SCOTTISH HOUSEHOLD SURVEY

Methodology and fieldwork outcomes 2007

Steven Hope

Ipsos MORI Scotland

Julia Hall

National Centre for Social Research

A National Statistics Publication for Scotland

Contents

List of Tables

1 Survey overview

1.1 Background to the SHS

The Scottish Household Survey (SHS) is a continuous, large-scale social survey of the composition and characteristics of Scottish households, the attitudes and behaviour of adults, and to a lesser extent, of children within these households. It was first commissioned by the Scottish Executive in 1998 and has been collecting data and reporting since February 1999.

The specific aims of the survey are as follows:

- to provide household and individual information previously unavailable in Scotland, particularly to support the work of the Scottish Government's transport, communities and local government policy areas and the work of the Scottish Parliament
- to permit disaggregation of such information both geographically and in terms of population sub-groups (such as families with children or the elderly)
- to allow the relationships between social variables within households to be examined, supporting cross-cutting analysis on a range of issues
- to allow early detection of national trends
- to allow detailed follow-up surveys of sub-samples from the main survey sample, if required.

The survey is funded by the Scottish Government and carried by a consortium of research organisations involving Ipsos MORI, TNS System Three and the Scottish Centre for Social Research.

Technical Reports have been published for each year of the survey covering the survey methodology, fieldwork outcomes and the questionnaire used. This report covers the methodology and fieldwork outcomes for the survey in 2007. On the whole, the methodology for the survey remains fixed for each two-year sweep and over the life of the survey, since 1999, there have only been minor changes to the methodology following refinements to the sampling assumptions. More substantial changes were introduced in 2007 and these are discussed below.

A second publication provides details of the survey questionnaire used in 2007, including details of changes between 2006 and 2007 and changes made in January 2008.

1.1.1 Reporting conventions

In tables showing percentages as whole numbers, zero values are displayed as a dash (-), values between 0% and 0.5% are displayed as 0% and values between 0.5% and 1% are rounded to 1%. Where percentages are shown to one or more decimal place, the final digit will have been rounded up or down.

As a result of rounding within tables, the sum of individual items may not equal the totals for rows or columns.

2 Sample design, selection and allocation

The sampling requirements of the Scottish Household Survey are specified in terms of providing data with a level of precision equivalent to what would be achieved by a simple random sample (SRS) of a particular size. Five requirements were identified in the survey specification and although the preference was for simple random sampling to be used across the survey, contractors were invited to propose design solutions that would most cost-effectively meet these requirements.

The sampling requirements were to provide the SRS equivalents of:

- 2,500 interviews in Scotland as a whole for each quarter in each year
- 500 interviews in each of the larger local authorities (those which have over 120,000 households) for each year
- 500 interviews in each local authority area (regardless of size) for each two-year period
- 500 interviews in each category of the six-fold urban rural classification in each year, and
- 2,000 interviews in the 15% most deprived areas of Scotland, taken together as a group, in each year.

These needs are met by a combination of:

- disproportionately stratifying the sample to ensure that each local authority and each other geographical area has enough interviews in each survey period
- modelling the impact of different sampling methods and combinations of clustered and unclustered sampling to assess the impact of the survey design on survey precision at different geographical levels and at different points in time
- using a combination of unclustered and clustered sampling to maximise the cost-effectiveness of fieldwork by minimising the impact of weighting and clustering on the survey's effective sample size
- allocating the selected sample appropriately over survey periods to provide the appropriate level of clustering at each period.

The survey has a number of other requirements that have been retained from previous sweeps such as:

- the sample should be fully national in character – while it is not uncommon for national surveys to exclude the area north of the Caledonian Canal or to be restricted to the Scottish mainland and the larger inhabited islands, the Scottish Household Survey includes all parts of Scotland, including small inhabited islands.
- the sample should be capable of producing data which are representative both of Scottish households and the adult (aged 16+) population resident in private households.

Each of these features of the sampling is discussed more fully below.

2.1 Stratification by local authority

In general, stratifying a sample by some known variable should improve the precision of survey estimates because structuring a sample in this way can be no worse than would be achieved by a random allocation. For example, selecting a national random sample **should** result in a geographical distribution that reflects the distribution of population but it may not and some random deviation from the known distribution would be expected. Stratification removes the element of chance by assigning sample to geographical areas and drawing smaller samples within each of the strata. In this way the sample distribution must reflect the known population distribution.

The sampling requirements for the SHS indicate a need for the sample to be allocated between local authorities that does not reflect the distribution of the population so that each can, after two years, have the SRS equivalent of 500 interviews. For a given sample size, meeting this need requires smaller local authorities to have more interviews that a proportionate allocation would give them and, as a result, larger local authorities have fewer interviews.

Analysis at a national level requires the data to be weighted so disproportionate stratification reduces the precision of survey estimates, making the gross sample equivalent to a smaller simple random sample. This impact needs to be considered in meeting the survey's sampling requirements.

2.1.1 Disproportionate stratification between local authorities

The underlying principle here is that the allocation of interviews by local authority area should be broadly proportionate to the number of households, except where the resulting sub-sample in any particular area would fall below a pre-determined accuracy threshold (the equivalent of a simple random sample of 500 interviews). The allocation was carried out in the following way.

1. The sample design needs to take account of the allocation of interviews to each local authority and for each allocation, the resulting impact on the number of interviews derived from the smallest category of the urban rural classification, the 15% most deprived areas and the number of interviews achieved each year in the five largest local authorities.
2. These estimates need to take account of the impact of clustering and weighting at each level and at each relevant time period to ensure that the sample will, after accounting for weighting and clustering, meet the SRS equivalent sample size requirements specified.
3. Having established that a particular design will meet the sampling objectives at each geographical level and at each time period, the total sample is built from the combinations of individual local authority samples.

2.1.2 Clustered and unclustered sampling within local authorities

In previous years, the SHS has used a combination of clustered and unclustered sampling but the tendering for the new contract allowed some reconsideration of the approach used. Although the survey specification expressed a preference for a wholly unclustered sample, it was necessary to test a number of designs to identify the extent to which a more cost-effective design could be achieved by retaining an element of clustering in areas where this would produce fieldwork efficiencies.

An unclustered sample involves, within each of the primary strata (the local authorities), either selecting a simple random sample of addresses (n) from all possible addresses (N) or selecting a systematic sample by identifying a random starting address and selecting addresses with a fixed interval equal to N/n. Systematic sampling has the advantage that by ordering the addresses by some characteristics (such as deprivation or postcode) it is possible to achieve a further level of implicit stratification as the ordering and fixed interval sampling ensures that the variables used to order the list are represented in proportion to their prevalence in the population.

Clustered sampling is a two-stage process of selecting a sample of geographical units within which a sample of individual addresses is selected. The benefit of clustered sampling is to reduce the mean distance between sampled addresses, improving the efficiency of survey fieldwork by reducing interviewer travel time between addresses and increasing the number of times interviewers can try to make contact at sampled addresses. Although clustering offers an administrative benefit, this comes at a statistical penalty because clustering increases the likelihood that the achieved sample will be more variable than the total population. The reason for this is that the number of sampled clusters is generally small relative to the total number of potential clusters and, within each cluster, the sampled addresses are likely to be more similar to each other than they are to the addresses sampled in other clusters. This increased variability can be estimated and compared with the level of variability of a simple random sample to give an estimate of the achieved sample's simple random sample equivalence.

The extent to which a sample should be clustered therefore requires some comparison of the precision of the sample and the cost of achieving that level of precision – a clustered sample is only more cost-effective than a simple random sample if it can achieve the same level of precision as the equivalent simple random sample at lower cost.

There are two key variables that influence whether a sample should be clustered

- Population density – the extent to which the population is naturally concentrated in a geographical area or spread across a number of small settlements separated by large distances.
- The sampling interval - the size of the survey sample in relation to the population from which it is selected – because for a given population, a larger sample will result in sampled addresses being closer together reducing the administrative gains from clustering compared with an SRS.

The implication of this is that there is likely to be no efficiency gain from clustering a large sample in an urban area whereas there is greater likelihood of efficiency gains from clustering a small sample in a dispersed or rural area.

This broad approach was used in the 2007 SHS sampling using the Scottish Government's urban rural classification to identify areas where sample should be clustered or unclustered. The general approach was that areas classified as 'large urban areas' or 'other urban areas' would use unclustered sampling while areas in the other four categories (accessible small towns, remote small towns, accessible rural and remote rural) would use clustered sampling.

This was applied **within** local authorities meaning that the sample was further stratified within each local authority using the urban rural classification and that each local authority potentially contains a combination of clustered and unclustered sampling.

In practice, this general approach is modified in two ways.

- Where more than 80% of households in a local authority fall into the 'urban' or 'non-urban' category, the whole local authority is treated as that category
- The three island authorities (Eilean Siar, Orkney and Shetland) use wholly unclustered sampling even though their urban rural classification suggests that they should use wholly clustered sampling. In these areas, the sampling interval is between 1 and 6 households and 1 in 8 households which means that clustered sampling would be no more efficient than unclustered sampling.

Table 2-1 shows the expected distribution of sample by local authority at the end of each two-year sampling period. It should be noted that this distribution does **not** meet the specified requirement of 2,000 interviews each year in the 15% most deprived datazones in Scotland. This was agreed as part of the post-tender negotiations because allowing this requirement to fall slightly below the target allowed an overall reduction in the survey sample of 3,000 interviews over the four-year contract period, achieving a significant reduction in the total survey cost.

Table 2-1: Projected two-year achieved sample size by local authority and SRS equivalent sample over target periods

			Sample size	
	Total households	**% of ints from unclustered sample**	**Gross in target period**	**SRS equivalent in target period**
Large local authorities – minimum of 500 SRS equivalent interviews each year				
Edinburgh, City of	209,502	100%	1,129	1,129
Fife	153,040	63%	914	825
Glasgow City	276,291	100%	1,489	1,489
North Lanarkshire	134,700	100%	726	726
South Lanarkshire	128,238	100%	691	691
Other local authorities – minimum of 500 SRS equivalent interviews after two years				
Aberdeen City	98,859	100%	1,065	1,065
Aberdeenshire	92,067	0%	1,344	992

Angus	47,861	63%	572	516
Argyll & Bute	41,864	0%	678	500
Clackmannanshire	20,876	55%	500	500
Dumfries & Galloway	65,487	30%	866	706
Dundee City	67,032	100%	722	722
East Ayrshire	51,345	37%	662	553
East Dunbartonshire	42,763	100%	500	500
East Lothian	38,757	25%	622	500
East Renfrewshire	35,388	100%	500	500
Eilean Siar	11,360	100%	500	500
Falkirk	63,684	100%	686	686
Highland	92,514	22%	1,255	997
Inverclyde	37,883	100%	500	500
Midlothian	33,229	66%	549	500
Moray	36,515	25%	622	500
North Ayrshire	60,027	70%	702	647
Orkney Islands	8,380	100%	500	500
Perth & Kinross	60,866	36%	788	656
Renfrewshire	75,867	100%	818	818
Scottish Borders	48,790	28%	648	526
Shetland Islands	9,287	100%	500	500
South Ayrshire	50,754	69%	595	547
Stirling	37,321	54%	568	500
West Dunbartonshire	41,112	100%	500	500
West Lothian	65,030	70%	761	701

	Target	Target period	Gross sample	SRS equivalent
National	2,500	Quarterly	3,552	3,052
Smallest category of urban rural classification	500	Annually	1,301	650
15% most deprived datazones	2,000	Annually	2,101	1,881

2.2 Allocating sample to different time periods

The consideration of clustering is complicated for the SHS because the sampling requirements are expressed in terms of the equivalent of a simple random sample at different points in time. Consideration needs to be given to the structure of the sample at these time points and the extent of clustering in the sample taken into account. For example, although the sample in Glasgow might be selected without clustering, in practice, the two-year sample is allocated to survey years, 'batched' into interviewer allocations and these are then assigned to months of the year – creating clusters of addresses. Thus, each month, the sample in Glasgow is made up of clusters of addresses, with the sample becoming progressively less clustered throughout the year. In practice, then, the sample in Glasgow is only completely unclustered after a full year. Even if the samples in all local authorities were sampled without clustering, the quarterly samples would be clustered and this needs to be considered in terms of the ability of the design to meet the quarterly target of an SRS equivalent of 2,500 interviews.

The way in which sample is grouped into clusters optimises the extent of unclustered sampling in appropriate areas to coincide with reporting requirements.

- Large local authorities identified as requiring separate reporting of results each year – with the exception of Fife, over two years, the sample is derived from wholly unclustered sampling. All of the unclustered sample is first randomly allocated to years, grouped into the most efficient fieldwork batches and then these batches are allocated to months within each year. In Fife, the clustered sample is treated in the same way as in all other local authorities.
- Other local authorities, which require separate reporting after two years – addresses in datazones classified for unclustered sampling are combined, sorted by deprivation indicator and a systematic sample selected. These addresses are batched, batches allocated to survey years and then to months within each year. Within each local authority, all of the datazones classified for clustered sampling are grouped and a sample of datazones selected with probability proportionate to size. Within the sampled datazones, a systematic sample of addresses is selected. Sampled datazones are randomly allocated to survey years and then to months within each year.

