



Census 2001 Review and Evaluation

November 2003

Disclosure Control: Executive Summary

| Content | Page |
|------------------------------------|------|
| Project Objective..... | 2 |
| Background..... | 2 |
| Methodology | 3 |
| Assessment and Lessons Learnt..... | 5 |
| Conclusions | 6 |

ONS is carrying out a review and evaluation of the 2001 Census in England and Wales which will culminate in a Data Quality report and a General Report being published.

Plans for individual reports on specific aspects of the Census operation and a timetable for release have been published.

Each report is written in isolation and is subject to amendments as processing progresses and further information comes to light.

Reports will be released on the ONS website in the form of a high level Executive Summary and a more detailed Evaluation Report.

Census Customer Services
ONS
Titchfield
Fareham
Hants PO15 5RR

Telephone: ++44 (0) 1329 813800
Fax: ++44 (0) 1329 813587
Minicom: ++44 (0) 1329 813669
E-mail: census.customerservices@ons.gov.uk
Website: www.statistics.gov.uk/census2001

Census 2001 Review and 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 Office for National Statistics (ONS) has made commitments publicly to ensure the protection of information collected from the 2001 Census.

These commitments have been 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 current terms of the Public Records Act 1958, the data will be treated as confidential for a period of 100 years

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*
- *modifying the data before the statistics are released.*

The White Paper can be found at www.statistics.gov.uk/census2001/pdfs/whitepap.pdf

By keeping to these commitments, ONS remains a responsible, trusted guardian of personal and confidential information. A robust disclosure control strategy is essential for respondents to cooperate with the Census and other government surveys and ensure that ONS has the highest possible quality statistics available.

ONS has to implement procedures that will protect information within all census output. This includes output in tabular form and within microdata (e.g. Samples of Anonymised Records). This paper describes the procedures, but does not evaluate the statistical attributes of disclosure protection.

Disclosure Protection in 1991

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

- minimum population thresholds of tables;
 - (i) 16 households and 50 persons for Census Area Statistics
 - (ii) 320 households and 1,000 persons for Standard Tables
- the design of tables where data were presented for small geographical areas, categories for some variables were often banded to protect the data; and
- the technique of Cell Perturbation. This method introduced uncertainty into all published census output by modifying cell counts by up to +2 and -2 in published tables. Unfortunately, a consequence of the Cell Perturbation method was that there were some inconsistencies of data within, and between, tables, through a loss of additivity.

Census 2001 Review and Evaluation

Methodology

The disclosure control strategy for the 2001 Census consisted of two main elements:

- (i) A review of 1991 disclosure methods and increased risk of disclosure since 1991
- (ii) A programme that researched possible disclosure options

A review of disclosure control for the 1991 Census and 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 was required due to inconsistencies that appeared in tabular output as a consequence of Cell Perturbation. ONS also needed to consider options of protection that would address any increased risk of disclosure since 1991 as a consequence 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 identified the following risks:

1. The 2001 Census results would be very widely disseminated via the internet. This means that users and the general public can acquire census data more readily and easily than ever before. The increased accessibility also increases the risk of misuse of census data.
2. The 2001 Census would have greater flexibility of census output and the production of 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, we needed to manage the risk of revealing information about any persons or households with unique characteristics in such small areas.

3. Census data users can obtain large volumes of census statistics freely and we would need to manage increased risk from attempts to break any confidentiality protection provided.
4. All questions from the 2001 Census would be fully coded. Previous censuses had coded only 10% of the responses for some key variables and that had added a level of uncertainty to published results.

The findings of the review concluded that there was a substantially increased risk of disclosure in the 2001 Census than previously and ONS 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 ONS would not be doing all that it could to fulfil its legal obligations of confidentiality and thus ensuring 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.

A research programme of possible disclosure protection options

Initial programme and conclusions

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.

ONS explored each disclosure option following a set of criteria:

1. The effectiveness of the method for disclosure protection
2. The impact of the method on the quality of census data
3. The practical aspects of implementing the method

Census 2001 Review and Evaluation

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 Cell Perturbation and the following pre-tabulation options were investigated:

- 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.

