

---

# SARS NEWSLETTER

NO. 19 – MARCH, 2003

Samples of Anonymised Records from the 1991 and 2001 Census

---

Website: <http://www.ccsr.ac.uk/sars/>

Email: [sars-helpdesk@man.ac.uk](mailto:sars-helpdesk@man.ac.uk)

SARs Helpline (0161) 275 4735

Fax: (0161) 275 4722

CCSR, University of Manchester, Manchester M13 9PL

---

## News from the SARs

This Newsletter presents a brief update on the 2001 SARs.

The **sampling design** of the SARs was agreed in October 2002 and the sample has been marked and is now being extracted. The sampling design is very similar to that for 1991 except that the sample is now drawn from the entire census population instead of a 10% sample.

ONS have agreed to do the **coding for the additional occupational information** that has not been coded for the main census (i.e. occupation/industry for 16-64s who last worked more than 5 years ago but less than 10, and 65-74s not currently working who worked in the previous 10 years). This is very good news because, without this information, we would have no occupation or social class information for people who have a long-term absence from the labour market (5-10 years), or who retired early or stopped work beyond normal retirement age. However, no occupational information will be released at unit level (353 categories). Occupational information for both files will be at the level of SOC minor (81 categories).

As with other Census outputs the SARs have been subject to a stricter confidentiality appraisal than in 1991 and this has delayed the completion of the specification. The original timetable assumed that all the detail would be agreed by September 2002. However, at the end of October 2002 ONS released their response to the ESRC request, specified by CCSR, and began their own consultation process. The paper released by ONS can be accessed from the SARs website: [www.ccsr.ac.uk/sars](http://www.ccsr.ac.uk/sars).

This has meant a great deal of re-thinking and the process is not yet complete. However, we are grateful to all the people who have attended meetings in London and Manchester, written a response to ONS and replied to our numerous requests for their views. After consulting users we have defined the key research priority of each file and tried to ensure that we maximise detail for the variables required to meet that priority. The implications for each data file are discussed below.

---

### **Individual file (3%)**

For the Individual file our starting point is the requirement for full ethnic coding and maximum detail on occupation, age and other related variables. This is needed to support the employment and labour market analysis that has been important with the 1991 SAR.

The Individual SAR will now be a 3% sample of individuals. The near final specification is on the CCSR web site at: ([www.ccsr.ac.uk/sars](http://www.ccsr.ac.uk/sars)). In order to retain sufficient individual detail (e.g. 16 categories of ethnic group) the lowest level of geography will be the Government Office Region. However, we expect to be able to compensate for this reduction in geography in two ways:

- By the addition of an indicator of the ranking on the Index Of Deprivation for 2001 (decile) for the ward of enumeration.
- By the addition of a pseudo-LAD indicator. This uses an imputation process in which distributions from the 100% Census tables are used as constraints to allocate individuals in the SAR to LADs (see insert).

ONS still hope that the final specification will be agreed by April with delivery of the Individual SAR to CCSR by September 2003. We will post developments on the SARs web pages: [www.ccsr.ac.uk/sars](http://www.ccsr.ac.uk/sars).

### **Household File (1%)**

In responding to ONS we have defined the research priority of this file as:

To provide demographic and household composition and information on household circumstances and living conditions of different groups, including minority ethnic groups, household projections and housing needs.

Key variables required: single year of age, sex and full ethnic group classification; relationships within families and household derived from household matrix. In order to retain as much individual detail as possible we have agreed to lose regional information.

The specification of the **Household SAR** is waiting on additional checks by ONS. In 1991 confidentiality concerns over large households led to individual records being suppressed for households of size 12 and over. This represented only 28 households. In their October consultation ONS expressed concern over households with more than 6 members – or more than 5 members if region was retained.

The current specification will have the following reduction in detail:

- Households with 7 members will have aged grouped in 5-year bands (except for children under 16 (single year) and young people to age 24 (2 year age groups)).
- Households with 8 members have age banded in 10 year groups and a reduction in detail on ethnic group (5 categories); marital status (2 categories).
- Households with 9 members have age in only 3 categories and reductions in marital status and ethnic group as above.
- For households with 10+ members there will be no individual information at all but the possibility of some derived variables. Based on 1991 figures this represents about 1,000 usual residents and about 100 households.

Whilst the actual numbers of individuals and households are relatively small, large households are not randomly distributed in the population. For example, the loss of this information would disproportionately affect Pakistani and Bangladeshi ethnic groups and would bias estimates of

---

overcrowding and various forms of deprivation. We are currently consulting over ways of imputing individual-level information for these large households.

