

Samples of Anonymised Records

Cathie Marsh Centre for Census and Survey Research

User guide to the SARs

Date: 9 November 2004

Version: 3

**Cathie Marsh Centre for Census
and Survey Research**

University of Manchester
Crawford House
Manchester
M13 9PL

www.ccsr.ac.uk/sars
sars-helpdesk@manchester.ac.uk
+44 (0) 161 275 4735

SARs
samples of
anonymised records

MANCHESTER
1824
The University of Manchester

JISC

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

Acknowledgements

Sources of material used in this Guide include:

- The Census 2001 Definitions Volume
- The 1991 Census Definitions and Concepts
- The National Statistics web site: www.statistics.gov.uk
- The General Register Office for Scotland web site: www.gro-scotland.gov.uk
- The Northern Ireland Statistics and Research Agency (NISRA): www.nisra.gov.uk

All these materials are Crown copyright and reproduced with the permission of the Controller of HMSO.

1. Getting Started with the SARs

1.1 What are the SARs?

The Samples of Anonymised Records are samples of individual level data drawn from the 1991 and 2001 Census. Identifying information has been removed to protect confidentiality. With a separate record for each individual the SARs are similar to the sort of data obtained from a sample survey. However the sample size is much larger than most surveys thus permitting analysis of small groups and sub-national areas. The SARs cover the full range of Census topics including, housing, education, health, transport, employment and ethnicity.

1.2 Why use the SARs?

The SARs are sets of microdata sampled from the Census. They allow flexible, multivariate analysis at the individual level. They are a unique data source for the investigation of a range of social issues including household composition, ethnic differences, education and employment. The SARs differ from traditional Census outputs in that they are not aggregated into pre-determined tables.

1.3 Which dataset do I want?

There are a range of SAR datasets available.

- The Individual SAR is a national sample of individual census returns.

The 1991 Individual SAR is a 2 per cent sample and contains over 1.1 million records. It contains a full range of census topics on individuals and summary information about households. In total 278 geographical areas are identified and include all those large local authority districts with a population of at least 120,000. Small local authorities have been grouped together.

The 2001 Individual SAR is a 3 per cent sample and contains over 1.5 million records. It contains a full range of census topics on individuals and summary information about households and the new information collected on qualifications and caring. Geographical information is given down to Government Office Region. Religion is not available on this file.

- The Household SAR is a sample of households and people within those households.

The 1991 Household SAR is a 1 per cent sample of the population and contains 216,000 households and 0.5 million people within these households. It is a hierarchical file that allows linkage between household and family members. The geographical base is the Registrar General's Standard Regions in 1991. The Household SAR contains about 40 variables though many derived variables have also been added.

The 2001 Household SAR is a 1 per cent sample of the population and contains over 200,000 households and 500,000 people within households. Again it allows linkage between household and family members.

All data files are available for England and Wales, Scotland and Northern Ireland.

- The 2001 Small Area Microdata. Further details of this dataset will be available on the SARs web site.

The following questions may help to guide you towards the most appropriate dataset for your research question.

- **What national geographic areas am I interested in?** Both the 1991 and 2001 SARs allow comparisons between England, Wales, Scotland and Northern Ireland. In 1991 England, Wales and Scotland are in the same file. Northern Ireland is a separate file. It is possible to construct a UK comparative file although this has less detail than the separate files. In 2001 all countries of the UK will be in a single file for each SAR. There are however, differences in questions between the four countries.
- **Will I be analysing individuals, households or individuals in a household context?** If you are interested only in individuals then the most suitable dataset is probably the Individual SAR. If you are interested in the relationships of people within households or in analysis at the household level then the Household SAR is the dataset for you. The 1991 the Census included 'visitors' in addition to 'usual residents'. To avoid double counting users of the 1991 file must specify a 'usually resident population' using RESIDSTA. The institutional population is included in the Individual SAR.

- **Do I want data with more detailed geography than region?** The 1991 Individual SAR provides information down to large local authority districts of 120,000 people. From 2001 onwards the Individual SAR does not go below Government Office Region, although Inner and Outer London are separately identified. The proposed 2001 Small Area Microdata SAR will have a more detailed geography (local authority level) but less detail.
- **What topics am I interested in?** The SARs contain a range of census topics on individuals and households. All the topics are covered in both files though the level of detail is different in 1991.
- **Am I interested in a sub-population?** The SARs are very large samples and allow analysis of different populations such as minority ethnic communities. Depending on the type and level of your analysis the Individual or Household SAR may be most appropriate.
- **What if my research needs more detail than is available in the SARs for 2001?** A more detailed version of the Individual and the Household SAR (Controlled Access Microdata Samples) will be available for analysis at ONS Titchfield for researchers who can demonstrate that the released files cannot meet their needs.

More information on all the data files is available from the SARs web site at www.ccsr.ac.uk/sars.

