

# SARs NEWSLETTER

November 2005

Samples of Anonymised Records from the 1991 and 2001 Census

CCSR, School of Social Sciences, University of Manchester, Manchester, M13 9PL  
www.ccsr.ac.uk/sars/ sars-helpdesk@manchester.ac.uk 0161 275 4735

## The SARs for 2001

**The delivery of the 2001 SARs is now complete**

Three microdata files have now been released from the 2001 Census:

- 3% Licensed Individual SAR
- 1% Special Licensed Household SAR
- 5% Licensed Small Area Microdata file

For the Individual and Household SARs there are much more detailed versions which can be accessed in a safe setting within the Office for National Statistics (ONS).

These data files are discussed in more detail in this newsletter.

### CONTENTS

The Licensed 3% Individual SAR.....	2
A more detailed version of the Individual SAR available as a Controlled Access Microdata Sample (CAMS).....	2
The Special Licensed 1% Household SAR.....	3
Household SAR available as a Controlled Access Microdata Sample.....	3
The Licensed 5% Small Area Microdata file (SAM).....	4
Documentation and user support.....	4
Access and costs.....	5
Overseas use.....	6
Key findings and publications.....	7
Seminars and workshops.....	7
CCSR Short Courses.....	7
Web sites.....	7
SARs support.....	8

## **The Licensed 3% Individual SAR**

The first version of the Licensed Individual SAR was available for download at the end of October 2004. This file did not contain religion and had restricted detail on occupation and country of birth. In March 2005 ONS supplied a second version of the file with more detail.

The key features of version 2 of the file are:

- 1.8 million records representing 3% of the enumerated population of England, Wales, Scotland and Northern Ireland
- About 85 variables covering all the topics in the census for the sampled individual and his/her family
- Ethnic group in 16 categories (14 in Scotland and 2 in Northern Ireland)
- Religion in 9 categories in England and Wales (11 in Scotland and 7 in Northern Ireland)
- Occupation in 81 categories (SOC Minor)
- NS-SEC in 40 categories
- Country of birth in 16 categories
- Age categories grouped between 16 and 74 (8 groups) and single years up to 16 and over 74

Variables are also present to indicate whether or not a response was imputed or changed through the perturbation process.

## **A more detailed version of the Individual SAR available as a Controlled Access Microdata Sample (CAMS)**

This file represents the same sample as the Licensed Individual SAR but contains much more detail. Access is controlled by ONS. Potential users need to complete a full application and, if accepted, they can analyse the data in a safe setting in one of four ONS offices: London, Newport, Titchfield and Southport.

The CAMS file contains geography at LA level, full occupational detail (unit SOC) and industry and more detail on many other variables. For example, the full coding on country of birth is available, which is about 200 categories. The 2004 Index of Deprivation scores have been added to SuperOutput Areas for England and Wales (and deciles for Scotland). In total the file contains about 200 variables and provides very considerable detail. A codebook giving a complete listing can be downloaded from CCSR at [www.ccsr.ac.uk/sars/2001/indiv-cams/](http://www.ccsr.ac.uk/sars/2001/indiv-cams/).

The CAMS version of the file has not had any perturbation applied to it.

## **The Special Licensed 1% Household SAR**

A 1% sample of households and all household members for England and Wales only has been prepared by CCSR and deposited at the UK Data Archive. The file allows linkage between individuals in the same family and in the same household. Access is under an ONS Special Licence in order to give additional confidentiality protection to the file.

The ONS assessment of the disclosure risk of the Household SAR resulted in proposals for very heavy grouping in key variables such as age and marital status. This loss of detail would have resulted in a file with very limited research value. The compromise solution was to produce a file with more detail but to add more protection through a Special Licence. The ONS Special Licence adds additional conditions to the standard End User Licence by requiring more tightly controlled conditions of use and heavier penalties for misuse. A copy of the ONS Special Licence can be viewed at [www.data-archive.ac.uk/orderingdata/specialLicence.asp](http://www.data-archive.ac.uk/orderingdata/specialLicence.asp) and a 'Good practice' guide to meeting the conditions imposed by the licence is linked to the same page.

The file contains:

- 225K households and 526K individuals representing 1% of households for England and Wales only
- Individual records in households of size 12 or more have been suppressed for confidentiality reasons
- About 60 variables covering all the topics in the census for the sampled household and individuals
- Ethnic group in 16 categories
- Religion in 9 categories
- Occupation in 81 categories (SOC Minor)
- Country of birth in 16 categories
- Age in two-year bands

Variables are also present to indicate whether or not a response was imputed or changed through the perturbation process. A set of derived variables will be added to the file during 2006.