This has implications for how much of the SHS sample is unclustered at any point in time. In each quarter, the whole sample is clustered. Each year, only the sample from four of the five large local authorities is unclustered. After two years, the five large local authorities and the unclustered samples from all other local authorities are unclustered, leaving only the sampled datazones as clustered samples.

After one year approximately 34% of the sample should be from areas of unclustered sampling but after two years this will increase to over 70%.

2.2.1 Allocating sample across the calendar year

As the fieldwork for the survey runs throughout the calendar year, it is important to ensure an even distribution of batches over time and to ensure that the allocation of batches is geographically and demographically representative. There are two main reasons for this: an uneven distribution would jeopardise the requirement for the sample to be representative of the national population on each quarter and some of the variables measured by the survey are likely to exhibit seasonal patterns – e.g. rates of economic activity, modes of transport.

The procedure for allocating PSUs to months of the year is derived from that developed by the Office for National Statistics (ONS) in managing the Family Expenditure Survey (FES)[1] and differs only in the need for the SHS sample to be spread evenly across 24 rather than 12 months.

[1] See C. Lound, 'Allocating primary sampling units for annual surveys to months of the year', *Survey Methodology Bulletin*, No. 39, July 1996.

Batches of addresses are allocated to survey years and within each year, sorted by local authority and by deprivation within each authority. The list of batches is then labelled with a random permutation of the numbers 1 to 12 representing the twelve months covered by the fieldwork. This permutation is generated with certain properties to avoid 'bunching' of interviews within particular quarters:

- the first four months are from different quarters
- every subsequent month is from the same quarter as the one four places before.

The example given by ONS (and used to allocate the 1996 FES) is as follows:

Table 2-2: Procedure for allocating PSUs by month of fieldwork

Position in list	Month	Quarter
1, 13, 25, etc.	10	4
2, 14, 26, etc.	8	3
3, 15, 27, etc.	5	2
4, 16, 28, etc.	1	1
5, 17, 29, etc.	11	4
6, 18, 30, etc.	7	3
7, 19, 31, etc.	4	2
8, 20, 32, etc.	2	1
9, 21, 33, etc.	12	4
10, 22, 34, etc.	9	3
11, 23, 35, etc.	6	2
12, 24, 36, etc.	3	1

As this sequence can be added automatically to the sampling procedures for the survey, no time is spent manually assigning batches to particular months.

2.3 Allocating sample between contractor organisations and questionnaire modules

Once all of the sampled addresses are batched for fieldwork, the batches are randomly assigned to one of the three contractor organisations in proportion to each contractor's fieldwork commitment (40% of interviews by TNS, 35% by Ipsos MORI and 25% by the Scottish Centre for Social Research).

The contract for 2007-2010 envisaged a requirement for greater modularisation of the SHS than had previously been the case. Modularisation requires the ability to only ask questions of a random sub-sample of respondents and for those sub-samples to be based on time periods or nationally representative sub-samples. One need identified in the survey specification was to create a module to measure participation in culture and sporting activities.

The allocation of sample to batches and these batches to months means that in theory, each month's sample is a random sub-set of the full sample, meeting any need for time-based modules. To meet the need for sub-sampling over the whole survey period, all sampled addresses were randomly assigned to one of 10 sub-samples or interview streams, which could be used as the basis for assigning samples of respondents to particular blocks of questions. For example, Culture and Sport module is intended to

provide representative data on adults' participation and this is achieved by assigning the module to streams 1 and 6 meaning that 1 in 5 addresses and (assuming no difference in response rates) 1 in 5 interviews will be directed through those questions.

Other smaller blocks of questions are asked of sub-samples at various points in the questionnaire and the published version of the script indicates where and at what point in time streaming is used.

2.4 Sampling from the Postcode Address File

The Small User File of the Postcode Address File (PAF) is now the standard sampling frame for general population surveys.[2] The principal advantages of the PAF are completeness (it is estimated to miss the addresses of only 2% of the adult population and is updated every three months) and lack of bias (those addresses which are missing from the PAF are not as likely to be concentrated among particular types of people). There are, however, a number of issues arising from its use.

2.4.1 Deadwood

The Small User File of the PAF, which forms the basis of the sample of addresses, contains a number of addresses that are not residential (usually small shops and offices) or which have been demolished or are unoccupied. In addition to PAF addresses that are out of scope for any household survey, there are also addresses that are deemed out of scope for this survey. These are mainly second home or holiday homes. In total, the extent of 'deadwood' in the PAF varies by area, but is usually estimated at around 10% in national samples. This is accounted for by drawing slightly more addresses than the response rate target would suggest. Thus, if the response rate target is 70% and deadwood is estimated to be 10% then for every 100 interviews to be achieved, 160 addresses are issued to interviewers rather than the 140 suggested by a response rate of 70% alone.

In practice, the number of additional addresses selected to allow for deadwood varies by local authority based on the contractors' experience of SHS fieldwork carried out between 1999 and 2005.

2.4.2 Accuracy and completeness

The sample for the survey is drawn for each two-year fieldwork period and so may exclude households in newly-built housing entering the PAF during the period of the survey. However, data suggests that new housing accounts for only around 1% of the housing stock in any year[3] and the impact of this is reduced by the fact that new properties are often entered onto the PAF some time before they are actually completed.

---

[2] Small users are postal addresses that receive fewer than 20 items of mail per day.

[3] http://www.scotland.gov.uk/Publications/2006/05/22160031/8 shows new housing completions of 24,581 in 2006 and a stock of dwellings of 2,407,000 at the end of 2005, making new housing 1.02% of the total 2005 stock.

2.4.3 Exclusions

Samples of the general population exclude prisons, hospitals and military bases. While prisons and hospitals do not generally have significant numbers of private households, the same may not be true of military bases. These are classified as Special EDs in the Census and account for just 0.5% of the population. Interviewing on military bases would pose fieldwork problems relating to access and security so they are removed from the PAF before sampling.

Specific accommodation types — The following types of accommodation are excluded from the survey if they are not listed on the Small User file of the PAF:

- nurses' homes
- student halls of residence
- other communal establishments (e.g. hostels for the homeless and old people's homes)
- mobile homes
- sites for travelling people.

Households in these types of accommodation are *included* in the survey if they are listed on the Small User file of the PAF and the accommodation represents the sole or main residence of the individuals concerned. People living in bed and breakfast accommodation are similarly included if the accommodation is listed on PAF and represents the sole or main residence of those living there.

Students' term-time addresses are taken as their main residence (in order that they are counted by where they spend most of the year). Since halls of residence are generally excluded, however, there will be some under-representation of students.

2.5 Multiple dwellings

There are potential problems associated with the fact that a single entry on the PAF may actually represent multiple dwellings or that a dwelling may contain multiple households. For example, an address listed as 14 Milton Street may consist of a tenement block containing 8 separate flats. Often, the existence of these additional addresses is indicated in the PAF in a field known as the Multiple Occupancy Indicator (MOI). To ensure that such households had an equal chance of inclusion, it is necessary to weight the address when drawing the sample. Thus 14 Milton Street would appear 8 times. In the address listings issued to interviewers, such addresses appear as '14 Milton Street - 3 of 8' etc., with interviewers given clear counting procedures for identifying the relevant selected dwelling.

Where the MOI is correct, this procedure is unproblematic. Sometimes, however, the MOI is incorrect or missing (in about 2% of cases) and the true number of dwellings at an address is only discovered once the survey is in the field.

Where an interviewer finds that the MOI is different from the actual number of dwellings observed (and there is more than one dwelling) he or she contacts the office where the correct details are used to randomly select one of the dwellings.

2.6 Respondent selection

As the survey is intended to collect information both about the structure and characteristics of Scottish households *and* about the people who occupy those households, the interview has a two-part structure. The respondent for the first part of the interview must be a householder – generally the Highest Income Householder or their spouse or partner[4]. For the second part of the interview, one adult (aged 16+) member of the household is selected at random by the CAPI script. If this person is not available at the time, the interviewer will call back to complete the interview at a later date if necessary.[5]

[4] This must be a person in whose name the accommodation is owned or rented or who is otherwise responsible for the accommodation. The Highest Income Householder (HIH) is taken as the household reference person for the first part of the interview. In households with more than one householder, the person with the highest income is taken as the household reference person. If householders have exactly the same income, the older or oldest is taken as the household reference person.

[5] The selection of the random adult is slightly more complex than this. The random adult needs to be one of the adult household members who is aged 16 years or over, is normally resident during term time (if a student) or has not been living outside of the household for 6 months or more.

3 Data collection methods and instruments

3.1 Use of Computer Aided Personal Interviewing (CAPI)

In common with many other large-scale government surveys, the SHS is carried out using Computer Aided Personal Interviewing (CAPI). This offers a number of important advantages over traditional pen-and-paper interviewing for a survey of this kind. These include the following.

- CAPI allows greater complexity in questionnaire design, since routing and 'loops' in the interview can be automated and thus effectively hidden from the interviewer. It also eliminates the need for complex selection procedures during an interview, since random selection can be built into the program.
- Overall data quality is improved because the need for a separate data entry stage is eliminated and because automatic routing and range and logic checks reduce the scope for interviewer error.
- Preliminary data are available at the end of each day's fieldwork and the lack of a separate data entry stage allows faster turnaround of results more generally.
- The CAPI system generates detailed information about the timing and duration of interviews, allowing fieldwork to be monitored more closely.

Until 2007, the SHS script had been based on In2itive software. The inclusion of the Scottish Centre for Social Research means that the interview is now also scripted using Blaise. The two versions of the script collect exactly the same information and although there are small differences in how the systems operate, these create no observable differences in the survey data.

3.2 Questionnaire development and changes

The original SHS questionnaire was developed between August and December 1998 by the contractor team, working in conjunction with the Technical Group for the survey from the Scottish Executive. This followed a broad consultation exercise in which interested parties from a range of policy areas, academia, voluntary organisations and other bodies were invited to propose topics or specific questions for inclusion in the survey.

The core of the SHS questionnaire is intended to remain constant, but there is also scope for the inclusion of different modules over time. A simplified version of the questionnaire for each survey year as well a detailed Topic List giving a broader look at past and present SHS topics can be found on the SHS website at www.Scotland.gov.uk/SHS.

This consultative approach was followed for the 2007 script. Also, the Scottish Government undertook a fundamental review of the Scottish Household Survey in 2006 to ensure that the SHS continues to meet Scottish Government needs and is appropriately focused on policy areas that reflect Government

priorities. As a result of this consultation and review process, the SHS questionnaire for 2007 is substantially changed compared with 2006. The accompanying volume on the survey questionnaire notes the changes made between 2006 and 2007, during 2007 and between 2007 and 2008.

3.3 Questionnaire structure, length and content

It was noted earlier that the questionnaire falls into two parts: the first collecting information about the composition and characteristics of the household from the Highest Income Householder or their spouse/partner; the second focusing mainly on the attitudes and experiences of a random adult member of the household. The former is intended to generate data representative of Scottish households and the latter data representative of the Scottish adult population resident in private households. It should be noted, however, that for reasons of space a handful of 'household' questions are also asked of the 'random adult'. These address household events or characteristics about which any adult member of the household would be likely to know.

A simplified version of the 2007 and 2008 questionnaire can be found on the SHS website (www.Scotland.gov.uk/SHS). The broad topic areas, however, are as follows.

In the first half of the interview (with the highest income householder or spouse/partner), respondents are asked about:

- household composition and characteristics of household members
- type of property/accommodation
- cars in household and access to public transport
- children in the household, childcare, satisfaction with schooling and travel to school
- employment status of the highest income householder
- household income from employment and other sources
- savings and household finances.

In the second half of the interview (with the 'random adult'), respondents are asked about their own:

- housing experiences, including homelessness
- education qualifications
- perceptions of the local area
- travel to work or education and use of private and public transport
- congestion
- travel on the previous day

- perceptions of services and local government
- experiences of neighbourhood disputes
- volunteering
- participation in sports and cultural activities
- health problems and caring responsibilities
- employment status
- individual income from employment and other sources.

3.4 Problems and errors in the survey scripts

Scripting the SHS for 2007 encountered a number of inter-connected problems related to the number and nature of changes to the questionnaire, the introduction of new survey modules, the time available for scripting and testing, and the sequence of scripting and testing of the In2itive and Blaise questionnaires.

