
SARS NEWSLETTER

NO. 23 – August 2004

Samples of Anonymised Records from the 1991 and 2001 Census

Census Microdata Unit School of Social Sciences University of Manchester Manchester M13 9PL

THE SARS FOR 2001

This newsletter provides an interim update from the May 2004 Newsletter. In September 2004 we aim to produce a consolidated newsletter, timed to coincide with the public meeting on the SARS on 30 September, which will provide answers to most outstanding questions.

The May Newsletter, available at <http://www.ccsr.ac.uk/sars/publications/newsletters/>, explained that there had been a reduction in detail in the 2001 SARS and that this reduction had, itself, been a major cause of the delay in release of the SARS – originally planned for September 2003. We also set out a range of options for accessing a more detailed version of the SARS that were being explored with ONS, GRO Scotland and NISRA. This newsletter provides an interim update.

Individual SAR

There will now be two versions of the Individual SAR.

- (1) A public-use file that should be released by ONS to CCSR by the end of September 2004
- (2) An in-house version of this file that contains more detail

The public-use SAR

Content

The public-use file will be a 3% sample of individuals that will provide no geography below region (although Inner and Outer London will be separately identified) and will contain only five ethnic groups. The file also groups age and work hours and contains only 15 categories of industry and 25 occupational groups. The file will also be subjected to post-randomised perturbation (PRAM). In effect this means that individual records which are unique in the population on a given set of characteristics will be subjected to some editing which will change values on some or all of those variables. The range of possible values that a particular variable can take will be constrained by related variables so that, for example, a child under 16 will not be changed to become married or self-employed.

Although much of the innovative research that was conducted on the 1991 SARS will not be possible with this dataset, the file will be of value in teaching and in providing a first step towards using a more detailed version of the file, only available to researchers who work within ONS. A full specification of the variables and categories in the file are now available at <http://www.ccsr.ac.uk/sars/2001/indiv-publicuse/>.

Access

Although CCSR will not hold exclusive distribution rights, it will, nonetheless, provide the main route for access. Academics will access the 2001 SARs from the CCSR web site. If not already registered with the Census Registration Service (CRS) they will need to register using an Athens ID. This can be done online in a matter of minutes at <http://census.data-archive.ac.uk/>.

Non-academics users will also be able to obtain the SARs from CCSR at a charge to reflect added value, training and user support. In addition, all data will be supplied with a copy of the analysis software package NSDTAT.

The in-house SAR

Content

During the development of the SARs it became clear that the public-use file would no longer meet many research needs and we have now got agreement from ONS, GROS and NISRA for in-house access to a more detailed version of the Individual SAR. This will be available to researchers who can demonstrate that the public-use file cannot meet their needs. The file will contain the same sample as the public-use SAR but more detail and some extra variables.

Of particular importance the file will identify local authorities (except the Isles of Scilly and the City of London). It will also contain the Index of Deprivation (the summary score and scores for four domains for each Super Output Area) and the full sixteen categories of ethnic group. Generally, the file will reflect the original request made to ONS in 2001. The specification of this file can be viewed at <http://www.ccsr.ac.uk/sars/2001/indiv-inhouse/>. Variables marked in red indicate those that extend the detail in the public use file.

Version 1 or interim version of the in-house file

The agreed specification of the in-house file will require some further work by ONS to replace previously collapsed variables. It will therefore require another extraction. In the interim we are using the version of the data initially extracted for the public use file but before final recodes (e.g. ethnic group) were made. This file has just had geography for local authorities added and the 2004 Index of Deprivation. A codebook for this data can be viewed at <http://www.ccsr.ac.uk/sars/2001/indiv-inhouse/codebook/>.

Access

Applying for access to the in-house SAR

ONS will be responsible for processing applications for access to the in-house SAR. Decisions will be made through a Census Research Access Board (CRAB). The Board will include representatives of GROS and NISRA as well as ONS. It will also contain an academic representative, currently Angela Dale. The Board will make recommendations for access to ONS data to the Microdata Release Panel. Comparable arrangements are in place for Scotland and Northern Ireland.

