

Data Management and Analysis Group

Review of the 2001 Census Small Area Microdata



October 2006

Review of the 2001 Census Small Area Microdata

This Briefing was produced under contract for CCSR SARs Team

For more information regarding content please contact:

Giorgio Finella or Rachel Leaser

Data Management and Analysis Group
Greater London Authority
City Hall (5 East)
The Queen's Walk
London SE1 2AA

Tel: 020 7983 4328

Email: giorgio.finella@london.gov.uk

Tel: 020 7983 4696

Email: Rachel.leaser@london.gov.uk

If you would like copies of other SARs related briefings contact

Tel: 0161 275 4735

Email: sars-helpdesk@manchester.ac.uk

Copyright © Cathie Marsh Centre for Census and Survey Research, 2006

Any Census data produced in this briefing has been reproduced with the permission of the Controller of Her Majesty's Stationery Office and the Queen's Printer for Scotland and is © Crown Copyright. An acknowledgement of the source is given at the end of each table.

Contents Page

1. Introduction to the Small Area Microdata	1
2. Summary Table	3
3. The 2001 Census Datasets	5
<i>Small Area Microdata (SAM)</i>	7
<i>Standard Tables</i>	9
<i>Census Area Statistics</i>	10
<i>Commissioned Tables</i>	11
<i>Sample of Anonymised Records Individual (Licenced) file</i>	13
<i>Sample of Anonymised Records Household (Licenced) file</i>	14
<i>Individual Controlled Access Microdata Sample (CAMS) file</i>	15
<i>Household Controlled Access Microdata Sample (CAMS) file</i>	16
<i>Longitudinal Study</i>	17
4. Non-Census datasets	19
<i>Annual Population Survey/Labour Force Survey</i>	20
<i>General Household Survey</i>	22
<i>Family Resources Survey/Households Below Average Income</i>	24
<i>Survey of English Housing</i>	26
<i>Housing Needs Surveys</i>	28
<i>Local Resident Surveys</i>	28
5. Uses for the SAM	29
6. Comparing figures from the SAM	35
7. Conclusions	45
Appendix 1 List of variables in SAM	47
Appendix 2 Accessing SAM	49
Appendix 3 Web links to further information	51

1. Introduction to the Small Area Microdata

Small Area Microdata (SAM) is the term used for an individual level sample of anonymised records from the 2001 Census with a geographical identifier at the local authority level. The SAM includes five per cent, or one in twenty, of all people in the UK Censuses, nearly three million people in all. The purpose of this guide is to provide a background and information on where and when local authority and other users can make best use of the SAM.

To understand the SAM, it is first necessary to know a little about the Census itself. The Census is a complete count¹ of residents in private households and in communal establishments, and has additional information about students living away from home. This is also available in the SAM.

The Census asks questions covering a broad range of topics, which are then processed to provide information on:

- place of residence
- demographics
- living arrangements
- country of birth
- ethnicity
- religion
- housing
- access to vehicles
- travel to work
- economic status
- employment
- educational attainment
- health
- caring
- migration

The SAM includes information from the full range of Census topics listed above, but not all variables within each topic. Some variables have reduced detail to prevent any potential disclosure of information leading to identification of an individual. For example, under the demographics topic, age is grouped into 13 bands, while under the employment category, industry and occupation are not included, but social classifications based on occupation are available. A range of derived variables and variables relating to the household, such as number of earners in the household are also provided. The full list of variables included in the SAM is given in Appendix 1.

One of the factors making the Census invaluable to local authorities and many other users is the detailed geography available for place of residence. The SAM combines some of this detail, giving the local authority of residence for each individual in the sample, with the individual attributes, allowing for flexibility in analysis and more complex statistical modelling, while the

¹ The actual response rate to the Census is believed to be around 94 per cent. Imputation of both missing people and missing data for people or households was carried out to make the Census a “complete” picture.

very large sample size makes this type of analysis robust for most purposes, even at this level of geography.

The SAM is available only through CCSR at the University of Manchester (see Appendix 2 for details) under licence arrangements allowing up to 10 users in an organisation (for local authorities), each of whom must sign a licence agreement.

The SAM is just one of the many ways in which Census data is available. This guide sets out information on the other Census datasets available and a selection of non-Census datasets readily available to local authority and other users, comparing them with the SAM in terms of uses and usability, strengths and limitations. These are summarised in the Summary Table on the next page.

More information on all the datasets can be found by following the web links in Appendix 3.

A local authority perspective on the uses and usability of the 2001 Census Small Area Microdata Summary Table

	lowest geography available	coverage	sample size	range of variables	detail of variables	ease of use	speed of access (inc. learning time)	cost	timeliness
2001 Census	CAS	Census Area Statistics	100%	core	☺	☺☺☺☺☺	☺☺☺	Free	2001
	ST	Standard Tables	100%	core	☺	☺☺☺☺☺	☺☺☺	Free	2001
	CT	Commissioned Tables	100%	core	☺☺☺	☺☺☺☺☺	☺☺☺	£150++	2001
	SAM	Small Area Microdata	5%	core	☺	☺☺☺☺☺	☺☺☺	£500	2001
	SARI	Individual Sample of Anonymised Records	3%	core	☺☺	☺☺☺☺☺	☺☺☺	£500	2001
	SARh	Household Sample of Anonymised Records	1%	core	☺☺☺	☺☺☺☺☺	☺☺☺	£500	2001
	CAMSi	Individual Controlled Access Microdata Sample	3%	core	☺☺☺	☺☺☺☺☺	☺☺☺	?	2001
	CAMSh	Household Controlled Access Microdata Sample	3%	core	☺☺☺	☺☺☺☺☺	☺☺☺	?	2001
	LS	Longitudinal Study	1%	+vital, health +labour	☺☺☺	☺☺☺☺☺	☺☺☺	?	2001
	LFS/APS	Labour Force Survey/ Annual Population Survey	1%	market - housing	☺☺☺☺☺	☺☺☺☺☺	☺☺☺	Free	3 months
Non-Census	GHS	General Household Survey	13,250 h/h	+ health, lifestyle	☺☺☺☺☺	☺☺☺☺☺	☺☺☺	Free	Annual
	FRS	Family Resources Survey/ Households Below Average Income	25,000 h/h	+ income + housing - labour	☺☺☺☺☺	☺☺☺☺☺	☺☺☺	Free	Annual
	SEH	Survey of English Housing	20,000 h/h	market	☺☺☺☺☺	☺☺☺☺☺	☺☺☺	Free	Annual
		Housing Needs Surveys		Housing- much of core				£+++	
		Local Resident Surveys		- much of core + attitude?				£+++	

Notes: Geography: OA = Output Area, ward = local authority ward, LA = Local Authority
 Variables: core variables are those included in the Census, covering household and family information, housing tenure and household accommodation, car availability, employment, education, health, demographic information about household members including migration

3. The 2001 Census Datasets

Coverage	UK residents and households
Type	full census
Size	59m people or 24m households in 2001
Time series	since 1801
Frequency	decennial
Data provider	ONS/GROS/NISRA

About the Census datasets

Datasets available from the 2001 Census range in detail, accessibility and format from the Key Statistics to complex tabular output to the Samples of Anonymised Records. The Key Statistics are designed mainly for the casual user, and comprise a small selection of summary variables giving, for example, numbers of people living in an area or percentage of owner occupiers. Tabular output is covered in more detail below. The Samples of Anonymised Records are individual or household level datasets, needing specialist skills and software for analysis, of which the SAM is one.

The majority of Census output is tabular, and these datasets, detailed later in this report are the Census Area Statistics (including Univariate and Theme Tables), the Standard Tables (also including Theme Tables) and specially commissioned tables. These datasets were designed to provide statistics for a range of geographies in England and Wales, Scotland and Northern Ireland and are available from a number of government websites including ONS Neighbourhood Statistics and Nomis. Information by single area can be downloaded into a spreadsheet; multiple areas prove a more challenging prospect. Data are also available from ONS by various electronic media. The Supertable software product is supplied by ONS to access the data. The academic community has its own software (CASWEB) to aid the manipulation of census data. Most local authorities and the health sector use SASPAC census software to manage and analyse the Census datasets. The advantage in using these Census datasets is that they provide a large amount of detail across a broad range of topics about people and households and, because the Census is (at least nominally) a complete count, the figures are robust.

The other Census datasets available are samples taken from the full Census database, each with its own characteristics. The advantage of taking a sample from the Census is that it reduces the non-response bias usually associated with sample surveys. These datasets are available as microdata or individual anonymised records. The SAM is one of these datasets. The others are:

- Individual Sample of Anonymised Records (SARi)
- Household Sample of Anonymised Records (SARh)
- Individual Controlled Access Microdata Sample (CAMSi)
- Household Controlled Access Microdata Sample (CAMSh)
- Longitudinal Study (LS)

Each of these is discussed in more detail later in this report. Specialist statistical software is needed to analyse these data. The SAM, SARi and SARh are available for use with software such as SPSS, STATA or NDSTAT at the user's workplace; whereas the CAMSi, CAMSh and LS are only accessible through a "safe setting" currently provided by ONS and output must be

checked to ensure that it does not breach confidentiality restrictions before it can be taken away.

The data

The topics covered by the Census are outlined in the introduction to this report, but the Census questions vary slightly between the countries of the UK² although the broad topics remain the same. For example, there is a question on Welsh language only in Wales and there are different ethnic group categories in Scotland and Northern Ireland to those in England and Wales. The number of questions and variables for output differ between the topics with a range of housing questions covering tenure, amenities, dwelling type, floor level and number of rooms, whereas there are just two questions relating to health; on general health and limiting long-term illness or disability. The amount of detail available also varies between the datasets, with greater detail for larger geographies, where the risk of identifying an individual is smaller. The degree of detail given for the smallest geographical areas is restricted. For example, only a limited number of individual countries of birth are identified for the smallest areas (called output areas, which average 125 households in England, Wales and Northern Ireland and 50 households in Scotland).

Censuses have been taken in Great Britain every ten years since 1801 (with the exception of 1941), but due to changes in society and technology, recent Censuses have been very different to those conducted earlier. The 2001 Census is not strictly comparable to earlier Censuses, even the 1991 Census, due to differences in the way the resident population is defined. However, many of the 2001 Census datasets are broadly similar to those produced from more recent Censuses, with comparable cross-tabulations produced for all areas. The release of Samples of Anonymised Records, allowing for analysis of individual records, was first undertaken in 1991, although the Longitudinal Study incorporates data back to 1971. This means that time series analysis of Census data is particularly challenging and must be carried out with caution.

Conclusions

The Census datasets are varied, but are all derived from the national Censuses which have the great benefit of giving very reliable statistics for the whole country. Further details of the SAM are given overleaf and each of the other Census datasets is then described and compared to the SAM.

² The Censuses for the various countries in the UK are the responsibility of different organisations. The Office for National Statistics (ONS) undertakes the Census in England and Wales, in Scotland it is the responsibility of the General Register Office for Scotland (GROS) and in Northern Ireland the responsibility of the Northern Ireland Statistics and Research Agency (NISRA).

Small Area Microdata (SAM)

Coverage	UK residents
Type	sample from census
Size	2.96m people (5 per cent sample)
Geography	Local Authority
Level of expertise	Experienced social researchers

About the dataset

The SAM is very large - a five per cent sample of individuals resident in each of the countries of the UK, with 2.96 million records in total, although great care has been taken to ensure that no information is included which would allow an individual to be identified. It is far larger than any other dataset from a single survey or year. It has 70 variables (listed in Appendix 1) across the range of Census topics, although not all variables are available, or the same, for all countries.