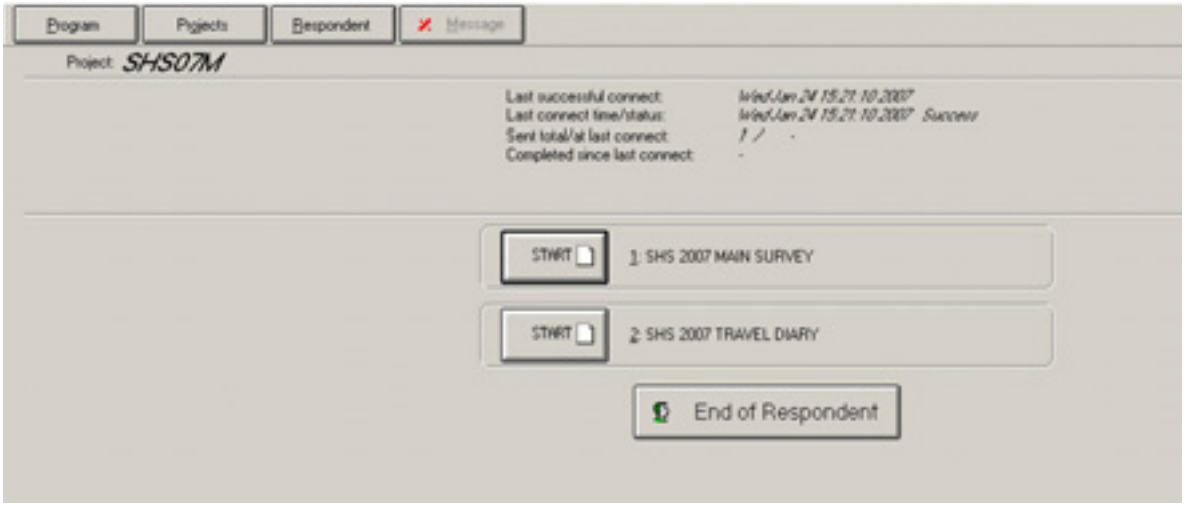
This caused a number of minor routing errors in the script – mainly in the culture and sport module – and the omission of some data from the main survey script. Two major problems arose with the In2itive software and a significant error in the Blaise script, which are detailed below.

3.4.1 In2itive software

Tendering for the SHS 2007-10 contract was on the assumption that Ipsos MORI and TNS would continue to use the same CAPI system as has been (very successfully) used since the survey started in 1999.

There were substantial revisions to the survey and for simplicity, the script was developed in three distinct parts reflecting the new requirements of the survey – the main SHS script, the Culture and Sport module and the revised travel diary. However, efforts to combine these into a single script were unsuccessful. We do not know precisely why it would not compile into one script. The best explanation is that the combined script breached some inherent programming limitation in the software such as the number of forms being processed or the number of routes involved in the script.

Attempts to compile the script and to investigate the problem delayed the start of fieldwork in 2007 and in the end, the need to get a working script into the field meant that an alternative solution had to be found. This was addressed by releasing the In2itive version of the SHS as two projects – one for Culture and Sport interviews and another for the main SHS. The main project was set up as two linked sub-scripts – the travel diary sits outside the main script as a linked module.



This set-up has very little impact on interviewers. The sample is pre-allocated to the streams required for all of the sub-sampling within the script and each script has been programmed to reject addresses allocated to the other script. The possibility of addresses being interviewed on the wrong script – main sample on the culture and sport script or vice versa – has therefore been removed.

Addresses are also identified as either Main sample or Culture and Sport on the contact sheets provided to interviewers. They therefore know which project to use, which showcards to use and how to introduce the survey.

In the main interview, where addresses are allocated to the travel diary stream, at the stage in the interview for completing the travel diary the interviewer sees a screen telling them to postpone the main interview and start the travel diary script. The transition is transparent to the respondent. However, there are instances of interviewers postponing the main script to undertake the travel diary and then being unable to restart the main interview.

Over the course of 2009, In2itive will be replaced by Nipo (for TNS interviewers) and SPSS Dimensions (for Ipsos MORI interviewers).

3.4.2 Errors in the In2itive and Blaise scripts

The problems with compiling the In2itive script had knock on effects on the quality of the scripts that were issued to interviews. Specifically, the intention of having a fully checked In2itive version of the questionnaire before the Blaise version was scripted proved impossible. As a result, both scripts were being checked and edited at the same time, within a timescale that was shorter than envisaged. This led, inevitably, to errors in both scripts. Most were minor and subsequent releases of the script corrected these.

A significant error in the Blaise script meant that the script went into the field without question HA3 which records the relationship of each household member to the Highest Income Householder. As well as omitting important information at this question, the information at this question is used in subsequent

routing to control the collection of the contribution to household income made by the HIH's spouse (where there is one). This error affected 631 households interviewed between the end of January and the end of March. Of these, we had telephone numbers for 561. These respondents were recontacted to collect data for HA3 and household income for some other variables where data was missing.[6] The information was successfully collected from 542 people.

3.5 Survey fieldwork

The main fieldwork for the survey has an on-going monthly cycle. Interviewers are required to make up to six calls at an address (an initial visit plus five 'call-backs'). In addition to the immediate reissue of contact sheets that have been wrongly completed or where the required number of call-backs has not been made, there is an on-going programme of reissuing 'non-contacts' in a bid to maximise the response rate. At the end of each fieldwork year a significant number of valid but 'non-contact' addresses remain 'live'.

The response rates for the SHS need to take account of the continuous nature of the survey. The data file for each year will contain a small proportion of interviews conducted on sample drawn the previous year. Similarly some of the addresses issued during any year will not be carried out until after the data file has been closed for analysis. These interviews are carried into the next data file. The response rates therefore report the outcomes for addresses sampled for a given period regardless of when the interview was carried out. Details of the most recent response rates are given at paragraph 4.2.1.

Because of the problems with the scripts, the fieldwork for the 2007 survey did not start until 25 January 2007 and over the first two months of the survey it became apparent that the main interview was substantially longer than the planned 45 minutes and this was having a detrimental effect on the interviewers' ability to meet their fieldwork targets. In June 2007, it was agreed to revise the fieldwork targets for 2007 downwards by 500 to reflect this.

A substantial number of changes were made to the script in June 2007 to reduce the average interview length to the target 45 minutes. These mainly involved introducing more streaming into the survey, reducing the proportion of the full sample that were asked questions, although some questions were deleted in June. The questions collecting the ethnicity and religious persuasion of all household members were replaced by questions asked only of the random adult.

[6] This additional missing information applied mainly to the Culture and Sport module where some core demographic questions (smoking, place of residence 1 year ago, employment details) had been omitted.

4 Fieldwork targets and outcomes

The requirements of the sample for the survey are expressed in terms of the ability of the sample to deliver the number of interviews required to meet particular requirements at particular points in time. There are also requirements in terms of the geographical scope of the interviews and their representativeness in terms of the types of people and areas included in the survey. Overall, the survey is expected to be an unbiased sample of Scottish households and adults and this is most likely to be achieved by maximising the level of participation in the survey.

With the sample designed to meet these objectives, these represent the key performance criteria for the survey. The survey's administration procedures are designed to minimise the impact of problems such as potential respondents not being at home or being unable to take part because of communication difficulties. Interviewers are required, for example, to make a minimum of 6 calls at each address on different days and at different times before it is considered 'no contact'. Even then, addresses will be reissued at a later stage in the fieldwork. Similarly 'soft' refusals such as 'too busy' or 'going out' are reissued.

Where interviewers are unable to conduct interviews in English or because the respondent is blind or partially sighted, these addresses can be revisited by an interviewer accompanied by an interpreter and, if appropriate, with showcards printed in a range of languages to minimise the extent to which language and communication barriers prevent people from taking part.

Nevertheless, participation in surveys is voluntary and some potential respondents refuse to take part. Similarly, no contact may be made at an address. This often reflects combinations of household types (single adults of working age), lifestyles (long working hours, active social lives), particular types of neighbourhoods and passive refusal (reluctance to open doors to strangers).

4.1 Sample performance

4.1.1 Ineligible addresses

The sampling for the survey makes assumptions about the proportion of addresses that will be ineligible for interview in each local authority. Ineligible addresses would include derelict, vacant and non-residential addresses. The allowance for ineligible addresses is based on the proportion that was actually identified in the course of the previous year's SHS fieldwork, the most recent two-year sweep when the sample was being designed in 2006. The extent to which these assumptions are accurate has an important bearing on the survey outcomes. If there are more 'deadwood' addresses, the interviewers have a smaller pool of addresses from which to achieve the target number of interviews. Conversely, a smaller proportion of 'deadwood' addresses should make it easier to achieve the target number of interviews but this target will be met with a lower response rate. Thus, overall, if the proportion of

deadwood differs from the sampling assumptions this might have some impact on achieving the interview target and the target response rate.

Table 4-1 shows the proportion of deadwood addresses assumed in each local authority when sampling for 2005/2006 and compares this with the proportion recorded by interviewers in the field.

Table 4-1: Deadwood rate assumptions and actual deadwood
Sorted in descending order of deviation (absolute value) between actual and assumption

	Average deadwood 1999-2006 (%)	**2007 actual deadwood (%)**	**Deviation** (% points)
Renfrewshire	12.2	6.9	-5.3
Orkney Islands	11.4	14.4	3.0
Eilean Siar	14.5	17.4	2.9
Highland	10.2	12.6	2.4
Shetland Islands	9.5	11.7	2.2
Argyll & Bute	22.0	19.9	-2.1
Falkirk	6.5	4.4	-2.1
Inverclyde	11.9	13.8	1.9
Aberdeen City	10.0	8.2	-1.8
Perth & Kinross	9.5	7.7	-1.8
Stirling	9.2	7.4	-1.8
Glasgow City	12.9	11.2	-1.7
Dumfries & Galloway	9.9	8.3	-1.6
Clackmannanshire	7.9	6.5	-1.4
Dundee City	12.0	10.6	-1.4
West Lothian	5.6	4.4	-1.2
Fife	8.2	7.1	-1.1
North Ayrshire	9.1	8.0	-1.1
Scottish Borders	9.8	8.7	-1.1
Midlothian	4.6	3.6	-1.0
East Dunbartonshire	4.2	3.3	-0.9
West Dunbartonshire	10.0	9.2	-0.8
East Lothian	6.5	7.3	0.8
Aberdeenshire	9.1	8.3	-0.8
Moray	9.3	8.7	-0.6
East Renfrewshire	4.4	4.9	0.5
North Lanarkshire	5.9	6.4	0.5
South Ayrshire	8.1	8.6	0.5
South Lanarkshire	7.4	6.9	-0.5
East Ayrshire	5.8	6.1	0.3
Angus	8.3	8.6	0.3
Edinburgh, City of	8.1	8.3	0.2
All areas	**9.4**	**8.9**	**-0.5**

This shows that overall, and in many local authorities, the level of deadwood recorded by interviewers was close to that used as the basis for the survey sampling. There is, of course, some deviation from the assumptions, reflecting sampling variability in the data used for sampling and the sampled addresses. In spite of the deviation from assumptions, using different deadwood assumptions in individual local authorities rather than previous practice of a uniform 10% in all areas improves the structure of the sample and should contribute to meeting fieldwork targets. Where the experience differs from the assumptions this is likely to reflect a combination of factors such as:

- housing regeneration and redevelopment, which leads to demolition and vacant properties (increasing deadwood) and properties being brought back into use (lowering deadwood).
- expansion of holiday properties and second homes, which are ineligible for inclusion in the survey, raising deadwood.

4.2 Fieldwork performance

The profile of the sample and the level of deadwood are primarily qualities of the sampling frame and the assumptions used to sample. Inaccuracy and bias in these can have a knock-on effect on fieldwork performance. The other elements of fieldwork performance reflect:

- survey administration procedures and interviewer performance
- the availability of members of the public to be interviewed
- the ability of members of the public to participate in the interview
- the willingness of members of the public to participate in the survey.

Performance on each of these elements (as well as deadwood) is recorded as part of interviewers' attempts to secure interviews although there is, inevitably, interaction between these different aspects of performance.

4.2.1 Survey response rates

Overall, performance is summarised in the survey response rate and this is shown below for the 2007 sample. This takes account of the continuous nature of the survey. The data file for each year may contain a small proportion of interviews conducted on samples drawn in previous years. Similarly some of the addresses issued during any year will not be carried out until after the data file has been closed for analysis. These interviews are carried into the next data file. The response rates therefore report the outcomes for addresses sampled for a given period regardless of when the interview was carried out.

Table 4-2: Summary of outcomes at issued addresses for 2007 sample

	Frequency	Percent	Valid Percent
Complete interview	17,093	59.7	65.7
Interview / partial interview achieved but data deleted	182	0.6	0.7
No contact with anyone at the address	3,083	10.8	11.9
Office refusal	648	2.3	2.5
Refusal by selected respondent	3,404	11.9	13.1
Refusal by proxy	333	1.2	1.3
Broken appointment, no recontact	294	1.0	1.1
Ill at home during survey period	319	1.1	1.2
Away/in hospital during survey period	230	0.8	0.9
Language	77	0.3	0.3
Other non-response	248	0.9	1.0
Total eligible for inclusion in the survey	**26,011**	**90.8**	**100.0**
Not yet built/under construction	47	0.2	
Demolished/derelict	236	0.8	

Vacant/empty	1,313	4.6
Non-residential address	271	0.9
Communal establishment/institution	35	0.1
Address out of scope	351	1.2
Insufficient address/no trace	164	0.6
Other ineligible	209	0.7
Total ineligible	**2,626**	**9.2**
Total issued addresses	**28,637**	**100.0**

This is the response rate for the whole sample issued in 2007 – both the main sample and the Culture and Sport module. Separately, the response rate for the main sample was 65% and 67% for the Culture and Sport module.

4.2.2 Trends in response rates

The response rate of 65.7% in 2007 is the lowest response rate achieved on the survey since 1999 and reflects the combined impact of the late start to 2007 fieldwork, the errors in the scripts and the excessive length of the interview at the start of the year.

The table below shows the response rates for each local authority for each two-year period between 1999 and 2006 and compares the 2007 rates with the average in each local authority in that period.

