CAMS User Guide

CAMS User Guide

This note is intended to provide some general guidance for users of the CAMS files. Users should seek clarification from the CAMS team at ONS if they need further advice. (email: sars@ons.gov.uk).

1. What are the Controlled Access Microdata Samples

The Controlled Access Microdata Samples (CAMS) are more detailed versions of the Samples of Anonymised Records (SARs). Because of the increased risk of disclosure, users can only access the CAMS in a safe-setting, within a Census Office site. Outputs from analysis will be assessed for disclosure before being released from the Census Offices.

2. Are the CAMS for me?

In order to obtain access to the CAMS, you will need to demonstrate that the analysis you wish to do cannot be done with the Licensed SAR or another dataset. You should consult the codebooks for both the Licensed SAR and CAMS before making this choice.
Both codebooks can be accessed from the Cathie Marsh Centre for Census and Survey Research's (CCSR) SARs website at: CCSR SARs Website.

2.1 Ethnicity

Version 2 of the Individual SAR (Licensed) contains 16 categories for England and Wales and 14 for Scotland. Accordingly it is more useful for serious research than the first release of the Licensed SAR. If your interest in the CAMS is based on an interest on ethnicity or religion, you may find that the new Licensed SAR contains sufficient detail.

2.2 Geography

Local authority geography is available on the CAMS but not the Licensed SAR. However, you should not expect to be able to produce detailed analyses on an LA by LA basis as this may prove disclosive. The CAMS also contain some contextual information about the Super Output Area in which a respondent resides and/or has migrated from. An urban/rural indicator and Index of Multiple Deprivation (IMD) Scores are available for England and Wales. Deciles are available for England & Wales and Scotland. No IMD data is available for Northern Ireland.

2.3 Household/Family Context

The Individual CAM contains a range of derived variables produced for the Office of the Deputy Prime Minister. These include variables on the number of dependent children living with a step-father or grandparent, the age gap between mother and child, and whether the sibiling is in the household. Information about these variables is in the codebook.

3. Management of the CAMS

The CAMS are housed within a safe setting at National Statistics Institute (NSI) establishments. The data are managed by the NSIs.

The CAMS are currently accessible from four ONS sites; London, Titchfield, Newport and Southport. There are plans to extend access to Belfast and Edinburgh. For details of how to find the ONS offices, follow this link: How to find ONS sites.

Queries concerning the CAMS, particularly those relating to access arrangements should be directed to the ONS. The ONS web page for the CAMS data: ONS CAMS website Queries should be sent to sars@ons.gsi.gov.uk.

CCSR (Cathie Marsh Centre for Census and Survey Research) are a group at the University of Manchester funded to support UK academic use of the SARs. As part of this role CCSR requested the production, and specified the content, of the CAMS and have been involved in setting up the data. However they do not have ownership or management control of the data.

4. Applications process

The official application form required is available on the ONS CAMS website provided above and should be submitted to the ONS. The application requires an account of what the data are to be used for. Your application will be considered by the Census Research Access Board (CRAB) which meets approximately every 6 weeks.

5. Visiting the Census Office

Upon visiting the Census Office, report to the reception and indicate to reception staff the contact name you have been supplied with. Your contact will meet you at reception and show you to the secure environment. You will be required to sign the Census Confidentiality Agreement and an Undertaking agreement before you access the data. Your contact will assist you with logging into the VML and answer any general queries about using the system.

6. Using the CAMS

You will be analysing data using the ONS' Virtual Microdata Lab (VML). The VML behaves very similarly to a normal windows environment. However, you will be working on a remote server that restricts your ability to move files. You will be given a personal username and password to provide you with access to your space on the server.

Make sure you have a note of your user name and password. If you are away from the PC for more than 2-3 minutes you will be locked out and will have to re-enter the user names and password.

You will not be able to work with a laptop or other portable media.
You need to get any command (syntax/do) files placed on the server for you. These files must be sent to sars@ons.gsi.gov.uk at least two working days prior to your visit.

Any other documentation you wish to use needs to be in hard-copy as you will not be able to access it on your laptop. You should find a hard copy of the CAMS codebook available for your use at each census office.

Your user area in the VML

The 'T' (Census_Users) drive.

All users are allocated their own folder on the The 'T' (Census_Users) drive. This drive is where you should save any working files. If you have sent any syntax files to ONS in advance of your visit, you will find these located here. Note that you are not able to delete any files that you create.

