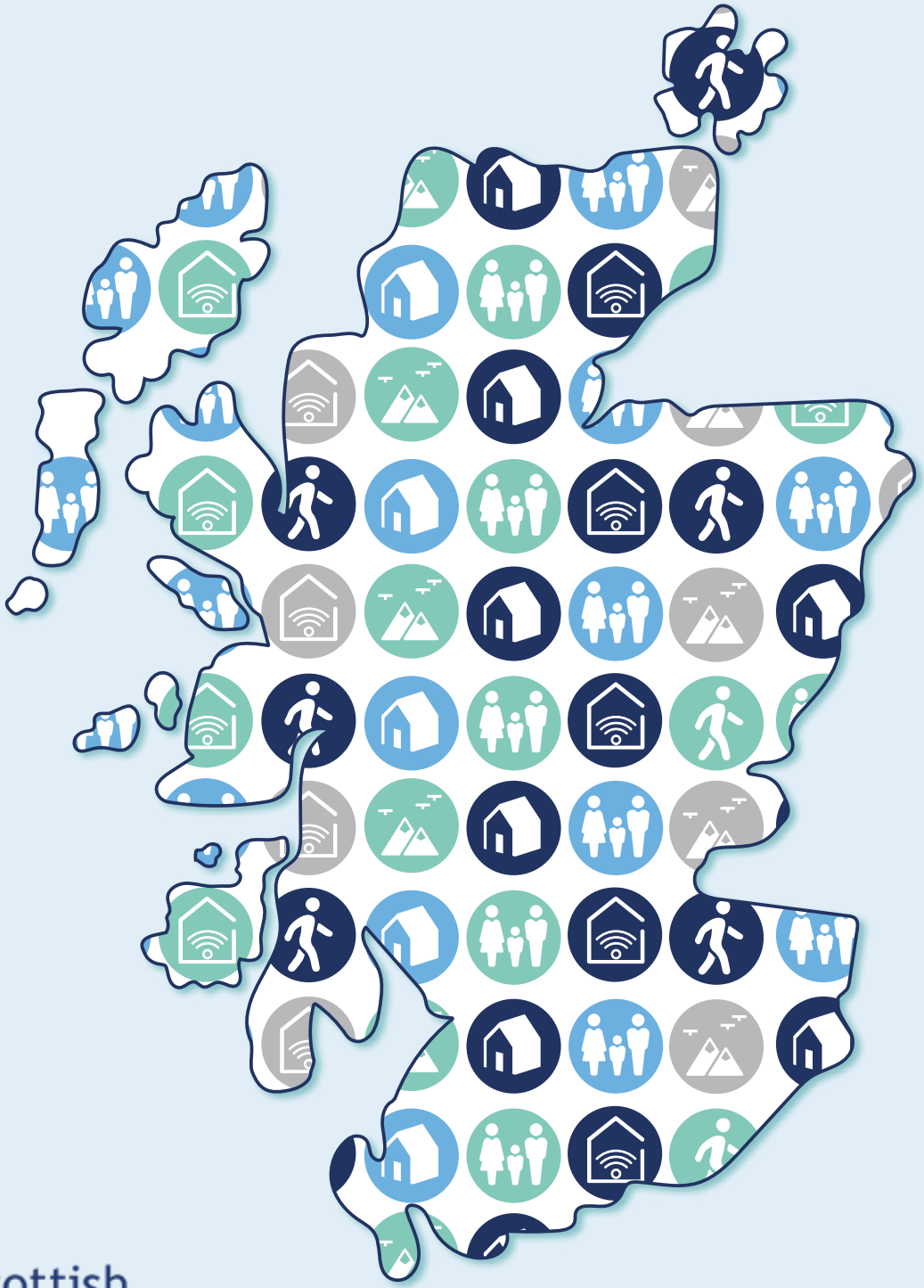
Methodology and Fieldwork Outcomes 2019

A National Statistics publication for Scotland



Scottish Household Survey

Help Shape Scotland

Scottish Government
Riaghaltas na h-Alba
gov.scot

Acknowledgements

Primarily, thank you to the 10,577 householders who agreed to fully participate in the survey. Without your goodwill and support the survey would not be possible. We would also like to express our thanks for the commitment and determination of the fieldwork team who collect the data. The Scottish Household Survey data is collected from homes all over Scotland by a team of around 100 interviewers, 50 surveyors and a small number of Regional Managers.

Thank you to all the people involved in the development, implementation and reporting of the Scottish Household Survey results.

Special thanks go to Ipsos MORI for their continued support and advice across all aspects of the survey management and equally to CA Design Services who organise the physical fieldwork and manage the telephone helpline and BRE who conduct modelling with SHS data.

Thanks also the Office of Chief Statistician team, Scottish Government, for their contributions to this report and all of the lead analysts in the Scottish Government for their contributions to the main report.

Comments and Suggestions

We are committed to continual improvement and would welcome any comments or suggestions on how the SHS Methodology and Fieldwork Outcomes report, or any other reports or releases, could be improved or adapted in the future. Similarly, if you have any enquiries on any aspects of the survey development then we would welcome your opinions or questions. Please contact the SHS Project Team.

