Annex 1: Using the Information in this Report

How Data is Displayed in Tables

Tables are generally presented in the format 'dependent variable by independent variable' where the independent variable is being used to examine or explain variation in the dependent variable. Thus, a table titled 'housing tenure by household type' shows how housing tenures vary among different household types. Tables generally take three forms within the report; column percentages (the dependent variable is in the rows), row percentages (the dependent variable is in the columns) and cell percentages which may show agreement or selection of a statement with one or a number of statements.

All tables have a descriptive and numerical base showing the population or population sub-group examined in it. While all results have been calculated using weighted data, the bases shown provide the unweighted counts, which have been rounded to the nearest 10 to comply with statistical disclosure control principles and the Code of Practice for Official Statistics. It is therefore not possible to calculate how many respondents gave a certain answer based on the results and bases presented in the report.

Reporting Conventions

In general, percentages in tables have been rounded to the nearest whole number. Zero values are shown as a dash (-), values greater than 0 per cent but less than 0.5 per cent are shown as 0 per cent and values of 0.5 per cent but less than 1 per cent are rounded up to 1 per cent. Columns or rows may not add to exactly 100 per cent because of rounding, where 'don't know/refused' answers are not shown[84] or where multiple responses to a question are possible.

In some tables, percentages have been removed and replaced with '*'. This is where the base on which percentages would be calculated is less than 50 and this data is judged to be insufficiently robust for publication.

[84] Missing responses are not included within the analysis. Similarly 'don't know/refused' options are not shown as a separate category in some tables.

Variations in Base Size for Totals

As the questionnaire is administered using computer assisted personal interviewing (CAPI), item non-response is kept to a minimum. Bases do fluctuate slightly due to small amounts of missing information (where, for example, the age or gender of household members has been refused and where derived variables such as household type use this information).

Some questions are asked of a reduced sample and the bases are correspondingly lower. From January 2012, the redesigned survey asked questions typically of full or one-third sample allocation. This concept of streaming was first introduced to the SHS in 2007, when some questions were streamed or changed in the course of the year and again the base size is lower. Further changes to streaming have been made in subsequent years.

Chapter 2 gives details of frequencies and bases for the main dependent variables.

Statistical Significance

All proportions produced in a survey have a degree of error associated with them because they are generated from a sample survey of the population rather than a survey of the entire population (e.g. Census). Any proportion measured in the survey has an associated confidence interval (within which the 'true' proportion of the whole population is likely to lie), usually expressed as ±x per cent. As a general rule of thumb, the larger the sample size for a given question, the smaller the confidence interval around that result will be (thus making it easier to detect real change year-on-year and differences between sub-groups.

It is possible with any survey that the sample achieved produces estimates that are outside this range. If the survey were to be run multiple times on the same population in the same year (i.e. under repeated sampling), the number of times out of a 100 surveys that the result achieved would be expected to lie within the confidence interval is also quoted; conventionally the level set is 95 out of 100, or 95 per cent. Technically, all results should be quoted in this way however, it is less cumbersome to simply report the percentage as a single percentage, the convention adopted in this report.

Where sample sizes are small or comparisons are made between sub-groups of the sample, the sampling error needs to be taken into account. There are formulae to calculate whether differences are statistically significant (i.e. they are unlikely to have occurred by chance) and Annex 3 provides a simple way to calculate whether differences are significant. Annex 3 also provides further explanation on statistical significance and on how confidence intervals can be interpreted. The local authority tables, published

alongside this report, incorporate a tool which highlights cells that are significantly different from the comparator figure - the default setting is to compare a local authority with national level data.

Annex 2: Glossary

This Annex includes a list of terms used within the report. Definitions for those terms and, in some cases, further explanation of the term are provided.

Current Economic Situation

The household respondent is asked to select which of the following categories best describes the current situation of each member of the household:

- Self-employed
- Employed full-time
- Employed part-time
- Looking after the home or family
- Permanently retired from work
- Unemployed and seeking work
- At school
- In further/higher education
- Government work or training scheme
- Permanently sick or disabled
- Unable to work because of short-term illness or injury
- Pre-school/not yet at school
- Other

SHS data on the economic situation 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. The SHS cannot provide estimates of unemployment that are comparable to official statistics of unemployment[85]. Therefore, the SHS cannot be used as a source of unemployment rates or average earnings. Please see the Scottish Government Statistics website[86] for details of Scottish Government contacts who deal with unemployment rates and average earnings statistics through the Labour Market topic.

Economic Activity, Qualifications and Training

The SHS is not directly comparable with the Labour Force Survey (LFS) which is the official source of employment, qualifications and training data in the UK. Compared with

[85] For further information, please see the SHS Methodology and Fieldwork Outcomes reports: www.gov.scot/Topics/Statistics/16002/PublicationMethodology

[86] www.gov.scot/Topics/Statistics

the LFS, the SHS under-estimates the level of employment and over-estimates both unemployment and economic inactivity. This is due to the fact that current economic situation in the SHS is asked in a single question whereas in the LFS it is determined by a selection of other questions.
The SHS also underestimates the number of people with a qualification of some sort, as the LFS covers all possible levels of qualifications. The LFS is the preferred source of estimates on employment, qualifications and training as it uses internationally agreed definitions and is used for international comparisons including OECD indicators.
It should be noted that SHS estimates of 16-64 year olds adults historically were based on the traditional 16-64 year olds definition (males aged 16-64, females aged 16-59). From 2011, these were replaced by estimates based on the population aged 16-64 to account for legislative changes in the state retirement age. Specifically the current female state pension age is changing dynamically to match the male state pension age. The Office for National Statistics (ONS) no longer publish rates using a 16-64 year olds definition, instead reports rates for all people aged 16 to 64.