The variables are a combination of:

- direct individual variables (e.g. sex of respondent)
- derived individual variables (e.g. generation indicator for respondent)
- variables relating to the household or family (e.g. social grade of household reference person)
- direct housing variables (e.g. tenure of household)
- derived household variables (e.g. whether or not there are dependent children in the family)

However, the amount of detail on an individual record in the SAM is less than in the Individual SAR (see later in this report) because the greater geographical detail in the former means that to ensure confidentiality of the data certain variables are not included and others are grouped or restricted.

Local Authority District is the lowest level of geography for England and Wales, Council Areas for Scotland and Parliamentary Constituencies for Northern Ireland. The only exceptions to this are the City of London and the Isles of Scilly, which are merged with City of Westminster and Penwith respectively, and Orkney and Shetland, which are merged.

The SAM includes those enumerated in a communal establishment and also full-time students who were enumerated at an address that was not their usual term-time residence. The sample includes people in wholly imputed households, people in a household where another person(s) is imputed, people who were imputed in a household that was counted as well as the vast majority, of people in wholly counted households. An alternate version of the SAM is available which includes flags for imputed data, although for most purposes, the "normal" SAM is recommended. Imputed items are not flagged in Scotland and Northern Ireland.

In terms of the practicalities of obtaining and using the dataset, there are three issues to consider:

- completing the licensing process, which is relatively straightforward,
- the cost, which is £500 to public sector users (see Appendix 2 for more detail), and
- the software. Free software called NDSTAT is available with the data, which allows users to access and analyse the data, but will be unfamiliar to most users. Commercial

software such as SPSS or STATA may be more familiar, but there is a further cost if these are not already available to the user.

The 2001 SAM is almost directly comparable with the 1991 SAR allowing comparison of microdata at Local Authority level over a ten year period. Users should be made aware of possible changes to definitions may have occurred between Censuses

Key strengths of the SAM

- Local Authority District level geography
- Versatility, allowing the user to define their own cross tabulations and explore complex relationships in more detail than predefined tables can offer
- The five per cent sample is significantly larger than other forms of survey data, therefore providing more robust estimates
- Data on a broad range of topics
- It is quicker than commissioning tables from the Census
- The SAM can be used as a precursor to conducting more detailed investigations using, for example, commissioned tables, or Controlled Access Microdata Sample data
- It is cheaper than carrying out a bespoke survey

Limitations of SAM

- No information on Industry or Occupation (other than the derived variables based on occupation - NS-SEC and social grade)
- Fewer variables than most other Census datasets and most other surveys
- Less detail than most other surveys, but covers a broader range of topics
- Less detailed variables than some other Census datasets

Conclusions

This dataset is useful for exploring more detailed information not available in predefined Census tables. With this dataset the user will be able to create queries relatively quickly to respond to changing information requirements where more complete data would take significantly longer. It is likely that Local Authorities would use this dataset to gain more detailed information for a variety of groups resident in their area; cross-tabulated with other variables not currently available as either a pre-defined table or a commissioned table.

In addition, it would be useful for exploring the relationships between a range of variables using multivariate analysis or for doing modelling work. An example of this might be in transport modelling, looking at car availability, economic position, distance and method of travel to work, age, sex and whether there are dependent children in the household.

Standard Tables

Geography Ward level to national
Level of expertise School children to experienced social researchers

About the Standard Tables

The Standard Tables consist of nearly 300 tables³ (including over 30 Theme Tables) providing very detailed information on a large number of subjects. The tables were designed to provide comparable data for different geographical areas, down to ward level (and Postcode Sectors in Scotland). The main tables were designed to give a two-way cross-tabulation of Census variables, but with over 100 variables available for cross-tabulation, it would be impossible to include all potential tables. In addition, there are Theme Tables, providing a range of summary information for a selected population group e.g. dependent children in 10 age groups

This large selection of quite detailed tables were assembled to cover the most common or widespread needs, and in most cases this means that researchers have good quality information for comparable areas. However, there are issues with producing consistent tables across a range of geographical areas in that some characteristics are not widespread. For example, certain population groups tend to be concentrated in some areas more than others. This is particularly true of people from minority ethnic groups who are more numerous in urban areas. Because the Standard Tables are consistent across all areas, the information available is restricted to preserve confidentiality. This means that the amount of detail included on a secondary variable, such as age, is compromised.

This dataset is available to the public from ONS or the Nomis Website, or to subscribers from the SASPAC helpdesk.

Key strengths of the SAM compared to Standard Tables

- The SAM is more versatile, allowing the user to define their own cross-tabulations and explore complex relationships in more detail than predefined tables can offer
- The SAM can be used with other statistical techniques such as multivariate analysis to understand the relationships between characteristics at an individual or household level.

Limitations of the SAM compared to Standard Tables

- The SAM has fewer variables
- The SAM has less detailed variables
- The SAM has less detailed geography; Standard Tables are available down to ward level.
- Standard Tables are available at no cost to the user
- The Standard Tables may be supplied as Excel spreadsheets or other formats as relevant, whereas the SAM needs more specialist software.

Conclusions

This dataset is probably the most widely used within local authorities, allowing analysis of the detailed information down to a ward level. Many investigations will be satisfied by this dataset, but there are likely to be occasions when the information required is not included in the Standard Tables.

³ Not all tables are available for all countries of the UK.

Census Area Statistics

Geography Output area level to national
Level of expertise School children to experienced social researchers

About the Census Area Statistics

The Census Areas Statistics (CAS) are a set of cross-tabulations produced from the Census with the great benefit of being available down to Output Area⁴ level geography. The tables follow a similar format to the Standard Tables but generally with less detail in each table to preserve confidentiality of the very small populations of Output Areas. Not all tables in the Standard Tables have an equivalent in the CAS, but there are still around 90 tables available for each of the 223,000 Output Areas. Additionally there are 61 Univariate Tables providing more detailed information on a single variable down to Output Area level (and a further two down to Local Authority level only).

This dataset is available to the public from ONS or the Nomis Website, or to subscribers from the SASPAC helpdesk.

Key strengths of the SAM compared to Census Area Statistics

- The SAM is more versatile, allowing the user to define their own cross-tabulations and allows the user to explore complex relations in more detail than predefined tables can offer.
- The SAM offers more detailed cross-tabulation of religion or ethnicity

Limitations of SAM compared to Census Area Statistics (CAS)

- The CAS are publicly available on the internet, whereas the SAM is only available under strict licensing arrangements
- The SAM has fewer variables
- The SAM has less detailed variables
- The SAM has less detailed geography; the CAS goes down to output areas.
- The CAS may be supplied as Excel spreadsheets or other formats as relevant, whereas the SAM needs more specialist software.

Conclusions

This dataset is useful for exploring most information requests down to a very small geography. The neighbourhood statistics uses this dataset in creating area profiles. Because of the very detailed geography, it is possible to aggregate this dataset to virtually any area needed, including areas that may cross local authority boundaries. Nevertheless, it will not always provide the detail needed for a particular analysis.

⁴ An Output Area is an area defined for Census purposes to have an average of 125 households in England, Wales and Northern Ireland and 50 households in Scotland

Commissioned Tables

Coverage	England and Wales residents and households ⁵
Geography	Output area level to national as requested by the user commissioning the table, but subject to confidentiality restrictions.
Data provider	ONS
Level of expertise	Experienced social researchers

About Commissioned Tables

Commissioned Tables are designed to provide data where requirements are not met by standard results in reports or local statistics. Once a user has identified an information gap, they can define and request a new table for a specified area or areas. These are often highly complex, and are only released subject to confidentiality criteria.

The user defines the table and the geographies for which it is required and pays for the cost of production. The tables can be expensive, with costs from £150 upwards for a simple table. More complex tables requiring specialist programming resource are more expensive, a straightforward derived variable can be expected to cost in the region of £1,000. Once a table has been produced, it is released at the discretion of the Disclosure Control Unit at ONS. It is usually possible to commission the same table for other areas, but not all commissioned tables are available for all countries in the United Kingdom. There is also a time factor in producing these tables. It can take a long time between specifying a table and the table delivery, the more complex the table, the longer it is likely to take.

Analysis of these tables, particularly comparisons between areas, is not easy in the format in which they are routinely produced.

Once commissioned tables have been delivered to the commissioning user, they are also made available to other users at no charge. In most cases files are available in Excel format, by e-mail, on CD or DVD as appropriate. By June 2006 more than 1000 tables had been commissioned. Most local authorities and health sector users use SASPAC census software to help them manage these tables. More than 300 of the commissioned tables have been converted into SASPAC format.

Key strengths of SAM compared to Commissioned Tables

- The SAM is more versatile, the user can define their own cross tabulations which allows the exploration of complex questions in more detail than pre-existing commissioned tables.
- The SAM is available to use immediately (once purchased), commissioned tables are constantly being released and therefore requires the user to continually update.
- The SAM allows the user an opportunity to see if the figures would merit further investigation before commissioning tables.
- With the SAM the user can easily make changes to a query, whereas with the commissioned table once the requested table has been delivered no other changes are possible without ordering another table, and there may be restrictions on what can be produced without breaching the confidentiality restrictions.

⁵ Tables relating to Scotland and Northern Ireland are dealt with in a similar manner, for the purposes of this report we will be concentrating on the tables commissioned for England and Wales.

- The SAM can be used with other statistical techniques such as multivariate analysis to understand the relationships between characteristics at an individual or household level.
- Once purchased there are no additional costs in using the SAM, the commissioning of tables is relatively expensive and there is a time lag between ordering and receipt.

Limitations of the SAM compared to Commissioned Tables

- The SAM has fewer variables
- The SAM has less detailed variables
- Existing commissioned tables are freely available on request from ONS
- Each table is commissioned for a particular request; geographies and variables vary according to the user definition.
- The commissioned tables dataset is very good at answering specific requests for information.
- Commissioned Tables are usually supplied as Excel spreadsheets, whereas the SAM needs more specialist software.

Conclusions

This dataset is useful for exploring research questions posed and planned in advance, allowing investigations of information down to a very small geography if required, but subject to disclosure control thresholds. Time constraints on supplying information may make this dataset unrealistic in many cases.

Sample of Anonymised Records Individual (Licenced) file

Coverage	UK residents
Type	sample from census
Size	2m residents
Geography	Government Office Region
Level of expertise	Experienced social researchers

About the Individual SAR

The Individual SAR (Licenced) is a three per cent sample of individual records from the full 2001 Census database, which translates to 1.84 million records. The data have been completely anonymised so that no individuals from the Census can be identified, and other measures have been taken where necessary to preserve confidentiality, such as grouping occupations. Residents in communal establishments are included as well as those in private households.

The Individual SAR contains 87 variables, including individual, housing and summary household variables, such as number of people aged 65 or over. Again, not all variables are available for all countries of the UK. Additionally, a number of variables derived from combinations of one or more other variables are provided to users. A second version of the file is also provided, containing flags for imputed data.

As with the SAM, the major strength of the individual level records is the flexibility this gives the user in terms of producing their own tables or exploring complex relationships between variables or other types of analysis, compared to pre-defined cross-tabulations.

It is possible to combine the data on the Individual SAR with data from the SAM to produce an eight per cent sample with reduced detail on some variables at Government Office Region level.

Licensing, cost, software and access arrangements for the individual SAR are the same as those for the SAM.