### **The Household SAR available as a Controlled Access Microdata Sample (CAMS)**

As with the Individual SAR, there is also a CAMS version of the Household file. This contains full detail on all variables in each file. It also includes data for Scotland and Northern Ireland. There has been no perturbation of the data. In addition, a large number of other derived variables have been added. A codebook can be downloaded from the CCSR web site at [www.ccsr.ac.uk/sars/2001/hhold-cams/codebook/](http://www.ccsr.ac.uk/sars/2001/hhold-cams/codebook/).

### **The Licensed 5% Small Area Microdata file (SAM)**

This file was designed in response to users' requests for more detailed geography. It was based on the assumption that there was a 'trade-off' between geographical detail and individual detail. It therefore took local authority geography as the fundamental requirement and assessed the sample size required for analysis at this level and the individual detail that would be safe to provide.

A paper developing the arguments for a Small Area Microdata file is published in the Journal of the Royal Statistical Society (Tranmer, M., A. Pickles, E. Fieldhouse, M. Elliot, A. Dale, M. Brown, D. Martin, D. Steel and C. Gardiner (2005) "The case for Small Area Microdata," JRSS(A), 168, 1, 29-49.).

The SAM is a 5% sample of individuals for all countries of the UK, with 2.96 million cases. Local Authority is the lowest level of geography for England and Wales, Council Areas for Scotland and Parliamentary Constituencies for Northern Ireland. The Scilly Isles have been merged with Penwith and the City of London with Westminster. For Scotland, Orkney and Shetland are merged into one area. All other areas are identified.

The file contains:

- About 40 variables covering all of the topics in the census for the sampled individual
- Ethnic group in 13 categories (8 in Scotland and 2 in Northern Ireland)
- Religion in 9 categories in England and Wales (11 in Scotland and 7 in Northern Ireland)
- Country of birth in 5 categories
- NS-SEC in 8 categories
- Age in 13 bands

There are no variables giving occupational detail.

## **Documentation and user support**

The SARs web site contains codebooks for all the above files, either as a PDF, html or both. There is also a *User Guide to the SARs*, which covers both 1991 and 2001 SARs. Again, this is available online at [www.ccsr.ac.uk/sars/guide/](http://www.ccsr.ac.uk/sars/guide/) and can be downloaded as a PDF. It provides an in-depth guide to both the 1991 and 2001 SARs and contains numerous links to related resources. It is updated as new data files and new resources become available.

A very useful *Getting Started with the Samples of Anonymised Records* guide can also be found on the SARs web site. It is valuable for new users and can also be used in teaching. In addition, overheads and exercises from workshops on the SARs, together with presentations from user group meetings, are available from [www.ccsr.ac.uk/sars/events/](http://www.ccsr.ac.uk/sars/events/). All these materials can be freely re-used, although we would appreciate it if the source is cited.

In the coming months we will be extending and revising our documentation.

## **Access and costs**

### *Licensed Individual SAR and Licensed Small Area Microdata file*

To access these files academic users need to register with the Census Registration Service (if not already registered) and can then download the files from the CCSR web site on supplying their Athens username and password. The files can also be accessed and downloaded using NESSTAR, an online data exploration and analysis package. A guide to using the SARs in NESSTAR can be downloaded from the SARs in NESSTAR web page at <http://nesstar.ccsr.ac.uk/index.html>. There is no cost for academic users.

The SARs are available for download in SPSS, SPSS portable, STATA and tab-delimited format. They can also be downloaded with NSDstat as an analysis package. Free download of NSDstat means that the data can be used by researchers who may not have access to a standard software package and forms an important part of our strategy to widen use of the data in the further education sector.

The data files are also available in two versions. One version contains imputation flags for each of the variables in the SAR files. This allows users to explore the extent and nature of item imputation and decide whether or not to include imputed information in their analysis. The other file leaves out these variables and is thus smaller and simpler. Both files contain a variable that flags whether a respondent has been imputed as part of the One Number Census. (To increase confidentiality the Scotland and Northern Ireland Small Area Microdata file does not contain imputation flags.)