Highest Level of Qualification

The highest level of qualification has been classified as follows:

- **O Grade, Standard Grade or equivalent** – Includes: School leaving certificate, NQ unit, O Grade, Standard Grade, GCSE, GCE O level, CSE, NQ Access 3 Cluster, Intermediate 1, Intermediate 2, Senior Certificate, GNVQ/ GSVQ Foundation or Intermediate, SVQ Level 1, SVQ Level 2, SCOTVEC/National Certificate Module, City and Guilds Craft, RSA Diploma or equivalent.
- **Higher, A Level or equivalent** – Includes: Higher Grade, Advanced Higher, CSYS, A Level, AS Level, Advanced Senior Certificate. GNVQ/ GSVQ Advanced, SVQ Level 3, ONC, OND, SCOTVEC National Diploma, City and Guilds Advanced Craft, RSA Advanced Diploma or equivalent.
- **HNC/ HND or equivalent** – Includes: HNC, HND, SVQ Level 4, RSA Higher Diploma or equivalent.
- **Degree, Professional qualification** – Includes: First degree, Higher degree, SVQ Level 5, Professional qualifications e.g. teaching, accountancy.
- Other qualification.
- No qualifications.
- Qualifications not known.

Please see the Scottish Government Statistics website[87] for details of Scottish Government contacts who deal with economic activity, qualifications and training statistics.

Household Economic Situation

Household economic situation refers to economic situation of the highest income householder (HIH) and/or their spouse or partner. The variable is derived from the question that asks about the economic activity of members of the household. Household economic situation variable includes the following categories:

- Single working adult
- Non-working single
- Working couple
- Couple, one works
- Couple, neither work

As mentioned previously (see Current Economic Situation), SHS data on the economic situation 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.

Household Income

The term net annual household income refers to income (i.e. after taxation and other deductions) from employment, benefits and other sources that 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 (e.g. dig money).
The definition is not the same as that used by other Government surveys such as the Family Resources Survey. These measure the income of all household members. Income data from the SHS should not, therefore, be compared with other sources without careful consideration of the methods used in compiling the data[88]. The SHS is not designed to provide reliable statistics on average income or average earnings. The current income information collected through the SHS is only intended to provide estimates by income band. The SHS asks for income only for use as a 'background' variable when analysing

[87] www.gov.scot/Topics/Statistics

[88] More information on household income can be found in Raab, G., MacDonald, C., and Macintyre, C. (2004) Comparison of Income Data between Surveys of Scottish Households: Research report for Communities Scotland. Further information on this report is available on the SHS website.

other topics, or for selecting the data for particular sub-groups of the population (such as the low paid) for further analysis[89].

Housing Lists

Housing lists are held by social landlords, local authorities and housing associations, individually or jointly as Common Housing Registers. They can include people who are already in social housing but are seeking a move and in some cases applicants will be on more than one landlord's list. Social landlords are responsible for allocating their housing, in line with their allocation policies and the legislative framework.

Calculating an estimate of the number of households on a housing list makes an assumption that the random adult response is valid for the entire household. This may however lead to a slight under-estimate because there may be a small number of multi-adult households where one adult is on a housing list but the remaining adults are not. In these cases, the SHS estimate will be influenced by which household member is selected as the random adult. In some cases, the household member on a housing list will be picked up, but in others cases they will not. This means that some households containing a household member who is on a housing list will not be identified in the survey. An example would be where a young adult is living with their parents but now wishes to form their own household separately from the existing household.

The weighting strategy for households is based on the 10,470 households responding to the household interview, rather than the 9,642 households with a complete random adult interview (providing responses to the housing lists question). This is likely to introduce a small level of non-response bias, because those households which do not complete a random adult interview are likely to be systematically different from those that do.

There is also the possibility, as with the majority of social survey questions, for a respondent to give an incorrect answer. In this case, a respondent may report being on a housing list when they are not as a result of local authorities refreshing lists and removing people from whom they have not had any contact. A respondent may report not being on a housing list when in fact they are, because some local authorities do not refresh lists and so somebody who no longer wishes to be on a housing list may still be on one that they signed up to many years previously.

Further to this, some households may not consider themselves to be on a housing list even though they are actively seeking social housing through other routes such as choice based lettings. Changes have been made to the 2017 SHS questions on housing lists with

[89] For further information, please see the SHS Methodology and Fieldwork Outcomes reports: www.gov.scot/Topics/Statistics/16002/PublicationMethodology

the aim to better capture households who are using choice based lettings when seeking social housing.

