Scottish Household Survey - Background information

A.1　The Scottish Household Survey (SHS) started in February 1999.　Its principal purpose is to collect information to inform policy on Transport, Communities and Local Government, but other topics are covered, such as household composition, amenities, employment or unemployment, income, assets and savings, credit and debt, health, disabilities and care, and other topics.　The SHS provides the first representative Scottish data on many subjects, such as access to the Internet, daily travel patterns, etc.

A.2　Where appropriate, the SHS uses the harmonised concepts and questions for government social surveys which have been developed by the Government Statistical Service, to facilitate comparison with the results of other government surveys.　However, differences in sampling and survey methods mean that SHS results will differ from those of other surveys.　The SHS is *not* designed to produce statistics on unemployment or income: it collects such information *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.

A.3　The SHS is intended to be a survey of private households.　For the purposes of the survey, a household is defined as one person or a group of people living in accommodation as their only or main residence and *either* sharing at least one meal a day *or* sharing the living accommodation.　A student's term-time address is taken as his/her main residence, in order that they are counted where they live for most of the year.

A.4　The sample was drawn from the Small User file of the Postcode Address File (PAF), which is a listing of all active address points maintained by the Post Office.　The Small User file excludes addresses where an average of more than 25 items of post is delivered per day. Blocks of flats etc, which have several dwellings at the same address, are *not* excluded from the Small User file: in such cases, the file's Multiple Occupancy Indicator is used to count each dwelling separately for the selection of the sample.

A.5　People in certain types of accommodation (such as nurses' homes, student halls of residence etc.) will be excluded from the SHS unless the accommodation is listed on the Small User file of the PAF and it represents the sole or main residence of the people concerned.　People living in bed and breakfast accommodation may be included, *if* it is listed in the Small User file of the PAF and if it is their sole or main residence.　Prisons, hospitals and military bases are excluded.

Interviewing, response rates and weighting

A.6　The survey interviews are carried out in respondents' homes using Computer Aided Personal Interviewing (CAPI). Each interview has two parts. The first part is carried out with the Highest Income Householder or their spouse or partner. This collects mainly factual information about the composition and characteristics of the household.　Some questions are asked in respect of each household member. The second part is with a randomly-chosen adult (aged 16+) member of the household. This focuses on individual attitudes and behaviours.

A.7　The data are weighted to take account of the unequal probabilities of selection inherent in the sample design: the over-sampling (relative to their numbers of households) of the Councils with smaller populations, in order to obtain a minimum number of interviews in each Council; and the under-sampling (relative to their share of the adult

population) of adults living in multi-adult households, because only one random adult is interviewed in each household.

A.8   In keeping with the main SHS, **these results use an improved weighting system for all years which better accounts for non response bias**. This was introduced in 2008 meaning time series figures will be the same as published last year but may differ slightly for years prior to this although the main trends are mostly not affected.

A.9   Totals may appear to differ slightly from the apparent sums of their component parts, in cases where they have been calculated by adding up the unrounded values of the components and then rounding each figure independently. Similarly, percentages may appear not to sum to 100 per cent.

A.10   In tables that analyse the results of questions for which multiple answers were allowed, the percentages may total more than 100 per cent.

A.11   The underlying sample numbers shown in different tables may not be the same. In some cases, this is because the tables relate to different populations. In addition some questions were not asked of all respondents as they only applied in certain circumstances (e.g. questions about children would not be asked in a household without any children). In some cases, the bases differ because some people were unable to, or did not want to, answer certain questions (e.g. some households did not wish to provide details of their income).

Highest Income Householder

A.12   This is the household reference person for the first part of the interview. This must be a person in whose name the accommodation is owned or rented, or who is otherwise responsible for the accommodation. In households with joint householders, the person with the highest income is taken as the household reference person. If householders have exactly the same income, the older is taken as the household reference person.

Adult

A.13   For the purposes of the SHS, an adult is someone who was aged 16 or over at the time of the interview; a *child* is someone who was aged 15 or under.

