

Statistics User Forum
Meeting 6th April 2006, Royal Statistical Society, 12 Errol St, London

**STATISTICS FOR SMALL AREAS – HOW CAN WE CONTROL
DISCLOSURE WITHOUT DESTROYING MUCH OF THE VALUE?**

This all-day meeting focussed on disclosure control for tables of statistics for small areas. Organized by Keith Dugmore (Demographic Decisions) and chaired by Philip Rees (University of Leeds), it considered:

- Users' experiences of practical problems with existing datasets such as the Census and Neighbourhood Statistics, and their advice to others
- Current policy and practice: case studies and alternative approaches
- A discussion of which options are best for users

Programme & Speakers:

- Cynthia Clark (Office for National Statistics) Setting the scene:

What problems have users encountered in practice?

- John Hollis (Greater London Authority) – the 2001 Census
- Wendy Pontin (Norfolk C.C) – a view from a shire county
- David McLennan (University of Oxford) – Neighbourhood Statistics
- Keith Dugmore (Demographics User Group) – commercial users' experiences

Current policy and practice; case studies and alternative approaches

- Peter Scrimgeour (General Register Office Scotland)
- Robert Beatty (Northern Ireland Statistics and Research Agency)
- Jane Longhurst and Paul Vickers (Office for National Statistics)

Discussion: how do users view the trade-off between different methods – which options are preferred?

- Led by Professor Philip Rees

Powerpoint presentations from this meeting are available at:

[http://www.rss.org.uk/main.asp?group=&page=1321&event=247&month=4&year=2006
&date=](http://www.rss.org.uk/main.asp?group=&page=1321&event=247&month=4&year=2006&date=)

NOTES ON DISCUSSION POINTS

These are a set of points that came up in questions and comments on each presentation and in the discussion session at the end of the afternoon.

Setting the scene, Cynthia Clark

Good practice at census data delivery: American Fact Finder

US Bureau's experience with the American Fact Finder, which gives users access to US 2000 Census tables may be useful for UK National Statistics Offices to investigate for guidance on developing new methods of access.

See

<http://factfinder.census.gov/home/saff/main.html>

American Fact Finder is impressively organised and yields information very quickly and usefully. There is a custom table facility which is very powerful and can be found at:

http://factfinder.census.gov/servlet/CTGeoSearchByListServlet?ds_name=DEC_2000_SF1_U&lang=en&ts=162444351450

This enables users to design their own tables from a standard set of variable definitions (categories) in a flexible way. I think it means that the user can customise their download of what are called Standard Area Statistics in the UK context rather than an online version of the commissioned tables facility.

The Commissioned Tables service needs radical upgrade

There was a clear message from users that the Commissioned Tables facility from the 2001 Census for England and Wales had been unsatisfactory in terms of delays (sometimes six months just to get to the top of the queue to discuss specifications). Experience in Scotland and in Northern Ireland had been much better. The questions to be addressed:

- (labour) resources available for this service in relation to demand
- software for speeding up disclosure control assessment of the requested outputs (*tau-Argus*) (see later)
- provision of software to users to speed the process of table design (e.g. that would indicate where user specifications were likely to be disclosive).

The legal framework for statistical disclosure control

Will this be improved in the new legislation being prepared that will make ONS an independent agency? (Jill Tuffnell, Cambridge) Can the New Act help in correcting out of date specification in old legislation (e.g. about bodies allowed to receive confidential statistics) (Alison Farland, City University).

The advice from ONS was to respond to the Treasury Consultation on the legislation governing the Office for National Statistics. The web link is:

http://www.hm-treasury.gov.uk/budget/budget_06/other_documents/bud_bud06_odstatistics.cfm

See also the RSS initial response

<http://www.rss.org.uk/main.asp?page=2614>

The 2001 Census, John Hollis

There was strong support for the points made by John Hollis.
SDC creates great inconsistency problems across tables and area levels
SDC makes the small area Origin-Destination Statistics unfit for use
Users want SDC to occur pre-tabulation so outputs can be used effectively.