A final point on the use of the Scottish Household Survey to estimate the number of households or adults on a housing list is that it is a sample of the general population living in private residences in Scotland, and therefore it may not pick up some people or households who are on a housing list but who are living in other types of accommodation such as hostels or bed and breakfast accommodation.

Household Members

For the purposes of the SHS, 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.

The respondent for the first part of the interview must be the household reference person, a person in whose name the accommodation is owned or rented or who is otherwise responsible for the accommodation.

In households that have joint householders, the **household reference person** is defined as the **highest income householder** (**HIH**), that is, the person with the highest income. If householders have exactly the same income, the older is taken as the household reference person.

Adult is used to refer to those aged 16 and over (except where otherwise stated). **Children** are aged under 16 years.

References to **16-64 year olds** population throughout the publication refer to the 16-64 year olds definition as discussed in the economic activity, qualifications and training section in the Annex 2: Glossary, i.e. those aged 16 to 64.

In each household, one of the eligible adult members of the household is randomly selected to take part in the second half of the interview. Eligible adults are adult household members who have not been living apart from the household continuously for the previous six months. This might include adults working away from home, in the Armed Forces or in prison. The person selected is referred to as the **random adult**. The household respondent is automatically the random adult in one-adult households and may be the same as the household respondent in households with more than one adult.

Household Type

The SHS uses eight household types defined as follows:

- A **single adult** household – contains one adult of 16-64 year olds and no children.
- A **single parent** household – contains one adult of any age and one or more children.
- A **single older** household - contains one adult of pensionable age and no children. Pensionable age is 65 for both women and men.
- A **small family** household – contains two adults of any age and one or two children.
- An **older smaller** household – contains one adult of 16-64 year olds and one of pensionable age and no children, or two adults of pensionable age and no children.
- A **large adult** household – contains three or more adults and no children.
- A **small adult** household – contains two adults of 16-64 year olds and no children.
- A **large family** household – contains two adults of any age and three or more children, or three or more adults of any age and one or more children.

Housing Tenure

The SHS collects information on the ways in which households occupy their accommodation and from which organisation or individual their accommodation is rented, where this is the case. These are combined into a housing tenure variable, which is shown in the annual report broken down into four categories, namely:

- **Owner occupied** – Includes households who own outright and those buying with a mortgage or loan.
- **Social rented** sector – Includes households renting from a local authority and all households renting from a Housing Association or Co-operative.
- **Private rented** sector – Includes households renting from an individual private landlord.
- **Other** tenure – Includes any other category of tenure such as living rent free.

Income Imputation

While in general the level of missing data throughout the SHS is minimal, one section of the questionnaire is substantially affected by missing information. In the section on household income, approximately one-in-three of respondents either refuse to answer the questions or are unable to provide information that is sufficiently reliable to report, for example, because there are no details of the level of income received for one or more components of their income.
Statistical analysis of data gathered in the survey on the characteristics of households where income is available, allows income data to be imputed for households where income data is missing. Income imputation is a process whereby complete information given by 'similar' households is used for respondents that have missing income information. Income is collected as a variety of different components, such as income from employment, benefits and other sources, which are summed to create total net household income. Income was imputed for each component using either Hot Deck imputation, where the sample is divided into subgroups based on relevant characteristics, or Predictive Mean, where a statistical model is constructed and the value is predicted using this model. After imputation, income data is unavailable for between 3 to 4 per cent of households. Please contact the SHS project team if you would like further information on the imputation process.

A more advanced income imputation project was undertaken by the Scottish Government Income and Poverty Statistics team in 2010 to impute income for adults in multi-adult households for which the SHS does not capture any information. Estimates from this project were released through the “Relative Poverty Across Scottish Local Authorities” publication in August 2010[90] as data being developed. These estimates were subsequently used in a project commissioned by the Improvement Service to develop improved measures of local incomes and poverty in Scotland at a small level published in March 2013[91].

Physical or Mental Health Problems and Disabilities

Random Adult

A two part question was introduced to replace the old question on long-standing illnesses. The new question asked, of the random adult respondent, to establish the prevalence of physical or mental health conditions among the adult population and the extent to which

[90] www.gov.scot/Publications/2010/08/26155956

[91] www.improvementservice.org.uk/income-modelling-project.html

such conditions reduce ability to carry out day-to-day activities[92]. The respondent's own assessment of what constitutes a physical or mental condition or illness was used rather than a medical assessment.

The current question was introduced in October 2012 and is split into two parts: 'Do you have a physical or mental condition or illness lasting or expected to last 12 months or more?' and if so then 'Does your condition or illness reduce your ability to carry-out day-to-day activities?'
It should be noted that these changes in the question mean the 2013 data is not directly comparable to reports relating to the period 1999-2012.

Household

In the household questionnaire, the household representative is asked whether anyone in the household (including children) has any physical or mental health condition or illness lasting or expected to last for twelve months or more. The current question was introduced in 2014 and has been designed to align it with the question asked of the random adult. The response options for this question are 'Yes', 'No', 'Don't know', and 'Refused'.