Finding the data

The 'V' (Census_Source) drive. This drive is where the CAMS data are stored. File names are:

Individual CAM: "ukindcams050419"
Household CAM: (To be added).

Storing your outputs

The 'Y' (Census_Output) drive.
This drive is where you should store any outputs you wish to take away from the safe setting. There will be a folder in this drive for you to store all of your outputs. It is important to note that you cannot save directly to the 'Y' drive. You must save your work to the 'T' drive and then copy files from 'T' to 'Y'.

Users will not be able to access other drives, e.g. CD or floppy disc drive.

Support for users
The VML allows you to work directly with the microdata. You will be doing your own analysis and should not assume that the staff at the centre at which you are accessing the data will know your statistics package or methods. You should therefore go well-prepared, having tested your methods as much as possible with the licensed data.

In contrast to the LS, there are no resources at ONS to run procedures for you.

Early users have found that working with the data typically takes longer than anticipated. Be conservative estimating the time a job will take you.

If you have any questions about the data during your visit you should phone either Tom Howe (73 2610) or Michelle Clift-Matthews (73 3598). Note that these numbers are internal to ONS.

Software on the VML

You will have access to SPSS 10.1, Stata 8.0 and Gauss. SPSS is to be upgraded to v.12 shortly. MLwin has also been identified as a package that would be useful in the VML, and will be added shortly.

Additionally the VML gives access to MS Word and MS Excel as well as a text editor called SuperEdit.

The VML uses a Windows XP environment. There are known issues to do with the stability of SPSS 10.1 in this environment. The upgrade of SPSS is expected to mitigate these problems.

You will not have access to the internet or email facilities.

Size of data files

It is a condition of using the CAMS that you create a subset containing only the variables and cases (i.e. countries) that you have stated on your application form to use the CAMS. This can be achieved easily in either SPSS or Stata by using the keep and drop (sub) commands. In particular, you may wish to drop the imputation flag variables. Almost half of the variables in the file are imputation flags. They are easily identifiable as they start with the letter 'z'. Outputs will be checked to ensure that only information relating to the variables and countries stated on the application form have been analysed.

If you save your commands in a syntax (or 'do') file, this will allow you to make modifications should you need to. However you should take caution when dropping variables that you think you *might* need. Re-running syntax will be time-consuming and you will want to get it right first time if at all possible.

The Individual CAMS contains 1,845,530 cases and over 300 variables. This is a huge dataset. Although the environment in which you will be analysing the data is reasonably powerful, the sheer size of the data is likely to cause processing problems unless you work with a subset.

Performance has been observed to improve dramatically when a subset of the CAMS is used, rather than the whole file.

Using syntax

We recommend that you use a syntax file when working with the CAMS. Not only is it good practice to produce a record of your work, but it will also save you time when doing substantial pieces of work. By working with the licensed SAR to produce a syntax file ahead of your visit to ONS you will also be able to save thinking time while at the Census Offices.

Use the licensed SAR as much as is possible to generate your syntax commands.

Some initial lines of syntax for STATA:

```
set mem 800m
use "V:\Individual CAM\ukindcams050419.dta"
log using "T:\Angela Dale\dalemay16.log", replace
keep if <selected cases>
keep <variable names>
```

Or for SPSS users:

*Open Individual CAM and keep age, distance moved and ethnicity for England and Wales only.

GET FILE=' V:\Individual CAM\ukindcams050419.sav '.
/KEEP = age0x dm ethewx.

*Select England cases only.

SELECT IF (country =1).

SAVE OUTFILE = 'T:\Tom Howe\subset100605.sav'.

EXE.

7. What checks are done on outputs?

When you leave the VML, your outputs will be checked by Census Office staff before being forwarded to you. They are checked to ensure that they do not breach confidentiality rules. These rules are laid out on the ONS CAMS site. See:

Confidentiality Guidelines

In general, you can minimise the risk of failing to pass disclosure checks by:

- Using modelling approaches in preference to large complex tables
- Avoid producing complex analyses for small sub-populations (such as local authority districts).
- Only produce outputs that you need.

8. Preparing publications

You must obtain clearance for publications based on outputs from the CAMS. A draft of conference overheads or papers and reports should be sent to ONS. Comments will be returned within 5-10 working days.