Table 4-3: Trends in SHS response rates 1999 to 2007

	Response rate 1999/2000 (%)	Response rate 2001/2002 (%)	Response rate 2003/2004 (%)	Response rate 2005/2006 (%)	Average 1999-2006	Response rate 2007 (%)
Aberdeen City	65	67	66	66	66	63
Aberdeenshire	68	70	74	73	71	74
Angus	67	73	75	73	72	69
Argyll and Bute	71	69	73	74	72	73
Clackmannanshire	66	62	71	77	69	67
Dumfries and Galloway	69	69	73	72	71	72
Dundee City	62	66	67	69	66	64
East Ayrshire	69	71	75	71	72	65
East Dunbartonshire	68	69	73	69	70	70
East Lothian	67	63	67	67	66	64
East Renfrewshire	59	63	66	63	63	61
Edinburgh, City of	64	60	63	66	63	56
Eilean Siar	79	81	79	78	79	81
Falkirk	66	65	74	72	69	69
Fife	65	65	75	76	70	76
Glasgow City	62	63	60	59	61	55
Highland	68	71	70	71	70	70
Inverclyde	68	69	73	69	70	76
Midlothian	66	66	68	67	67	64
Moray	72	72	76	76	74	72
North Ayrshire	70	63	69	66	67	64
North Lanarkshire	61	64	67	70	65	63
Orkney Islands	70	80	80	77	77	80
Perth and Kinross	70	68	67	71	69	64
Renfrewshire	64	65	71	62	65	59
Scottish Borders	68	71	78	78	74	71
Shetland Islands	70	78	80	76	76	70

South Ayrshire	67	68	71	70	69	66
South Lanarkshire	64	65	67	67	66	64
Stirling	68	71	77	80	74	76
West Dunbartonshire	63	64	67	67	65	69
West Lothian	65	65	71	71	68	57
Total	**66**	**67**	**69**	**69**	**68**	**65**

The most notable differences between the 1999-2006 average and the 2007 response rate are in Shetland, Renfrewshire and Glasgow (down 6 percentage points), East Ayrshire and Edinburgh (down 7 percentage points) and West Lothian, where the 2007 response rate is 11 percentage points lower than the average for 1999-2006. Some areas have higher response rates than the 1999-2006 average, although fewer and with more modest differences. The largest comparable positive difference is in Fife, where the 2007 response rate is 5 percentage points higher than the average.

In 2007, the highest response rates were achieved in island and largely rural authorities. Response was highest in Eilean Siar (81%), Orkney (80%), Stirling and Fife (76%). Inverclyde – an urban authority - also achieved a response rate of 76% The lowest response rates were, as they generally have been, in Glasgow City (55%), Edinburgh (56%), West Lothian (57%), Renfrewshire (59%) and East Renfrewshire (61%).

4.2.3 Achieved interviews and fieldwork target

Compared with the total target of 17,977 interviews that the sampling for the survey was designed to yield, the 2007 total of 17,093 represents a total shortfall of 884 or 5% of the interview target. As mentioned earlier, in response to the longer interview at the start of 2007, the Scottish Government accepted that the interview target should be reduced by 500 to reflect the lower productivity of fieldwork in the first quarter of the year. This means that the achieved interviews represent a shortfall of 384 interviews or 2% of this revised target.

The revised target was not related to specific local authority targets so the percentage achieved in each local authority shown below relates only to the original target that the sample was designed to achieve.

4.2.4 Achieved interviews compared with targets – household interviews

The number of interviews compared with the target, and the corresponding response rates, are the principal measures of survey performance although issues of data quality and bias also need to be considered. The table below compares interview targets and achievement in each local authority.

Table 4-4: Household interview targets and numbers achieved in each local authority, 2007

	Target	Achieved	% of target achieved	Over / under achieved
Aberdeen City	630	600	95%	-30
Aberdeenshire	825	817	99%	-8
Angus	384	376	98%	-8
Argyll & Bute	484	492	102%	8

Clackmannanshire	360	337	94%	-23
Dumfries and Galloway	545	549	101%	4
Dundee City	460	455	99%	-5
East Ayrshire	445	395	89%	-50
East Dunbartonshire	350	328	94%	-22
East Lothian	429	413	96%	-16
East Renfrewshire	350	312	89%	-38
Edinburgh City	1,340	1,174	88%	-166
Eilean Siar	340	340	100%	0
Falkirk	460	420	91%	-40
Fife	1,016	1,088	107%	72
Glasgow City	1,750	1,612	92%	-138
Highland	776	758	98%	-18
Inverclyde	340	346	102%	6
Midlothian	382	360	94%	-22
Moray	429	404	94%	-25
North Ayrshire	442	422	95%	-20
North Lanarkshire	840	793	94%	-47
Orkney Islands	350	345	99%	-5
Perth & Kinross	520	505	97%	-15
Renfrewshire	510	449	88%	-61
Scottish Borders	428	398	93%	-30
Shetland Islands	340	288	85%	-52
South Ayrshire	422	393	93%	-29
South Lanarkshire	810	773	95%	-37
Stirling	396	390	98%	-6
West Dunbartonshire	340	361	106%	21
West Lothian	484	400	83%	-84
Total	**17,977**	**17,093**	**95%**	**-884**

The highest levels of under-achievement were in Glasgow City, West Lothian, Renfrewshire and the City of Edinburgh. Taken together, these authorities account for 45% of the total shortfall (i.e. the sum of all the authorities where the interviews achieved is short of target). The highest percentage shortfalls were in West Lothian (83% achieved), Shetland Islands (85%), Edinburgh (88%) and Renfrewshire (88%).

4.2.5 Achieved interviews – the random adult

The two-part structure of the SHS interview requires the selection of a random adult within the household who completes the second half of the interview. This represents a second opportunity for potential respondents to withdraw from the interview either because they refuse to take part or are unable, unavailable or not contactable for interview.

There is inevitably a degree of attrition between the household and random adult sections of the interview, especially where the person selected is not the same as the household respondent. This aspect of the survey has deteriorated since 1999/2000 when a random adult interview was achieved in 94% of households in which a household interview was completed. In 2007, random adult interviews were achieved at 92% of households where a household interview was completed. The participation rate varied from 86% in Dundee and Aberdeen to 97% in Orkney and 96% in Inverclyde. A participation rate of 92% means that while the overall household response rate was 65%, the random adult response rate was 60%.

The combination of lower response rates to the household survey and attrition at the random adult stage means that the random adult response rates varied from 49% in West Lothian and 51% in both Glasgow and Edinburgh to 78% in Orkney and 73% in Eilean Siar.

Table 4-5: Random adult (RA) response rates, 2007

	Valid addresses	Household interviews	Random adult interviews	RA interviews as % of valid addresses	RA interviews as % of household interviews
Aberdeen City	971	600	515	53	86
Aberdeenshire	1,099	817	754	69	92
Angus	530	376	349	66	93
Argyll and Bute	669	492	457	68	93
Clackmannanshire	502	337	313	62	93
Dumfries and Galloway	772	549	510	66	93
Dundee City	698	455	391	56	86
East Ayrshire	598	395	373	62	94
East Dunbartonshire	471	328	306	65	93
East Lothian	647	413	375	58	91
East Renfrewshire	527	312	291	55	93
Edinburgh City	2,084	1,174	1,056	51	90
Eilean Siar	426	340	313	73	92
Falkirk	614	420	375	61	89
Fife	1,438	1,088	1,010	70	93
Glasgow City	2,943	1,612	1,489	51	92
Highland	1,056	758	708	67	93
Inverclyde	469	346	333	71	96
Midlothian	561	360	331	59	92
Moray	574	404	385	67	95
North Ayrshire	657	422	392	60	93
North Lanarkshire	1,249	793	716	57	90
Orkney	428	345	333	78	97
Perth and Kinross	786	505	445	57	88
Renfrewshire	754	449	433	57	96
Scottish Borders	568	398	369	65	93
Shetland	401	288	270	67	94
South Ayrshire	583	393	357	61	91
South Lanarkshire	1,198	773	688	57	89
Stirling	522	390	362	69	93
West Dunbartonshire	522	361	338	65	94
West Lothian	694	400	338	49	85
Total	**26,011**	**17,093**	**15,675**	**60**	**92**

The random adult response rate for the main SHS interview was 59.9% and for the Culture and Sport module was 62.4%.

5 Weighting

Two types of weighting are potentially necessary with a sample of this kind. The first is intrinsic to the survey design and represents weights necessary to compensate for unequal probabilities of selection for individuals, households or other units of analysis. The second may be necessary to counteract the effects of non-response bias. Although these represent two distinct rationales for weighting, in terms of analysis the different weights are combined into a single weighting variable for each unit of analysis.

In the SHS, there are nine weights that can be used – four in the main survey dataset, two specific to the Culture and Sport module, two for analysing core questions common to the main and Culture and Sport samples and one specific to the travel diary. However, LA_WT and IND_WT are used for most analyses, with the others used for smaller specific subsets of the sample.

- LA_WT which is used for analysis of data about the household and data collected from or about the HIH and spouse in the main SHS sample. This includes all variables asked in the first part of the interview, apart from the questions about the random schoolchild and the random child receiving childcare.
- IND_WT which is used for analysis of data in derived variables about the random adult or collected from the random adult. This includes all variables in the second part of the interview.
- LA_CS and IND_CS are equivalent to LA_WT and IND_WT but apply to the Culture and Sport sample.
- LA_ALL and IND_ALL are equivalent to LA_WT and IND_WT but only apply to questions common to the main sample and the Culture and Sport sample.
- KID_WT which is used for analysis of questions related to the random schoolchild – HE9 to HE17N inclusive (see ***Questionnaire***).
- RANKIDWT which is used for questions about childcare. One child is selected at random from all the children in the household.
- TRAV_WT, contained in the travel diary data, which is used for analysing that data.

5.1 Design weighting

5.1.1 Weighting for analysis based on household data

The weight for analysis of household data, LA_WT, has two main elements. Firstly, it is necessary to 'weight up' those local authorities which were under-sampled and 'weight down' those which were over-sampled (this is a weight of the first type mentioned above, which adjusts for unequal probabilities of selection). Secondly, the weight addresses any additional disproportionality introduced by response rates differing from the target for each local authority. The combination of these two elements is shown in Table

5-1. (The weights for some local authorities vary between one quarter and the next because the number of achieved interviews fluctuates between quarters.) The final weighted sample profile across the two years should, therefore, correctly reflect the distribution of households across Scotland's local authorities.

Weights are calculated for each local authority so that in each quarterly data file, the data is nationally representative. This should allow any published findings to be reproduced by selecting the relevant quarter's data. In practice, however, it may not be possible to reproduce exactly some of the results from earlier publications if the data for that quarter were subsequently changed (e.g. to correct errors that were identified later) and because there is some overlap between the quarter in which interviews take place and the quarter's data with which it is processed. For example, the data processed as Q4 2007 contained data from interviews carried out in January 2008 so although they were weighted as Q4, they have a value of 1 for the Quarter variable.

Table 5-1: Weights to account for disproportionate sampling and differences in household response rates by local authority and quarter, 2007

	Main sample (LA_WT)				Culture and sport (LA_CS)			
	Q1	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**
Aberdeen City	1.35	1.15	1.16	1.13	2.23	1.50	1.16	1.13
Aberdeenshire	0.68	0.89	0.87	1.30	0.96	1.53	0.87	1.30
Angus	1.03	0.93	1.22	1.63	0.81	0.63	1.22	1.63
Argyll and Bute	0.61	0.61	0.72	1.12	0.39	0.53	0.72	1.12
Borders	0.84	1.11	1.09	0.95	0.75	0.93	1.09	0.95
Clackmannanshire	0.55	0.45	0.43	0.46	0.43	0.31	0.43	0.46
Dumfries and Galloway	0.74	0.89	1.12	1.34	0.92	0.99	1.12	1.34
Dundee City	1.14	1.13	1.00	1.01	1.02	0.97	1.00	1.01
East Ayrshire	1.09	1.01	0.86	1.31	0.82	0.74	0.86	1.31
East Dunbartonshire	0.88	1.46	1.35	1.02	0.79	0.89	1.35	1.02
East Lothian	0.86	0.76	0.92	1.18	0.64	0.50	0.92	1.18
East Renfrewshire	0.85	1.22	1.00	0.92	0.74	0.83	1.00	0.92
Edinburgh City	1.16	1.38	1.18	0.99	1.57	1.85	1.18	0.99
Falkirk	1.10	1.28	1.14	1.01	1.30	1.03	1.14	1.01
Fife	1.00	1.00	0.97	0.98	1.68	1.54	0.97	0.98
Glasgow City	1.30	1.17	1.22	0.99	1.97	2.00	1.22	0.99
Highland	0.83	0.87	1.00	1.11	1.28	1.02	1.00	1.11
Inverclyde	0.95	1.17	1.16	0.82	0.47	1.24	1.16	0.82
Midlothian	0.87	0.67	0.88	1.00	0.66	0.50	0.88	1.00
Moray	0.81	0.68	0.68	1.03	0.55	0.40	0.68	1.03
North Ayrshire	1.37	1.25	1.25	1.06	1.01	0.94	1.25	1.06
North Lanarkshire	1.41	1.21	1.16	1.00	1.67	1.69	1.16	1.00
Orkney	0.20	0.18	0.15	0.25	0.11	0.12	0.15	0.25
Perth and Kinross	0.93	0.89	1.35	1.03	0.99	0.71	1.35	1.03
Renfrewshire	1.42	1.34	1.05	1.19	1.51	1.20	1.05	1.19
Shetland	0.31	0.29	0.22	0.22	0.15	0.20	0.22	0.22
South Ayrshire	1.16	0.84	1.14	1.24	0.99	0.85	1.14	1.24
South Lanarkshire	1.40	1.23	1.03	1.16	1.76	1.60	1.03	1.16
Stirling	0.76	0.72	0.76	0.92	0.56	0.63	0.76	0.92
West Dunbartonshire	0.85	1.11	1.26	0.94	0.57	0.70	1.26	0.94
West Lothian	1.35	1.37	1.30	0.90	1.01	1.45	1.30	0.90
Western Isles	0.29	0.24	0.27	0.27	0.19	0.19	0.27	0.27

No other weight is applied across all cases in order to compensate/adjust for the unequal probabilities of selection. Strictly speaking, however, a corrective weight should be applied in those cases in which the Multiple Occupancy Indicator (MOI) on the Postcode Address File (PAF) is found to be inaccurate. The

reason for this is that a property-type bias might otherwise be introduced. For example, if tenement properties were consistently found to contain multiple dwellings when the MOI had indicated that they contained just one, each achieved interview at such an address should be given a weight proportional to the actual number of dwellings, to compensate for the reduced probability of selection for each dwelling at that address. All properties within that local authority area should then be weighted back down slightly in order that the actual and weighted sample sizes remain the same.