Key strengths of the SAM compared to 2001 Individual SAR (Licenced)

- The SAM is a larger sample (five per cent compared with three per cent), therefore providing more reliable estimates.
- The SAM goes down to Local Authority District level whereas the Individual SAR (Licenced) has a more restricted geography, only government office regions

Limitations of the SAM 2001 compared to Individual SAR (Licenced)

- The SAM has fewer variables.
- The SAM has less detailed variables.

Conclusions

This dataset is useful for exploring research questions at a regional level. As with the SAM queries can be created relatively quickly to investigate if further analysis would be of benefit. Often used as a precursor to using either commissioning data or using the Individual Controlled Access Microdata Sample (CAMS). In many cases this is of little use to local authorities or health sector users due to the lack of low level geography.

Sample of Anonymised Records Household (Licenced) file

Coverage	England and Wales households
Type	sample from census
Size	225,000 households containing 526,000 individual records
Geography	England and Wales only
Level of expertise	Experienced social researchers

About the Household SAR

A one per cent Special Licence Household SAR has been released for England and Wales only from the 2001 Census. It includes 525,715 individuals living in 225,436 households. The dataset is hierarchical, that is, it contains information about households and all the individuals within those households⁶ and any family relationships between those individuals.

All the information included in the SAM and the Individual SAR is available in the Household SAR, although many of the household summary variables are not given, as the information is already on the file. The level of detail on the variables is comparable to that in the Individual SAR, other than the geographical detail.

The major value of this dataset is in looking at variables in the family or household context, such as investigating economic position of parents in relation to age and number of children or issues around ethnicity in a household context. However, the lack of geographical information makes this of limited value to researchers in local authorities.

Access is under an ONS Special Licence from the UK Data Archive. Users must agree to a higher standard of data stewardship to protect the confidentiality of the data.

Key strengths of SAM compared to 2001 Household SAR (Licenced)

- The SAM is a five per cent sample versus a one per cent for the Household SAR
- The SAM has a more detailed geography. The Household SAR has no geographical identifiers

Limitations of the SAM compared to the 2001 Household SAR (Licenced)

- The Household SAR (Licenced) has more detail as the name suggests regarding the inter-relationships between household members
- The SAM has fewer variables
- The SAM has less detailed variables

Conclusions

This dataset is useful for exploring research questions regarding households. As with the SAM queries can be created relatively quickly to investigate if further analysis would be of benefit. It is most likely to be useful to local authorities as a precursor to commissioning tables for more detailed geographies but in many cases this is of little use to local authorities or health sectors due to the lack of geographical detail.

⁶ There is no information provided for any individuals in households larger than 11.

Individual Controlled Access Microdata Sample (CAMS) file

Coverage	UK residents
Type	sample from census
Size	2m residents
Geography	Local Authorities
Level of expertise	Experienced social researchers

About the Individual CAMS

The Individual Controlled Access Microdata Sample (Individual CAMS) provides a more detailed version of the Individual Licensed SAR and can be accessed only in a safe setting at ONS offices. The Individual CAMS contains the same records as the Individual SAR, but with an additional 90 variables. The main differences to the Individual SAR are more detailed:

- Geography, down to Local Authority
- Age
- Country of birth
- Status in communal establishment
- Count of dependent children
- Economic activity of family reference person
- Full occupational detail (unit SOC)
- Industry
- 2004 Index of Multiple Deprivation information

The file is managed and retained by ONS. To access the Individual CAMS it is necessary to complete a complex application form with a well-defined project and tables you will be attempting to produce. If access is granted, it is necessary to go to the ONS data “laboratory” to perform the analysis. Any output will need to be checked by ONS for confidentiality reasons and forwarded to the user at a later date. Any errors or omissions will require another visit. Changes to the research may require a re-submission of the research proposal. This process is likely to take some time (4-8 weeks to process the application alone), as well as the cost of undertaking the research and the expenses involved in visiting the data laboratory.

Key strengths of the SAM compared to Individual CAMS

- The SAM is a five per cent sample versus a three per cent sample for the CAMS
- The relatively simple licensing arrangements compared to the CAMS
- Once obtained, access to the SAM is quick
- Analysis of the SAM can be done at the user’s pace in the user’s workplace and making changes to the research as it proceeds.

Limitations of the SAM 2001 compared to Individual CAMS

- The SAM has fewer variables.
- The SAM has less detailed variables.

Conclusions

This dataset is potentially useful for exploring planned research questions. Any use of this is likely to be following initial research using the SAM, which may identify where greater detail is required, but the time and procedural requirements to access the data may mean that the Individual CAMS is of limited use to local authorities or the health sector.

Household Controlled Access Microdata Sample (CAMS) file

Coverage	UK households
Type	sample from census
Size	256,000 households containing 594,000 individual records
Geography	Local Authorities
Level of expertise	Experienced social researchers

About the Household CAMS

The Household Controlled Access Microdata Sample (Household CAMS) is a one per cent sample of households from the 2001 Census, providing a large amount of detail on all variables and includes data for Scotland and Northern Ireland. As with the Household SAR, the file includes relationship information for people within families and households. The main differences are:

- more detailed geography, down to Local Authority level, also includes Council Areas in Scotland and Parliamentary Constituencies in Northern Ireland
- Full country of birth
- Information is included for all individuals in all households, regardless of household size

Access arrangements are the same as for the Individual CAMS, with strict controls over what data can be extracted, and only in the safe setting of ONS offices. Time and expenses can also be an issue.

Key strengths of SAM compared to Household CAMS

- SAM is a five per cent sample versus a one per cent for the household CAMS.
- The relatively simple licensing arrangements compared to the CAMS
- Once obtained, access to the SAM is quick
- Analysis of the SAM can be done at the user's pace in the user's workplace and making changes to the research as it proceeds.

Limitations of the SAM compared to the Household CAMS

- The Household CAMS has more detail as the name suggests regarding the inter-relationships between household members.
- The SAM has fewer variables
- The SAM has less detailed variables

Conclusions

This dataset is potentially useful for exploring planned research questions. It is the only source to give this detailed household information combined with local authority level geography, other than using tabular form. However, due to small numbers, the level of detail is likely to lead to issues of confidentiality and restrictions on what can be taken away from ONS offices. This means that a large amount of effort and time may provide little in the way of results, although where the numbers prove to be too small, the Household CAMS might still be useful in helping with requesting commissioned tables from the full Census database.

Longitudinal Study

Coverage	people in England and Wales
Type	sample from censuses linked to administrative records
Size	540,000 individuals in 2001 or over 1m in total
Time series	since 1971
Frequency	decennial
Data provider	ONS
Geography	Ward level
Level of expertise	Experienced social researchers

About the Longitudinal Study

Longitudinal Studies deal with change over time among individuals or groups. The ONS Longitudinal Study (LS) contains linked census data from 1971, 1981, 1991 and 2001 and vital event data (births, deaths and cancer registrations) for around one per cent of the population of England and Wales. It also includes snapshot Census data on individuals living in the same households as an LS "member". At each census, data on slightly more than 500,000 sample members are included (540,082 in 2001). During the 30 years of the study, around 1 million people have been recorded in the sample at some point.

Studies that make the fullest use of LS data are those that link social, occupational and demographic information at successive censuses to data on vital events. Examples include studies of mortality, cancer incidence and survival, and fertility patterns, including fertility by ethnic group. The LS has also been useful in looking at changing ethnic identity and how this has been captured in the different Censuses.

Access to the Longitudinal Study requires completion of a detailed application form, similar to the process for accessing the Individual and Household CAMS, but normally any direct access to the LS database is restricted to ONS staff and a specific academic support group. Analyses need to be specified and these will then be run by the LS team and results given to the user after checking for confidentiality thresholds. Obtaining results from the LS is likely to take some time.

Key strengths of SAM compared to Longitudinal Study

- The SAM is a five per cent sample versus a one per cent for the Longitudinal Study.
- Once obtained, access to the SAM is quick
- Analysis of the SAM can be done at the user's pace in the user's workplace, make changes to the research as it proceeds.

Limitations of the SAM compared to the Longitudinal Study

- The Longitudinal Study has linked data going back to 1971.
- The Longitudinal Study also has information on co-residents.
- The SAM has fewer variables.
- The SAM has less detailed variables.
- The Longitudinal Study requires the filling out of a complex form, with a well defined project and tables to be produced

Conclusions

Linked census data for members of the Longitudinal Study allow researchers to examine change between censuses by investigating the same people at two or more censuses. The kind of studies possible are detailed and unlikely to be used on a day-to-day basis by local authorities, but could be used for studies of particular topics relevant to the local authority, examples of which are given below.

- Studies of social mobility have examined changing class position by age. Information on co-residents of Longitudinal Study sample members has been used to study inter-generational mobility. It is now possible to compare class mobility from 1971 to 1991 with that from 1981 to 2001.
- The size of the Longitudinal Study makes it suitable for the study of ageing. Studies have used the information collected on the co-residents and family status of Longitudinal Study sample members to examine changes to household and family arrangements that come with age.
- Census forms ask about addresses one year ago. The linked census data in the Longitudinal Study have been used to study ten-year migration patterns between censuses.
- The addition of 2001 Census information to the Longitudinal Study has meant that individual-level ethnic identity changes and changes in limiting, long-term illness status between 1991 and 2001 can be studied for the first time.

4. Non-Census datasets

There is a range of other datasets available, which local authority users might consider for obtaining information or undertaking analysis. Some of the largest and most relevant are considered here. They can be divided into two groups – those produced by government departments (e.g. ONS or the Department for Work and Pensions (DWP)) and bespoke surveys commissioned or carried out by the local authority.

All are individual or household level datasets and need specialist statistical software for analysis, although local authorities may commission academics or others to undertake the analysis on their behalf. As before, a brief description of each dataset is given followed by a comparison in terms of the strengths and limitations of each compared to the SAM.

These datasets vary in size and topics. All are sample surveys (although a local authority may from time to time carry out a complete survey, but even these do not achieve the level of response of a national Census). A technical issue arising from this is that of weighting. All these datasets come with weights to be used for grossing up the data to estimate the complete population, making allowances to compensate for sample and response biases. This is not the case with the SAM. As it is a five per cent sample from the Census, weights are not needed other than the simple multiplication by 20 to get back to an approximation of the full population.

Annual Population Survey/Labour Force Survey

Coverage	UK households
Type	sample survey of households
Size	APS is 235,000 households/500,000 individuals
Time series	since 1973
Frequency	every 3 – 6 months depending on the dataset.
Data provider	ONS
Geography	Local Authority ⁷
Level of expertise	Some knowledge of statistical software, very complex dataset

About the APS and LFS

The Labour Force Survey (LFS) is the largest regular household survey carried out by the ONS. The LFS has been carried out in various guises since 1973 and since 1992 has been run on a quarterly basis. The survey questionnaire is large and collects a wide range of data about people and their labour market circumstances.

A number of datasets are produced from this survey:

- The quarterly LFS (This is the core component of all LFS derived datasets.)
- The household LFS
- The Annual Population Survey (APS)
- Various longitudinal datasets (up to five quarters)

The APS is an annual version of the quarterly LFS dataset, which provides a far larger sample to facilitate local authority level analysis. It was introduced in 2004 and replaced the old annual local area Labour Force Survey (LFS) dataset. The APS comprises three key elements:

- Data from the core Labour Force Survey quarterly surveys
- Annual boosts to the LFS sample for England, Wales and Scotland
- APS boost: An additional boost, for a core set of topics. These topics are a sub-set of the existing range of LFS questions. The boost is temporary and is being dropped in 2006 though the dataset will continue to be called the APS.

