
SARS NEWSLETTER

NO. 22 – May, 2004

Samples of Anonymised Records from the 1991 and 2001 Census

Census Microdata Unit Faculty of Social Sciences and Law University of Manchester Manchester M13 9PL

THE SARS FOR 2001

This newsletter provides a further update on recent developments in the production of the SARs for 2001.

Timetable for the SARs

The last newsletter warned that the Individual SAR was unlikely to be delivered to CCSR before April 2004 and that all details in the specification were subject to revision following a further round of confidentiality checks. We have just been told by ONS that the estimated release date for the public Individual SAR file is September 2004 and for the Household SAR and SAM public files is December 2004.

The Individual SAR was extracted by ONS according to the October 2003 specification on the CCSR web site. This had already lost detail by comparison with the original specification (September 2001) - for example, all geography below region, as well as detail on country of birth.

The disclosure control analysis has now been conducted and, based on this ONS have further reduced the detail available in the file. There is every possibility that more recodes will be added and that the 2001 Household SARs will have greater restrictions.

Reduction on detail on the Individual SAR (May 2004)

The table below gives the reduction in detail to the Individual SAR *additional* to that in the October 2003 specification. It is clear that the consequences of this loss of detail will be to remove much of the research value from the released file. Although disclosure work on the Household SAR has not yet been completed there is every reasons to suppose that it will be subject to similar, if not greater, reductions in detail.

Employment related variables

Work hours – reduced from a continuous variable with top coding at 80 hours to:

1 - 15
16 - 30
31 - 37
38 - 48
48 +

Industry – reduced from 60 Divisions to:

15 categories, A - O, P, Q

Occupation – reduced from 81 categories (SOC Minor) to:
SOC Sub Major - 25 groups

The international status classifications, ISEI and SIOP, will be dropped from the public use data set.

ISCO is reduced to the major categories, 0-9

Demographic variables

Age – reduced from single years with top-coding at 95 to:

0 - 15 single years
16 - 18
19 - 29
30 - 44
45 - 54
55 - 64
65 - 74
75 - 94 single years
95+

Country of Birth, reduced from 16 categories to:

England
Scotland
Northern Ireland
Republic of Ireland
Rest of Europe
Rest of World
N/A

Ethnicity, reduced from 16 categories to:

White
Mixed
Asian or Asian British
Black or Black British
Chinese or other ethnic group

This is only applicable to England, Wales and Scotland. For Northern Ireland there will be no change to the ethnicity coding of white/non-white.

Housing and location

Area-level classification

There is, as yet, no agreement on this

Tenure, reduced from 10 categories to:

1 Owns outright
2 Owns with a mortgage or loan
3 Pays part rent and part mortgage (shared ownership)
4 Rents from Council (Local Authority)/Scottish Homes/NI Housing Executive
5 Rents from Housing Association/Housing Co-operative/Charitable Trust/Registered Social Landlord)
6 Private rented or living rent free

Scotland

In the last newsletter we reported particular difficulties with the release of SARs for Scotland because published tables contain unique values, unlike output for England, Wales and Northern Ireland where small numbers have been adjusted to 0 or 3. For this reason GROS require special precautions before a safe SAR can be released for Scotland. This entails ensuring that there is no possibility of making a unique match between existing tables and the proposed SAR microdata and also ensuring that *there is no possibility of such a match with any future commissioned tables*. At the moment all the evidence suggests that SAR data for Scotland will have a level of detail that is the same as that for England and Wales.

The disclosure analysis of the Household SAR has not yet begun. The data will not be released until November 2004 and there are severe concerns over the research value of the final product. For example, ONS plan to suppress individual records for large households (over 9 household members) despite the demonstrated bias that this will cause in the sample.

Why has there been such a delay in the SARs?

The Individual SAR is timetabled for release in September 2004 – one year later than the date that ONS gave us in March 2003. Reasons for the delay are closely related to the reduction in detail in the publicly released file. They include:

- a much fuller assessment of disclosure risk than in 1991
- this work has been commissioned externally and is based on new methods that ONS are now validating
- the perturbation methods to be used on the records identified as risky have only just been developed by ONS. These will be Quality Assured by Statistics Netherlands before they are accepted
- work on the SARs started later than initially planned because of delays to the other census outputs
- the fact that other work, for example the ONS Modernisation agenda, has priority over the SARs – reflected in the number of staff employed on key areas such as Information Management.

Are there any alternative routes of access for researchers planning to use the SARs?