Previously, the question had asked the household representative whether anyone in the household had any long-standing illness, health problem or disability that limits daily activity. The response options were 'Disability', 'Long-term illness', 'Both', 'Neither' and 'Refused'.

As noted in Chapter 2, this figure is likely to under represent the true value as the household representative may not know about the health conditions of other household members.

The above changes in the question mean that the 2014 results are not directly comparable with previous years' data.

Marital Status

The random adult is asked to confirm their legal marital status using the following categories:

- Single – never married or never formed a legally recognised same sex civil partnership
- Married and living with husband/wife
- A civil partner in a legally recognised same sex civil partnership

[92] For further details, please see questions RG5A and RG5B in the 2013 SHS questionnaire and RG5 in previous years: www.gov.scot/Topics/Statistics/16002/PublicationQuestionnaire

- Married and separated from husband/wife
- In a legally recognised same sex civil partnership and separated from your civil partner
- Divorced
- Formerly a civil partner – the same sex civil partnership now legally dissolved
- Widowed
- A surviving same sex civil partner – your partner having since died

It should be noted that this question was changed from October 2012 to remove references to “single” and to simplify the wording of the other status types. Whilst two different variables have been created in the datasets to reflect the different questions being asked, a combined derived variable was produced.
Where these have been used in the report to analyse results, these categories have been combined as:

- Single/never been married
- Cohabiting/living together
- Married/civil partnership
- Separated/divorced/dissolved civil partnership
- Widowed/bereaved civil partner

Participation, Attendance and Engagement in Sports and Physical Activity

Participation in “any sporting activity” means that people do at least one activity from the available list asked of respondents in the survey (rather than each and every sporting activity). The activities are listed as follows:

- Walking – at least 30 minutes for recreational purposes
- Swimming
- Football
- Cycling – at least 30 minutes for recreational, health, training or competition purposes
- Keep Fit / Aerobics
- Multigym use / Weight Training
- Golf
- Running / Jogging
- Snooker / Billiards / Pool
- Dancing
- Bowls

- Other (specified) – e.g. Angling, Badminton, Judo, Horse-riding, Skiing, Sailing, Yoga
- + Angling, bird-watching
- + Racket/ball sports
- + Field sports – shooting, archery
- + Water sports
- + Winter sports – curling, skating, skiing
- + Boxing, martial arts
- + Riding
- + Pilates, Yoga, Tai-Chi
- + Climbing, hillwalking
- None of these

Note, that activities prefixed above with a '+' indicate that these are backcoded following data collection based on the open text responses to the 'Other' category. This means that these activities will have been coded as 'Other' at point of collection but then moved out during the post-data processing to be assigned against the more detailed variables, and the number of responses within the 'Other' category thus lowered. The analysis presented in this report groups these additional activities back under the 'Other' category though.

Scottish Index of Multiple Deprivation

The Scottish Index of Multiple Deprivation (SIMD)[93] is a relative measure of deprivation across small areas in Scotland. It is the Scottish Government's official tool for identifying those places in Scotland suffering from multiple deprivation. It incorporates several different aspects of deprivation, combining them into a single index.

SIMD16 has been used throughout this report. It divides Scotland into 6,976 small areas, called data zones, each containing around 350 households. The Index provides a relative ranking for each data zone, from 1 (most deprived) to 6,976 (least deprived). By identifying small areas where there are concentrations of multiple deprivation, SIMD can be used to target policies and resources at the places with greatest need.

SIMD16 uses seven domains to measure the multiple aspects of deprivation: income; employment; health; education; skills and training; housing; geographic access to services; and crime. In the tables, the data zones are grouped as quintiles (from the 20 per cent

[93] www.gov.scot/Topics/Statistics/SIMD

most to the 20 per cent least deprived data zones)[94]. Occasionally deciles (from the 10 per cent most deprived data zones to 10 per cent least deprived)[95] are used.
There are also time series charts in the annual report, comparing certain characteristics over time using SIMD. SIMD was updated in 2006, 2009, 2012 and, most recently, 2016. The time series charts use the most relevant version of SIMD for each year; SIMD 2006 for 2006 – 2008 data, SIMD 2009 for 2009 – 2011 data, SIMD 2012 for 2012 – 2015 data, and SIMD16 for 2016 data. Therefore, upon each update, some areas will shift away from being the most deprived. This creates "breaks" in the data which are represented by the dotted lines in the figures.

A2 1: Number of households by Scottish Index of Multiple Deprivation 2016

2016 data, Frequency rounded to base 10

	Unweighted Frequency	Weighted Frequency	Weighted Per cent
1 - 10% most deprived	992	1,123	10.7
2	1,062	1,156	11.0
3	1,029	1,057	10.1
4	1,111	1,110	10.6
5	1,178	1,095	10.5
6	1,069	989	9.5
7	1,184	1,029	9.8
8	966	915	8.7
9	958	1,007	9.6
10 - 10% least deprived	921	989	9.4
Total	*10,470*	*10,470*	*100*

The classificatory variable used in the analysis contained in the report is based on the 2016 version of SIMD.