APS datasets are produced every three months with data relating to the previous 12 month period. Quarterly LFS datasets still exist alongside the APS dataset but the APS is the preferred source of local authority level data due to its larger sample size and better reliability.

The LFS and the datasets produced from it are continually evolving, although there are core elements that do not change. This means that the availability of consistent time series data depends on the nature of analysis required.

⁷ ONS LFS data service sometimes release smaller area data on request but this has very high confidence intervals attached and is not widely used.

The data

Topic coverage on the APS/LFS is vast and includes:

- Household, family and individual characteristics
- Demographics
- Employment and self-employment details
- Unemployment and jobsearch activities
- Occupation and Industry sector
- Qualifications and training
- Educational participation
- Health problems and disabilities
- Earnings & benefits

Special weights are applied to the data depending on whether the topic is a core APS topic or a non-core LFS topic (i.e. not included in the APS boost interviews).

Key strengths of SAM compared to APS/LFS

- The SAM is larger than the APS, so more robust. APS estimates for local authorities can have large confidence intervals attached making interpretation and application of the data difficult.
- The SAM includes individuals in communal establishments (and information for students away from home) while LFS coverage of these groups is partial.
- The SAM includes a broader range of information on housing topics, (although fewer variables on employment topics and less detail).
- Any non-response bias in the SAM is likely to be much less than in the APS/LFS.

Limitations of the SAM compared to APS/LFS

- The number of variables on the APS/LFS is far greater than in the SAM (APS dataset has 568 variables compared with around 74 on the SAM)
- APS datasets at local authority level are available only under special licence. This provides maximum detail under each variable (e.g. single year of age, individual country of birth). This level of detail is not available on the SAM, which therefore limits the analytical potential of the dataset massively.
- The SAM is only available for 2001, whereas APS/LFS local authority level data is available every three months.
- APS/LFS estimates, which are downloaded from Nomis⁸ for local authorities, are accompanied by confidence intervals to give users a clear view about statistical reliability. These are not currently available from the SAM.

Conclusions

The APS is a widely used source of inter-censal data on labour market and demographic topics. Its wide topic coverage, timeliness and flexible datasets offer excellent detail and good analytical possibilities. However, estimates for local authorities are subject to large confidence intervals and the data fail to provide robust data about sub-groups within local authority populations. The SAM is in many ways complementary as its larger sample offers more robust estimates, albeit for a much smaller number of variables.

⁸ Nomis is the official website for ONS labour market statistics (www.nomisweb.co.uk)

General Household Survey

Coverage	GB households
Type	sample survey of households available as household microdata
Size	19,000 individuals in 8,000 households in 2001/2
Time series	since 1971 (except 1997/8 and 1999/2000)
Frequency	annual
Data provider	ONS
Geography	region
Level of expertise	Some knowledge of statistical software, complex dataset

About the survey

The General Household Survey (GHS) is a multi-purpose continuous survey collecting information on a range of topics from people living in private households in Great Britain. The GHS has documented the major changes in households, families and people that have occurred over the last 30 years. One of the particular strengths of the General Household Survey is the detail on health and lifestyle issues combined with the broader range of topics similar to those in the Census.

Changes to the GHS are in place, with a move to a longitudinal design and a greater range of questions on social exclusion topics.

The data

The main aim of the survey is to collect data on a range of core topics, comprising:

- household and family information
- housing tenure and household accommodation
- consumer durables including vehicle ownership
- employment
- education
- health and use of health services
- smoking and drinking
- family information including marriage, cohabitation and fertility
- income
- demographic information about household members including migration.

The information is used by government departments and other organisations for planning, policy and monitoring purposes, and to present a picture of households, families and people in Great Britain.

Additional data is collected from time to time on topics such as:

- drug use in the last 12 months
- sport and leisure
- experience of crime
- language
- hearing
- contraception

Key strengths of SAM compared to GHS

- The SAM is a much larger sample than the GHS, so more robust.

- The level of geography in the SAM is more detailed, giving local authorities compared to only regions available from the GHS
- The SAM includes individuals in communal establishments (and information for students away from home)
- Any non-response bias in the SAM is likely to be much less than in the GHS
- Missing data has been imputed on SAM, but can be identified (and therefore taken out if required)

Limitations of the SAM compared to GHS

- The range of variables in the GHS is wider than in the SAM, and the detail available is greater.
- The SAM is only available for 2001, whereas the GHS is conducted annually, so more up-to-date information is available.
- Time-series analysis is available from the GHS
- The full household information, including relationships is available for the GHS, whereas only summary household variables are included in the SAM.

Conclusions

The GHS is useful for looking at change over time and household relationships and particularly for health-related topics, but the restricted geography and small sample size mean that it cannot be used in research on a particular local authority. The SAM has all the flexibility of microdata and more detailed geography, but fewer variables and less detail on some of them.

Family Resources Survey/Households Below Average Income

Coverage	GB households (extended to UK from 2002/3)
Type	sample survey of households available as household microdata
Size	44,700 adults in 25,300 households in 2001/2
Time series	since 1993/4
Frequency	annual
Data provider	Department for Work and Pensions
Geography	UK Regions
Level of expertise	Some knowledge of statistical software, very complex dataset

About the survey

The Family Resources Survey (FRS) is voluntary and covers people in households, but not those in communal establishments. It uses a sample of addresses stratified by geographic area and Census figures on socio-economic group, economic activity and male unemployment rates. Computer assisted personal interviewing helps to achieve a response rate of 66 per cent fully co-operating households (i.e. those where information is collected for all adults in the household). The fieldwork for the FRS is currently contracted out to the Office for National Statistics (ONS) Social Survey Division and the National Centre for Social Research.

The data

The main purpose of the FRS, as the name suggests, is to collect very detailed information on

- incomes
- sources of income (including details of state benefits)

of individuals, families and households, along with other characteristics such as (This bit looks messy and it is not totally clear)

- tenure,
- caring needs and responsibility,
- housing costs,
- employment,
- savings,
- age,
- sex,
- ethnic group
- relationships of all people in the household.

Additional data is collected from time to time on topics such as:

- NHS treatment
- travel to work
- household durables
- vehicle ownership

Families (and households) are the primary focus of this survey and therefore a critical element for analysis. The data are therefore essentially hierarchical, i.e. individuals make up families or family units, which in turn make up households, although in the case of one-person households, these would clearly all be the same. This has an impact on the complexity of the dataset.

The datasets available from this survey are the FRS itself and another dataset called Households Below Average Income. This is a more straightforward dataset (but with a variety of weights) that is particularly useful in looking at household income distributions and characteristics of people at different levels of income.

Key strengths of SAM compared to FRS

- The SAM is much larger than the FRS, so more robust.
- The level of geography in the SAM is more detailed, giving local authorities compared to only regions available from the FRS
- The SAM includes individuals in communal establishments (and information for students away from home)
- Any non-response bias in the SAM is likely to be much less than in the FRS
- Missing data has been imputed on SAM, but can be identified (and therefore taken out if required)
- The SAM is easy to use, whereas the FRS is generally supplied as a series of separate files, so is difficult to use. The reason for this is that a record represents a different unit on each file, so that on one file, the unit is a household, on another the unit is a job, on another the unit is the individual (adult or child), on another the unit might be the pension or the investment account etc. Thus there may be several records relating to the same individual on one file, and none for other individuals in the survey.

Limitations of the SAM compared to FRS

- The range of variables in the FRS is greater than in the SAM, and there is more detail available.
- The SAM has no information on income or sources of income.
- The SAM is only available for 2001, whereas the FRS is conducted annually, so more up-to-date information is available.
- Time-series analysis is available from the FRS

Conclusion

The FRS is useful for looking at change over time and particularly for income-related topics, but the restricted geography and small sample size mean that it cannot be used in research on a particular local authority. The SAM has all the flexibility of microdata and more detailed geography, but fewer variables (and nothing at all on income) and less detail on some of them. It is considerably easier to get started using the SAM than the FRS.

Survey of English Housing

Coverage	England households
Type	sample survey of households
Size	20,000 households
Time series	since 1993/4
Frequency	annual
Data provider	Department for Communities and Local Government
Geography	English Regions
Level of expertise	Some knowledge of statistical software, complex dataset

About the survey

The Survey of English Housing (SEH) is a continuous household interview survey with a sample of 20,000 responding households each year. It is a multi-purpose housing survey, which provides a comprehensive range of basic information on households and their housing, and full information on the private rented sector. From autumn 1994 questions on housing attitudes have been included, mostly on a rotating basis. The survey covers only England and data are available for the former Standard Statistical and the current Government Office Regions, but the dataset also includes rural/urban indicators and Acorn codes.

The data

Data are collected on

- type of accommodation
- household and personal characteristics
- tenure
- second homes
- moves
- repossessions
- satisfaction with the accommodation and area
- waiting lists for council or housing association housing
- owner occupation
- social sector tenants
- private renters.

Key strengths of SAM compared to SEH

- The SAM is much larger than the SEH, so more robust.
- The level of geography in the SAM is more detailed, giving local authorities compared to only regions available from the SEH
- The SAM includes individuals in communal establishments (and information for students away from home)
- The SAM includes a broader cross-section of topics
- Any non-response bias in the SAM is likely to be much less than in the SEH
- Missing data has been imputed on SAM, but can be identified (and therefore taken out if required)

Limitations of the SAM compared to SEH

- The range and detail of variables in the SEH is greater than in the SAM.
- The SAM is only available for 2001, whereas the SEH is conducted annually, so more up-to-date information is available.
- The SAM does not include any attitudinal questions.
- Time-series analysis is available from the SEH

Conclusions

The SEH is useful for looking at housing and migration issues and for change over time, but the restricted geography and small sample size mean that it cannot be used in research on a particular local authority. The SAM has all the flexibility of microdata and more detailed geography, but fewer variables and less detail on some of them.

Housing Needs Surveys

Coverage	local
Type	sample survey of households
Size	variable
Time series	variable
Frequency	ad hoc
Data provider	local
Geography	local
Level of expertise	Some knowledge of statistical software

About the surveys

Housing Needs Surveys are carried out from time to time by some local authorities. Their content, size and detail vary according to local priorities and needs. They are clearly an alternative source of information for some local authorities, but will not include the comparability or coverage offered by the SAM. It is not possible to compare them in any detail because of their variability.

Local Resident Surveys

Coverage	local
Type	sample survey or, more rarely, complete survey
Size	variable
Time series	variable
Frequency	ad hoc
Data provider	local
Geography	local
Level of expertise	Some knowledge of statistical software

About the survey

Local Resident Surveys are extremely variable and cover everything from Electoral Roll Enhancement Surveys, usually designed to collect minimal information but with a very large response, to small sample surveys or panel surveys designed to collect information on attitudes and local issues. As with the Housing Needs Surveys, it is not possible to compare them in any detail with the SAM, but they are a valuable source of information to those local authorities that carry them out.

5. Uses for the SAM

The SAM will have many uses to local authority researchers and others in the public sector. Given below are some of the ways in which the SAM will be valuable. Some of these are based on uses made of the Individual SAR from the 1991 Census; others are based on more speculative assessments. These illustrations are based on different geographies to provide a variety of examples. The geographies used are: UK, London (the most diverse region in England and Wales) and Camden (a London borough with a good cross section of residents of different ages, ethnicity, wealth etc.).

Using the SAM to create a simple univariate table.