SHS Project Team
Communities Analysis Division
Victoria Quay
Edinburgh
EH6 6QQ

shs@gov.scot

https://www.gov.scot/collections/scottish-household-survey

Table of Contents

1 Survey Overview 7

1.1 Technical reports 9

2 Sample Design 10

2.1 Requirements 11
2.2 Sample design and assumptions 12
2.3 Sample selection 15
2.4 Exclusions 18

3 Data Collections Methods and Instruments 20

3.1 Use of Computer Aided Personal Interviewing (CAPI) 21
3.2 Questionnaire structure 21
3.3 Survey fieldwork 26

4 Physical Fieldwork and Physical Survey Form 28

4.1 Physical survey team 28
4.2 Types of physical survey 29
4.3 Physical survey administration 31
4.4 Surveyor variability 33
4.5 Physical survey form 37

5 Data Processing 39

5.1 Social data processing 40
5.2 Physical survey data validation 44

6 Survey Response 46

6.1 Introduction 46
6.2 Scotland-Level Summary 48
6.3 Drop in response rate 50
6.4 Local authority performance 51
6.5 Monitoring and reducing the respondent burden 52
6.6 Physical survey 54

7 Survey Weighting 56

7.1 Introduction 57
7.2 Household weights 58
7.3 Random Adult weights 59
7.4 Random Schoolchild weights 60
7.5 Travel Diary weight 60
7.6 Physical survey weight 62
7.7 Summary of weighting variables to apply in analysis 63

8 Limitations of the Data 65

9 Bias and Data Quality 69

9.1 Comparisons with Scotland's Census 2011 and mid-2018 Population Estimates 71
9.2 Total survey error 73

List of Tables and Figures

Tables

Table 2.1: Target sample sizes and selected addresses 13
Table 3.1: Topics covered in SHS 2019 Household component 23
Table 3.2: Topics covered in SHS 2019 Random Adult component 25
Table 4.1: Full physical surveys by surveyor and dwelling type 35
Table 6.1: Fieldwork outcomes, Scotland 49
Table 6.2: Response rates over time 50
Table 6.3: Response rate and eligibility by local authority 51
Table 6.4: Conversion to full physical survey 55
Table 7.1: Physical Survey calibration targets 63
Table 7.2: Survey weighting variables 64
Table 9.1: Comparison of tenure of household between Census 2011 and SHS 2019 estimates 72
Table 9.2: Comparison of age of adults between Census 2011, mid-2018 population estimates and SHS 2019 estimates 73

Figures

Figure 3.1: Representation of multi-year core and modular design 22
Figure 4.1: Relationship between social outcomes and type of physical survey required 30
Figure 4.2: Physical survey form example 38
Figure 5.1: SHS social survey data processing procedures 40
Figure 5.2: Physical data validation screens 45
Figure 9.1: The lifestyle of a survey from a quality perspective 75

1 Survey Overview

Summary

- The Scottish Household Survey (SHS) has run continuously in homes in Scotland since in 1999.
- A sample of the general population are randomly chosen to take part every year.
- From 2012 onwards, the survey was substantially redesigned to include elements of the Scottish House Condition Survey (SHCS)[1] including the follow-up Physical Survey component.
- The new survey uses a fully unclustered core and modular structure with some questions asked of the full sample and others of a one-third sub-sample. Some questions are only asked every other year.
- While the overall sample size of the survey reduced from 2012 onwards from around 14,000 household interviews to about 10,000, the survey design improvements have meant that the precision of estimates has not been affected significantly.
- From 2012, it is possible to obtain local authority estimates on an annual basis where sample sizes produce robust estimates. Up to 2011 the data was collected over two years and the local authority level data was available only after the two-year cycle was completed.

[1] https://www.gov.scot/collections/scottish-house-condition-survey

The Scottish Household Survey (SHS) is a continuous survey based on a sample of the general population in private residences in Scotland.

The SHS is designed to provide reliable and up-to-date information on the composition, characteristics, attitudes and behaviour of Scottish households and individuals, both nationally and at a sub-national level, and to examine the physical condition of Scotland's homes. It covers a wide range of topics. The specific aims of the survey are:

- Meet central and local Government needs for priority policy relevant data across a broad range of topics (including needs for continuing time-series of data collected by the SHS and SHCS previously);
- Be understandable and useful to stakeholders and so lead to a high level of buy-in and use of the SHS;
- Have built in flexibility to respond to different data needs regarding geography and frequency (e.g. to provide some data annually at Local Authority level, and some biennially at national level), and changes to these requirements over time;
- Align with other surveys and data vehicles (in particular the Scottish Health Survey and Scottish Crime and Justice Survey) to allow for Scottish Surveys Core Questions data to be produced;
- Produce high quality data in accordance with the Code of Practice for Official Statistics[2] so as to provide data that is suitable for the production of National Statistics publications in a cost effective way;
- To permit disaggregation of information both geographically and in terms of population sub-groups (such as families with children or households in the social rented sector);
- To allow the relationships between social variables within households to be examined. This will support cross-analysis on a range of issues;
- To allow detailed follow-up surveys of sub-samples from the main survey sample, if required.