Self-identified Sexual Orientation

The question on self-identified sexual orientation, presented in Chapter 2, was introduced to the SHS in 2011 to provide statistics to underpin the equality monitoring responsibilities of public sector organisations and to assess the disadvantage or relative discrimination experienced by the lesbian, gay and bisexual population. Despite this positive step in collecting such information, it is felt that the figures are likely to under-report the percentage of lesbian, gay or bisexual (LGB) people within society due to a number of reasons, including the following:

- Asking about sexual orientation/identity is a new development in national surveys and such questions can be seen as intrusive and personal.

[94] Numbered 1 (most deprived) to 5 (least deprived).

[95] Numbered 1 (most deprived) to 10 (least deprived).

- There is still significant prejudice and discrimination against LGB people in society. In a context where some LGB people will not have told friends and family about their sexual identity, there is a real question about whether LGB people generally would want to be open with an interviewer.
- The default option for being uncertain about one's sexual orientation may be to respond 'straight/heterosexual' rather than to say 'Don't know / not sure'.
- Particular LGB people are still less likely to be open where they belong to groups or communities where an LGB identity is less acceptable.

Despite the uncertainties of the data, it does make sense to collect statistics on sexual orientation to start to make this a more standard element within data collection. This does not mean that data will necessarily become reliable over the short term, but they may still be able to offer useful insights into the experience of some LGB people in particular areas of policy interest. The Scottish Government is looking at how it can improve its data collection on these issues going forward.

Socio-economic Classification (NS-SEC)

National Statistics Socio-economic Classification (NS-SEC)[96] is an occupationally-based classification which, in line with all official statistics and surveys, is used in the SHS. The eight-fold analytic version of NS-SEC has been used.
Respondents' occupations and details of their employment status (whether an employer, self-employed or employee; whether a supervisor; number of employees at the workplace) have been used to create the following classifications:

- Higher managerial and professional occupations.
- Lower managerial and professional occupations.
- Intermediate occupations.
- Small employers and own account workers.
- Lower supervisory and technical occupations.
- Semi-routine occupations.
- Routine occupations.

Urban Rural Classification

The Scottish Government six-fold urban/rural classification of Scotland is used throughout this report. This classification is based on settlement size and remoteness (measured by drive times) allowing more detailed geographical analysis to be conducted on a larger

[96] More information on the definition of NS-SEC can be found at - www.ons.gov.uk/ons/guide-method/classifications/current-standard-classifications/index.html

sample size. The classification being used in this report is the 2013-2014[97] version based on 2011 datazone codes[98]
The areas in which respondents live have been classified as follows:

- **Large urban areas** – settlements of over 125,000 people.
- **Other urban areas** – settlements of 10,000 to 124,999 people.
- **Accessible small towns** – settlements of between 3,000 and 9,999 people and within a 30 minute drive of a settlement of 10,000 or more.
- **Remote small towns** – settlements of between 3,000 and 9,999 people and with a drive time of over 30 minutes to a settlement of 10,000 or more.
- **Accessible rural** – settlements of less than 3,000 people and within 30 minute drive of a settlement of 10,000 or more.
- **Remote rural** – settlements of less than 3,000 people with a drive time of more than 30 minutes to a settlement of 10,000 or more.

Table A2 2 shows the percentage of households in each area type.

A2 2: Number of households by Scottish Government 2013-2014 Urban Rural Classification
2016 data, Frequency rounded to base 10

	Unweighted Frequency	Weighted Frequency	Weighted Per cent
Large urban areas	3,139	3,761	35.9
Other urban areas	3,563	3,640	34.8
Accessible small towns	1,001	938	9.0
Remote small towns	587	376	3.6
Accessible rural	1,148	1,140	10.9
Remote rural	1,032	615	5.9
Total	*10,470*	*10,470*	*100*

Volunteering

This section of the questionnaire was revised for the 2006 survey in order to gather greater information on individuals' experience of volunteering and barriers that may prevent them from participation. Respondents were asked to give a 'yes' or 'no' response to a question on whether they had given up any time to help clubs, charities, campaigns or organisations in the last 12 months. This question was followed up by a question asked of those who said no to the first, which gave a list of types of groups and organisations and asked for which, if any, the respondent had undertaken any work or activities on a

[97] More information on the six-fold urban/rural classification of Scotland is available at - www.gov.scot/Topics/Statistics/About/Methodology/UrbanRuralClassification

[98] Previous SHS reports were based on 2001 datazone codes

voluntary basis. The list of options was revised substantially in 2007. The third question asked if there were any other types of organisations not on the list for which respondents had given up their time. Respondents who did not answer 'yes' to the first question, or who answered 'none' to the first question but 'yes' to the second or third question were classed as having taken part in voluntary activities.

A series of follow-up questions are asked to determine the frequency and types of activities adults volunteer within, if it is clear from their responses to the first three questions that they have indeed volunteered within the previous 12 months. Similarly, for those that haven't volunteered a follow-up question is asked on what might encourage them to volunteer in the future.

In 2012, it was noticed that in some cases during post-data processing, respondents that have been subsequently identified as volunteers from their answers to the second and third questions, may not have been asked the follow up questions during the actual interview. As such the number of people asked the follow-up questions might not have matched the total number of volunteers identified in the final dataset.