This example is a table for a particular population subgroup, which is not available from the standard Census output. This example is produced for a region, which means that even the smallest groups have a large sample. This could be further increased by the addition of figures from the SAR, but this would not be expected to make a significant difference.

Why is this a good idea?

The flexibility of the SAM means that it is possible to produce a table such as this quickly and easily (and at no additional cost once the SAM has been obtained). In producing a report or piece of research, issues often arise that have not been foreseen, even in the most well-planned research. The SAM allows these to be explored.

Table 5.1 Percentage of children aged under 16 born in the UK by ethnic group London, 2001

	% born in the UK	Sample size
White British	99	37,657
White Irish	98	699
Other White	55	2,242
Caribbean (Black/Black British)	93	3,532
African (Black/Black British)	69	4,219
Mixed: White & Black Carib/Black African/Black Other	94	3,668
Indian (Asian/Asian British)	92	4,231
Pakistani (Asian/Asian British)	85	1,815
Bangladeshi (Asian/Asian British)	88	2,570
Other Asian (Asian/Asian British)	73	1,130
Chinese	84	556
Other	56	670
Mixed: White and Asian/Other Mixed/Other	88	2,242
Total	90	65,231

Source: 2001 Census, Small Area Microdata

Using the SAM to create a cross-tabulation for different geographies.

This can be useful as a first look at whether there are local differences where a table does not exist in the standard output. Again the major advantages of using the SAM are speed, ease of use and cost.

Table 5.2 Marital Status and Limiting Long Term Illness, 2001

United Kingdom	Limiting Long Term Illness			% with	% of all
	Yes	No	Total	LLTI (row %)	those with LLTI (column %)
Marital Status					
Single (never married)	116,380	1,185,243	1,301,623	8.9	21.4
Married/re-married	252,756	937,666	1,190,422	21.2	46.5
Separated (but still legally married)/divorced/widowed	174,124	270,216	444,340	39.2	32.1
Total	543,260	2,393,125	2,936,385	18.5	100.0

Great Britain	Limiting Long Term Illness			% with	% of all
	Yes	No	Total	LLTI (row %)	those with LLTI (column %)
Marital Status					
Single (never married)	111,940	1,148,642	1,260,582	8.9	21.3
Married/re-married	244,800	912,507	1,157,307	21.2	46.5
Separated (but still legally married)/divorced/widowed	169,453	264,884	434,337	39.0	32.2
Total	526,193	2,326,033	2,852,226	18.4	100.0

London	Limiting Long Term Illness			% with	% of all
	Yes	No	Total	LLTI (row %)	those with LLTI (column %)
Marital Status					
Single (never married)	16,040	173,858	189,898	8.4	28.9
Married/re-married	21,446	97,901	119,347	18.0	38.6
Separated (but still legally married)/divorced/widowed	18,096	30,882	48,978	36.9	32.6
Total	55,582	302,641	358,223	15.5	100.0

Camden	Limiting Long Term Illness			% with	% of all
	Yes	No	Total	LLTI (row %)	those with LLTI (column %)
Marital Status					
Single (never married)	675	5,408	6,083	11.1	42.1
Married/re-married	393	2,075	2,468	15.9	24.5
Separated (but still legally married)/divorced/widowed	534	825	1,359	39.3	33.3
Total	1,602	8,308	9,910	16.2	100.0

Source: 2001 Census, Small Area Microdata

Following on from this, the user can quickly look at other potential relevant factors, such as age, sex or change the marital status variable for family type, such as in the following example. Note that only two areas are shown.

Table 5.3 Family type, age and marital status, 2001

Great Britain	Limiting Long Term Illness			% of all	
	Yes	No	Total	% with LLT1 (row %)	those with LLT1 (column %)
Family Type under 50					
Lone parent	28,439	261,539	289,978	9.8	18.9
Married /cohabiting couple - no children	17,855	200,244	218,099	8.2	11.9
Married/ cohabiting couple -children	74,121	1,061,024	1,135,145	6.5	49.2
Ungrouped individual (not in a family)	30,116	195,398	225,514	13.4	20.0
Total	150,531	1,718,205	1,868,736	8.1	100.0
Over 50					
Lone parent	20,362	26,884	47,246	43.1	5.8
Married /cohabiting couple - no children	155,392	293,187	448,579	34.6	44.4
Married/ cohabiting couple -children	42,454	127,272	169,726	25.0	12.1
Ungrouped individual (not in a family)	131,450	135,003	266,453	49.3	37.6
Total	349,658	582,346	932,004	37.5	100.0

Camden	Limiting Long Term Illness			% of all	
	Yes	No	Total	% with LLT1 (row %)	those with LLT1 (column %)
Family Type under 50					
Lone parent	125	1,109	1,234	10.1	20.6
Married /cohabiting couple - no children	54	1,032	1,086	5.0	8.9
Married/ cohabiting couple -children	162	2,285	2,447	6.6	26.7
Ungrouped individual (not in a family)	265	2,059	2,324	11.4	43.7
Total	606	6,485	7,091	8.5	100.0
Over 50					
Lone parent	110	118	228	48.2	12.1
Married /cohabiting couple - no children	169	414	583	29.0	18.6
Married/ cohabiting couple -children	107	293	400	26.8	11.8
Ungrouped individual (not in a family)	523	617	1,140	45.9	57.5
Total	909	1,442	2,351	38.7	100.0

Using the SAM to explore Census data prior to commissioning a table.

The SAM has great potential to help specify a table for commissioning. It is quick and can identify whether the data is likely to show anything useful, as well as get an idea of any potential disclosure issues. The SAM may not always give the required level of detail, but will help to see whether a planned table is viable. It is useful to explore different options and layouts. It could be used to replicate a Standard Table, but with an additional variable for cross-tabulation, such as replicating a section of a Theme Table, but cross-tabulated by sex.

This example is a speculative look at whether housing need is related to religion and/or whether a person is born in the UK. A "Household housing indicator" is a derived variable included on the SAM based on overcrowding and access to amenities. The population base used for this example is all people, but could be restricted to household representatives to give estimates for numbers of households. Note that brackets are used to denote small numbers or percentages based on small numbers (under 50).

Table 5.4 People in households by housing need, born in the UK and religion, 2001

				Christian	Buddhist	Hindu	Jewish	Muslim	Sikh	Other	No religion	Religion not stated	Total
Not overcrowded or lacking amenities	Born in UK	Camden		2,132	(29)	(33)	332	163	(5)	107	943	427	4,171
		UK		134,692	626	3,366	5,445	6,075	1,868	3,478	34,535	17,528	207,613
	Born outside UK	Camden		804	(47)	50	141	261	(9)	(31)	292	184	1,819
		UK		25,243	1,101	6,152	1,176	8,761	1,763	961	5,452	4,257	54,866
Overcrowded/lacks bath/shower, wc or heating	Born in UK	Camden		1,032	(13)	(15)	52	236	(4)	55	453	182	2,042
		UK		32,587	342	1,669	584	5,939	845	997	9,396	5,982	58,341
	Born outside UK	Camden		539	(32)	(20)	(38)	452	(6)	(20)	174	129	1,410
		UK		12,821	591	3,188	253	9,557	701	381	2,549	2,640	32,681
				Christian	Buddhist	Hindu	Jewish	Muslim	Sikh	Other	No religion	Religion not stated	Total
Camden	% of born in UK in "housing need"			33	(31)	(31)	14	59	(44)	34	32	30	33
	% of born outside UK in "housing need"			40	(41)	(29)	(21)	63	(40)	(39)	37	41	44
	% of total in "housing need"			35	(37)	(30)	16	62	(42)	35	34	34	37
UK	% of born in UK in "housing need"			19	35	33	10	49	31	22	21	25	22
	% of born outside UK in "housing need"			34	35	34	18	52	28	28	32	38	37
	% of total in "housing need"			22	35	34	11	51	30	24	23	28	26

Source: 2001 Census, Small Area Microdata

Carrying out more complex statistical analysis using the SAM.

It is one of the strengths of microdata that investigations of relationships between characteristics can be undertaken at an individual level using statistical techniques such as **multivariate analysis**. The example shown earlier for limiting long term illness, taking into account age, marital status, family type, sex and other individual characteristics such as ethnic group can be combined with external characteristics such as housing to explore issues relevant to particular policies.

Using the SAM in modelling for scenario building, policy planning or impact assessments.

An example of this might be in transport planning, where car availability, mode of transport to work, employment status, age, sex, whether there are children in the household and dwelling type might be used in creating a model to look at the transport implications of a proposed development site based on the characteristics of individuals within the local authority.

Another example, in the education sphere, might be looking at ethnic differences in education, attainment, occupation and lifestyle, incorporating ethnic group, NS_SEC, level of highest qualification, employment status and family type.

6. Comparing figures from the SAM

This section looks at a number of tables and compares the results from the SAM with other datasets. Wherever possible two different geographies, generally one regional and one district have been used. To identify whether differences relate to the complexity of the query or to the geographies used the areas selected were the same as those used in the previous section. These examples look at:

- proportions from SAM samples compared with other datasets,
- absolute differences in percentages,
- the difference in proportion.

The sample size has been included to assist in this review. Clearly, areas with smaller populations are likely to have a higher degree of variability. A number of strategies are available to aid with presentation and interpretation of statistics where sample size, and therefore variability, may be an issue:

- rounding of percentages,
- suppressing figures,
- using brackets or italics,
- providing confidence intervals.

It may be appropriate to use a combination of methods, for example, suppressing figures based on very small samples (eg under 20) and using brackets for figures based on a slightly larger sample size (eg under 50). In this guide, no suppression is used, as the purpose is to illustrate the SAM, but figures based on a cell count of 50 or less are in brackets. This is a purely arbitrary figure and says little about the reliability of the data, although estimates based on very small samples should always be treated with caution. A better approach is to calculate confidence intervals, which are used to indicate how reliable an estimate might be. Examples of confidence intervals are calculated for some of the figures in the tables below. These use the formula⁹:

$$\text{Standard Error (SE)} = \sqrt{(p*(100-p)/n)}$$

Where p is the percentage calculated and n is the sample size

For example

In Table 6.2, the SAM estimates that **2.1** per cent of the population of Camden are from an Asian or Asian British **Indian** ethnic group. This is based on a total sample size for Camden of 9910. The standard error is therefore calculated as:

$$SE = \sqrt{(2.1*97.9/9910)} = \sqrt{0.02} = 0.14$$

Confidence intervals can then be calculated, giving the probability that the true estimate lies within a range, as follows (note that calculations are performed with all decimal places, but are presented here to just one decimal place):

1. There is a 68 per cent probability that the true estimate lies in the range from one standard error below the estimate to one standard error above the estimate.
In this example, there is a 68 per cent probability that the true estimate lies in the range 2.1 ± 0.1 or 2.0 to 2.2 per cent.

⁹ This formula assumes that the sample is a simple random sample. There may be additional design effects that have an impact on the estimates that are not taken account of here. Further work would need to be carried out to investigate whether there are any such effects for the SAM.

2. There is a 90 per cent probability that the true value lies within the range 1.6 SE below to 1.6 SE above the estimate. In this example, this is 2.1 ± 0.2 or 1.9 to 2.3.
3. There is a 95 per cent probability that the true value lies within the range two SE below to two SE above the estimate. In this example, this is 2.1 ± 0.3 or 1.8 to 2.4.

As we are comparing with a figure from a Standard Table, which contains the complete Census count, we have the true figure, which is in this case 2.3. This figure (with full decimal places) lies outside the 68 per cent confidence interval, but within the 90 per cent confidence interval.

