risks since 1991. It specifically examined the use of

disclosure methods similar to the 1991 Census and new options of disclosure control for the additional measures that needed to be applied in 2001. Census data users were consulted on an ongoing basis through advisory groups and roadshows in the inter-censal period.

The Census Offices explored each disclosure control option following the criteria set out below:

- The effectiveness of the method for disclosure protection;
- The impact of the method on the quality of Census data; and
- The practical aspects of implementing the method.

It concluded that the design of tables and the population thresholds (in 1991) were effective measures that did not affect the quality of the data and could be repeated in 2001. However, the increased risk of disclosure meant that the population thresholds would need to be increased. The programme investigated an alternative to Barnadisation and the following pre-tabulation options:

- record swapping - swapping a household record with a similar record in the same geographic area;
- data switching - swapping the values of one or more variables in one record with the values for the same variables in another record; and
- over-imputation - randomly deleting variables in existing records and imputing the variables using the Edit and Donor Imputation System.

Record swapping was chosen as it added uncertainty to the data, was easily implemented and did not substantially damage the quality of the data. However, it was recognised that this measure had limitations as (i) it would not be apparent to a user that methods of disclosure control had been implemented and (ii) there would be a perception that people and households were identifiable (particularly for a single count) and the observer may act upon the information as if it were true. Accordingly, two options of further protection based upon cell modification were also considered, namely the rounding of all counts to a multiple of 3 and small cell adjustment.

During 2002 a consultation exercise led by ONS sought to establish what users felt about each of these approaches. The issue was controversial and a large number of users would have preferred to have no additional disclosure control protection measures. Where users indicated a preference, small cell adjustment was the preferred choice. This was largely due to the advantage that the method allowed tables to be internally additive and only adjusted small cells. The disadvantage of the method was that knowledge of the adjustment method had the risk of allowing cells containing a single observation to be deduced.

## **Measures finally implemented**

The disclosure measures implemented for the 2001 Census represented a combined approach based upon a set of judgements. While it was recognised that each method alone did not provide adequate protection, NISRA concluded that, for Northern Ireland, the combination of the chosen methods offered the protection that was needed. The final set of disclosure measures are outlined below:

### ***Increased thresholds***

The thresholds were increased from those used in 1991 to:

- 40 households and 100 persons for Census Area Statistics
- 400 households and 1,000 persons for Standard Tables

Where areas fell below these thresholds desirable Census output was examined on a case-by-case basis.

The smallest area for which commissioned output will be released is a Census Output Area which contains approximately 125 households. All other areas will be amalgamations of these areas.

### ***Design of tables***

All output tables have been assessed for disclosure risk and data utility. Hence, some tables have been collapsed or are unavailable at lower geographical levels.

A general principle of making the average cell count in a table greater than or equal to one has been applied to the design of all 2001 Census output.

### ***Record swapping***

This procedure adds uncertainty to data by swapping the geographical location of a small sample of households with that of another household in the same District Council (or group of District Councils). The procedure was designed such that the integrity of swapped data was not substantially different among key variables from that of unswapped data. The percentage of records swapped and the basis on which they are swapped must remain confidential.

### ***Small Cell Adjustment***

This method adjusts small counts in tables to add uncertainty to tabular output in which individual information could be identified. The definition of a small count must remain confidential so that the protection provided by the adjustment is maintained. Totals and subtotals are calculated from the adjusted data, thus

ensuring consistency within tables that was not present in 1991. However, the same totals (e.g. All persons) appearing in different tables may vary slightly.

### ***Conditions of Use***

Conditions of use included in all end user licences stress that the Census material shall not be used to attempt to derive information relating to an identified person or household nor shall a claim be made that such information has been obtained or derived.

Further information of the disclosure methodology can be found [HERE](#).  
Further details of the rationale behind the chosen methods can be seen [HERE](#).

It is noted that the suite of disclosure control methods employed in Northern Ireland were similar to those employed by ONS for England and Wales, while GROS employed a slightly different suite in Scotland.

### **Assessment and Lessons Learnt**

The main lessons learnt from the project were that some elements of the disclosure risk assessment should have been carried out much earlier. It was less than one year before Census day, in 2000, that it was concluded that extra precautions would need to be taken to protect information as a result of the increased risk since 1991. The risks for Census confidentiality, from the increased amount of small area statistics that were to be published had to be reassessed. Consultation should have been carried out earlier and more time allowed to research and develop different options of disclosure control. It must be recognised, however, that the Census Offices has an obligation to continually assess the risk of disclosure as things change, (for example with advances in technology or with the new increased flexibility of the new geography), and to review and amend procedures as necessary.

There are some lessons to be learnt relating to the consultation with users, which worked well for the project as a whole but which could have been better for specific aspects of the process. For example, consultation regarding the options of rounding and small cell adjustment was considered to be too late in 2001 and 2002. The Census Offices recognises that it could have made users explicitly aware that the disclosure procedures were under a further review and prepared users for the real possibility of further changes.

The consultation highlighted that many users were concerned about the impact of small cell adjustment on data quality. These concerns emphasised a need to provide users with information about the uncertainty that exists within Census data introduced by many of the processes aimed at improving data quality such as the One Number Census and Edit & Imputation.

The Disclosure Control project has achieved its objective. While each disclosure measure alone would not provide adequate protection, the combination of all measures give sufficient protection to meet the commitments that NISRA has made. The record swapping, small cell adjustment, table design and threshold constraints have been successfully implemented.

## **Conclusions**

The disclosure control project aimed to design and implement procedures to protect information within all Census output and this aim was achieved. The project began with a review of the procedures for disclosure protection in 1991, identifying problems that occurred with these methods and assessing increased risk resulting from the increased use of electronic resources. The review showed that an alternative to Barnadisation was needed and that additional protection was required. A set of measures was developed to offer the level of protection that was needed. The Census Offices recognise that the research and development of some disclosure options should have taken place two to three years earlier, particularly the consultation on small cell adjustment and rounding of Census output. Valuable lessons have been learnt, mainly about the timing of the assessment of disclosure risk and the timing of consultation with users. With this knowledge, and the successful implementation of the measures, there is a good basis on which to build future disclosure control strategies. For future Censuses, final decisions must be reached earlier in the planning cycle.