The first meeting of the Census Research Access Board took place on 29 July and agreed terms of reference, criteria for proposals, application forms and the process for approval. Once all these documents have been finalised they will be available on the web and the full details will be included in the next newsletter. It is envisaged that CRAB will meet – either physically or by email - every six weeks or so.

How and where will data be accessed?

Researchers who access the in-house SARs will be required to sign legally binding Census Confidentiality undertakings. At the moment access is only available at Titchfield using a stand-alone PC held in secure conditions. Software packages available are SPSS and STATA.

Longer term there are plans to use a 'virtual microdata laboratory', based at ONS Drummond Gate and originally set up to provide access to business microdata. This will mean that the SARs can be accessed on a secure server (sited in London) from one of a number of ONS sites (Titchfield, Drummond Gate, Newport and Southport) and also from Belfast (NISRA) and Edinburgh (GROS). However, this cannot move forwards until the completion of a security review, commissioned by ONS and due to report in September.

Researchers using the in-house SARs will be limited in the amount of detail they can take away from the safe setting. Basically output that may be disclosive will not be allowed out. We hope to be able to provide clear guidance on what exactly this means by September. All output will be vetted by ONS staff.

Charges for access

At the moment it is envisaged that there will be a charge of about £50 a day for use of the safe setting.

Further information and questions

The SARs meeting on 30 September (<http://www.ccsr.ac.uk/sars/events/2004-09-30>) is planned to provide as much information as possible about all the developments outlined above. It will also provide an opportunity for you to come along and ask questions and raise issues of concern. See below for more details.

Household SAR

As with the Individual SAR we expect there to be two versions, a released file with less detail than requested and an in-house file with full detail. The disclosure work for the Household SAR is currently being done and it is not yet possible to provide more information on the outcomes of this. ONS have a target date of December for release of this file to CCSR.

Small Area Microdata

We made a request for a 5% sample of individual records from the 2001 Census at which the lowest level of geography was the local authority. ONS cannot yet tell us what level of individual detail will be available for this file. The fact that there will now be LA-level data, with extensive individual detail, available in-house at ONS raises the question of whether this file will now have enough additional value to justify the costs. We expect more information on the level of detail that will be permitted to be available for September.

SARs Open meeting: 30 September, RSS, London

An Open Meeting on the Samples of Anonymised Records from the 2001 Census, future census-type activities and procedures for access to microdata will be held at the Royal Statistical Society on Errol Street, London on Thursday 30 September 2004.

The meeting will provide an overview of the ONS protocol on data access and confidentiality and how it has been implemented for the SARs. A range of different methods of data access will be reviewed and discussed and users will be consulted on proposals for in-house access and remote access to a virtual lab.

Most documents will be available on the web some weeks before the meeting. Participants are strongly recommended to read them so that the amount of discussion time can be maximised.

Programme

10.00 Coffee and registration

10.30 Welcome and Introduction to the day
Angela Dale, University of Manchester

10.40 The role of microdata in national statistics
Len Cook, National Statistician
Norman Caven, Registrar General for Northern Ireland
Frank Thomas, General Register Office for Scotland

Discussion and questions

11.15 A framework for implementing ONS's Protocol on Data Access and Confidentiality
Richard Laux, ONS

Discussion and questions

12.00 An overview of the content of the 'public use version' of the 2001 SARs
Jo Wathan, CCSR, University of Manchester

12.30 Lunch

1.15 Methods used to ensure confidentiality in the SARs

How was disclosure risk assessed? How well has it worked? How has perturbation been applied? What is the effect on data quality?

Bill Gross, Katherine Merritt and Philippe Guiblin, ONS
Mark Elliot, CCSR, University of Manchester

Discussion opened by *Chris Skinner, University of Southampton*

2.15 The in-house version of the SARs: what detail will be available, what are the procedures for access?
Keith Whitfield, ONS

Discussion

3.00 Tea

3.30 Demonstration of access to a virtual microdata lab
Felix Ritchie, ONS

Discussion

4.00 Balancing the requirement for confidentiality with research access: A panel discussion

What are the requirements for confidentiality? What are the options for meeting them? What guarantees can HEIs make? What is the role of the ESRC?

Speakers from NSIs and ESRC will give short introductions and then the topic will be thrown open for discussion

To book for this meeting please go to www.ccsr.ac.uk/sars/events.