John Hollis referred to work on the impact of disclosure control by Eileen Howes and Paul Williamson reported at BSPS 2004. See <http://www.lse.ac.uk/collections/BSPS/annualConference/2004/localGovernmentCensusIssues.htm>

Users were also referred to advice to a Population Trends paper by Phil Rees, John Parsons and Paul Norman on "Making an estimate of the number of people and households for Output Areas in the 2001 Census", available at http://www.statistics.gov.uk/downloads/theme_population/PopTrends122v1.pdf

A key issue for ONS was to make sure that future Origin-Destination Statistics were designed to yield useful and consistent information. There is software for enabling users to build their own origin and destination geographies (to obtain sensible flow numbers) provided by the ESRC/JISC Census Programme through the Census Interaction Data Service at <http://cids.census.ac.uk/> (John Stillwell and Oliver Duke-Williams).

A view from a shire county, Wendy Pontin

Considerable concern was expressed about the impact of SDC practice on the quality of Neighbourhood Statistics tables derived from administrative sources. Local government and health authorities were often holding data not subject to the procedures employed in NeSS but unable to "expose" the errors in NeSS because of confidentiality agreements. NeSS tables are subject to different SDC procedures than census statistics: these include suppression and rounding to multiples of five.

Neighbourhood Statistics, David McLennan

There was discussion on the issue raised and analysed by David that Index of Deprivation component indicators published on NeSS differed, because of SDC, from the indicators used to construct the IMD and its components. Gregg Philpotts (ONS) made the point, with respect to the unemployment counts it was the Department of Work and Pensions which had imposed different SDC requirements on ODPM and its contractor, the University of Oxford, and on ONS as publisher of NeSS.

Commercial users' experiences, Keith Dugmore

There was debate about whether there should be further development of privileged access (special licences or safe laboratories) or better released safe datasets. Both routes are necessary but the former should not detract from delivering better datasets to users via CD/DVD/Web for use on their own desk/lap tops. He also challenged the definition of "Disclosure". Until very recently this had been interpreted as finding out something new about a person, rather than simply identifying them, and this had remained the interpretation for the 2001 Census in Scotland. With this in mind, there should be no problem in releasing univariate tables.

Disclosure control at GROS, Peter Scrimgeour

GROS policy with respect to SDC, mainly pre-tabulation record swapping, table design and thresholds, received much praise from users. The small area statistics for Scotland were consistent across tables and area hierarchies, which made analysis much easier for users. For 2011 GROS were investigating “over-imputation as an additional, pre-tabulation SDC measure.

Later discussion revealed that ONS were carrying out research on the impact of record swapping on research analyses.

GROS had decided, however, not to release a Household Sample of Anonymised Records from the 2001 Census as potentially too disclosive. Access to equivalent data in a safe setting via the Scottish Longitudinal Study would be offered soon, see <http://www.lscs.ac.uk/>.

Disclosure control at NISRA, Robert Beatty

In his talk, Robert Beatty discussed the SDC investigation which resulted in a decision to release univariate statistics for grid cells, a product from the 2001 Census which will shortly be available.

For details of the Northern Ireland grid cell data described in Robert Beatty’s presentation, see:

<http://www.nisranew.nisra.gov.uk/census/Census2001Output/GSspec.pdf>

Grid cells constitute a second small area geography but are protected from disclosure by thresholds and releasing only univariate statistics. They enable statistics to be defined for areas which are comparable over time. NISRA will be collaborating with Queens University Belfast in analysis of population trends at the small area scale from 1971 to 2001 using grid cell data.

In response to a question about whether any claims had been made in Northern Ireland of disclosure, Robert Beatty confirmed that no claims had been made, to NISRA’s knowledge.

Phil Rees commented that NISRA’s investigation of the risks of disclosure of a second small area set of census statistics had been innovative in using simple database queries (in SPSS) to look at the “slivers” (inner and outer haloes). It was not necessary to undertake complex GIS analysis, though GIS was useful to display the geography of the slivers.

Robert Beatty provided information about Northern Ireland’s “Pointer” project which aimed to develop an accurately georeferenced address register for Northern Ireland. The project was a collaboration between NISA, Ordnance Survey Northern Ireland and the Royal Mail. There was a lot of work to do to improve the accuracy of the georeferencing but increased accuracy would lead to further concerns about possible disclosure. Keeping geographical relationships “fuzzy” (as in the past) might be a useful technique for SDC.