In practice, the MOI has been found to be inaccurate in only about 2% of cases. The impact of weighting to correct for these would have been negligible so it was decided not to weight by the MOI in order to avoid additional complexity in the weighting scheme for the survey.

Similarly, in theory an additional weight should be applied in cases where a dwelling contains more than one household, only one of which is interviewed, in order to adjust for the lower probability of selection for each of the households in that dwelling. In practice, however, as only a very small number of dwellings were found to contain more than one household, the use of such a weight would make very little difference to the overall results, and it was therefore felt that it was not worthwhile introducing further complication to the weighting calculations.

LA_CS and LA_ALL are calculated in the same way for the Culture and Sport sample and the combined samples.

5.1.2 *Weighting for analysis based on individual (random adult) data*

Using the Postcode Address File produces a sample of households, so for analysis of individual level data it is also necessary to weight the responses of the random adult by the number of adults resident in the household who were eligible for interview.[7] The reason for this is that individuals living in larger households have a lower probability of selection than adults in, for example, single adult households where that one person must be sampled.

As a result of this, the *unweighted* profile of 'random adult' respondents will tend to be skewed towards those sections of the population most likely to live in households with fewer adults (older people and older females in particular) and away from those likely to live in households with larger numbers of adults (younger people). Once the data are weighted by the number of eligible adults in the household, however, one should see the profile correct itself significantly. In most surveys of this kind, however, some under-representation of younger people and males, and over-representation of older people and females, is likely to remain because of the effects of non-response bias. Depending on the extent of the remaining skew, it may be necessary to adopt further corrective measures but this has not been the case so far.

[7] This weight incorporates the local authority weight described earlier. This is necessary for all analyses (whether of households or individuals) if the Scottish population resident in private households is to be represented accurately. The way in which weights are combined is further described later in this section.

Analysis of data based on the random adult also requires a further weight to take account of differences between the number of such interviews completed in each local authority area and the actual adult population of such areas. Like the element of the household data weight which adjusts for differences in fieldwork outcomes by local authority, this is intended not to compensate for unequal probabilities of selection but to ensure that the final profile of 'individual' data correctly reflects the relative populations of the different local authority areas once variations in fieldwork outcomes have been assessed. This is not identical to the weight described for analysis of household data, since variation in response rates for the second part of the interview may have produced a slightly different distribution from that of 'householder' interviews. The weights required for each local authority (which are then multiplied by the number of adults in the household to create the weight for each case, which is then scaled so that the number of weighted cases is the same as the total number of random adult interviews) are summarised below.

Table 5-2: Weights to account for disproportionate sampling and differences in random adult response rates by local authority and quarter, 2007

	Main sample (IND_WT)				Culture and sport (IND_CS)			
	Q1	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**
Aberdeen City	1.45	1.18	1.10	1.10	2.63	1.52	1.10	1.10
Aberdeenshire	0.67	0.84	0.87	1.14	0.99	1.44	0.87	1.14
Angus	0.97	0.99	1.18	1.54	0.72	0.65	1.18	1.54
Argyll and Bute	0.62	0.63	0.72	0.94	0.37	0.57	0.72	0.94
Borders	0.78	1.05	1.08	1.04	0.80	1.06	1.08	1.04
Clackmannanshire	0.50	0.42	0.41	0.42	0.51	0.31	0.41	0.42
Dumfries and Galloway	0.67	0.84	1.07	1.12	0.94	1.09	1.07	1.12
Dundee City	1.44	1.53	1.09	1.06	1.25	1.17	1.09	1.06
East Ayrshire	1.04	0.98	0.81	1.27	0.82	0.65	0.81	1.27
East Dunbartonshire	0.81	1.49	1.39	0.94	0.91	0.80	1.39	0.94
East Lothian	0.92	0.83	1.02	1.33	0.70	0.54	1.02	1.33
East Renfrewshire	0.94	1.12	1.06	1.17	0.75	0.67	1.06	1.17
Edinburgh City	1.22	1.56	1.32	1.07	1.59	2.17	1.32	1.07
Falkirk	1.29	1.39	1.23	1.08	1.37	1.08	1.23	1.08
Fife	1.01	0.99	1.00	1.00	1.78	1.31	1.00	1.00
Glasgow City	1.28	1.13	1.28	1.07	1.74	1.93	1.28	1.07
Highland	0.83	0.85	0.96	1.04	1.36	0.92	0.96	1.04
Inverclyde	0.91	1.03	1.10	0.88	0.44	1.39	1.10	0.88
Midlothian	0.79	0.65	0.91	0.97	0.60	0.50	0.91	0.97
Moray	0.73	0.63	0.65	1.02	0.56	0.47	0.65	1.02
North Ayrshire	1.29	1.24	1.14	0.94	0.93	0.90	1.14	0.94
North Lanarkshire	1.37	1.28	1.18	1.05	1.58	1.48	1.18	1.05
Orkney	0.18	0.15	0.13	0.22	0.09	0.13	0.13	0.22
Perth and Kinross	1.04	0.86	1.47	1.11	1.14	0.74	1.47	1.11
Renfrewshire	1.23	1.15	0.91	0.99	1.38	1.08	0.91	0.99
Shetland	0.31	0.28	0.18	0.19	0.14	0.19	0.18	0.19
South Ayrshire	1.32	0.89	1.26	1.27	1.05	0.83	1.26	1.27
South Lanarkshire	1.73	1.30	1.00	1.20	2.23	1.80	1.00	1.20
Stirling	0.87	0.68	0.76	0.82	0.55	0.63	0.76	0.82
West Dunbartonshire	0.71	1.21	1.28	0.91	0.53	0.82	1.28	0.91
West Lothian	1.48	1.94	1.41	0.99	1.17	1.89	1.41	0.99
Western Isles	0.29	0.20	0.27	0.26	0.19	0.16	0.27	0.26

5.1.3 Weighting for analysis based on the 'random schoolchild'

Data relating to the information collected about a 'random schoolchild' needs to be weighted so that this information will represent correctly the population of schoolchildren resident within households. If not, it

will proportionately over-represent the characteristics and experiences of 'only' children and under-represent those of children from larger families. The weight for the random schoolchild case is created by combining the number of schoolchildren in the household and the relevant local authority weight, and scaling the result so that the number of weighted cases is the same as the total number of random schoolchildren about whom the questions were asked.

5.1.4 Weighting for the selection of a random child

In households with more than one child using some form of childcare, one child is selected randomly by the CAPI script and questions about the use of childcare are asked in relation to that person. This data needs to be weighted to account for the lower probability of each child being selected in households with multiple children. The weight for the random child is created by combining the number of children in the household using childcare and the relevant local authority weight, and scaling the result so that the number of weighted cases is the same as the total number of children about whom the questions were asked.

5.1.5 Weighting for analysis based on the Travel Diary

Examination of the SHS data suggests that significantly fewer interviews take place on Fridays, Saturdays and Sundays than on other days of the week. As differences in the proportions of adults interviewed on each day of the week will affect the Travel Diary data's representativeness of travel patterns for the week as a whole, it was decided to introduce a weight to compensate for this. This simply 'up-weights' interviews carried out on days of the week on which fewer than one-seventh of all interviews have taken place and 'down-weights' those carried out on days on which more than one-seventh of all interviews have been completed.

It is also apparent that the distribution of interviews by the day of the week differs for certain sub-sections of the adult population. For example, disproportionately more adults in full-time employment are interviewed at the weekend (due to their greater availability then), thus yielding an inaccurate picture of the travel patterns of those in full-time employment. The Travel Diary weighting factor is therefore refined to compensate for this.

The weight created for any analysis of the Travel Diary combines the above weighting factors and the existing 'random adult' weights. Further information about the Travel Diary, including a comparison to the National Travel Survey, is available in the Travel Diary User Guide.[8]

5.2 No additional corrective weighting

The weighting scheme for the SHS is intentionally simple. This reflects, in part, a desire to keep the processes of the survey straightforward so that the data can be made available for analysis as quickly as

[8] http://www.scotland.gov.uk/Topics/Statistics/16002/4712

possible. It also reflects the limited extent to which the SHS data differs substantially from comparator data, as shown below. Thus, no additional corrective weighting has ever been applied to the data beyond that required to account for sample design and differential response rates between local authorities.

This aspect of the survey has been subject to review by the Office for National Statistics as part of a major study comparing non-respondents to the SHS with Census data.[9] This study concluded that while comparison with the Census showed some bias in the SHS, this was not substantial, although some corrective weighting would be recommended. Further work to develop a new weighting scheme for the SHS is being undertaken and will be applied to the 2007-2008 and previous datasets.

[9] Freeth, S and Sparks, J (2004) *The Scottish Household Survey: Report of the 2001 Census-linked study of survey non-response*. Full report available at http://www.scotland.gov.uk/topics/statistics/16002/22861

6 Limitations of the data

There are a number of important methodological and data issues that users need to be aware of when using the SHS data.

Like all sample surveys, the SHS can only produce estimates and these estimates are limited by a number of factors.

- Sample coverage – although there are no geographical exclusions to the survey, the sampling frame does not cover the whole population because of a combination of inherent limitations and administrative errors and delays.
- Sampling variability – all samples can differ from the population by chance. This is often referred to as sampling error.
- The number of cases that analysis is based on – estimates based on large samples are more accurate than those based on small samples.
- Bias in the achieved sample – if a sample under-represents sections of the population or if a large proportion of people do not answer some questions, the estimates may differ substantially from the population for reasons that are not a result of chance. For example, in 2007, the unweighted sample of adults is 58% female and even after weighting 55% of the sample is female, but the true figure in the population is only 52%.[10] This is an example of bias caused by young males, in particular, being difficult to contact or refusing to take part in the survey.

The SHS is also limited in the amount of detail it can collect about some topics. For example, it was not designed to provide reliable "economic" statistics (e.g. employment/unemployment rates and average earnings).

The SHS's information about the **economic status** of members of the household reflects the view of the respondent to the "household" part of the interview, and so may not conform to official definitions of employment and unemployment, for example. As a result, the SHS cannot provide estimates of unemployment that are comparable to official statistics of unemployment.

There are several reasons why the SHS data on **income** may not be reliable.

- The SHS only collects information from, or about, the Highest Income Householder and, if there is one, their spouse or partner.

[10] General Register Office for Scotland mid-year population estimates 30 June 2007 available at http://www.gro-scotland.gov.uk/files1/stats/07mype-cahb-t1.xls

- Information is provided "off the top of the head" as part of an interview on many other topics. There is no requirement to refer to pay slips or bank statements to check the figures.
- Some people may not know the correct figure (particularly in the case of the income of a spouse/partner), and may just provide a guess, perhaps based on a level that they remember from some time ago.
- Other interviewees may under-state their income because they do not want to reveal how much they really earn.
- Because about a third of the households in the sample are unwilling or unable to provide income information, values for some or all of the main components of income have to be imputed.[11]

In 2004, researchers commissioned by the Scottish Executive and Communities Scotland compared the income data collected by the SHS and the Scottish House Conditions Survey (SHCS) with the income statistics produced from the Family Resources Survey.[12] Their main conclusions were:

- the SHS (and SHCS) under-estimate total household income, due to collecting only the income of the highest income householder and any spouse/partner
- when households with one adult or two adults who are spouses/partners are compared, there is good agreement between the SHS/SHCS and FRS income distributions for such households
- SHS (and SHCS) greatly under-estimate investment income and interest payments compared to FRS
- uncorrected bias in the SHS (and SHCS) age and sex distributions affects income distributions, particularly for one person households
- overall income from benefits agrees well between the surveys, but the individual benefits may be less accurately classified in the SHS (and SHCS).