The survey is funded by the Scottish Government with contracts awarded for the fieldwork periods 1999 to 2002, 2003 to 2006, 2007 to 2011, 2012 to 2017 and from 2018 for 4 years with a possibility to extend the contract for a further 2 years. The survey is run through a consortium led by Ipsos MORI.

[2] https://code.statisticsauthority.gov.uk

The survey started in 1999 and up to 2011 followed a fairly consistent survey design. The data was collected over two years and the local authority level data was available only after the two-year cycle was completed. Due to the re-design of the survey, and the 2011 fieldwork being only a half of that cycle, local authority level data is not available for 2011.

From 2012 onwards, the survey was substantially redesigned to include elements of the Scottish House Condition Survey (SHCS)[3] including the follow-up Physical Survey component. The new SHS went in to the field with a substantially restructured sample design, integrating the previous SHCS. The new survey uses a fully unclustered core and modular structure with some questions asked of the full sample and others of a one-third sub-sample. The overall sample size has reduced from around 14,000 household interviews to about 10,000 though improvements in efficiency of the survey design mean it is possible to obtain local authority estimates on an annual basis where sample sizes produce robust estimates. This also means that any set of years can be combined to create larger samples, where necessary. While the overall sample size of the survey has reduced, the survey design improvements have meant that the precision of estimates has not been affected significantly.

1.1 Technical reports

Technical Reports have been published for each year of the survey covering the survey methodology, fieldwork outcomes and the questionnaire[4] used. This report covers the methodology and fieldwork outcomes for the 2019 survey.

[3] https://www.gov.scot/collections/scottish-house-condition-survey

[4] https://www.gov.scot/collections/scottish-household-survey/#questionnaires

2 Sample Design

Summary

- From 2012 the three Scottish Government interviewer-led population surveys (SHS, Scottish Health Survey (SHeS) and the Scottish Crime and Justice Survey (SCJS) have coordinated sample designs.
- The SHS sample has been designed to allow annual publication of results at Scotland level and for local authorities.
- From 2012, the physical survey of the Scottish House Condition Survey (SHCS) has been incorporated into the SHS.
- Just less than half of the main SHS sample has been allocated to the physical survey, which has a required sample size of 3,004 for Scotland and a minimum of 80 for each local authority.
- In order to provide annual local authority results without specifying an excessive overall sample size, the sample was disproportionately stratified by local authority (smaller local authorities have a higher sample proportion relative to their populations than the larger local authorities).
- For the physical survey, the minimum sample size over each rolling three year period for each local authority is 240 giving a minimum of 80 per annum. There is also a minimum annual sample for Scotland of 3,004.
- Samples of the general population exclude prisons, hospitals and military bases.
- The Royal Mail's small user Postcode Address File (PAF) was used as the sample frame for the address selection.
- Addresses selected for any of the surveys (SHS, SHeS, SCJS) are removed from the sample frame for a minimum of 4 years so that they cannot be re-sampled for another survey. This reduces respondent burden.

- In order to collect information both about the structure and characteristics of Scottish homes and about the people who occupy them, the interview has a two-part structure, generally the Highest Income Householder or their spouse or partner answers the first part, and one adult (aged 16+) member of the household is selected at random by the CAPI script to answer the second part.

2.1 Requirements

The sample for the 2019 Scottish Household Survey (SHS) was designed by the Scottish Government. From 2012 the sample design has been coordinated with the sample designs for the Scottish Health Survey (SHeS) and the Scottish Crime and Justice Survey (SCJS) as part of a survey efficiency project and to allow the samples of the three surveys to be pooled for further analysis[5].

The SHS sample has been designed to allow annual publication of results at Scotland level and for local authorities. To meet these requirements the target sample size for Scotland was 10,450 household interviews with a minimum local authority target of 250. From 2012, the physical survey of the Scottish House Condition Survey (SHCS) has been incorporated into the SHS. A subsample of just less than half of the main SHS sample has been allocated to the physical survey, which has a required sample size of 3,004 for Scotland and a minimum of 80 for each local authority.

[5] Further information on the sample designs and the methodology used is available here: http://scotland.gov.uk/Topics/Statistics/About/SurveyDesigns201215

2.2 Sample design and assumptions

2.2.1 Main sample

The 2019 Scottish Household Survey has a single-stage unclustered sample design. In order to provide annual local authority results without specifying an excessive overall sample size, the sample was disproportionately stratified by local authority (smaller local authorities have a higher sample proportion relative to their populations than the larger local authorities).

To deliver the required local authority precision the minimum sample size for each local authority was set at 250.

In order to estimate the annual target achieved sample size for each local authority, analysis of design effects from the 2007-08 survey was undertaken, since:

$$\text{Effective sample size} = \frac{\text{Achieved sample}}{\text{Design effect}}$$

As rural areas of local authorities were clustered in the 2007-08 survey, for the 2019 unclustered sample the median design effect from a range of variables for the unclustered parts of local authority samples were assumed for the entire areas in 2019. This allowed the calculation of the target achieved sample size for each local authority, as shown in Table 2.1.

