

POPULATION AND MIGRATION STATISTICS COMMITTEE (SCOTLAND)

UK Statistical Disclosure Control Policy for 2011 Census Output

Background

The Registrars General of Scotland, England and Wales and Northern Ireland have agreed to aim for a common Statistical Disclosure Control (SDC) methodology for 2011 Census outputs. This will help achieve the aim of harmonising the 3 Censuses where that is in the interest of Census users.

Adoption of a common SDC methodology across the UK will only be possible if there is an agreed SDC policy position across the Census Offices i.e. agreement about what constitutes a disclosive risk in a census context and tolerable risk thresholds. This paper sets out the SDC policy position that has been agreed by the Registrars General.

UK SDC policy position

The UK 2011 Census SDC policy position is based on the principle of protecting confidentiality set out in the National Statistics Code of Practice which includes the guarantee that 'no statistics will be produced that are likely to identify an individual unless specifically agreed with them'.

Because the key strength of the Census is its complete coverage, and its ability to generate statistics about very small groups of people (as is necessary to ensure that Government and other policies take account of the needs of small groups and communities), it is impracticable entirely to remove the risk of disclosure, without harming the utility of the data. With that in mind, the Registrars General have concluded that the above Code of Practice statement can be met in relation to Census outputs if no statistics are produced that allow the identification of an individual (or information about an individual) with a high degree of confidence. The Registrars General consider that, as long as there has been systematic perturbation of the data, the guarantee in the Code of Practice would be met.

It is considered that "attribute disclosure" (i.e. learning something about an individual) as opposed to "identification" of an individual is the key disclosure risk, because identification reveals no new information to the user. 'Attribute disclosure', however, involves a user discovering something new from the census data that was not previously known to them.

In a Census context, where thousands of tables are generated from one database, the risk of attribute disclosure occurring can be addressed by introducing uncertainty about the true value of small cells.

In order to meet the agreed interpretation of the Code of Practice, it has thus been agreed that small counts (0s, 1s, and 2s) could be included in publicly disseminated Census tables provided that

- a) uncertainty as to whether the small cell is a true value has been systematically created; and
- b) creating that uncertainty does not significantly damage the data.

The exact threshold of uncertainty required has not been decided. The Registrars General will make this judgement at a later stage in the context of results from methodological research into the balance of protection afforded, and damage caused, by various SDC methods.

Different levels of disclosure control are applied to Census outputs according to the mode of access. In general the aim will be to make as much as possible of the Census tabular output publicly accessible. However, if tabular outputs are likely to be seriously compromised by SDC (for example Origin/Destination tables at low geographical levels) then these could be released under other access arrangements (licence or safe setting) where restrictions on access to the data allow less stringent levels of SDC to apply, in order to protect the utility of the data.

As a result of the Chancellor's decision to legislate for ONS independence, the current NS Code of Practice: Protocol on Data Access and Confidentiality will be replaced. But the obligation to preserve the confidentiality of census outputs is likely to be heavily informed by the current Code of Practice.

Implications of the proposed SDC policy position for SDC methodology

The decision to allow small cells in publicly disseminated tables means that no methods of SDC (pre-tabular, post-tabular or combinations of the two) have been ruled out and all methods will be evaluated. The Registrars General have expressed a preference for pre-tabular methods, provided there is not undue damage to the data.

To ensure that the public, and expert audiences, are confident that confidentiality will be preserved by the measures taken to avoid disclosure, clear explanations would be given on the protection afforded by the SDC strategy, and other steps which protect confidentiality, that had been applied.

The choice of SDC methodology for 2011 Census outputs will be based on evaluation of the risk and utility of the various possible methods. Methods will be recommended that afford an acceptable level of protection and preserve the highest level of utility of outputs. Consistency and additivity across tabular output is a priority for users and these will be given a high priority in the assessment of the utility of SDC methods.

Next steps

The principles outlined in this paper provide a basis for consultation with users of Census data and for a two year period of methodological research. The research will assess pre and post tabular SDC methods in terms of the protection they afford together with their impact on the integrity of the data (a risk/utility framework). Because of the interdependence between disclosure control of (pre-defined) census tabular data and disclosure control for other types of census outputs (microdata samples and flexible user defined tabular outputs), SDC methods for all types of

Census output will be assessed concurrently and a key consideration in evaluating SDC methods for tabular data will be the potential impact on these other types of census output.

Users will be updated and consulted during the research period. There will also be an independent review through the UK Census Design and Methodology Advisory Committee.

The Census White Paper (and parallel documents relating to Scotland and Northern Ireland) are timetabled to be published in October 2008 and will formalise the agreed policy position of the Registrars General by the inclusion of an SDC policy statement. Recommended SDC methods for all types of 2011 Census outputs will be published in Autumn 2008 for consultation and finalised in Spring 2009.

Office for National Statistics
General Register Office for Scotland
Northern Ireland Statistics and Research Agency
November 2006