As a multi-purpose survey of households, the SHS is not designed to provide the kinds of information about economic activity and household income that can be obtained from more specialised surveys such as the Labour Force Survey and the Family Resources Survey, which have questions and procedures which are designed to obtain much more reliable information on those matters than the SHS can collect. The SHS has questions on such topics *only* for selecting the data for particular groups of people (such as the unemployed or the low-paid) for further analysis, or for use as "background" variables when analysing other topics (such as the means of travel or the frequency of driving).

---

[11] Details of the imputation process can be found in the SHS Annual Report *Scotland's People: results from the 2007 Scottish Household Survey*

[12] Raab, G., MacDonald, C. & Macintyre, C. (2004) *Comparison of income data between surveys of Scottish households*. Research commissioned by Communities Scotland. Copies are available on the SHS website or via the SHS Team.

Although the SHS has a large sample that covers the whole of Scotland, it has some geographical limitations because of the sample sizes in small local authorities and because it is designed to be representative only at national and local authority level. This means:

- users need to be mindful of the sampling errors for analysis but especially when this is based on breakdowns within a single local authority
- it is not appropriate to undertake geographical analysis below local authority level since the sampling techniques used in some local authorities cannot guarantee representativeness in smaller areas.

6.1 Quarterly data for Scotland as a whole

The SHS was designed to provide results which are representative for Scotland as a whole for each quarter of the year. Although based on a large sample (nearly 4,000 households per quarter), they are still subject to sampling errors, so may well fluctuate from one quarter to the next. Therefore, apparent quarter-to-quarter changes should be interpreted cautiously, as they may well be due to sampling variability rather than representing genuine change.

This can be seen if one looks at the apparent quarter-to-quarter changes in some figures which one would expect to change only gradually from one quarter to the next - especially figures which show trends that one would *not* expect to be subject to short-term reversals. The SHS's quarterly Statistical Press Notices[13] provide a set of quarterly tables and charts. Examples of two of these appear on the following pages. The first example shows, quarter-by-quarter since the survey started, the (weighted) percentages of households in the sample with various numbers of cars available for private use; the second shows quarter-by-quarter figures for household tenure.

In both cases, the quarterly charts and tables show the kinds of long-term trends that one would expect (e.g. a gradual increase in two-car households) - but with some apparent "wobbliness" in the lines. Given the nature of car ownership and household tenure, one would not expect sudden short-term departures from the long-term trend (such as a sharp fall in the percentage of homes which are owned outright). However, the survey results sometimes suggest very surprising quarter-to-quarter changes. For example, the table below the first chart shows that the (weighted) percentage of households with 3+ cars appears to vary from quarter to quarter. The cause cannot be any real change in car ownership across Scotland: it must just be sampling variability (the "luck of the draw" regarding which households were included in the sample in each quarter, and which of them agreed to take part in the survey). In the Annual Report's Appendix on confidence intervals and statistical significance, Table A3.1 indicates that the 95% confidence limits for an estimate of 5% based on a sample of 4,000 cases are about +/- 0.8%. The apparent fluctuation in the percentage of households with 3+ cars in the sample in the first three quarters of 2004 is a good illustration of such sampling variability.

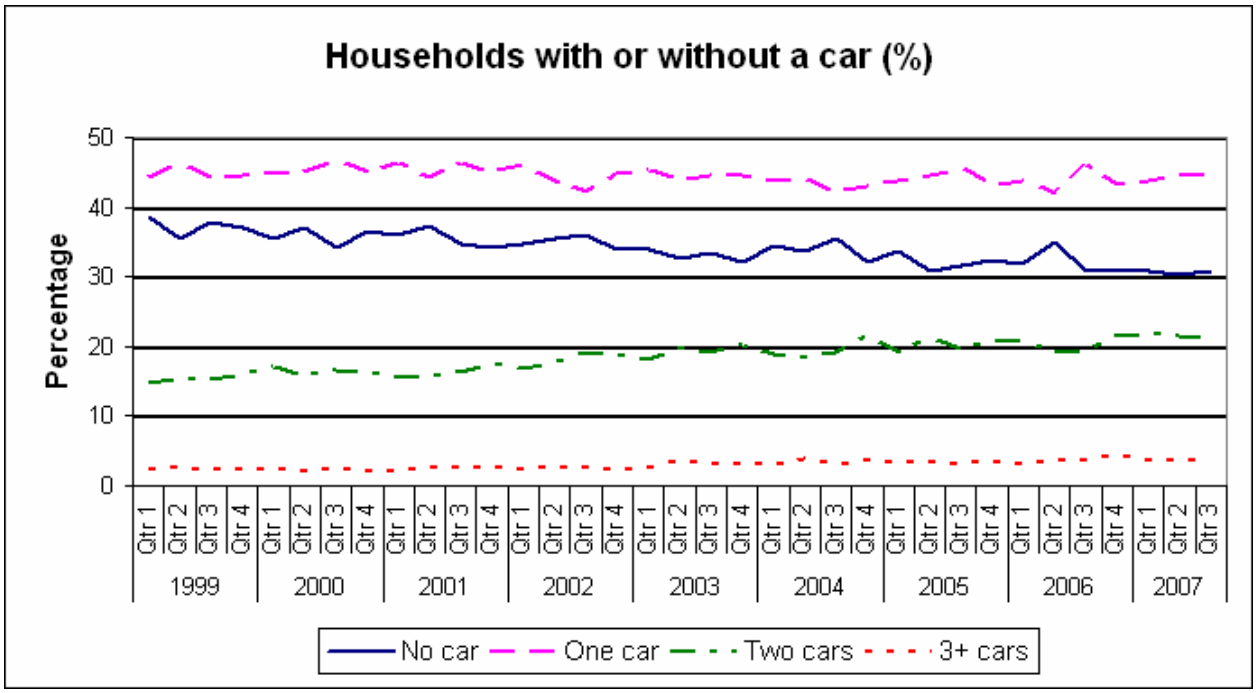
[13] Available at http://www.scotland.gov.uk/topics/statistics/16002/14050

The quarterly charts and tables also cover the following topics:

- rating of the neighbourhood as a place to live
- people who hold a full driving licence
- employed adults who work at or from home
- usual method of travel to work
- adults who make personal use of the internet
- adults who have given up their time to help as an organiser or a volunteer
- whether the household respondent/partner/spouse has a bank/building society account
- households with individuals who need regular help or care.

They can all be found on the SHS Web site: www.scotland.gov.uk/shs under "Publications".

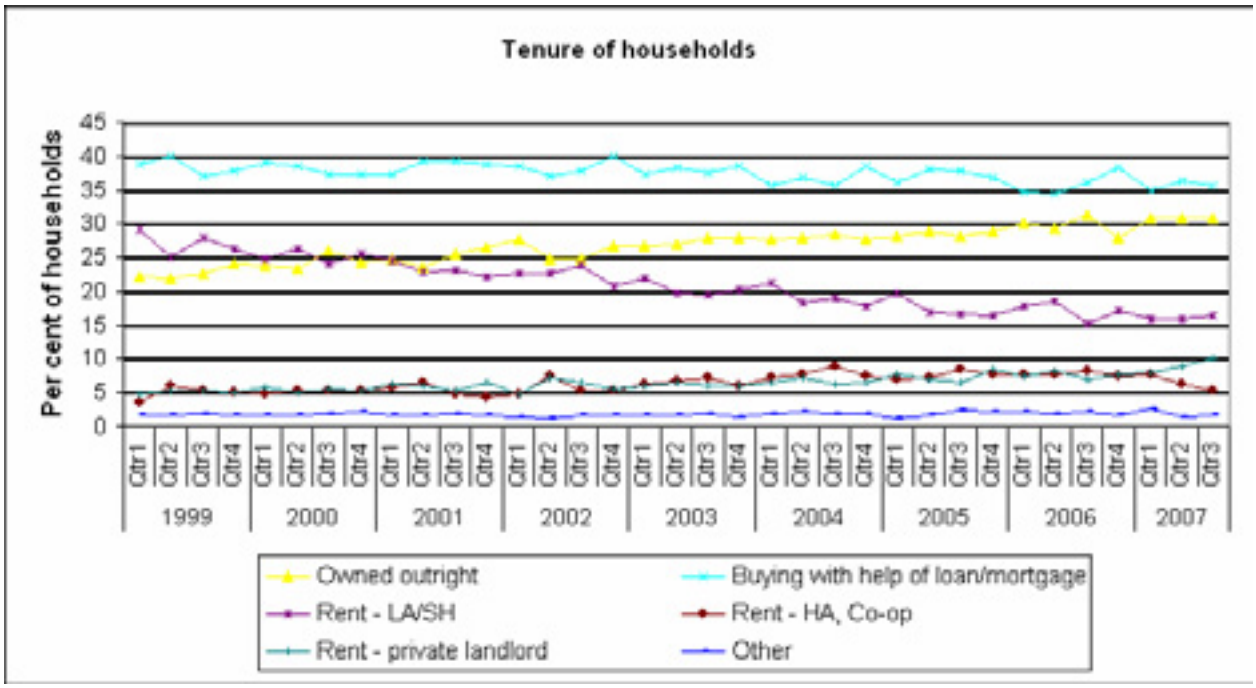
Examples of charts and tables showing quarterly figures



Year	Qtr	None	1	2	3 or more	All households
1999	Qtr 1	38.7	44.3	14.7	2.3	100
	Qtr 2	35.5	46.5	15.5	2.5	100
	Qtr 3	37.9	44.4	15.4	2.4	100
	Qtr 4	37.1	44.5	16.1	2.3	100
2000	Qtr 1	35.5	44.9	17.2	2.4	100
	Qtr 2	37.1	45.0	15.7	2.2	100
	Qtr 3	34.2	46.6	16.5	2.7	100
	Qtr 4	36.4	45.1	16.4	2.0	100
2001	Qtr 1	36.1	46.4	15.5	2.0	100
	Qtr 2	37.2	44.3	15.9	2.6	100
	Qtr 3	34.6	46.5	16.2	2.7	100
	Qtr 4	34.2	45.1	17.7	2.9	100
2002	Qtr 1	34.7	46.0	16.8	2.4	100
	Qtr 2	35.5	43.9	18.0	2.6	100
	Qtr 3	35.9	42.3	19.2	2.6	100

	Qtr 4	33.9	44.9	18.8	2.4	100
2003	Qtr 1	34.0	45.4	18.2	2.5	100
	Qtr 2	32.7	44.0	19.8	3.5	100
	Qtr 3	33.3	44.5	19.1	3.1	100
	Qtr 4	32.1	44.6	20.3	3.0	100
2004	Qtr 1	34.5	43.7	18.8	3.0	100
	Qtr 2	33.7	44.1	18.4	3.9	100
	Qtr 3	35.4	42.2	19.3	3.1	100
	Qtr 4	32.0	43.1	21.5	3.5	100
2005	Qtr 1	33.6	43.9	19.2	3.3	100
	Qtr 2	30.9	44.6	21.2	3.3	100
	Qtr 3	31.6	45.5	19.8	3.1	100
	Qtr 4	32.4	43.3	20.8	3.6	100
2006	Qtr 1	31.9	43.8	21.0	3.2	100
	Qtr 2	35.1	41.9	19.3	3.7	100
	Qtr 3	30.9	46.0	19.4	3.7	100
	Qtr 4	30.7	43.2	21.4	4.7	100
2007	Qtr 1	30.8	43.9	21.7	3.6	100
	Qtr 2	30.3	44.6	21.4	3.7	100
	Qtr 3	30.7	44.5	21.1	3.7	100

Apparent quarter-to-quarter changes in the figures must be interpreted with caution: they could well be the result of sampling variability.



Year	Qtr	Owned outright	Buying with help of loan or mortgage	Rent - LA/SH	Rent - HA, Co-op	Rent - private landlord	Other	All Tenures
1999	Qtr1	22.1	38.7	29.1	3.5	4.9	1.6	100
	Qtr2	21.8	40.0	25.1	6.0	5.2	1.8	100
	Qtr3	22.7	37.1	28.0	5.2	5.2	1.9	100
	Qtr4	24.0	37.8	26.3	5.0	5.1	1.8	100
2000	Qtr1	23.9	38.9	24.9	4.9	5.7	1.7	100
	Qtr2	23.3	38.5	26.2	5.4	5.1	1.6	100
	Qtr3	25.9	37.4	24.0	5.3	5.6	1.9	100
	Qtr4	24.3	37.2	25.6	5.4	5.4	2.1	100
2001	Qtr1	24.7	37.2	24.5	5.8	6.2	1.6	100
	Qtr2	23.6	39.3	22.9	6.6	6.0	1.6	100
	Qtr3	25.6	39.2	23.0	4.8	5.3	2.0	100
	Qtr4	26.4	38.8	22.2	4.4	6.4	1.8	100
2002	Qtr1	27.7	38.5	22.5	4.8	4.9	1.5	100
	Qtr2	24.8	37.0	22.5	7.4	7.2	1.2	100

	Qtr3	24.9	37.8	23.8	5.2	6.6	1.7	100
	Qtr4	26.8	39.9	20.6	5.3	5.5	1.8	100
2003	Qtr1	26.7	37.4	21.8	6.3	6.1	1.7	100
	Qtr2	27.0	38.2	19.8	6.7	6.4	1.8	100
	Qtr3	27.8	37.6	19.6	7.2	5.9	1.9	100
	Qtr4	27.9	38.4	20.2	6.0	5.9	1.5	100
2004	Qtr1	27.6	35.6	21.1	7.2	6.4	2.0	100
	Qtr2	27.8	36.9	18.2	7.7	7.2	2.1	100
	Qtr3	28.3	35.5	19.1	8.9	6.2	2.0	100
	Qtr4	27.7	38.6	17.7	7.5	6.6	2.0	100
2005	Qtr1	28.1	36.1	19.8	7.0	7.6	1.3	100
	Qtr2	28.8	38.1	16.9	7.3	7.0	1.8	100
	Qtr3	28.1	37.8	16.7	8.5	6.6	2.4	100
	Qtr4	28.8	36.8	16.3	7.7	8.4	2.1	100
2006	Qtr1	30.1	34.6	17.8	7.7	7.5	2.2	100
	Qtr2	29.4	34.4	18.5	7.7	8.1	1.9	100
	Qtr3	31.2	36.2	15.2	8.3	6.9	2.2	100
	Qtr4	27.8	38.2	17.0	7.5	7.7	1.8	100
2007	Qtr1	30.9	34.9	15.9	7.7	8.0	2.6	100
	Qtr2	30.8	36.3	16.0	6.3	9.0	1.5	100
	Qtr3	30.9	35.5	16.3	5.3	10.1	1.7	100

Apparent quarter-to-quarter changes in the figures must be interpreted with caution: they could well be the result of sampling variability.