Household types

- *Single pensioner* household consists of one adult of pensionable age (60+ for women, and 65+ for men) and no children
- *Single parent* household contains an adult and one or more children.
- *Single adult* household consists of an adult of non-pensionable age and no children.
- *Older smaller* household contains *either* (a) an adult of non-pensionable age and an adult of pensionable age and *no* children *or* (b) two adults of pensionable age and *no* children.
- *Large adult* household has three or more adults and *no* children under 16.
- *Small adult* household contains two adults of non-pensionable age and *no* children.
- *Large family* household consists of *either* (a) two adults and three or more children *or* (b) three or more adults and one or more children.

- **Small family** households consist of two adults and one or two children.

Annual net household income

This is the total annual *net* income (i.e. after taxation and other deductions) from employment, benefits and other sources, which is brought into the household by the highest income householder and/or their spouse or partner. This includes any contribution to household finances made by other household members. Due to refusals or don't knows, full information for the main components of household income was not collected from all households. Subsequently, SHS contractors impute the missing components of income for almost all of these households, using information that was obtained from other households that appeared similar.

The Scottish Index of Multiple Deprivation (SIMD)

A.14   The Scottish Index of Multiple Deprivation (SIMD) is used to rank the data zones used for the production of Scottish Neighbourhood Statistics in order of deprivation. More information can be found at the SIMD website (**http://www.scotland.gov.uk/simd**).

A.15   Households in the SHS sample have been allocated the SIMD value of the data zone that contains the postcode of the residence. In the small number of cases where a postcode is split between more than one data zone, the SIMD value used is that of the data zone into which the largest number of dwellings in that postcode falls. The SIMD values have further been assigned to one of 5 quintiles, with quintile 1 containing the most deprived 20 per cent of data zones in Scotland, and quintile 5 the least deprived 20 per cent.

The SHS urban/rural classification

A.16   The urban/rural classification is based on settlement sizes and (for the less-populated areas) the estimated time that would be taken to drive to a settlement with a population of 10,000 or more. The classification is based on postcodes. Six categories were then defined:

- **Large urban areas** - settlements with populations of 125,000 or more.
- **Other urban areas** - other settlements of population 10,000 or more.
- **Accessible small towns** - settlements of between 3,000 and 9,999 people, which are within 30 minutes drive of a settlement of 10,000+ people
- **Remote small towns** - settlements of between 3,000 and 9,999 people, which are *not* within 30 minutes drive of a settlement of 10,000+ people
- **Accessible rural areas** - settlements of less than 3,000 people, which are within 30 minutes drive of a settlement of 10,000+ people
- **Remote rural areas** - settlements of less than 3,000 people, which are *not* within 30 minutes drive of a settlement of 10,000+ people

A.17   The urban/rural classification used for the SHS data is based on the Settlement file maintained by the General Register Office for Scotland (GROS).

Sampling variability and confidence limits

A.18   Although the SHS's sample is chosen at random, the people who take part in the survey will not necessarily be a representative cross-section of the people of Scotland.

Purely by chance, the sample could include disproportionate numbers of certain types of people, in which case the survey's results would be affected.

A.19 The likely extent of sampling variability can be quantified, by calculating the standard error associated with the estimate of a quantity produced from a random sample. Statistical sampling theory states that, on average only about one sample in three would produce an estimate that differed from the (unknown) true value of that quantity by more than one standard error; only about one sample in twenty would produce an estimate that differed from the true value by more than two standard errors; only about one sample in 400 would produce an estimate that differed from the true value by more than three standard errors. By convention, the 95 per cent confidence interval for a quantity is defined as the estimate plus or minus about twice the standard error (from sampling theory, the interval is plus or minus 1.96 times the standard error), because there is only a 5 per cent chance (on average) that a sample would produce an estimate that differs from the true value of that quantity by more than this amount.

A.20 Table 36 shows the 95 per cent confidence limits for estimates of a range of percentages calculated from sub-samples of a range of sizes (NB: the confidence limits for estimates of *x* per cent and for *(100-x)* per cent are the same). The table was produced in the same way as the tables of 95 per cent confidence limits in the Annual Report volumes of *Scotland's People,* but has a more detailed breakdown of the smaller sample sizes.