Registration for business users of the SARs costs £1000 for each file. Local Authorities, Health Authorities, Central Government and charities are entitled to a discounted non-academic rate of £500 per file. VAT is payable on these amounts. This will entitle the organisation to use the data and software for up to ten people, however, more users can be added using an additional Licence Agreement and with a supplementary charge of £50 per user. An application form can be downloaded from [www.ccsr.ac.uk/sars/access/nonacademic/](http://www.ccsr.ac.uk/sars/access/nonacademic/). Two copies must be completed, signed by someone legally competent on behalf of the organisation and sent to CCSR. Once this has been authorised by the University of Manchester, data are supplied on CD-ROM. (A web download can be arranged for users who prefer that mode of access). All non-academic users are provided with a copy of NSDstat which can be used to analyse the data.

### *Household SAR*

As discussed above, the Household SAR is available only under an ONS Special Licence. An application for a Special Licence (SL) is made via the UK Data Archive (UKDA). Access to data must be for statistical research purposes only and applicants need to justify why the more detailed version of the data is needed. The ONS Special Licence sets out standards and methods to ensure that outputs avoid any inadvertent disclosure of information that could identify a respondent. Researchers are also required to supply the UKDA with bibliographic details of any published work. The licence also contains clauses relating to data and site security (for example the data cannot be used at home or transported on a laptop), destruction of the data on completion of the project and the sanctions to be applied if any conditions are breached.

After submitting a completed application, the form is screened by UKDA and ONS. Data are only provided to those researchers who can justify their need for the data for statistical research purposes and agree to meet the terms of the SL. Users who obtain access under the SL are also required to follow the *Guide to good practice: microdata handling and security*, which provides practical guidance on how to meet the conditions of security imposed by the licence.

The Household SAR is not confined to academic use. Providing the conditions of the Special Licence are met, the file can be used by researchers in any sector.

### *Controlled Access Microdata files (CAMS)*

Access to the CAMS is handled by ONS. Applications need to explain the research purposes for which the data will be used and why the licensed files do not give the required information. The application forms, a description of the application process, the criteria used to assess applications and the confidentiality guidelines are available on the web at [www.statistics.gov.uk/census2001/sar\\_cams.asp](http://www.statistics.gov.uk/census2001/sar_cams.asp).

Applications to use the CAMS are reviewed by the Census Research Access Board (CRAB), led by ONS and chaired by Simon King. Membership includes representatives of ONS, the General Register Office for Scotland (GROS), Northern Ireland Statistical Research Agency (NISRA) and Cathie Marsh Centre for Census and Survey Research (CCSR).

The application process should take 1-2 months. Although researchers can access the full detail on the microdata file, only non-disclosive results are allowed out of ONS. Results are checked for confidentiality and then emailed to the researcher. There is a charge of £52 per day + VAT. There is some ESRC funding available to support academic access and this is administered by CCSR. Details of how to apply are available on the SARs web site at [www.ccsr.ac.uk/sars/2001/indiv-cams/access/](http://www.ccsr.ac.uk/sars/2001/indiv-cams/access/).

There is a database of CAMS users on the SARs web site at [www.ccsr.ac.uk/sars/2001/indiv-cams/users/](http://www.ccsr.ac.uk/sars/2001/indiv-cams/users/), so that potential applicants can see work that has been done so far. Researchers must submit all publications, conference papers and overheads to ONS for clearance. A very useful CAMS User Guide can be downloaded from the ONS web site at [www.statistics.gov.uk/census2001/downloads/cams\\_user\\_guide.pdf](http://www.statistics.gov.uk/census2001/downloads/cams_user_guide.pdf). Researchers are asked to indicate on the application form that they have read the CAMS User Guide.

As for the Household SAR, the CAMS can be used by researchers in any sector as long as the conditions of use are met.

### *The working environment for the CAMS*

Access to the CAMS is through the Virtual Microdata Laboratory, a server held at Drummond Gate, London, to which users have access remotely from dumb terminals. Each user has an allocated space on the server in which they can work using SPSS, STATA or MLWin. The user interacts directly with the data. Syntax files must be prepared beforehand and sent to ONS for reading onto the server before the user arrives. Outputs are stored on a secure separate area of the server and are emailed to the user after screening by ONS staff. All users need to make a subset of variables and cases before beginning analysis.

### **Overseas users**

ONS has recently approved access to the Licensed 3% Individual SAR and the Licensed 5% Small Area Microdata file by overseas users. The Special Licensed Household SAR is not available for overseas use. Details of access conditions will be posted on the SARs web site as soon as they are available.