7 Bias and data quality

The issue of bias arises in every survey of the population. There are a number of sources of bias, some of which reflect aspects of the survey design (such as the sampling frame or who is deemed eligible for interview). However, bias is also a reflection of those aspects of fieldwork outcomes mentioned above:

- the quality of survey administration procedures
- whether potential respondents can be found at home at times when interviewers call
- whether they are able to participate in the interview i.e. not restricted by ill health, disability or communication barriers
- the willingness of members of the public to participate in the survey.

A high response rate is generally viewed as one of the key measures of data quality and, all other things being equal, a high response rate and a large sample should ensure accurate estimates. However, to the extent that non-response to the survey is not spread evenly, either geographically or between sub-groups of the population, the resulting bias will limit the accuracy of the survey's estimates. The question of bias is considered by comparing key results from the SHS with comparator data. Since the publication of the 2001 Census, this source is the most accurate comparator for population data and in spite of being a few years behind the current SHS, population measures such as age distribution and household types change little from year-to-year.

7.1.1 Household type, property type, tenure and number of bedrooms

Single adult and large adult households are under-represented, and single pensioner and older smaller households over-represented, when household types in the 2005/2006 SHS are compared with the Census (Table 7-1).

Table 7-1: Comparison of household types in the 2001 Census and the 2007 SHS

	2001 Census	**2007 SHS ***
	% (n=2,192,246)	% (n=17,060)
Single adult	17.9	16.2
Small adult	16.9	17.2
Single parent	5.6	5.6
Small family	13.3	13.5
Large family	7.1	6.6
Large adult	11.2	9.0
Older smaller	13.0	15.9
Single pensioner	15.0	16.1

* SHS data full sample, weighted by local authority size only

As Table 7-2 shows, the sample appears robust in terms of the variables associated with accommodation/property characteristics. Compared with the 2001 (which is six years older than the data in the SHS) there a higher proportion of houses and a lower proportion of flats. Outright ownership appears to be over-represented compared with owning with a mortgage and in total, owner-occupation is over-represented in the SHS compared with rented tenures.

Table 7-2: Comparison of housing variables in the 2001 Census and the 2007 SHS

	2001 Census (n= 2,192,246)	**SHS** (n=13,414)
	%	%
Property type* ‡		
House or bungalow	64	67
Flat, Maisonette or Apartment	35	32
Other	1	1
Tenure* †		
Own outright	23	31
Own with mortgage	39	35
Rent	35	32
Local authority	*22*	*16*
Housing Association/Co-operative	*6*	*6*
Private rented	*7*	*10*
Other	4	2

* SHS main sample, data weighted by local authority size only
‡ includes households in shared dwellings
† Pays part rent and mortgage (shared ownership) included in 'Own with mortgage'

7.1.2 Age and sex profile of the 'random adult' sample

When a single adult is randomly selected within households, the unweighted sample of adults always under-represents those living in multi-adult households, since they have a smaller chance of selection for interview. As Table 7-3 shows, weighting to equalise probabilities of selection generally has the effect of bringing the profile of the 'random adult' sample closer to that of the adult population. The SHS data shown have been weighted both by the number of adults resident in the household and by the local authority weight described in the previous section. These two weights tend to act in the same direction, since those larger local authority areas which are 'weighted up' also tend to be ones with a higher average household size.

Table 7-3: Comparison of weighted and unweighted age and sex profile of 2007 SHS data with 2001 Census estimates

	Census estimates for 2001	SHS random adults Unweighted	SHS random adults weighted*	SHS all adults weighted**
	%	%	%	%
Male				
16 – 24	7.0	3.4	5.0	6.2
25 – 59	29.3	24.4	25.8	27.5
60 plus	11.0	15.0	13.8	13.2
Total	**47.3**	**42.8**	**44.5**	**46.9**
Female				
16 – 24	6.9	4.3	5.4	6.9
25 – 59	30.7	31.2	32.3	30.1
60 plus	15.1	21.7	17.8	16.1
Total	**52.7**	**57.2**	**55.5**	**53.1**
All adults		(n=15,630)	(n=15,630)	(n=31,145)
16 – 24	13.9	7.7	10.3	13.0
25 – 59	60.1	55.6	58.1	57.7
60 plus	26.1	36.6	31.5	29.3
Total	**100.0**	**100.0**	**100.0**	**100.0**

* SHS full sample. Weighted by number of adults and local authority size

** Weighted by local authority size

However, even after this design weighting has been applied, the weighted random adult sample for 2007 still does not match the profile of the adult population suggested by the Census estimates with, as expected, under-representation of younger people in general and 16-24 year olds in particular. Consequently, older people are over-represented in the random adult sample.

7.1.3 Driving and transport

In relation to driving and transport, the survey results also look broadly in line with what one might expect from other sources such as the National Travel Survey and the differences which exist are, again, comfortably within the confidence intervals associated with the two surveys. Mode of travel comparisons with other sources are less conclusive, though methodological or classification differences may be playing a part here.

Table 7-4: Comparison of key variables relating to driving and transport

	2004/2005 National Travel Survey (n= 1,594 households)	2001 Census (n= 2,192,246 households)	2007 SHS
	%	%	%
% adults with full driving licences			(n=15,515) ***
Males aged 17 +	78		77
Females aged 17 +	58		60
Total	67		68
Mode of travel to school†			(n=2,519) **
Walking	54	51	54
Car	23	20	22
Bus	20	25	21
Other	3	3	4
% households with regular use of cars††			(n=17,060) ***
No car	31	34	31
1 car	43	43	44
2 or more cars	22	22	25

	2001 Census	2007 SHS*
	%	%
Mode of travel to work or study		
incl. those who work at / from home		(n=6,271)
Car or motorcycle	64	62
Bus, minibus, coach or taxi	13	11
Train, underground	3	3
Other means (e.g. walking and cycling)	14	13
Working at or from home	6	11

	2006	2007	2007
	Labour Force Survey, Oct-Dec 2006 and Oct-Dec 2007		SHS*
	%	%	%
Mode of travel to work			(n=6,625)
excl. those who work at / from home			
Car, van, minibus, works van	69	69	69
Bicycle	1	2	2
Bus, coach, private bus	12	12	11
Rail (incl Underground)	4	4	3
Walk	11	11	13
Other (incl Taxi)	2	2	2

* SHS full sample, weighted by number of adults and local authority size

** SHS main sample, weighted by local authority size and number of school children in household

*** SHS full sample, weighted by local authority size only

† Census figures are for method of travel to place of study, age 5-17

†† the National Travel Survey figures relate to 2004 alone, and were produced from the combined Scottish results of the NTS, the General Household Survey and the Expenditure and Food Survey. The Census figures relate to cars and vans available for private use.

7.1.4 Urban/rural classification

Analysis of the Scottish Household Survey makes extensive use of the Scottish Government's classification of areas into different degrees of urbanity and rurality. This classifies settlements according to their size and for settlements with a population of less than 10,000, their proximity to a settlement with a population of 10,000 or more.[14]

Table 7-5 compares the urban/rural classification of the SHS sample for 2007 with the profile of all addresses sampled for the survey, the profile of eligible addresses and participating households. This shows that the addresses sampled in 2007 (column 2) under-represent urban areas and over-represent rural areas but when disproportionate sampling is taken into account by weighting, the profile matches the population.

Table 7-5: Comparison of all Scottish households, all sampled households, all eligible households and participating households by urban rural classification

	All Scottish addresses*	All sampled addresses (unweighted)	All sampled addresses**	All eligible households**	All participating households***
Large urban areas	40	36	40	40	40
Other urban	30	24	31	30	30
Small accessible towns	9	8	8	9	9
Small remote towns	4	4	4	4	4
Accessible rural	11	17	11	11	11
Remote rural	6	11	6	6	6

* Weighted by number households within each unit postcode

** Weighted to reflect disproportionate sampling across local authorities

*** Weighted to reflect disproportionate sampling and non-response across local authorities

Comparison of the households at which SHS interviews were achieved and the classification of all households sampled at a local authority level shows that there is a good match between the two within local authorities. Table 7-6 compares the proportion of households in each local authority in each type of area.

[14] Full details available in Scottish Executive (2006) *Scottish Executive Urban Rural Classification 2005-2006* available at http://www.scotland.gov.uk/Publications/2006/07/31114822/0

Table 7-6: Urban rural classification of eligible addresses and participating households

Row percentages, all eligible addresses shown in bold, participating households in plain text

	Large urban areas	Other urban areas	Accessible small towns	Remote small towns	Accessible rural	Remote rural	Total
Aberdeen City	**93.7**		**4.0**		**2.3**		**100.0**
Aberdeen City	92.0		5.1		2.9		100.0
Aberdeenshire		**27.9**	**6.1**	**13.5**	**37.2**	**15.3**	**100.0**
Aberdeenshire		27.4	5.4	13.6	36.7	16.8	100.0
Angus	**6.6**	**57.3**	**14.0**		**21.5**	**0.5**	**100.0**
Angus	7.1	58.9	11.8		21.6	0.5	100.0
Argyll and Bute		**16.1**		**33.8**	**3.2**	**46.9**	**100.0**
Argyll and Bute		14.7		32.0	4.2	49.0	100.0
Scottish Borders		**28.7**	**13.4**	**5.5**	**40.3**	**12.1**	**100.0**
Scottish Borders		28.7	12.1	5.4	40.8	13.1	100.0
Clackmannanshire		**56.6**	**28.9**		**14.5**		**100.0**
Clackmannanshire		55.8	25.2		19.0		100.0
Dumfries and Galloway		**28.3**	**15.3**	**9.0**	**24.2**	**23.2**	**100.0**
Dumfries and Galloway		27.6	15.1	9.5	24.9	22.9	100.0
Dundee City	**99.9**				**0.1**		**100.0**
Dundee City	100.0						100.0
East Ayrshire		**37.7**	**40.6**		**16.0**	**5.8**	**100.0**
East Ayrshire		35.8	42.5		16.3	5.4	100.0
East Dunbartonshire	**59.5**	**28.6**	**7.2**		**4.7**		**100.0**
East Dunbartonshire	61.1	28.2	6.9		3.8		100.0
East Lothian	**27.7**		**26.0**	**15.3**	**13.8**	**17.2**	**100.0**
East Lothian	26.1		23.1	16.4	13.4	21.1	100.0
East Renfrewshire	**85.6**		**11.4**		**3.0**		**100.0**
East Renfrewshire	84.2		10.9		4.9		100.0
Edinburgh City	**96.5**		**2.4**		**1.1**		**100.0**
Edinburgh City	95.9		2.8		1.3		100.0
Eilean Siar				**25.0**		**75.0**	**100.0**
Eilean Siar				25.3		74.7	100.0
Falkirk		**89.4**			**10.6**		**100.0**
Falkirk		89.9			10.1		100.0
Fife		**67.9**	**14.0**		**18.0**		**100.0**
Fife		67.5	14.1		18.4		100.0
Glasgow City	**99.9**				**0.1**		**100.0**
Glasgow City	99.9				0.1		100.0
Highland		**23.9**	**6.0**	**23.3**	**12.2**	**34.5**	**100.0**
Highland		19.2	5.6	22.7	13.7	38.8	100.0
Inverclyde		**89.0**	**8.4**		**2.6**		**100.0**
Inverclyde		88.6	8.1		2.9	0.4	100.0
Midlothian		**67.0**	**14.4**		**18.6**		**100.0**
Midlothian		67.7	15.1		17.1		100.0
Moray		**27.6**	**22.3**	**17.6**	**25.8**	**6.7**	**100.0**
Moray		26.9	19.1	19.8	27.6	6.7	100.0
North Ayrshire		**75.2**	**18.1**		**3.2**	**3.5**	**100.0**
North Ayrshire		71.7	20.5		3.1	4.6	100.0
North Lanarkshire	**67.4**	**15.2**	**10.8**		**6.6**		**100.0**
North Lanarkshire	65.3	14.9	12.4		7.4		100.0
Orkney				**33.0**		**67.0**	**100.0**
Orkney				30.3		69.7	100.0
Perth and Kinross	**3.0**	**33.5**	**13.4**	**14.6**	**30.1**	**5.5**	**100.0**
Perth and Kinross	3.4	27.2	12.4	15.6	34.9	6.4	100.0
Renfrewshire	**78.0**	**9.7**	**8.2**		**4.1**		**100.0**
Renfrewshire	78.9	9.5	7.6		4.0		100.0
Shetland				**29.9**		**70.1**	**100.0**
Shetland				25.7		74.3	100.0
South Ayrshire		**68.8**	**2.4**	**3.0**	**23.2**	**2.6**	**100.0**
South Ayrshire		68.6	2.1	3.5	22.6	3.2	100.0
South Lanarkshire	**23.3**	**56.7**	**9.5**		**9.3**	**1.2**	**100.0**
South Lanarkshire	20.8	56.9	10.6		10.6	1.1	100.0
Stirling		**58.1**	**12.6**		**24.2**	**5.2**	**100.0**
Stirling		57.2	12.7		23.9	6.2	100.0
West Dunbartonshire	**49.8**	**49.4**			**0.9**		**100.0**
West Dunbartonshire	43.8	54.9			1.3		100.0
West Lothian		**82.4**	**8.7**		**8.9**		**100.0**
West Lothian		81.1	8.8		10.1		100.0
Scotland	**40.4**	**30.6**	**8.5**	**4.0**	**10.8**	**5.8**	**100.0**
Scotland	39.9	30.0	8.5	4.0	11.3	6.3	100.0

Rows may not always add to 100% because of rounding.