In 2014, the routing of the questionnaire was changed so that the maximum number of suitable people were asked the follow up questions. This means that, although it will only affect a small proportion of the sample, the 2014 results to the follow up volunteering questions are not directly comparable with previous years.

Participation, Attendance and Engagement at Cultural Events

Cultural engagement is defined as those adults who have either participated in a cultural activity or who have attended at least one type of cultural place in the previous 12 months.

A number of changes were made to the questions in 2012. The ordering of questions changed from asking about "attendance" then "participation", in 2011 for example, to asking about "participation" first from January 2012. The types of activities or events were also reworded (e.g. 'Dance' became 'Dance – e.g. ceilidh, salsa, Highland dancing, ballet') as well as switching the order of the activities and events also. More detailed information on the changes can be found in the SHS 2011 and 2012 Questionnaire publications.

Attendance at "a cultural event or place of culture" can cover any one of the following:

- Cinema
- Library – Includes: mobile and online
- Classical music performance or opera
- Live music event – e.g. traditional music, rock concert, jazz event (not opera or classical music performance)
- Theatre – e.g. pantomime / musical / play

- Dance show / event – e.g. ballet
- Historic place – e.g. castle, stately home and grounds, battle or archaeological site
- Museum
- Gallery
- Exhibition – Includes: art, photography and crafts
- Street arts – e.g. musical performances or art in parks, streets or shopping centre
- Culturally specific festival (e.g. mela /Feis/ local Gala days)
- Book festival or reading group
- Archive or records office – e.g. Scotland's Family History Peoples Centre
- None

Participation in "any cultural activity" means that people do at least one activity from the available list asked of respondents in the survey (rather than each and every cultural activity). The activities are listed as follows:

- Read for pleasure – not newspapers, magazines or comics
- Dance – e.g. ceilidh, salsa, Highland dancing, ballet
- Played a musical instrument or written music
- Took part in a play / sang in a choir or other performance – not karaoke
- Painting, drawing, printmaking or sculpture
- Photography / making films or videos as an artistic activity – not family or holiday 'snaps'
- Used a computer / social media to produce creative work of any kind
- Crafts such as knitting, wood, pottery, etc.
- Creative writing - stories, books, plays or poetry
- Other cultural activity
- None

Youth Activities

The Scottish Government is interested in the extent to which young adults and children are involved in a range of activities. Those households for which there is someone aged between 8 and 21 are asked a series of questions within the SHS on whether they take part in a series of activities regularly. These activities are:

- *Any music or drama activities* such as playing in a band or a theatre group;
- *Any other arts activities* such as a photography or art club including classes;
- *Any sports or sporting activity* whether played competitively or not;
- *Any other outdoor activities* such as walking, angling, bird-watching, etc.;

- *Any other groups or clubs* such as a youth club or youth group, scouts, chess club, bridge club, etc.;
- *Representing young people's views* or involvement in youth politics (e.g. Youth Forum or Dialogue Youth);
- Mentoring or peer education; and,
- None

Annex 3: Confidence Intervals and Statistical Significance

The Representativeness of the Scottish Household Survey

Although the Scottish Household Survey (SHS) sample is chosen at random, the people who take part in the survey will not necessarily be a representative cross-section of the population. Like all sample surveys, the results of the SHS are estimates for the whole population and these results might vary from the true values in the population for three main reasons:

1. The sample frame does not completely cover the population because accommodation in hospitals, prisons, military bases, larger student halls etc. are excluded from the sampling frame. The SHS provides a sample of private households rather than all households. The effect of this on the representativeness of the data is not known.
2. Some people refuse to take part in the survey and some cannot be contacted by interviewers. If these people are systematically different from the people who are interviewed, this represents a potential source of bias in the data. Comparison of the SHS data with other sources suggests that for the survey as a whole, any bias due to non-response is not significant[99].
3. Samples always have some natural variability because of the random selection of households and people within households. In some areas where the sample is clustered, the selection of sampling points adds to this variability.

Each of these sources of variability becomes much more important when small sub-samples of the population are examined. For example, a sub-sample with only 100 households might have had very different results if the sampling had, by chance, selected four or five more households with children, rather than households including one or two adults of pensionable age and no younger adults.

[99] For further information, please see the SHS Methodology and Fieldwork Outcomes reports - www.gov.scot/Topics/Statistics/16002/PublicationMethodology

Confidence Intervals

The likely extent of sampling variability can be quantified by calculating the 'standard error' associated with an estimate produced from a random sample. Statistical sampling theory states that, on average:

- Only about one sample in three (33 per cent) would produce an estimate that differed from the (unknown) true value by more than one standard error;
- Only about one sample in twenty (5 per cent) would produce an estimate that differed from the true value by more than two standard errors;
- Only about one sample in 400 (0.25 per cent) would produce an estimate that differed from the true value by more than three standard errors.
    - By convention, the '95 per cent confidence interval' is defined as the estimate plus or minus about twice 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.