In the last Newsletter we made four recommendations to the Office for National Statistics, all of which received support from users. These were based on concerns over the timing of the release of the SARs and the recognition that the data to be released were unlikely to meet many research needs. These recommendations were:

- i. that ONS provides access to a more detailed version of the SARs in a safe-setting
- ii. that ONS implements a ‘research’ version of the SARs that can be used in safe settings in universities. This would represent an intermediate position, mid-way between safe data released for public use and very detailed data accessible only within the statistical office
- iii. that remote access to microdata in a safe setting is urgently explored
- iv. that ONS provide immediate in-house access to the Individual SAR (extracted in mid-November 2003) for researchers who have an urgent and pressing need.

Progress so far:

1. We held a round-table meeting with senior staff at ONS in November 2003, following which a SARs Project Board was set up. This now meets monthly and provides a forum where all parties concerned with the SARs track progress and identify problems.

2. On 8 January 2004, we submitted to the ONS Microdata Release Committee a request for immediate in-house access to the Individual SAR, before disclosure control, for four researchers with a pressing need (recommendation iv). This request was accepted by that Committee but the procedures to allow access for these researchers have not yet been implemented. They should, however, be in place by the end of May.

3. In February 2004 a meeting between academics and ONS staff responsible for disclosure control discussed academic research needs and ONS security requirements and tried to reach some resolution. This meeting was attended by Len Cook (the National Statistician) and Ian Diamond (Chief Executive, ESRC). One of the conclusions of the meeting was that we should recognise a continuum from safe data to a safe setting and that we should explore setting up a 'middle-way' that allowed relatively safe data to be used in universities where there were secure institutional controls to protect the data (recommendation ii).

4. Census division have prepared a paper making the case for recognising a range of modes of access, including in-house access to a more detailed version of the SARs and, in the longer term, the possibility of a 'middle-way' (see above). This paper has been accepted by the ONS Statistical Policy Committee.

5. As a result we are now able to work with ONS to establish the procedures needed for in-house access to the SARs. Initially this will be in Titchfield and the data file will be that specified in October 2003.

6. We will also be pressing to regain the geographical information and other detail that was lost at this point, including an area-level classification. The safe setting within ONS should enable access to significantly more detail than envisaged for released data.

7. We want to ensure that other statistical offices will be able to provide access to the SARs, for example, ONS in London, Newport and Southport and in Scotland (Edinburgh) and Northern Ireland (Belfast). For many researchers this form of access will be very difficult with significant costs. It is therefore important that there are as many locations as possible.

8. Finally, we will start to discuss the protocols and procedures that will be needed to implement the 'middle-way' – access to fairly safe data in a safe setting in a university (recommendation iv).

Although these steps are helpful they are going to take time to implement and, meanwhile, there are urgent requirements for access to data. The difficulties over the production of 2001 SARs have had serious consequences for much research. Some projects have been abandoned altogether and researchers are facing contractual difficulties. For projects with an immediate and pressing need we suggest making an application for in-house access, as set out below.

Applications to analyse the Individual SAR at Titchfield

Researchers who need immediate access to the Individual SAR should write a statement supplying the following information:

1. Who is the researcher and for what institution do they work?
2. What are the aims of the research and why, particularly, do they need to use 2001 SARs data for the work?
3. Do they need Scotland and Northern Ireland data in addition to England and Wales data?
4. What analyses do they intend to perform on the data?
5. What outputs do they expect to take away from the microdata laboratory?
6. Who is the intended audience for the work?
7. How long (days) do they expect to be working with the data?

Please email this information to: Gillian.meadows@man.ac.uk. We will then send these applications to ONS for submission to the Microdata Release Committee.

It would be helpful if these applications are received by 31 May.

2001 SARs: Balancing risks - a one-day meeting in September (Date and venue to be confirmed)

We are planning an open meeting at which we invite potential SAR users to:

- find out more about the released 2001 SARs, e.g. the disclosure control methods used, the methods used to perturb risky records and assessment of the resulting bias
- debate the balance of risks between ensuring safe data and supporting scientific research
- find out more about how safe settings will work and give us your views on the likely uptake of this mode of access, both at ONS and within safe settings at universities
- discuss the opportunities for remote access to a safe setting.

We hope that staff from ONS, NISRA and GRO(S) will be able to take part in this meeting, to provide information and explain future developments.

Dates are not yet available but information will be posted on the web and emailed to the SARs User Group.

SARs Contact Details

Web site: <http://www.ccsr.ac.uk/sars/>

Email: sars-helpdesk@man.ac.uk

SARs Helpline (0161) 275 4735

Fax: (0161) 275 4722

Professor Angela Dale
(0161) 275 4876
Angela.Dale@man.ac.uk