From the same table (6.2), the figure with the largest percentage point difference is the percentage of White British women. The SAM estimate is 49.7 of women are from a White British ethnic group. The total number of women in Camden in the SAM is 5,147. The SE is calculated as:

$$SE = \sqrt{(49.7 * 50.3 / 5147)} = \sqrt{0.49} = 0.7$$

The confidence intervals are therefore:

68 per cent confidence interval is 49.0 to 50.4

90 per cent confidence interval is 48.6 to 50.9

95 per cent confidence interval is 48.3 to 51.1

The true figure (from the Standard Tables) lies within the 95 per cent confidence interval¹⁰.

Comparing SAM with a Standard Table

A table with highly different cell values (ethnic group by gender) from the Standard Tables has been reproduced from the SAM. The ethnic groups in the SAM have seen a small amount of consolidation; Mixed White and Black Caribbean, Mixed White and Black African and Black Other have been grouped. Mixed White and Asian and Other Mixed have also been aggregated.

Table 6.1 Ethnic group by gender for London

The figures appear nearly identical – there are no problems with using either source up to one decimal place. However if a Standard Table includes the required information, it should always be used in preference, as it gives 100 per cent coverage. One figure from the SAM (percentage of men that are Black or Black British Caribbean) has a 95 per cent confidence interval that does not cover the true figure at the second decimal place.

Table 6.2 Ethnic group by gender for Camden

All cells except one are within one percentage point difference. This table suggest there is slightly more difference between the two datasets, as would be expected with ever smaller samples. Nevertheless, there are only two SAM estimates giving a 95 per cent confidence interval which does not include the true figure.

¹⁰ In social research it is customary to quote just the 95 per cent confidence interval.

Table 6.1 Comparing SAM with a Standard Table, London

Ethnic Group for London	SAM			ST102			percentage difference		
	Male	Female	Total	Male	Female	Total	Male	Female	Total
White British	60.5	59.2	59.8	60.4	59.2	59.8	0.1	-0.1	0.0
White Irish	3.0	3.2	3.1	3.0	3.2	3.1	0.0	0.0	0.0
Other White	7.9	8.7	8.3	8.0	8.6	8.3	0.0	0.1	0.0
Mixed: White & Black Carib/Black African/Black Other	2.2	2.3	2.3	2.3	2.4	2.3	0.0	0.0	0.0
Mixed: White and Asian/Other Mixed/Other	1.7	1.7	1.7	1.7	1.7	1.7	0.0	0.0	0.0
Indian (Asian/Asian British)	6.2	5.9	6.1	6.2	6.0	6.1	0.0	-0.1	0.0
Pakistani (Asian/Asian British)	2.1	1.9	2.0	2.1	1.9	2.0	0.0	0.0	0.0
Bangladeshi (Asian/Asian British)	2.2	2.1	2.1	2.2	2.1	2.1	0.0	0.0	0.0
Other Asian (Asian/Asian British)	2.1	1.7	1.9	2.1	1.7	1.9	0.0	0.0	0.0
Caribbean (Black/Black British)	4.3	5.2	4.8	4.4	5.2	4.8	-0.1	0.0	0.0
African (Black/Black British)	5.1	5.4	5.3	5.2	5.4	5.3	0.0	0.0	0.0
Chinese	1.1	1.1	1.1	1.1	1.1	1.1	0.0	0.0	0.0
Other	1.5	1.7	1.6	1.5	1.7	1.6	0.1	0.0	0.0
Total	100.0	100.0	100.0	100.0	100.0	100.0			

Ethnic Group for London	difference as a proportion			Sample Size		
	Male	Female	Total	SAM Total	ST102 Total	
White British	0.2	-0.1	0.0	214,208	4,287,861	
White Irish	1.5	0.9	1.2	11,143	220,488	
Other White	-0.4	1.0	0.4	29,821	594,854	
Mixed: White & Black Carib/Black African/Black Other	-2.2	-1.7	-1.9	8,105	165,460	
Mixed: White and Asian/Other Mixed/Other	-0.1	2.9	1.5	6,133	121,002	
Indian (Asian/Asian British)	0.3	-1.2	-0.5	21,726	436,993	
Pakistani (Asian/Asian British)	-1.1	0.7	-0.2	7,113	142,749	
Bangladeshi (Asian/Asian British)	0.6	-0.4	0.1	7,693	153,890	
Other Asian (Asian/Asian British)	0.8	1.4	1.0	6,714	133,056	
Caribbean (Black/Black British)	-2.5	0.7	-0.7	17,040	343,561	
African (Black/Black British)	-0.6	-0.6	-0.6	18,818	378,931	
Chinese	-1.4	0.0	-0.6	3,980	80,203	
Other	3.4	-0.2	1.5	5,729	113,033	
Total				358,223	7,172,081	

95 per cent confidence interval does not include true figure

Table 6.2 Comparing SAM with a Standard Table, Camden

Ethnic Group for Camden	SAM			ST102			percentage difference		
	Male	Female	Total	Male	Female	Total	Male	Female	Total
White British	54.1	49.7	51.8	54.5	51.0	52.7	-0.4	-1.3	-0.9
White Irish	4.4	5.0	4.7	4.7	4.6	4.6	-0.3	0.4	0.1
Other White	15.1	17.3	16.2	14.8	16.8	15.8	0.3	0.5	0.4
Mixed: White & Black Carib/Black African/Black Other	2.2	2.0	2.1	1.9	2.0	1.9	0.3	0.0	0.1
Mixed: White and Asian/Other Mixed/Other	2.0	2.4	2.2	2.1	2.4	2.3	-0.1	-0.1	-0.1
Indian (Asian/Asian British)	1.9	2.3	2.1	2.4	2.2	2.3	-0.5	0.1	-0.2
Pakistani (Asian/Asian British)	0.7	0.6	0.7	0.7	0.6	0.6	0.0	0.0	0.0
Bangladeshi (Asian/Asian British)	6.6	5.9	6.3	6.6	6.1	6.3	0.1	-0.2	-0.1
Other Asian (Asian/Asian British)	1.3	1.2	1.3	1.1	1.0	1.1	0.2	0.2	0.2
Caribbean (Black/Black British)	1.7	2.1	1.9	1.6	2.0	1.8	0.1	0.1	0.1
African (Black/Black British)	6.3	6.4	6.3	5.7	6.2	6.0	0.6	0.1	0.4
Chinese	1.3	1.9	1.6	1.7	1.8	1.8	-0.3	0.1	-0.1
Other	2.4	3.3	2.9	2.2	3.1	2.7	0.2	0.2	0.2
Total	100.0	100.0	100.0	100.0	100.0	100.0			

Ethnic Group for Camden	difference as a proportion			Sample Size		
	Male	Female	Total	SAM	ST102	Total
White British	-0.7	-2.6	-1.7	5,137	104,389	5,137
White Irish	-6.0	8.8	1.6	465	9,147	465
Other White	1.8	3.0	2.5	1,608	31,364	1,608
Mixed: White & Black Carib/Black African/Black Other	18.4	-1.4	7.8	206	3,820	206
Mixed: White and Asian/Other Mixed/Other	-6.1	-3.1	-4.4	218	4,558	218
Indian (Asian/Asian British)	-21.3	3.3	-9.2	208	4,576	208
Pakistani (Asian/Asian British)	2.3	5.1	3.6	65	1,254	65
Bangladeshi (Asian/Asian British)	0.8	-3.6	-1.4	620	12,570	620
Other Asian (Asian/Asian British)	14.9	15.1	15.0	124	2,155	124
Caribbean (Black/Black British)	3.5	3.3	3.4	188	3,634	188
African (Black/Black British)	10.5	2.1	6.0	626	11,803	626
Chinese	-20.2	3.5	-7.2	161	3,468	161
Other	7.8	6.6	7.1	284	5,298	284
Total				9,910	198,036	9,910

95 per cent confidence interval does not include true figure
 Over one percentage point difference

Comparing SAM with a Commissioned Table

Table 6.3 shows figures on people by age, with/without a limiting long term illness and whether or not they live alone. This information was not part of the standard output and was commissioned for Local Authority Districts in England and Wales.

At a London level results are within one percentage point. Looking at Camden the differences, as might be expected, are slightly larger and more than half the cells are more than one percentage point different. Three cells are over five percentage points different. The value with the biggest discrepancy has one of the lowest SAM sample sizes in this table and relates to people aged over 85 without a limiting long term illness. However, confidence intervals calculated for this table show that none of the figures gives a confidence interval which does not include the true figure.

Table 6.3 Comparing SAM with a Commissioned Table, London

London	Commissioned Table % Living Alone		Small Area Microdata % Living Alone		SAM Vs CT	
	Without		Without		LLTI	Without
	LLTI	LLTI	LLTI	LLTI	diff.	LLTI
16-19	3.1	1.8	3.1	1.7	0.0	-0.1
20-24	10.4	5.9	9.7	5.7	-0.7	-0.2
25-29	18.5	11.1	18.8	11.1	0.3	0.0
30-39	25.9	15.7	25.8	15.8	0.0	0.1
40-49	26.3	15.2	26.9	15.2	0.6	0.0
50-59	28.4	17.4	28.3	17.4	-0.1	0.0
60-64	28.2	22.4	28.0	22.6	-0.2	0.2
65-74	32.4	30.2	31.5	30.1	-0.9	-0.1
75-84	46.3	45.6	47.0	45.3	0.6	-0.3
85+	62.2	62.1	63.1	61.9	0.9	-0.1
Total	32.6	15.5	32.6	15.5	0.0	0.0

London	CT proportional difference		SAM Sample Size		CT Sample Size	
	Without		Without		LLTI	Without
	LLTI	LLTI	LLTI	LLTI	LLTI	LLTI
16-19	-1.4	-6.2	808	14,998	15,928	305,731
20-24	-7.2	-3.0	1,261	24,303	25,005	487,346
25-29	1.8	-0.4	1,725	32,133	35,502	648,040
30-39	-0.1	0.4	5,575	60,276	111,751	1,208,515
40-49	2.4	0.0	6,739	39,706	134,268	788,947
50-59	-0.3	-0.2	8,736	27,988	172,616	556,900
60-64	-0.6	1.0	4,739	9,241	95,653	185,301
65-74	-2.8	-0.2	9,241	13,846	186,108	276,952
75-84	1.4	-0.7	8,161	6,864	163,099	137,028
85+	1.4	-0.2	3,448	1,444	68,613	29,156
Total	0.1	-0.2	53,477	300,024	1,008,543	4,623,916

Source: 2001 Census, Commissioned Table, C0680 / SCT247

Table 6.4 Comparing SAM with a Commissioned Table, Camden

Camden	Commissioned Table		Small Area Microdata		SAM Vs CT	
	% Living Alone		% Living Alone			Without
	LLTI	Without LLTI	LLTI	Without LLTI	LLTI diff.	LLTI diff.
16-19	4.8	2.4	6.3	3.8	1.4	1.4
20-24	18.5	9.8	23.1	9.5	4.6	-0.2
25-29	26.6	14.0	26.2	13.5	-0.5	-0.5
30-39	42.4	25.1	42.6	25.2	0.2	0.1
40-49	45.8	29.2	46.2	30.5	0.4	1.2
50-59	50.3	30.3	55.4	33.3	5.1	3.0
60-64	43.0	33.9	49.2	32.1	6.2	-1.8
65-74	43.1	39.8	41.7	35.8	-1.4	-3.9
75-84	51.1	50.9	48.5	50.0	-2.6	-0.9
85+	63.1	60.0	61.6	75.7	-1.5	15.7
Total	44.9	23.1	46.4	23.4	1.5	0.3

Camden	CT proportional difference		SAM Sample Size		CT Sample Size	
	LLTI	Without LLTI	LLTI	Without LLTI	LLTI	Without LLTI
	diff.	diff.				
16-19	29.8	57.3	(16)	372	457	7,062
20-24	24.8	-2.3	(39)	735	779	15,054
25-29	-1.8	-3.4	65	1,134	1,280	23,406
30-39	0.5	0.3	197	1,775	4,011	34,641
40-49	1.0	4.2	212	909	4,535	18,779
50-59	10.2	9.9	278	681	5,176	13,752
60-64	14.4	-5.4	128	221	2,317	4,410
65-74	-3.1	-9.9	211	335	4,394	6,575
75-84	-5.1	-1.8	206	168	3,741	3,303
85+	-2.4	26.1	86	(37)	1,608	760
Total	3.3	1.3	1,515	7,927	28,298	127,742

Source: 2001 Census, Commissioned Table, C0680 / SCT247

 Over one percentage point difference

Comparing SAM with the Individual SAR

Table 6.5 looks at commuting patterns by religion.