7.1.5 Economic activity

One area where the results of the SHS indicate the potential for significant differences from other sources is in relation to indicators of economic activity. As the following table shows, the most recent results from the Labour Force Survey (LFS) in Scotland suggest that the SHS may be under-representing people in employment, and over-representing the economically inactive. It should be emphasised, however, that the information from the SHS shown here is based on the respondent's own classification of their economic activity (collected at the start of the interview)[15], rather than on the full International Labour Organisation definition, which is not classified by the respondent and is the basis for official estimates of unemployment. The SHS is not an official source of statistics on employment (see above on limitations of the data).

Table 7-7: Comparison of economic activity variables among adults of working age

	2007 Annual Population Survey	**2007 SHS ***
	%	%
Males	(n=14,991)	(n=4,923)
Employed	79.1	74.6
Unemployed	4.3	5.2
Economically inactive	16.6	20.2
Females	(n=15,232)	(n=5,518)
Employed	72.7	67.3
Unemployed	3.4	3.2
Economically inactive	23.9	29.4
All adults	(n=30,223)	(n=10,441)
Employed	76.0	70.8
Unemployed	3.9	4.2
Economically inactive	20.1	25.0

* weighted by number of adults and local authority size

Figures in this table have been calculated using all working age people as the denominator, headline unemployment statistics are not calculated on this basis. Annual Population Survey data are sourced from quarterly Labour Force Survey data and the annual Labour Force Survey boost data.

[15] This further complicated by the fact that where the household respondent is not subsequently selected as the random adult, the classification is given by the household respondent and attributed to the random adult.

8 Survey design factors and complex standard errors

Data collected in surveys are always an estimate of the true proportions in the population. The accuracy of these estimates – the sampling error – can be calculated for any estimate in the survey using information about the proportion of people giving the response and the number of people in the sample (or sub-sample). The sampling error can be expressed as a 'confidence interval', which can be added to and subtracted from the survey estimate to give a range within which it is fairly certain that the true value lies.

Since the SHS is not a simple random sample (SRS) design, the confidence intervals need to take account of the impact of clustering and stratification. The SHS, therefore, has what is known as a 'complex standard error'. While for some variables the design of the sample improves the precision of the survey estimates compared with a simple random sample, the overall effect of the survey design is to reduce the precision of the estimates. The relationship between the complex standard error and the theoretical simple random sample standard error for a sample of the same size is summarised in the 'design factor'.

The Taylor Expansion Method was used to calculate the complex standard errors for a series of results in the study. This is a well-established technique for working through the effects of stratification and clustering. As can be seen from Table 8-1, these ranged from 0.92 to 1.18. The overall average is 1.22, but that should not be taken as a 'typical' value, given the distribution of values across different variables. However, it suggests that using a value of 1.2 as a 'rule of thumb' for adjusting the standard errors of the survey data would account for the design factors associated with most variables in the survey.

The 95% confidence intervals shown are based on complex standard errors.

Table 8-1: Design factors and confidence intervals for key variables in 2007 data

		95% Confidence interval		SRS error for the same size of sample	SHS Complex Standard Error	Design Factor
	Estimate	Lower	Upper			
Tenure						
Owner-occupied	66.1	65.2	67.0	0.41	0.46	1.12
Social-rented Sector	22.2	21.4	23.1	0.36	0.41	1.15
Privately rented	9.6	9.1	10.2	0.25	0.28	1.09
Property type						
House	67.3	66.4	68.2	0.41	0.48	1.18
Flat/maisonette	31.8	30.9	32.8	0.40	0.47	1.17
Economic status of working age adults						
Full time employee	48.2	47.0	49.4	0.55	0.62	1.13
Part time employee	13.5	12.6	14.3	0.38	0.42	1.11
Self-employed	7.3	6.6	7.9	0.29	0.33	1.16
Unemployed	4.1	3.6	4.5	0.22	0.24	1.10
HIH or partner has a bank/ building society account	92.3	91.7	92.8	0.25	0.28	1.11
Marital status of all adults						
Married/cohabiting	59.9	58.9	60.9	0.44	0.5	1.14
Separated/divorced	8.3	7.8	8.8	0.25	0.24	0.96
Single/never married	22.7	21.9	23.6	0.38	0.44	1.17
Widowed	9.0	8.6	9.5	0.26	0.24	0.92
Access to the internet	56.8	55.8	57.7	0.47	0.5	1.08
Travel to work in a car	67.6	66.3	68.9	0.65	0.68	1.04
Require regular care or help	11.8	10.7	12.9	0.55	0.57	1.04
Reporting long-standing illness, disability or health problem	25.1	24.2	26	0.39	0.44	1.12
Rating of area as a place to live						
Very/fairly good	93.0	92.4	93.5	0.25	0.29	1.16
Not good	7.0	6.5	7.6	0.25	0.29	1.16
Smoking Status of all adults						
Smokes	24.7	23.7	25.6	0.42	0.48	1.15
Does not smoke	75.2	74.2	76.1	0.42	0.49	1.15

HIH = Highest income householder

SCOTTISH GOVERNMENT STATISTICIAN GROUP

OUR AIM

To provide relevant and reliable information, analysis and advice that meet the needs of government, business and the people of Scotland.

OBJECTIVES

1. To produce statistics and analysis relevant to user needs by
   - Developing our understanding of customer requirements to ensure statistics are kept relevant and analysis is well targeted;
   - Developing the range of statistics and analysis we produce;
   - Where practicable improving timeliness;
   - Providing more statistics disaggregated by age, gender and ethnicity;
   - Developing more data for small areas through the Neighbourhood Statistics project;
   - Contributing to production of comparable statistics across the UK and internationally.

2. To ensure effective use of our statistics by
   - Contributing more directly to policy processes inside and where possible outside government;
   - Improving access to and presentation of data and analysis;
   - Improving the advice provided on statistics.

3. To work effectively with users and providers by
   - Maintaining arrangements to consult and involve users and providers;
   - Involving users and providers in planning developments in outputs and processes;
   - Minimising the burden on data providers through dropping or streamlining collections as appropriate, to ensure the benefits of the information justify the costs of collection.

4. To develop the quality of statistics by
   - Assuring and improving quality as an integral part of data collection and analysis and through regular reviews in line with National Statistics quality strategy;
   - Developing statistical methods, systems and classifications;
   - Working with the rest of the Government Statistical Service to develop joint approaches/solutions where appropriate.

5. To assure the integrity of statistics by
   - Maintaining and promoting integrity through implementation of the National Statistics Code of Practice and related protocols;
   - Safeguarding the confidentiality of data subjects.

6. To ensure the efficient and effective delivery of statistics products and services by
   - Making best use of all sources including administrative sources;
   - Working with other analysts to maximise the contribution of our own and other analysts' work;
   - Ensuring value for money;
   - Making best use of Information and Communications Technology;
   - Ensuring effective communication within the Statistician Group.

7. To develop our workforce and competences
   - Ensuring recruitment of staff with the necessary skills and potential;
   - Ensuring development of expertise amongst existing staff;
   - Promoting and upholding the standards of the statistics profession.

This is a National Statistics publication

"This is a National Statistics publication. It has been produced to high professional standards set out in the National Statistics Code of Practice Protocol - see **www.statistics.gov.uk/about_ns/cop/default.asp**

These statistics undergo regular quality assurance reviews to ensure that they meet customer needs. They are produced free from any political interference."

Details of pre-release access will be provided in the Scottish Government Statistics Website under 'Forthcoming Releases'

Correspondence and enquiries

Enquiries on the SHS Methodology and fieldwork outcomes 2007
should be addressed to:

The SHS Team
Housing and Household Surveys
Communities Analytical Services
1F(N) Victoria Quay
Edinburgh EH6 6QQ
Telephone: 0131 244 8420; Fax: 0131 244 7573
e-mail: shs@scotland.gsi.gov.uk

General enquiries on Scottish Government statistics can be addressed to:

Office of the Chief Statistician
Scottish Government
3 Floor West Rear
St Andrews House
EDINBURGH EH1 3DG
Telephone: (0131) 244 0442; Fax: (0131) 244 2223
e-mail: statistics.enquiries@scotland.gsi.gov.uk

Advice on specific areas of Scottish Government statistical work can be obtained from staff at the telephone numbers given below:

Scottish Government Statistics contacts

Agricultural census and labour force	(0131) 244 6150
Business	(0141) 242 5446
Community Care	(0131) 244 3777
Courts and law	(0131) 244 2227
Environment	(0131) 244 0445
Equality	(0131) 244 0324
Fisheries	(0131) 244 6441
Further and Higher Education	(0141) 242 0273
Health	(0131) 244 3432
Housing	(0131) 244 7236
Income, Tax and Benefits	(0131) 244 2583
Labour market	(0141) 242 5446
Local government finance	(0131) 244 7033
Planning	(0131) 244 0439
Prisons	(0131) 244 2147
Recorded crime	(0131) 244 2635
Schools – pupils and teachers	(0131) 244 1689
Schools – qualifications	(0131) 244 0315
Scottish Government personnel	(0131) 244 3926
Scottish Neighbourhood Statistics	(0131) 244 0442
The Economy	(0131) 244 2234
Transport	(0131) 244 7255

Other contacts for Scottish statistics

Forestry Commission	(0131) 314 6337
General Register Office for Scotland - Vital statistics and publications - Population statistics, census statistics or digital boundary products	(0131) 314 4243 (0131) 314 4254
The Scottish Funding Councils for Higher and Further Education	(0131) 313 6575

For **general enquiries about National Statistics** in the United Kingdom Government contact the National Statistics Public Enquiry Service on

020 7533 5888
minicom: 01633 812399
Email: info@statistics.gov.uk
Fax: 01633 652747
Letters: room DG/18, 1 Drummond Gate,
LONDON SW1V 2QQ

You can also find National Statistics on the internet - go to **www.statistics.gov.uk**

If you would like to be consulted about new or existing statistical collections or receive notification of forthcoming statistical publications, please register your interest on the Scottish Government ScotStat website at **www.scotland.gov.uk/Topics/Statistics/scotstat**

Current contact points, e-mail addresses and the publications listed below as well as a range of other statistical publications can be found on the Scottish Government website at **www.scotland.gov.uk/stats**

Further information on the General Register Office for Scotland is available on the website **www.gro-scotland.gov.uk**

Most recent Statistical Publications relating to the Scottish Household Survey

Ref no.	Title	Last published	Price
	Scotland's People: Annual Report: Results from 2007 Scottish Household Survey	August 2008	
	Scottish Household Survey: Questionnaire: January 2006 to December 2008	August 2008	
	Headline Results from the 2007 Scottish Household Survey	June 2008	
	Scotland's People: Results from the 2005/2006 Scottish Household Survey	August 2007	
TRN/2008/1	Scottish Household Survey Travel Diary 2005/2006	April 2008	
TRN/2007/6	Transport across Scotland in 2005 and 2006: some Scottish Household Survey results for parts of Scotland	December 2007	£2.00
Trn/2007/5	Household Transport in 2006: some Scottish Household Survey results	October 2007	£2.00

Publications with no price indicated are available free from the SHS Team at the address above or on the SHS website www.Scotland.gov.uk/SHS

Complaints and suggestions

If you are not satisfied with our service, please write to the Chief Statistician, Mr Rob Wishart, 3R.01, St Andrews House, Edinburgh, EH1 3DG, Telephone: (0131) 244 0302, e-mail rob.wishart@scotland.gsi.gov.uk. We also welcome any comments or suggestions that would help us to improve our standards of service.

ISBN 978 0 7559 1796 9 (web only publication)