Table 2.1: Target sample sizes and selected addresses

Local Authority	Main sample		Physical survey sub-sample	
	Target interviews	Selected addresses	Target interviews	Selected addresses
Aberdeen City	357	677	91	298
Aberdeenshire	372	620	89	256
Angus	250	472	80	261
Argyll and Bute	250	457	80	260
Clackmannanshire	250	372	80	179
Dumfries and Galloway	250	431	80	207
Dundee City	250	411	80	215
East Ayrshire	250	488	80	260
East Dunbartonshire	250	468	80	263
East Lothian	250	431	80	224
East Renfrewshire	250	457	80	255
Edinburgh City	783	1,489	190	657
Falkirk	250	442	80	266
Fife	559	911	137	400
Glasgow City	985	1,987	246	885
Highland	362	648	88	275
Inverclyde	250	435	80	243
Midlothian	250	457	80	254
Moray	250	431	80	217
Na h-Eileanan Siar	250	383	80	203
North Ayrshire	250	481	80	271
North Lanarkshire	506	885	124	377
Orkney	250	367	80	175
Perth and Kinross	250	428	80	231
Renfrewshire	283	478	80	229
Scottish Borders	250	432	80	232
Shetland	250	376	80	175
South Ayrshire	250	449	80	222
South Lanarkshire	486	862	119	376
Stirling	250	376	80	194
West Dunbartonshire	250	426	80	224
West Lothian	257	402	80	221
Scotland	10,450	18,429	3,004	9,005

2.2.2 Physical survey sub-sample

For the physical survey, the minimum sample size over each rolling three year period for each local authority is 240 giving a minimum of 80 per annum. There is also a minimum annual sample for Scotland of 3,004. An iterative approach was taken to allocate the physical surveys across local authorities. Firstly, the overall sample of 3,004 was allocated to local authorities proportionate to the number of occupied dwellings. Where the allocated number of interviews was below 80, the allocation was increased to 80. The remaining sample was then allocated across the local authorities which had an initial allocation of more than 80.

Completion of the physical survey requires that selected households respond to the main social survey and agree to a follow-up visit for the physical survey to be completed. Therefore, in order to achieve the sample targets a conversion rate from household interview to physical survey is required. For each local authority, assumptions for conversion from household interview to physical survey were based on the average conversion rate from the three most recent SHCS with information available. Additional conditions were added to the conversion rate assumptions setting upper and lower limits of 90 per cent and 60 per cent, respectively.

In order to calculate the total number of addresses in the sample to assign to the physical sample, the number of responding households required to yield the physical survey responses is calculated using the conversion rates. The response rate and ineligible address assumptions are then applied.

Table 2.1 shows the target sample size and the number of selected addresses for the main sample and physical survey by local authority.

2.3 Sample selection

The Royal Mail's small user Postcode Address File (PAF) was used as the sample frame for the address selection. The advantages of using the small user PAF are as follows:

- It has previously been used as the sample frame for Scottish Government surveys so previously recorded levels of ineligible addresses can be used to inform assumptions for 2019 sample design;
- It has excellent coverage of addresses in Scotland; and
- The small user version excludes the majority of businesses

The PAF does still include a number of ineligible addresses, such as small businesses, second homes, holiday rental accommodation and vacant properties. A review of the previous performance of individual surveys found that they each recorded fairly consistent levels of ineligible address for each local authority. This meant that robust assumptions could be made for the expected levels of ineligible addresses in the sample size calculations.

As the samples for the SHS, SHeS and SCJS are all being selected by the Scottish Government from 2012 onwards, addresses selected for any of the surveys are removed from the sample frame so that they cannot be re-sampled for another survey. This will help to reduce respondent burden and facilitate the development of the pooled sample. The addresses are removed from the sample frame for a minimum of 4 years.

The sample design specified above was implemented as follows:

1) Systematic random sampling was used to select the addresses from the sample frame with the addresses ordered by urban-rural classification, SIMD rank and postcode.

2) Once the overall sample was selected systematic random sampling was used to select the subsample for the physical survey.

2.3.1 Selecting households at addresses with multiple dwellings

A small number of addresses have only one entry in the PAF but contain multiple dwelling units. Such addresses are identified in the PAF by the Multiple Occupancy Indicator (MOI). To ensure that households within MOI addresses had the same probability of selection as other households, the likelihood of selecting the addresses were increased in proportion to the MOI. For addresses which are flagged as having multiple dwellings in the PAF the dwelling to interview was randomly selected as part of the sample selection process.

Where the MOI is correct, this procedure is unproblematic. Sometimes, however, the MOI is incorrect or missing (in about 2 per cent 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.

2.3.2 Selecting individuals within households

As the survey is intended to collect information both about the structure and characteristics of Scottish homes and about the people who occupy them, 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[6]. For the second part of the interview, one adult (aged 16+) member of the household is selected at random by the CAPI script (see section 3.1). If this person is not available at the time, the interviewer will call back to complete the interview at a later date if necessary[7].

2.3.3 Allocation of sample to different time periods

Finally, addresses were grouped into batches for effective fieldwork. This was done by minimising the distance required to visit each address in a batch. Batches were then allocated to a particular fieldwork quarter. All quarters had, as far as possible, the same number of batches in each local authority to help ensure that the fieldwork was carried out throughout the year.