This information was not previously available; the only sources are the SAR, SAM or commissioning a table.

Due to the ability of the SAM to provide data at a lower geographical level the numbers this cross-tabulation produces are quite small. Some of the differences between the SAM and SAR may be explained by the smaller sample of the SAR, three per cent compared to the SAM's five per cent. Nevertheless, the resulting differences are all within five percentage points.

Table 6.5 Comparing SAM with the Individual SAR percentages

SAR	Camden		Inner London		Outer London		East of England		South East		London	
	Jewish	Total	Jewish	Total	Jewish	Total	Jewish	Total	Jewish	Total	Jewish	Total
Work mainly at or from home	-	-	15	9	13	9	15	9	14	10	14	9
Train, inc. Underground, metro, light rail, tram etc	-	-	30	38	26	27	14	7	12	6	27	31
Bus, minibus, coach	-	-	9	15	3	9	2	4	4	4	5	11
Motor cycle, scooter or moped//taxi	-	-	3	2	3	2	2	2	1	2	3	2
Car/(car pool NI)	-	-	27	21	51	45	60	65	59	65	43	36
Bicycle	-	-	3	4	1	2	2	4	2	3	1	2
On foot/other	-	-	13	11	3	7	5	9	7	10	6	9
SAM												
Work mainly at or from home	17	11	16	9	14	9	16	9	14	10	15	9
Train, inc. Underground, metro, light rail, tram etc	33	38	27	38	22	26	17	7	10	6	24	31
Bus, minibus, coach	6	12	7	14	5	9	2	4	3	4	6	11
Motor cycle, scooter or moped//taxi	2	2	4	2	3	2	2	2	2	2	3	2
Car/(car pool NI)	32	16	32	21	52	45	57	65	62	65	46	36
Bicycle	3	3	3	3	1	2	2	4	2	3	1	2
On foot/other	6	18	10	12	3	7	4	10	8	10	5	9
diff SAM-SAR												
Work mainly at or from home	-	-	1	0	2	0	2	0	0	0	1	0
Train, inc. Underground, metro, light rail, tram etc	-	-	-3	0	-4	0	3	0	-2	0	-4	0
Bus, minibus, coach	-	-	-1	0	1	0	0	0	-1	0	0	0
Motor cycle, scooter or moped//taxi	-	-	1	0	0	0	-1	0	0	0	0	0
Car/(car pool NI)	-	-	5	0	1	-1	-3	0	2	0	2	0
Bicycle	-	-	1	0	0	0	0	0	0	0	0	0
On foot/other	-	-	-3	0	0	1	-1	0	1	0	-1	0

Percentages in ***Bold Italics*** are based on less than 50. Values and percentages based on less than 20 should be suppressed.

	Camden		Inner London		Outer London		East of England		South East		London	
	Jewish	Total	Jewish	Total	Jewish	Total	Jewish	Total	Jewish	Total	Jewish	Total
proportional difference SAR												
Work mainly at or from home	-	-	3.8	-1.8	12.8	-0.6	10.5	-0.8	-1.1	0.3	9.4	-1.1
Train, inc. Underground, metro, light rail, tram etc	-	-	-11.0	0.5	-14.4	-1.8	21.1	0.6	-18.7	0.4	-13.4	-0.7
Bus, minibus, coach	-	-	-12.5	-1.3	37.3	1.6	-0.9	1.9	-16.5	-0.4	9.5	0.2
Motor cycle, scooter or moped//taxi	-	-	30.7	-1.1	3.4	5.9	-30.1	5.7	39.2	-4.0	11.2	2.9
Car/(car pool NI)	-	-	19.2	-0.1	1.5	0.4	-4.4	-0.1	4.0	0.2	5.6	0.2
Bicycle	-	-	27.3	-5.7	-10.0	-0.2	-11.8	-2.2	-16.5	-0.6	13.0	-3.2
On foot/other	-	-	-22.7	3.7	0.9	1.6	-19.9	0.3	7.4	-0.9	-16.6	2.7
SAR Sample Size												
Work mainly at or from home	-	-	107	3,417	184	5,541	65	7,572	(38)	12,010	291	8,958
Train, inc. Underground, metro, light rail, tram etc	-	-	207	14,644	375	17,120	63	5,537	(33)	6,991	582	31,764
Bus, minibus, coach	-	-	59	5,631	(49)	5,841	(11)	3,169	(10)	5,286	108	11,472
Motor cycle, scooter or moped//taxi	-	-	(19)	923	(47)	1,239	(11)	1,211	(3)	1,909	66	2,162
Car/(car pool NI)	-	-	188	8,300	734	28,648	265	51,732	160	77,968	922	36,948
Bicycle	-	-	(19)	1,399	(9)	1,048	(8)	3,126	(5)	3,754	(28)	2,447
On foot/other	-	-	93	4,350	(42)	4,602	(20)	7,579	(20)	12,597	135	8,952
Total	-	-	692	38,664	1,440	64,039	443	79,926	269	120,515	2,132	102,703
SAM Sample Size												
Work mainly at or from home	(48)	486	170	5,432	331	8,852	112	12,077	63	19,381	503	14,284
Train, inc. Underground, metro, light rail, tram etc	91	1,716	282	23,813	515	27,026	119	8,960	(45)	11,296	797	50,839
Bus, minibus, coach	(18)	544	79	8,997	108	9,538	(17)	5,193	(14)	8,474	187	18,535
Motor cycle, scooter or moped//taxi	(5)	108	(38)	1,478	78	2,109	(12)	2,060	(7)	2,950	116	3,587
Car/(car pool NI)	90	742	343	13,419	1,196	46,231	395	83,135	279	125,794	1,539	59,650
Bicycle	(9)	159	(37)	2,135	(13)	1,682	(11)	4,920	(7)	6,005	50	3,817
On foot/other	(17)	800	110	7,301	68	7,516	(25)	12,224	(36)	20,088	178	14,817
Total	278	4,555	1,059	62,575	2,311	102,954	691	128,569	451	193,988	3,370	165,529

Source: 2001 Small Area Microdata and 2001 Individual Sample of Anonymised Records

Comparing SAM with the LFS/APS

This compares the earlier simple example from the SAM of a univariate table for a population sub-group with the “same” table from the 2001 Annual LFS. As shown below, the SAM has a more detailed breakdown of the white ethnic group whilst the LFS has more detailed mixed ethnic categories. The SAM sample size is almost ten times the LFS for London in 2001 leading to much smaller confidence intervals for the SAM estimates. For example, the SAM 95 per cent confidence interval for the percentage of Chinese children born in the UK is 81 to 87 per cent, whereas the 95 per cent confidence interval for the LFS estimate of the same measure is 83 to 99 per cent.

Table 6.6 Percentage of children aged under 16 born in the UK by ethnic group London, 2001

	% born in the UK	Sample size
White British	99	38,097
White Irish	98	710
Other White	55	4,098
Caribbean (Black/Black British)	93	3,812
African (Black/Black British)	69	6,155
Mixed: White & Black Carib/Black African/Black Other	94	3,898
Indian (Asian/Asian British)	92	4,600
Pakistani (Asian/Asian British)	85	2,136
Bangladeshi (Asian/Asian British)	88	2,930
Other Asian (Asian/Asian British)	73	1,553
Chinese	84	660
Other	56	1,197
Mixed: White and Asian/Other Mixed/Other	88	2,536
Total	90	72,382

Source: 2001 Census, Small Area Microdata

Table 6.7 Percentage of children aged under 16 born in the UK by ethnic group London, 2001

	% born in the UK	Sample size
British	99	3,959
Other White	55	498
White and Black Caribbean	97	255
White and Black African	93	69
White and Asian	90	99
Other Mixed	(79)	38
Indian	94	438
Pakistani	88	244
Bangladeshi	89	388
Other Asian	73	187
Black Caribbean	96	501
Black African	70	751
Other Black	98	98
Chinese	91	57
Other	63	199
Total	90	7,781

Source: 2001 Labour Force Survey

7. Conclusions

This report has shown a range of advantages and disadvantages to using the SAM compared to other datasets. Because the SAM is taken from the national Censuses, which are very reliable, it does not suffer from the issues of non-response bias in the same way as sample surveys. In addition, it is significantly larger than any of the non-Census survey data, providing more robust figures for lower level geography than even the largest of these can provide. It has the flexibility to explore more detailed information not available in the predefined Census tables, while the speed of access makes it more open to changing information requirements where more complete data would take significantly longer.

The SAM is, however, fixed in time, whereas more up-to-date information is available from the non-Census datasets. It also has a smaller number of variables than most surveys can offer, since the Census asks only a few questions, whereas surveys conducted by interviewers can cover topics in more depth.

As with most things, the decision whether to use the SAM is a case of looking at when it adds greatest value. As shown in section 5, there are numerous applications of different types for the SAM within the public sector research domain. It's flexibility, ease of use and low cost are its chief assets, but it will not always contain the appropriate detail to answer a particular query. It should be used selectively and appropriately – for example, it would be unwise to use it to produce information that is available from a Standard Table. Nevertheless, the SAM is a valuable additional tool available for use by local authorities and others.

In summary, it is more flexible than the standard datasets from the Census; it gives more detailed geography than the other microdata from the Census available for use in a timely fashion. It is more accessible than the more detailed microdata; it has a broader range of topics than some of the national surveys and is larger, and therefore more reliable, than all of them. Inevitably, these strengths are offset by compromises on the number and level of detail given on some variables to preserve the confidentiality of Census respondents.