ONS current policy and practice, Jane Longhurst

The discussion arising from Jane Longhurst's paper focussed on the software tool "tau-argus (τ -argus)" and its planned use with ONS statistical outputs.

Tau Argus is a Statistical Disclosure Control package for table statistics developed by Statistics Netherlands as part of the Computational Aspects of Statistical Confidentiality (CASC) project directed by Anco Hundepool (Statistics Netherlands) with EU Framework 4 and 5 funding. For details see the project home page: <http://neon.vb.cbs.nl/casc/>. The software incorporates an algorithm (CRP-Controlled Rounding Program) developed by Salazar-González (2004, 2006). See

Salazar-González, J.J. (2004) Controlled Rounding and Cell Perturbation: Statistical Disclosure Limitation Methods for Tabular Data. Technical paper, University of La Laguna, Tenerife, Spain.

and

Salazar-González, J.J. (2006) Controlled rounding and cell perturbation: statistical disclosure limitation methods for tabular data

Math. Program., Ser. B 105, 583–603.

Digital Object Identifier (DOI) 10.1007/s10107-005-0666-4.

ONS has adopted τ -argus and CRP as a corporate set of methods for use across a range of table statistics. For a paper describing this adoption see

Philip Lowthian, Giovanni Merola (no date) The application of controlled rounding for tabular data with particular reference to the Tau-Argus software. Office for National Statistics, London.

www.statistics.gov.uk/events/downloads/SessionF2.doc.

This software has the potential for delivering rounded tables which are consistent for a single small area and consistent between small areas and the larger areas they nest within. However, it was not clear from the discussion whether the software had been demonstrated to achieve this user-defined specification or whether there existed a mathematical proof that it could be achieved and under what conditions.

General discussion

The general discussion returned to some of the issues raised earlier in order to achieve clarification.

Disclosure Control through use of derived statistics

There was some confusion about whether ratio statistics (e.g. percentages) could be used as a SDC measure. Although counts could be recovered given knowledge of the base population for a table, there would be a fuzzy range of counts for each cell, dependent on the precision of the percentage.

Aggregation to user geographies

Facilities existed for easy aggregation from OA building bricks to user defined areas (e.g. SASPAC, NOMISWEB, SPSS with OA datasets). However, the errors could be very large when such aggregation was done. There needed to be a way to generate better tables for such bespoke areas (as in the standard hierarchy). If NISRA had demonstrated that geographical differencing did not pose a significant risk and a set of small area univariate statistics for a second geography could be published, then could not this practice be generalised?

Need for delivered tables not analysis laboratories

Jill Tuffnell (Cambridge LA) made the point that LA users did not have time to engage in complex analysis and sophisticated table specification. Simple standard statistics for small areas fit for purpose was what was required.

Maintenance of confidence in National Statistics

A pointed message from a user was as follows: the heavy handed application of SDC to small area statistics presented a considerable risk that users would lose confidence in National Statistics. This risk should be included along with the risks with confidence loss associated with disclosure in determining policy and action on SDC.

Practice in the USA

Cynthia Smith reported that the 2000 USA Census used pre-tabulation record swapping, that tables included 1s & 2s, and that this proved to be a successful solution. This approach had considerable support from users of small area statistics.

Disclosure Control Action Group

The Statistics User Forum agreed to setup a Disclosure Control Action Group, which will continue to discuss SDC issues with the National Statistics offices. The following members agreed to serve on this group: Phil Rees (U of Leeds), John Hollis (GLA), Wendy Pontin (Norfolk CC), Les Taylor (Taylor Nelson Sofres), Jill Tuffnell (Cambs C.C.), Nigel Godfrey (Derbyshire CC), Steve Brown (Competition Commission) and Keith Dugmore (Demographic Decisions). The Group would discuss matters via email and documents, propose meetings with ONS/GROS/NISRA when appropriate and raise SDC matters in the sequence of Advisory Group meetings already organised by the National Statistics offices on which Group members serve.