2.3.4 Allocation of sample to questionnaire modules

To meet the need for modularisation, all sampled addresses were randomly assigned to one of 12 sub-samples or interview streams, which could be used as the basis for assigning samples of respondents to particular blocks of questions.

For example, the Physical Survey module is intended to provide representative data on the physical condition of Scotland's homes and this is achieved by assigning the module to streams 1 to 4. It should be noted that given difficulties in achieving the target number of physical surveys (see section 6.3), a fifth stream (stream nine) was opened up to the physical survey but not for the complete fieldwork year. Households who were in stream nine were asked to participate in the physical survey in quarters 2 and 4.

[6] 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.

[7] 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.

Other smaller blocks of questions are asked of sub-samples at various points in the questionnaire and the published version of the questionnaire[8] indicates where and at what points in time streaming is used.

2.4 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 enumeration districts (EDs) in the Census and account for just 0.5 per cent 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.

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; and
- 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[9].

[8] https://www.gov.scot/collections/scottish-household-survey/#questionnaires

[9] The target population of the old SHCS excludes "dwellings without foundations". So, for example, mobile homes would not be eligible for the physical survey. However, these dwellings are eligible for the social survey. In order to work around this, such dwellings are still included in the social interview but are then streamed past questions on surveyor appointments, thus avoiding surveyors being called out to static caravans or houseboats.

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

3 Data Collections Methods and Instruments

Summary

- The SHS social interview is carried out using Computer Aided Personal Interviewing (CAPI).
- The social questionnaire is in two parts; Household and Random Adult.
- The SHS (between 2012 and 2019) has a core and modular design which rotates and replicates across to subsequent years. A "core" set of 20 questions is included which have been designed to be asked in consistent ways with other surveys, such as age and gender.
- The Household Composition part of the questionnaire covers topics on Accommodation, Household Services, Driving and Transport, Children and Young People, Health and Disability, House Condition Survey, Household Employment, Household Income, Household Finances and Mortgages and Rent.
- The Random Adult part of the questionnaire covers topics on Adult Characteristics, Accommodation, Neighbourhoods and Communities, Education and Training, Internet, Travel and Transport, Volunteering, Local government & services, Culture & Sport, Environment, Health, Disability & Caring and Employment.
- There was a follow-up component comprising of a 'Physical Survey' of the dwelling. Such surveys are conducted by professional surveyors through a visual inspection of the dwelling.
- Before the first interviewer visit, households were sent an advance letter and leaflet outlining the purpose of the survey and the importance of participation. Interviewers post the advance letters themselves.
- The main fieldwork for the survey runs continuously throughout the year.

- The fieldwork for the 2019 SHS was completed in February 2020 and was not impacted by the coronavirus pandemic.
- Interviewers were required to make up to six calls at an address (an initial visit plus five 'call-backs').

3.1 Use of Computer Aided Personal Interviewing (CAPI)

In common with many other large-scale government surveys, the SHS social interview 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.

CAPI programming is integral to ensuring high quality data. The main parameters of the data are defined within the programme, such as; the acceptable range of responses to a question; the acceptable relationships between questions – or, the routing; and the relationships between responses given at different questions.

Between 2012 and 2019, the SHS script was scripted using Quancept. This was the software used for both the SHS and the SHCS in 2011.

3.2 Questionnaire structure

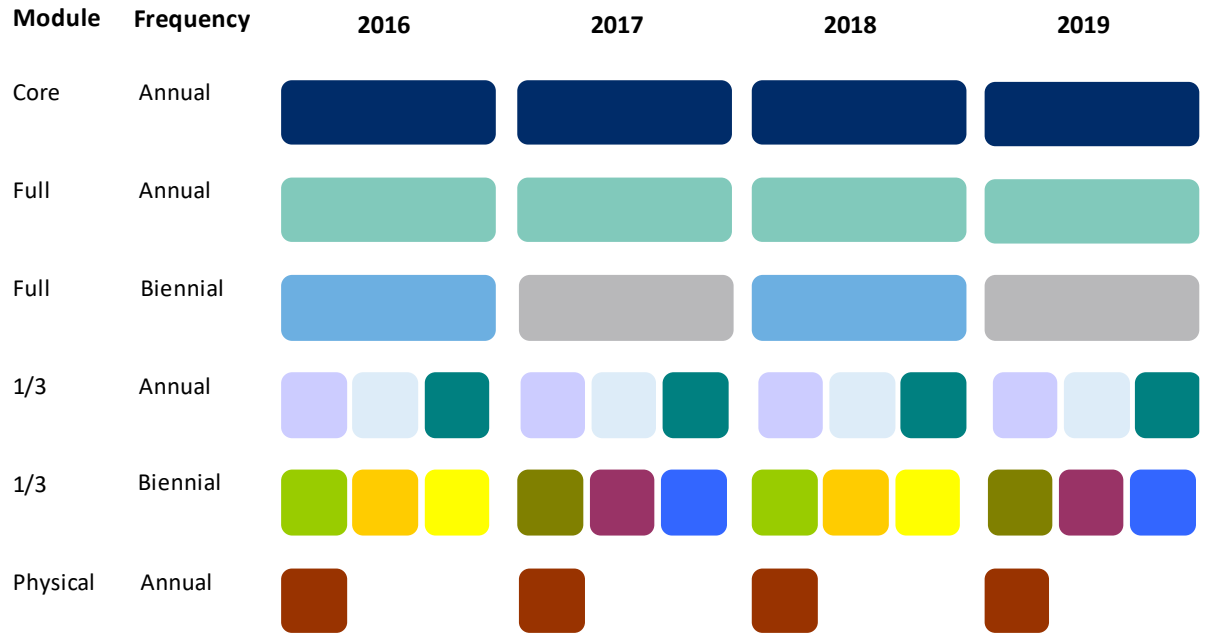
The social questionnaire is in two parts:

- Household – Information is collected about the composition and characteristics of the household from the Highest Income Householder or their spouse/partner – this allows the SHS to produce information representative of Scottish Households;
- Random Adult – Information is collected focusing on the attitudes and experiences of a randomly selected adult member of the household. This produces data which is representative of the Scottish adult population resident in private households.

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, e.g. age, gender, religion.

Figure 3.1 provides a visual representation of the core and modular design of the SHS (between 2016 and 2019) and how this rotates and replicates across to subsequent years also. This includes a "core" set of 20 questions which have been designed to be asked in consistent ways with other surveys, such as age and gender[10]. The subsequent "modules" of questions have been designed to be flexible in terms of topic, frequency and geography. For example, questions asked of the "full" sample and asked on an "annual" basis would be able to provide local authority level estimates on an annual basis. Similarly, questions might only be asked of "1/3" of the sample on a "biennial" basis (i.e. asked every second year). Such questions could only get national level estimates.

Figure 3.1: Representation of multi-year core and modular design



A simplified version of the questionnaire can be found in the publications section of the SHS website[11]. This includes details on how questions are allocated to the different streams. The broad topic areas, however, are described in the following section.

[10] Scottish Government Survey Harmonisation (Core Questions): http://www.scotland.gov.uk/Topics/Statistics/About/SurveyHarm

[11] https://www.gov.scot/collections/scottish-household-survey/#questionnaires

3.2.1. Household

The household reference person, who is the Highest Income Householder (HIH) or their spouse/partner completes part 1 of the interview ('Household'). Details of all members of the household, including children, are collected during the household interview. This includes questions related to the composition and characteristics of the household, and involves capturing basic demographic information of all members of the household, such as gender, age and economic situation at this stage, as well as detailed information on dwelling characteristics as captured through the old SHCS. The topics covered in the Household section of the survey are presented in Table 3.1.

Table 3.1: Topics covered in SHS 2019 Household component

Household Composition	People living in household, basic demographics
Accommodation	Property type, Tenure & Length of tenure, Ownership of property, Previous home
Household Services	Number of bedrooms, Internet access, Food waste/recycling
Driving and Transport	Cars, Fuel spend, Bicycles
Children and Young People	Schools and travel to school, childcare costs
Health and Disability	Long-term health condition/illness
House Condition Survey	Repairs, Satisfaction with accommodation, Adaptations, Heating (including heating patterns, control, cost and suitability), Energy efficiency and Insulation
Household Employment	Employment details including status, working patterns, type of work
Household Income	Householder/Spouse paid/self-employed/other jobs, Benefits, Other sources

Household Finances	Household financial management and material deprivation
Mortgages and Rent	Mortgage and rent, Household costs including service charges and council tax
Childcare and childcare costs	Types of childcare used, amount of childcare used and costs, reasons, satisfaction with funded childcare

Subsequently a child is selected from all household members under 16 (the 'Random Child') and the household respondent is asked questions about childcare for that child. A child who is at school is also selected (the 'Random School Child') and the household respondent answers questions about the school that child attends and the journey they make to go there[12].

3.2.2 Random Adult

Once the composition of the household has been established, one of the adults in the household is randomly selected by the computer to complete part 2 ('Random Adult')[13].

This covers the behavioural and attitudinal type questions, such as satisfaction with local services, and captures further demographic information on the random adult. This element also covers the 'Travel Diary' component which asks about travel behaviours on the day previous to that of the interview day. In all households with a single adult the same person completes both the household and the random adult sections. As the number of adults in the household increases, the probability of the random adult being the same as the household respondent declines[14].

[12] The random school child may be the same as, or different from, the random child.

[13] Adults who are household members but have been living away for the previous six months are excluded from the selection of the random adult. Children and students living away during term time are counted as household members but are excluded from the random adult and random school child selection.

[14] Where the same person completes both parts one and two (i.e. they are both the household respondent and selected as the random adult) the CAPI (Computer Assisted Personal Interviewing) script does not repeat the questions common to both sections. This means that these respondents are not asked for the same information twice.

The topics covered in the Random Adult section of the survey are presented in Table 3.2.