The standard error of the estimate of a percentage depends upon several things:

- The value of the percentage itself;
- The size of the sample (or sub-sample) from which it was calculated (i.e. the number of sample cases corresponding to 100 per cent per cent);
- The sampling fraction (i.e. the fraction of the relevant population that is included in the sample); and
- The 'design effect' associated with the way in which the sample was selected (for example, a clustered random sample would be expected to have larger standard errors than a simple random sample of the same size).

Figure A3 1 at the end of this Annex shows the 95 per cent confidence limits for a range of estimates calculated for a range of sample sizes, incorporating a design factor of 1.15[100] to account for the complex survey design. To estimate the potential variability for an estimate for the survey you should read along the row with the value closest to the estimate until you reach the column for the value closest to the sub-sample. This gives a value which, when added and subtracted from the estimate, gives the range (the 95 per cent confidence interval) within which the true value is likely to lie. Where the exact value is not given in the table, we recommend using the closest value in the table. Otherwise, you may also derive more precise estimates through using standard formulas for confidence intervals from survey estimates, incorporating a design factor of 1.15.

[100] The design factor is calculated as an overall average across a number of variables, and should not be taken as a 'typical' value across all variables. For further information, please see the SHS Methodology and Fieldwork Outcomes reports - www.gov.scot/Topics/Statistics/16002/PublicationMethodology

For example, if the survey estimates that 18.0 per cent of households in Scotland are ‘single adult’ households and this has a confidence interval of ±0.9 per cent, it means that, we could be 95 per cent confident that the true value for the population lies between 17.1 per cent and 18.9 per cent.

However, smaller sample sizes have wider confidence intervals. So, for example, looking at household type might show that in, say, Edinburgh, 28.0 per cent of households are ’single adult’ households. However, if there were 780 households in Edinburgh interviewed, this estimate would have a 95 per cent confidence interval of approximately ±3.7 per cent. This suggests that the true value lies between 24.3 per cent and 31.7 per cent. Clearly, the estimate for any single area is less reliable that the estimate for Scotland as a whole.

Statistical Significance

Because the survey's estimates may be affected by sampling errors, apparent differences of a few percentage points between sub-samples may not reflect real differences in the population. It might be that the true values in the population 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-sample and a low estimate for the other.

A difference between two areas is significant if it is so large that a difference of that size (or greater) is unlikely to have occurred purely by chance. Conventionally, significance is tested at the 5 per cent level, which means that a difference is considered significant if it would only have occurred once in 20 different samples. Testing significance involves comparing the difference between the two samples with the 95 per cent confidence limits for each of the two estimates.

For example, suppose the survey estimates that there are 14 per cent ‘single adult households’ in Stirling (±4.1 per cent), 10 per cent in Aberdeenshire (±1.7 per cent), 15 per cent in Fife (±2.0 per cent), and 24 per cent in Edinburgh (±2.5 per cent). Assuming that the estimates' values are 'exact' (i.e. that the figure underlying 10 per cent is 10.0 per cent), we can say the following:

- The difference between Stirling and Fife is not significant because the difference between the two (1 per cent) is smaller than either of the confidence limits (at least ±2.0 per cent). In general, if the difference is smaller than the larger of the two limits, it could have occurred by chance and is not significant;
- The difference between Stirling and Edinburgh is significant because the difference (10 per cent) is greater than the sum of the limits (4.1 + 2.5 = 6.6 per cent). In general, a difference that is greater than the sum of the limits is significant.

If the difference is greater than the larger of the two confidence limits, but less than the sum of the two limits, the difference might be significant, although the test is more complex.

Statistical sampling theory suggests that the absolute value of the difference between the two estimates $|p1 - p_2|$ is significant if it is greater than the square root of the sum of the squares of the limits for the two estimates, as explained by the following formula:

$$|p_{1-}p_2| > \sqrt{[(CI_1)^2 + (CI_2)^2]}$$

The difference of 5 per cent between Aberdeenshire and Fife is greater than the largest confidence limit (±4.1 per cent) but it is less than the sum of the two limits (4.1 per cent + 2.0 per cent = 6.1 per cent) so it might be significant. In this case 4.12 = 16.81 and 2.02 = 4 giving a total of 20.81. The square root of this is 4.56, which means that the difference of 5 per cent is significant (although only just). Similar calculations will indicate whether or not other pairs of estimates differ significantly.

It should be noted that the estimates published in this report have been rounded, generally to the nearest whole number, and this can affect the apparent significance of some of the results. For example:

- If the estimate for Aberdeenshire was 10.49 per cent (rounded to 10 per cent) and the estimate for the Fife was 14.51 per cent (rounded to 15 per cent) the difference would be calculated as 4.02 per cent rather than 5 per cent. This is below the calculated 'significance threshold' value of 4.56 per cent;
- If, however, the estimate for the Lothians was 10.51 per cent (rounded to 11 per cent) and the estimate for Fife was 15.49 per cent (rounded to 15 per cent) the difference would be calculated as 4.98 per cent rather than 5 per cent. This is higher than 4.56 per cent.

For this reason, caution should be exercised where differences are on the margins of significance. In general, we would suggest that differences should only be considered significant where the difference is clearly beyond the threshold of significance.