A.21 The interpretation of an entry in Table 41 is best explained by an example:

- The value in the cell at the intersection of the 45 per cent or 55 per cent column and the 800 row is 4.5
- This means that the 95 per cent confidence limits for an estimate of 55 per cent which is produced from a sub-sample of 800 are +/- 4.5 percentage-points
- The 95 per cent confidence interval for the estimate is 55 per cent +/- 4.1 percentage-points (i.e. from about 50.5 per cent to around 59.5 per cent, assuming that the value of the estimate is 55.0 per cent)

A.22 As the survey's estimates may be affected by sampling errors, apparent differences of a few percentage points between the figures for two sub-groups of the population may not be significant: it could be that the true values for the two sub-groups are similar, but the random selection of households for the survey has, by chance, produced a sample which gives a high estimate for one sub-group and a low estimate for the other.

A.23 One way of assessing significance at the 5% level involves comparing the difference with the 95 per cent confidence limits for the two estimates. Suppose that these are +/- 3.0 percentage-points and +/- 4.0 percentage-points, respectively. Clearly a difference which is *less* than the magnitude of the largest limit (4.0 percentage-points) is *not* significant; and a difference which is *greater* than the *sum* of the magnitudes of the limits (3.0 percentage-points + 4.0 percentage-points = 7.0 percentage-points) *is* significant. Statistical sampling theory suggests that a difference whose magnitude is between these values is significant *if* it is greater than the square root of the sum of the squares of the magnitudes of the limits for the two estimates - in this case, $(3.0^2 + 4.0^2)^{0.5}$=5.0. So, in this case, a 5.0 percentage-point difference would be considered statistically significant (at the conventional 5% level). However, one may well find some apparently significant results that are actually just the result of sampling variability, having arisen by chance.

A.24   The above information relates only to sampling variability. The survey's results could also be affected by non-contact/non-response bias: the characteristics of the people who should have been in the survey but who could not be contacted, or who refused to take part, could differ markedly from those of the people who were interviewed. If that is the case, the SHS results will not be representative of the whole population. Without knowing the true values (for the population as a whole) of some quantities, one cannot be sure about the extent of any such biases in the SHS. However, comparison of SHS results with information from other sources suggests that they are broadly representative of the overall Scottish population, and therefore that any non-contact or non-response biases are not large overall. The *Fieldwork Outcomes* and *Methodology* volumes of *Scotland's People* provide more information on these matters.

Published results, and anonymised data

A.25   SHS results are also included in other Scottish Government publications, such as
- *Scottish Transport Statistics*
- *Main Transport Trends*
- *Scottish Household Survey Travel Diary results*
- *Bus & Coach Statistics*
- *SHS Annual Report – Scotland's People*
- *Local Authority tables – available as web tables*

A.26   These publications are available on the Scottish Government Transport Statistics website at www.scotland.gov.uk/transtat/latest.

A.27   Anonymised copies of the survey data are deposited at the UK Data Archive.

Enquiries and further information

A.28   General enquiries about the SHS should be addressed to the survey's Project Manager:

> SHS Project Manager
> Communities Analytical Services
> Scottish Government
> Victoria Quay
> Edinburgh, EH6 6QQ
>
> Tel: 0131 244 8420
> Fax: 0131 244 7573
> E-mail: **shs@scotland.gsi.gov.uk**

A.29   Enquiries about the statistics in this bulletin should be addressed to:

> Transport Statistics
> Transport Scotland
> Scottish Government
> Victoria Quay
> Edinburgh, EH6 6QQ
>
> Tel: 0131 244 1457
> E-mail: **transtat@transportscotland.gsi.gov.uk**

A.30   Further information about the survey can be found on the SHS *website* at **http://www.scotland.gov.uk/shs**

A.31   This website provides some background to the survey, information about the progress of the survey, and the published results. Copies of the Transport Statistics bulletins can be found on the Transport Statistics website at www.scotland.gov.uk/transtat/latest.

A.32   Please use the SHS Web site to register your interest in Population and Household Surveys if you wish to be added to an *e-mail mailing list* to be kept informed of SHS news and developments. The Project Manager will also, on request, distribute paper copies of information about the survey, and about significant developments when they occur, to people who are unable to access the website.

A.33   To keep informed with changes to Scottish statistics, please register your interest with ScotStat at www.scotland.gov.uk/scotstat.