Appendix 1 List of variables in SAM

agea - Age of Respondents
cobirta - Country of Birth
combgn - Community Background - Religion or Religion Brought Up In (Northern Ireland)
country - Country
distmova - Distance of Move for Migrants (km)
distwrka - Distance to Work (Including Study in Scotland)
econach - Economic Activity (last week)
ethewa - Ethnic Group for England and Wales
ethn - Ethnic Group for Northern Ireland
ethsa - Ethnic Group for Scotland
everwork - Ever Worked
genind - Generation Indicator
health - General Health Over the Last Twelve Months
hourspwg - Hours Worked Weekly
lancode - local authority (GB) or parliamentary constituency (NI)
lastwrka - Year Last Worked
liti - Limiting Long Term Illness
marstata - Marital Status
miginda - Migration Indicator
migorgn - region of origin
nssec8 - NS-SEC 8 classes
popbasea - Population Base qualifier
profqual - Professional Qualification (England and Wales)
provcare - Number of Hours Care Provided per Week
qualvwn - Level of Highest Qualifications (Aged 16-74, EWN)
qualvs - Level of highest qualifications (16-74)
regiona - Region of usual residence
relgew - Religion (England and Wales)
relgn - Religion (Northern Ireland)
relgs1 - Religion Belongs to (Scotland)
sex - Sex
student - Schoolchild or Student in Full-Time Education
suprvsr - Supervisor/Foreman
termtima - Term time Address of Students or Schoolchildren
tranwrka - Transport to Work, UK (Including to Study in Scotland)
workforc - Size of Work Force
wrkplcea - Workplace
acctypa - Accommodation Type
bathwc - Use of Bath/Shower/Toilet
carsh - Cars/Vans Owned or Available for Use
cemtyp - Type of communal establishment
cenheat0 - Central Heating
ceststat - Status in Communal Establishment
densitya - No. of Residents per Room
furn - Accommodation Furnished (Scotland)
hedind - Household Education indicator

hempind - Household Employment indicator
hhsind - Household housing indicator
hhtlhind - Household health & disability indicator
hmptpuk - Hhd headship (ODPM)
hncarers - Number of Carers in the Household
hnearnra - Number of Employed Adults in Household
hnllti - Number in Household with Limiting Long-term Illness
hnrhlth - Number of Household Members with Poor Health
hnrresida - Number of Usual Residents in Household
hrsocgrd - Social Grade of Household Reference Person
lowflora - Lowest floor level of household living accommodation
occupncy - Occupancy Rating of Household
reltohra - Relationship to HRP
roomsflr - Number of Floor Levels (Northern Ireland)
roomsnum - Number of Rooms Occupied in Household Space
selfcont - Accommodation Self-Contained
stahuka - Household with Students Away During Term Time
tenurewa - Tenure of Accommodation, England and Wales,
tenursna - Tenure of Accommodation, NI, Scotland
famtypa - Family Type
fndepcha - Dependent Children in Family
freconac - Economic Position of Family Reference Person
frnssec8 - NS-SEC Social-Economic Classifications of Family Reference Person
frsex - Sex of Family Reference Person

Appendix 2 Accessing SAM

The 2001 Small Area Microdata is available after completing an End User Licence form. Registration for business users of the SAM will cost £1000. Local Authorities, Health Authorities, Central Government and charities will be entitled to a discounted non-academic rate of £500 per file. VAT is payable on these amounts. The Licence entitles up to ten people within an organisation to use the data and software supplied. All organisations must complete the End User Licence Agreement. Schedule 2 must name all individuals who will access the data. Any additional users must be notified to CCSR using an additional Licence Agreement. All ten users in each organisation will receive support backup and a free place on a training course. There will be a supplementary charge of £50 per additional user.

Academic users (based at institutions of further or higher education) can obtain the SARs free of charge for the purposes of research or learning and teaching. To do this, users will need to register with the Census Registration Service, which will permit them to use a range of Census products beyond the SARs. Any students using data in class must be individually registered.

The SAM can be used on an organisation's own PC without the need to check outputs to ensure that they do not breach respondents' confidentiality.

For details of how to access the SAM:

see: www.ccsr.ac.uk/sars/access/

email: sars-helpdesk@manchester.ac.uk

or complete the query form at: www.ccsr.ac.uk/sars/contact.html

or call the SARs helpline on 0161 275 4735

Appendix 3 Web links to further information

Census 2001

www.statistics.gov.uk/census/
www.gro-scotland.gov.uk/census/censushm/index.html
www.nisranew.nisra.gov.uk/census/start.html
www.neighbourhood.statistics.gov.uk/
www.nomisweb.co.uk/

Samples of Anonymised Records and Small Area Microdata

www.ccsr.ac.uk/sars/
www.nisranew.nisra.gov.uk/census/Census2001Output/sars.html

Controlled Access Microdata Samples

www.statistics.gov.uk/census2001/sar_cams.asp

Longitudinal Study

www.celsius.lshtm.ac.uk/

Census quality issues:

www.statistics.gov.uk/census2001/methodology.asp
www.gro-scotland.gov.uk/grosweb/grosweb.nsf/pages/cencr102
www.ccsr.ac.uk/sars/guide/2001/quality.html
www.nisranew.nisra.gov.uk/census/censusmethodology/census2001methodologypaper.html

Labour Force Survey

www.statistics.gov.uk/ssd/surveys/labour_force_survey.asp
www.esds.ac.uk/government/lfs/
www.data-archive.ac.uk/findingData/lfsTitles.asp
www.nomisweb.co.uk/

Annual Population Survey

www.esds.ac.uk/government/aps/
www.statistics.gov.uk/StatBase/Product.asp?vlnk=10855&More=Y
www.data-archive.ac.uk/findingData/apsTitles.asp

General Household Survey

www.statistics.gov.uk/ssd/surveys/general_household_survey.asp
<http://www.esds.ac.uk/government/ghs/>
www.data-archive.ac.uk/findingData/ghsTitles.asp

Family Resources Survey

www.dwp.gov.uk/asd/frs/
www.statistics.gov.uk/ssd/surveys/survey_family_resources.asp
www.esds.ac.uk/government/frs/
www.data-archive.ac.uk/findingData/frsTitles.asp

Survey of English Housing

www.statistics.gov.uk/STATBASE/Source.asp?vlnk=326
www.communities.gov.uk/index.asp?id=1154759
www.esds.ac.uk/government/SEH/
www.data-archive.ac.uk/findingData/sehTitles.asp

Regular Briefings from the GLA Data Management and Analysis Group

Recent *DMAG Briefings*:

DMAG 2005/37	Benefits Data for London: No 3 Children in Benefit Claiming Families	Lovedeep Vaid
DMAG 2005/38	Patterns of Ethnic Segregation in London	Baljit Bains
DMAG 2005/39	London's changing population	Eileen Howes (editor)
DMAG 2005/40	GLA 2005 Round Interim Ward Population Projections	Georgia Hay
DMAG 2005/41	Statistics of Schools in London: Key Facts 2001-2005	Karen Osborne
DMAG 2005/42	Model-based income estimates (2001/02)	Lovedeep Vaid
DMAG 2005/43	Census Information Note 2005-2	Gareth Piggott
DMAG 2005/44	2001 Census: London County Report	Gareth Piggott
DMAG 2005/45	Greater London Demographic Review 2004	Demography Team
DMAG 2006/1	Census Information Note 2006-1	Eileen Howes
DMAG 2006/2	Simpson's diversity indices for wards 1991 and 2001	Gareth Piggott
DMAG 2006/3	2001 Census: The health of a diverse population	Gareth Piggott
DMAG 2006/4	London borough residents country of birth	Giorgio Finella
DMAG 2006/5	GLA Resident Labour Force Projections	John Hollis
DMAG 2006/6	Parents and Work in London	Lorna Spence
DMAG 2006/7	Claimant Count Model: Technical Note 2006	Lorna Spence/ Georgia Hay
DMAG 2006/8	Demography Team Workplan 2006-07	John Hollis
DMAG 2006/9	Benefits Data for London: No. 4: Housing and Council Tax Benefits	Lovedeep Vaid
DMAG 2006/10	Household Representative Rates: Technical Report	Georgia Hay
DMAG 2006/11	Borough and Sub-regional Demographic Profiles, 2006	Georgia Hay
DMAG 2006/12	Interim Household Projections	John Hollis/ Georgia Hay
DMAG 2006/13	Social Exclusion Team Workplan 2006-07	John Hollis
DMAG 2006/14	Benefits Data for London: No 5 Pension Benefits	Lovedeep Vaid
DMAG 2006/15	Census Information Note 2006-2	Giorgio Finella
DMAG 2006/16	2001 Census: Quality of origin-destination data	Eileen Howes, Giorgio Finella, Bill Armstrong
DMAG 2006/17	Sub-Regional Demographic Profiles 2006	Georgia Hay
DMAG 2006/18	Independence for Statistics	John Hollis
DMAG 2006/19	Child Poverty in London	Lorna Spence (ed)
DMAG 2006/20	Benefits Data for London: 6 Tax Credits	Lovedeep Vaid
DMAG 2006/21	Family and Children's Study 2004	Lovedeep Vaid
DMAG 2006/23	A 2001 Census Ward Atlas of London	Bill Armstrong
DMAG 2006/24	National Insurance Registrations by overseas nationals	Gareth Piggott
DMAG 2006/26	A Profile of Londoners by First Language	Lorna Spence
DMAG 2006/27	2001 Census Profile: Jewish population of London	Gareth Piggott/Rob Lewis
DMAG 2006/28	Review of the 2001 Census Small Area Microdata	Rachel Leeser / Giorgio Finella

A full list of DMAG Briefings is available to internal customers through the GLA Intranet; otherwise please contact dmag.info@london.gov.uk A CD containing PDF versions of the Briefings, or hard copies, can be provided.

Contact details for the Data Management and Analysis Group are as follows:

Rob Lewis (020 7983 4652) is **Head of the Data Management and Analysis Group**. rob.lewis@london.gov.uk

Bill Armstrong (020 7983 4653) works in the **Census Team** with particular responsibilities for **commissioned tables, workplace data** and **mapping**. bill.armstrong@london.gov.uk

Baljit Bains (020 7983 4613) works in the **Demography Team** and is responsible for **ethnic demography**, including **ethnic group projections**. baljit.bains@london.gov.uk

Gareth Baker (020 7983 4965) is the **GIS Manager**. gareth.baker@london.gov.uk

Shen Cheng (020 7983 4889) works in the **Education Team** and is responsible for **school roll projections**. shen.cheng@london.gov.uk

David Ewens (020 7983 4656) works in the **Education Team** and is responsible for **research and data analysis**. david.ewens@london.gov.uk

Giorgio Finella (020 7983 4328) works in the **Census Team**. giorgio.finella@london.gov.uk

John Hollis (020 7983 4604) is responsible for the work of the **Demography, Education** and **Social Exclusion Teams**, and particularly for **demographic modelling**. john.hollis@london.gov.uk

Eileen Howes (020 7983 4657) is responsible for the work of the **Census, SASPAC** and **General Statistics Teams** and particularly for **census analysis**. eileen.howes@london.gov.uk

Ed Klodawski (020 7983 4694) works in the **Demography Team**. His post is joint with the **London Health Observatory** and specialises in **ethnic** and **health** issues. edmund.klodawski@london.gov.uk

Rachel Leeser (020 7983 4696) works in the **Social Exclusion Team** with particular responsibilities for **surveys, income data** and the **Social Exclusion Data Users Group**. rachel.leeser@london.gov.uk

Alan Lewis (020 7983 4348) works on the **SASPAC** project. alan.lewis@london.gov.uk

Gareth Piggott (020 7983 4327) works in the **Census Team**. gareth.piggott@london.gov.uk

Kelly Rump (020 7983 4655) is DMAG's **Senior Coordinator**. kelly.rump@london.gov.uk

Lorna Spence (020 7983 4658) is a member of the **Social Exclusion Team**, with particular responsibilities for **labour market data**. lorna.spence@london.gov.uk

Lovedeep Vaid (020 7983 4699) works in the **Social Exclusion Team** with particular responsibilities for **benefits, indicators, income data** and the **Social Exclusion Extranet**. lovedeep.vaid@london.gov.uk

Please use the above descriptions in deciding whom to contact to assist you with your information needs.