Later consideration and additional measures

The chosen method of record-swapping, however, did have its limitations and ONS became more and more concerned about these. It would not be apparent to a person using the Census data that any methods of disclosure protection would have been implemented. There would be a perception that persons and households were identifiable (particularly for a single count) and the observer may act upon the information as if were true. Therefore two options of further protection based upon cell count modification were also considered:

- Rounding of all counts to a multiple of 3
- Small Cell Adjustment

A key consultation with users in 2002 involved these options concerning the tabular modification of cell counts. The issue was controversial and a large number of users would have preferred to have no additional

disclosure 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.

The response to the consultation can be seen at www.statistics.gov.uk/census2001/cn_22.asp

Measures finally implemented

The disclosure measures implemented for the 2001 Census were a combined approach that were based upon a set of judgements. Each method alone did not offer adequate protection, but ONS concluded that the combination of the chosen methods offered the protection that was needed. The final set of disclosure measures were :

Increased thresholds and design of tables

The thresholds were increased from those used in 1991 to:

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

Where areas fell below these thresholds either summary statistics were produced, or the areas were amalgamated with contiguous areas in consultation with the local authorities concerned.

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

Census 2001 Review and Evaluation

Record swapping

This procedure adds uncertainty to data by swapping a small proportion of records with similar records in other small geographical areas. 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 adjusted data, thus ensuring consistency within tables that was not present in 1991. However, the same totals appearing in different tables may be different.

Conditions of Use

A condition of use included in all end user licences is 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 at www.statistics.gov.uk/census2001/discloseprotect.asp

Further details of the rationale behind the chosen methods can be seen at www.statistics.gov.uk/census2001/pdfs/ag0203.pdf

to protect information as a result of the increased risk since 1991. We have had to reassess the risks for census confidentiality, from the increased amount of small area statistics that were to be published by the Neighbourhood Statistics project. 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 ONS 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 has worked well for the project as a whole but which could have been better for specific aspects of the process. The consultation regarding the options of rounding and small cell adjustment was considered to be too late in 2001 and 2002. ONS 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. Each disclosure measure alone would not provide adequate protection, but the combination of all measures provide sufficient protection to meet the commitments that ONS has made. The record swapping, small cell adjustment and threshold constraints have been successfully implemented.

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 we would need to take extra precautions

Census 2001 Review and Evaluation

Conclusions

The disclosure control project aimed to design and implement procedures to protect information within all census output and this aim has been 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 Cell Perturbation was needed. Later consideration found that additional protection was required. A set of measures were developed to offer the level of protection that was needed. ONS recognises 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. We must reach final decisions much earlier, for a future census.

Census 2001 Review and Evaluation

| Census Topics | Target Dates for Release |
|--|----------------------------------|
| Legislation | Published |
| Non-Compliance (Executive Summary Only) | Published |
| Data Needs | Published |
| Geography | Published (Executive Summary) |
| Publicity | Published |
| Data Collection Development | Published |
| Data Collection Support | Published |
| Census Coverage Survey | Published |
| Processing | Published |
| Annex: Quality of Data Capture and Coding | Published |
| Downstream Processing | Published (Executive Summary) |
| Data Quality | |
| - Question non-response rates | Published |
| - Disclosure Control (Executive Summary only) | Published |
| - Data Validation (Executive Summary only) | Published |
| Edit & Imputation | Published |
| One Number Census | |
| - Quality Assurance | Published |
| - Lessons learnt (Executive Summary only) | Published |
| Output Policy | Published (Executive Summary) |
| Output Production | |
| - Part 1:Review of Output Released to date | Published (Executive Summary) |
| - Part 2:including Sample of Anonymised Records (SARs)/Origin Destination Matrices | Published |
| Census Access | Published |
| Programme Management | Published (Executive Summary) |
| Quality Report | Published |
| General Report | Published |

Please note that the dates for release of individual evaluation reports noted above are target dates, and therefore subject to change. For the latest information please visit www.statistics.gov.uk/census2001/reviewevaluation.asp