Table 3.2: Topics covered in SHS 2019 Random Adult component

Adult Characteristics	Basic demographics, Country of birth and date of entry
Accommodation	Housing experiences
Neighbourhoods and Communities	Rating, Belonging, Crime & Police performance, Greenspace, Anti-social Behaviour, Safety, Discrimination and Harassment, Involvement with Neighbours
Education and Training	Qualifications
Internet	Use, Methods, Public sector, Non-users, Cyber crime
Travel and Transport	Licence, Driving, Electric cars, Travel to work/education, Congestion, Car sharing, Air travel, Public transport & Incidents, Journey planning, Road accidents, and Travel diary
Volunteering, Local government & services, Culture & Sport, Environment	Vounteering, Perceptions of local government, Perceptions of local services, Culture, Sport, Views on Climate change, Access to outdoors
Health, Disability & Caring	Self-assessed health, Disability, Caring responsibilities, Concessionary travel
Employment	Employment status & Government work scheme

3.2.3 Physical survey

At the end of the Household component of the survey, the HIH is asked if they would be willing to have the follow-up component 'Physical Survey' of the dwelling arranged. Such surveys are conducted by professional surveyors through a visual inspection of the dwelling. The surveyor will assess the condition, design and energy efficiency of the home, with much of their time spent surveying the outside, but they will ask to see all the rooms inside. Results from the Physical Survey are reported on separately, for the 2019 data please see the 2019 SHCS Key Findings Report[15].

3.3 Survey fieldwork

Before the first interviewer visit, households were sent an advance letter and leaflet outlining the purpose of the survey and the importance of participation. Interviewers were given the advance letters to post themselves in order that the letter would arrive a day or two before their first call. This helped to ensure that householders were likely to be aware of the letter and leaflet when the interviewer first visited.

The main fieldwork for the survey has an on-going monthly cycle. Interviewers were 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 had been wrongly completed or where the required number of call-backs had not been made, there was an on-going programme of reissuing 'non-contacts' in a bid to maximise the response rate.

On occasion, when an interviewer arrived at a sampled address, they would find that an address comprised more than one dwelling or household. In these cases, interviewers were required to select a household/dwelling at random using a Kish grid.

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.

[15] Scottish house condition survey: 2019 key findings: https://www.gov.scot/publications/scottish-house-condition-survey-2019-key-findings

The social survey fieldwork for the 2019 sample began in January 2019, with the aim of completing re-issues and 2019 fieldwork as a whole in February 2020. The response rate target was set at 65%. The social survey fieldwork was completed in February 2020. Physical survey fieldwork usually finishes one month after social survey fieldwork. The 2019 physical fieldwork was completed in early march 2020 and was not impacted by the coronavirus pandemic.

4 Physical Fieldwork and Physical Survey Form

Summary

- There were three different types of Physical Survey: full surveys, dwelling descriptions, and abbreviated dwelling descriptions.
- All surveyor appointments made by interviewers were allocated for a full physical survey.
- Vacant dwellings and dwellings where an interviewer had not made contact with a householder were allocated to surveyors for a dwelling description.
- For addresses where a respondent had refused to undertake a social interview, surveyors were asked to undertake an abbreviated dwelling description. This type of survey only collected information on the age of the dwelling and the type of dwelling. If this information could not be collected from a public road, they were instructed not to complete any information at all and return a “non-survey”.
- The SHS physical survey is a dwelling-based survey of the home and surrounding area and uses a 14 page paper form formatted for use with digital pens.

4.1 Physical survey team

The physical survey team comprised of 60 surveyors and 4 Regional Managers. The Regional Managers also acted as surveyors.

Overall, 38 of the surveyors and all Regional Managers had worked on the SHCS under the previous contract, and many had been with the team since the 2002 survey. A one-day refresher briefing with the returning surveyors was undertaken in February 2018. This focused on a number of areas including the development of the SAP methodology, the physical survey’s use in policy development and workshops on maintaining the conversion rate.

In early 2018, 24 surveyors were recruited to the physical survey team. Surveyors were required to be fully professionally qualified. They were recruited from a variety of different dwelling-related professions: chartered surveyors, architects, civil and structural engineers, environmental health officers and building control officers.

New recruits were required to attend a five-day residential training course, which incorporated fieldwork practise, so that all were fully proficient with the methodology used in the SHCS. The training was led by representatives by Ipsos MORI with the Scottish Government providing support and input.

The role of the Regional Manager was to ensure the quality of the surveyor data. This included: the completion of the physical inspections; the use of the surveyor appointment system; return of all work and expenses; the validation of the physical survey forms; and that the contractual obligations of the surveyors were being met. They oversaw the work of each of their surveyors, provided technical advice, attended surveyor briefings, and ensured that surveyors maintained quality and timeliness of output throughout the period of the survey. Regional Managers accompanied all new surveyors on at least two inspections.