Statistical Significance and Representativeness

Calculations of confidence limits and statistical significance only take account of sampling variability. The survey's results could also be affected by non-contact/non-response bias. If the characteristics of the people who should have been in the survey but who could not be contacted, or who refused to take part, differ markedly from those of the people who were interviewed, there might be bias in the estimates. 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, we 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. However, such biases could, of course, be more significant for some sub-groups of the population or in certain council areas, particularly those that have the highest non-response rates.

In addition, because it is a survey of private households, the SHS does not cover some sections of the population - for example, it does not collect information about students in halls of residence. Please refer to the companion technical reports[101] for a comparison of SHS results with information from other sources.

[101] For further information, please see the SHS Methodology and Fieldwork Outcomes reports: www.gov.scot/Topics/Statistics/16002/PublicationMethodology

Figure A3 1: Estimated sampling error associated with different proportions for different sample sizes

	100	200	300	400	500	700	1,000	2,000	3,000	4,000	5,000	6,000	7,000	8,000	9,000	10,000	11,000
5%	4.9%	3.5%	2.8%	2.5%	2.2%	1.9%	1.6%	1.1%	0.9%	0.8%	0.7%	0.6%	0.6%	0.5%	0.5%	0.5%	0.5%
10%	6.8%	4.8%	3.9%	3.4%	3.0%	2.6%	2.1%	1.5%	1.2%	1.1%	1.0%	0.9%	0.8%	0.8%	0.7%	0.7%	0.6%
15%	8.0%	5.7%	4.6%	4.0%	3.6%	3.0%	2.5%	1.8%	1.5%	1.3%	1.1%	1.0%	1.0%	0.9%	0.8%	0.8%	0.8%
20%	9.0%	6.4%	5.2%	4.5%	4.0%	3.4%	2.9%	2.0%	1.6%	1.4%	1.3%	1.2%	1.1%	1.0%	1.0%	0.9%	0.9%
25%	9.8%	6.9%	5.6%	4.9%	4.4%	3.7%	3.1%	2.2%	1.8%	1.5%	1.4%	1.3%	1.2%	1.1%	1.0%	1.0%	0.9%
30%	10.3%	7.3%	6.0%	5.2%	4.6%	3.9%	3.3%	2.3%	1.9%	1.6%	1.5%	1.3%	1.2%	1.2%	1.1%	1.0%	1.0%
35%	10.8%	7.6%	6.2%	5.4%	4.8%	4.1%	3.4%	2.4%	2.0%	1.7%	1.5%	1.4%	1.3%	1.2%	1.1%	1.1%	1.0%
40%	11.0%	7.8%	6.4%	5.5%	4.9%	4.2%	3.5%	2.5%	2.0%	1.7%	1.6%	1.4%	1.3%	1.2%	1.2%	1.1%	1.1%
45%	11.2%	7.9%	6.5%	5.6%	5.0%	4.2%	3.5%	2.5%	2.0%	1.8%	1.6%	1.4%	1.3%	1.3%	1.2%	1.1%	1.1%
50%	11.3%	8.0%	6.5%	5.6%	5.0%	4.3%	3.6%	2.5%	2.1%	1.8%	1.6%	1.5%	1.3%	1.3%	1.2%	1.1%	1.1%
55%	11.2%	7.9%	6.5%	5.6%	5.0%	4.2%	3.5%	2.5%	2.0%	1.8%	1.6%	1.4%	1.3%	1.3%	1.2%	1.1%	1.1%
60%	11.0%	7.8%	6.4%	5.5%	4.9%	4.2%	3.5%	2.5%	2.0%	1.7%	1.6%	1.4%	1.3%	1.2%	1.2%	1.1%	1.1%
65%	10.8%	7.6%	6.2%	5.4%	4.8%	4.1%	3.4%	2.4%	2.0%	1.7%	1.5%	1.4%	1.3%	1.2%	1.1%	1.1%	1.0%
70%	10.3%	7.3%	6.0%	5.2%	4.6%	3.9%	3.3%	2.3%	1.9%	1.6%	1.5%	1.3%	1.2%	1.2%	1.1%	1.0%	1.0%
75%	9.8%	6.9%	5.6%	4.9%	4.4%	3.7%	3.1%	2.2%	1.8%	1.5%	1.4%	1.3%	1.2%	1.1%	1.0%	1.0%	0.9%
80%	9.0%	6.4%	5.2%	4.5%	4.0%	3.4%	2.9%	2.0%	1.6%	1.4%	1.3%	1.2%	1.1%	1.0%	1.0%	0.9%	0.9%
85%	8.0%	5.7%	4.6%	4.0%	3.6%	3.0%	2.5%	1.8%	1.5%	1.3%	1.1%	1.0%	1.0%	0.9%	0.8%	0.8%	0.8%
90%	6.8%	4.8%	3.9%	3.4%	3.0%	2.6%	2.1%	1.5%	1.2%	1.1%	1.0%	0.9%	0.8%	0.8%	0.7%	0.7%	0.6%
95%	4.9%	3.5%	2.8%	2.5%	2.2%	1.9%	1.6%	1.1%	0.9%	0.8%	0.7%	0.6%	0.6%	0.5%	0.5%	0.5%	0.5%