## **Key Findings and publications**

Key findings from the 1991 SARs and a list of SARs-related publications and downloadable newsletters (since 1993) provide a very full source of information on the research uses of the SARs, as well as a record of how the data have developed over time. We ask all users to respond to the annual publication trawl by sending us details of their new publications. In the coming months we will be adding publications and key findings from the 2001 SARs.

## **Seminars and workshops**

The SARs team is very happy to come and run a workshop or give a seminar at your home institutions. Please contact Gill Meadows in the first instance at [gillian.meadows@manchester.ac.uk](mailto:gillian.meadows@manchester.ac.uk).

## **CCSR short courses**

CCSR run a programme of short courses on analysis methods relevant to users of the SARs. The next courses to run include Sampling Theory, Introduction to STATA and Qualitative Comparative Analysis. Courses are offered at introductory, intermediate and advanced level. The complete list of short courses is available at [www.ccsr.ac.uk/courses/](http://www.ccsr.ac.uk/courses/).

Bursaries are available for all courses but are limited and awarded on a first-come, first-served basis. Details can be found at [www.ccsr.ac.uk/courses/bursary.html](http://www.ccsr.ac.uk/courses/bursary.html).

### *December 2005 to February 2006*

5 December:	Sampling Theory
9 December:	Introduction to STATA
19 December:	Qualitative Comparative Analysis
9 January:	SPSS for Social Scientists
10 – 12 January:	Structural Equation Modelling in M-plus
13 January:	Introduction to Data Mining
18 January:	Introduction to STATA
25 – 27 January:	Design and Analysis of Complex Surveys
3 February:	Computerised Qualitative Analysis
15 February:	Computerised Qualitative Analysis
22 February:	Computerised Qualitative Analysis
22 February:	Aspects of Statistical Modelling

For further information about these or any other CCSR courses, please contact Kathryn Clements on [kathryn.clements@manchester.ac.uk](mailto:kathryn.clements@manchester.ac.uk) or 0161 275 4736.

## **Web sites**

The SARs web site at CCSR: [www.ccsr.ac.uk/sars](http://www.ccsr.ac.uk/sars)

Census Registration Service: <http://census.data-archive.ac.uk>

Controlled Access Microdata files: [www.statistics.gov.uk/census2001/sar\\_CAMS.asp](http://www.statistics.gov.uk/census2001/sar_CAMS.asp)

# ***SARs support at CCSR***

## **Contact us at:**

Cathie Marsh Centre for Census and Survey Research (CCSR)  
School of Social Sciences  
University of Manchester  
Crawford House  
Manchester  
M13 9PL

If you have any queries or comments about the SARs please direct them to the SARs helpdesk by one of the methods below. The helpdesk is staffed from 9:00am to 5:00pm Monday to Friday.

*Email:* sars-helpdesk@manchester.ac.uk  
*Helpline:* +44 (0) 161 275 4735  
*Fax:* +44 (0) 161 275 4722  
*Web site:* [www.ccsr.ac.uk/sars](http://www.ccsr.ac.uk/sars)

## **The SARs team**

Angela Dale, Ed Fieldhouse and Mark Brown are responsible for the overall direction of the SARs programme.

Angela Dale has responsibility for consultation and negotiations over the content of the SAR files. The day-to-day work to support and develop the SARs is provided by research fellow Jo Wathan and research associate Reza Afkhami, who work half-time on the SARs and half-time on other research projects. They have a great deal of experience of large-scale datasets, including the GHS, LFS and the SARs.

Sam Smith is the SARs web interface developer, with particular responsibility for downloads and access using Nesstar. Gill Meadows provides administrative support including organising events, preparing newsletters and updating web pages.

## **Resources for the SARs**

The SARs web site contains important resources to facilitate your analysis of the data. These include:

- Frequently Asked Questions
- Getting Started with the SARs
- User Guide to the SARs - essential documentation for all users of the SARs
- Introductions to common software packages
  - Analysing the SARs in SPSS
  - Analysing the SARs with STATA

## **Publications**

A list of SAR-related publications and past copies of SARs newsletters are available at [www.ccsr.ac.uk/sars/publications/](http://www.ccsr.ac.uk/sars/publications/). To submit a publication based on research conducted with the SARs you can email [gillian.meadows@manchester.ac.uk](mailto:gillian.meadows@manchester.ac.uk) or use the web form at [www.ccsr.ac.uk/sars/publications/submit.html](http://www.ccsr.ac.uk/sars/publications/submit.html).