1.4 Consulting the documentation

It will not be possible to analyse the SARs without reference to the documentation. You can access documentation on the web before you register to use the data. Documentation can also be downloaded with the data.

1.5 Exploring the data online

An easy way of starting to explore the data is to look at the data in NESSTAR. Go to <http://nesstar.ccsr.ac.uk/nesstarlight/index.jsp>.

Nesstar is a data exploration and analysis tool that allows you to look at univariate frequencies and variable descriptions without needing to register for the data. This software makes it easy to specify and make your own tables or graphs from the raw data. This is an excellent way of finding out whether the data are likely to be suitable for your purposes. You can access Nesstar from the SARs datasets web page. Once in Nesstar, click on the Help tab in the top right hand corner of the screen to view a user guide. Once you have registered you will be able to use Nesstar to its full potential, undertaking more complex analyses and downloading subsets of data in a wide variety of formats.

1.6 Registering for the data

Once you have decided to use the SARs for your research you will need to register to obtain the data. Academic users can register through the Census Registration Service at www.census-registration.ac.uk. Non-academic users should contact us at the above address. The data are free to academic users.

CCSR will supply the 2001 SARs to public sector bodies for a charge of £500 per file and to the business sector for £1000 per file. The data includes NSDstat. This is a desktop statistical analysis package developed around the concepts of user-friendliness and the power to process large amounts of data in minimal time. Features of NSDstat include univariate, bivariate and multivariate analysis, graphical representation of data, regression, weighting and recoding.

The licence will entitle the organisation to receive the data, which may then be accessed by ten people, all of whom will have to sign a user undertaking. All ten users in each organisation will receive support backup and a free place on a training course. Further copies of the CD and registration of additional users will be available for an additional fee. Please contact the SARs support desk for more information.

1.7 Downloading the data

Academic Users: once registered you can download data from the SARs datasets page. You may either download the data directly as a portable SPSS file or as a tab de-limited file or as a STATA file. If you are unable to use these methods, you can request a CD-ROM containing SARs files from CCSR via our contact form at www.ccsr.ac.uk/sars/contact.html. Obtaining the data in this medium will cost £30.

Non-Academic Users: once registered the data will be made available on CD with NSDSTAT. Please contact the SARs support desk for more information. Email sars-helpdesk@man.ac.uk.

1.8 Analysing the data using an appropriate analysis package

Because the SARs are microdata you will need to analyse them using an appropriate package. You can undertake straightforward analysis, including linear regression, on the whole dataset or a subset using Nesstar. However, it is not possible to produce derived variables or do more advanced analyses using Nesstar. To do this you will need to use a package like SPSS, STATA or SAS.

Several organisations offer introductory SPSS and STATA courses. CCSR offers one-day courses although there is a charge. Interested parties may currently apply for a bursary to cover this cost. Details of these courses can be obtained through the CCSR Short Course Pages at www.ccsr.ac.uk/courses/.

Or you could use NSDstat, as described in section 6. Academics can download a copy of NSDSTAT for the SARs free of charge. For non-academic use, purchase of the SARs includes a copy and license for NSDstat for up to 10 users. More information can be found at www.ccsr.ac.uk/sars/access/modes.html.

1.9 Does the data provide you with the detail you need?

Access to a more detailed version of the Individual and Household SARs is possible in an ONS safe setting. For more information on the Controlled Access Microdata Sample (CAMS) go to the SARs web site.

1.10 A teaching resource

You may also use the teaching/learning materials provided by the Centre for Census and Survey Research, which are freely available online and include direct access to SAR data, at www.ccsr.ac.uk/rschproj/chcc/. The materials include both methods learning (from tables to logistic regression) and substantive topics (for sociology, demography, social policy and other disciplines).

1.11 Communicating with other users and obtaining support

SARs queries are logged and monitored on a helpdesk system. Email queries should be sent to sars-helpdesk@man.ac.uk to ensure the fastest possible response. A dedicated telephone helpline is available on 0161 275 4735 from 9.00am to 5.00pm, Monday to Friday.

CCSR provides a range of support services at www.ccsr.ac.uk/sars/support/services.html, including provision of customised tables and analyses, for those who use or are interested in using the SARs. CCSR also offers workshops, seminars and downloadable training documentation for the SARs in general and for analysing the SARs in supported statistical packages.

There is a regular SARs newsletter available at www.ccsr.ac.uk/sars/publications/newsletters. To join the SARs email list go to www.jiscmail.ac.uk/lists/sars.html.

1.12 Publishing

We maintain a record of all publications which use the SARs. For a list of publications from the SARs, or to add your publication to the list go to www.ccsr.ac.uk/sars/publications/.

Data collectors and distributors should be acknowledged in publications. The data remain Crown copyright and this should be acknowledged in publications along with an acknowledgement in the form set out below: "This work is based on the SARs provided through the Centre for Census and Survey Research of the University of Manchester with the support of ESRC and JISC."