We hope to have a specification of the Household file on our web site by early May. Whilst the intention is that this file should be released on September with the Individual file there is a greater risk that it will be delayed.

### **Small Area Microdata**

ONS has accepted the principle of a Small Area Microdata file but the specification for this will not be developed until the Individual and Household files have been resolved. However, we hope that this file will restore some information at LAD level although we do not yet know the extent of individual detail that will be available.

### **A safe setting**

ONS are working on plans for a microdata lab that would allow access to much more detail than on the released files. We hope that this facility may be available by the autumn but no decisions have yet been taken. We have emphasised to ONS that this is not a replacement to access to data on the desktop.

### **Web-based access to the 1991 SARs**

We now have the 1991 SARs available for web-browsing and download. The software and modes of access are those that will be used with the 2001 SARs. We therefore urge prospective 2001 SAR users to try out two new software packages with the 1991 data. As well as the ability to download the entire files in SPSS or STATA format, users can now access the data via their web-browser using NESSTAR or download to their desktop with a tailored analysis package, NSDstat.

Users will be asked to provide their ATHENS user name and password –quick, easy and linked from the CCSR site. If a user tries to log in without registering for access to Census data, then they are automatically prompted to register, and can immediately afterwards access the data.

### **NESSTAR**

The Nesstar Light is accessed via a web browser: just go to [www.ccsr.ac.uk/sars](http://www.ccsr.ac.uk/sars)). NESSTAR has been developed at the UK Data Archive at the University of Essex and the Norwegian Social Science Data Services. You can use it to browse detailed information about the SARs, interactively, to carry out simple data analysis (e.g. tabulations and graphical displays) and then download data, in whole or part, in one of a number of popular formats. . One of the principal strengths of Nesstar is its ability to quickly and easily generate a multivariate table or graph and then provide the data for download in a selection of formats for desktop analysis. The software enables sophisticated data browsing tools to be seamlessly integrated into the user's desktop. The generic data browser operates on the Web and is not tied to a particular dataset. Nesstar shares its core statistical engine with NSDstat – also available to SAR users.

### **NSDSTAT**

The 1991 SARs can also be downloaded direct to a PC with their own analysis package. NSDstat 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. It can be downloaded from the SARs website together with the 1991 SARs.

---

NSDSTAT is supplied under licence by Norwegian Social Science Data Services (NSD) and freely available for academic users - thanks to licences purchased through our ESRC/JISC grant. This allows any academic users to explore and analyse the SARs without the requirement for pre-existing software such as SPSS.

The NSDstat interactive help includes descriptions of how and why an analysis can be conducted, in addition to how you would conduct that analysis in NSDstat. NSDstat is suitable for use by those who require some degree of support for irregularly performed statistical operations, as well as for teaching.

#### **Access to the 2001 SARs**

All these methods of access will be available for the 2001 SARs. However, **non-academic users** will register direct with CCSR rather than through the census registration system. Instead, we will supply 2001 SARs on CD-ROM with NSDstat – and in a range of other formats. If non-academic users wish to download direct from the web we will supply them with a password to allow this. Please contact CCSR and we will provide detailed information on the costs of acquiring the 2001 SARs.

## **Teaching and learning materials for the SARs Contemporary and Historical Census Collection (CHCC)**

The value of SAR data has been considerably enhanced by the development of an extensive set of web-based teaching and learning resources. This work has been part of a wider project funded by the Joint Information System Committee (JISC), the aims of which are to increase use of historical and contemporary census data (CHCC) in learning and teaching by improving accessibility to the primary data resources and developing an integrated set of learning and teaching materials.

The SAR-based materials developed by CCSR are packaged in a series of stand alone units, covering both methodological and substantive topics. Their design and content has been informed by active consultation with potential users of the resource, and pilot units are currently under evaluation. All units include materials that can be downloaded and customised for use in the classroom, as well as on-line versions for private study. All units incorporate practical exercises based on a teaching subset of SAR data. Crucially, new modes of web-based data visualisation and analysis (NESSTAR and NSDSTAT – see above) means these are accessible to those without access to, or experience in using, packages like SPSS and STATA. The CHCC initiative should considerably broaden the appeal and accessibility of SARs to a wider community of users.

For latest progress on the SAR teaching and learning materials, including access to latest pilot units visit the project website at: [www.ccsr.ac.uk/rschproj/chcc](http://www.ccsr.ac.uk/rschproj/chcc). Current units are based on 1991 data but will be updated to incorporate 2001 data as this becomes available.