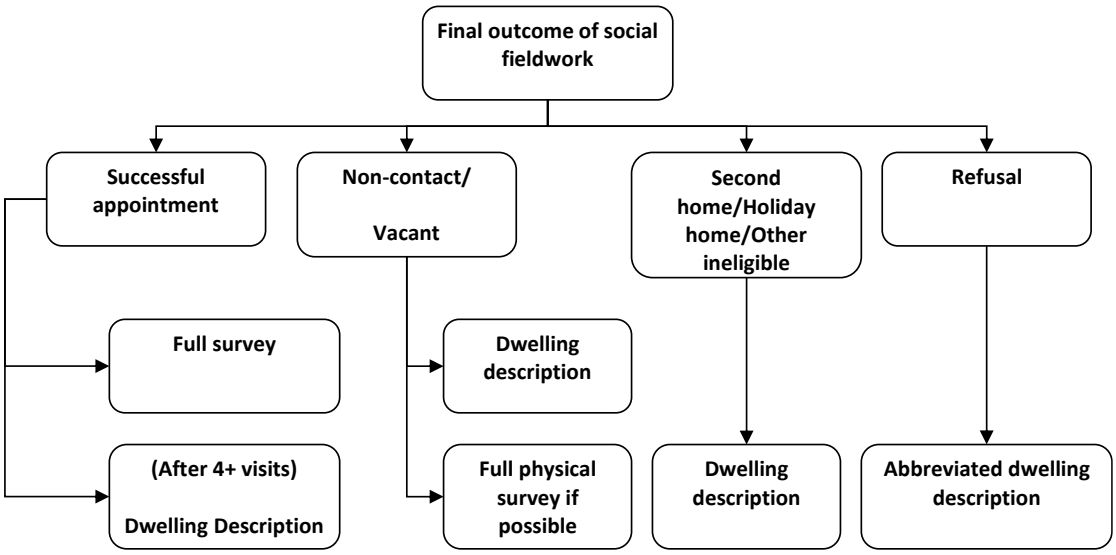
4.2 Types of physical survey

There were three different types of Physical Survey:

- full surveys,
- dwelling descriptions, and
- abbreviated dwelling descriptions.

The type of survey required by the surveyors was determined by the outcome to the social interview (see Figure 4.1).

Figure 4.1: Relationship between social outcomes and type of physical survey required



All surveyor appointments made by interviewers were allocated for a full physical survey. Only a completed social survey interview with a full physical survey constituted a paired case.

A full physical survey is a visual inspection of both the inside and outside of a property. The surveyor is required to complete all parts of the physical survey form. Surveyors were required to take four photographs to accompany each full physical survey. The photographs are of the front and the back of the property and two of the surrounding area. The photographs were used in the data validation process (see Chapter 5).

In a small percentage of cases, the appointment made for the surveyor visit was broken by a respondent. In these instances, surveyors were required to make a further 3 visits, with at least one visit during a weekend and one in the evening, in order to try to obtain a full survey. After 4 unsuccessful attempts to obtain a full survey, surveyors were required to complete a short dwelling description survey. This was a short physical survey that provides a summary of the property only.

Vacant dwellings and dwellings where an interviewer had not made contact with a householder were allocated to surveyors for a dwelling description. Surveyors were required to take one photograph of the property for a dwelling description and only required to make 1 visit to these addresses. On occasions, however, surveyors would make contact with a householder at these addresses. In these instances, they were asked to attempt to gain agreement for a full physical survey and pass contact information on to the fieldwork department for a social survey to be organised. An interviewer would then return and undertake the social survey, thereby completing a paired case.

Addresses out of scope of the survey, such as second homes and holiday homes, were also allocated for a dwelling description. This information is used to help derive the calibration totals which are based on the findings from the survey such as household type and age. For these addresses, surveyors were not required to attempt to try to obtain a full physical survey.

For addresses where a respondent had refused to undertake a social interview, surveyors were asked to undertake an abbreviated dwelling description. This type of survey only collected information on the age of the dwelling and the type of dwelling. If this information could not be collected from a public road, they were instructed not to complete any information at all and return a “non-survey”.

4.3 Physical survey administration

The administration of the physical survey was as follows:

- At the end of the social interview, interviewers attempted to arrange a firm appointment for the surveyor inspection. Appointments were generally made for between 7 and 14 days after the interview date. Interviewers were asked to make appointments in batches, as far as possible, at intervals of one hour plus travel time between addresses. Interviewers left an appointment card with respondents that gave the appointment time and the telephone number of CA Design Services in case they wished to reschedule the appointment.

- When a respondent was unable to commit to a firm appointment time, interviewers were instructed to put in a dummy appointment time, collect the respondent’s contact details and indicate that this was not a firm appointment. CA Design Services would then attempt to arrange a surveyor appointment.

- Following download of the CAPI data, details of the appointments were automatically transferred to CA Design Services secure web-based surveyor appointment system. Information sent included the date and time of the appointment, contact details, whether it was a firm appointment, and any other information that the interviewer deemed helpful to the surveyor (such as directions to the property).
- Details of addresses that did not result in a social interview were communicated to the CA Design Services website for allocation for an appropriate type of survey.
- CA Design Services staff then allocated appointments to surveyors. In advance of each of the fieldwork periods, surveyors were required to supply details of their general availability through CA Design Services' web-based surveyor appointment system to help with the allocation.
- In cases where the initial appointment was not met, surveyors were required to make a further three repeat visits.
- Completed surveys were uploaded onto the SHS physical survey validation system, checked by the surveyor, and then sent to their Regional Manager for sign-off (see Chapter 5).

Staff at CA Design Services' Edinburgh office managed the day-to-day fieldwork process for the physical survey. Helpdesk staff managed communication between respondents and surveyors, booking or re-arranging appointments as necessary. Respondents, social survey interviewers and surveyors were able to contact CA Design Services using a dedicated telephone helpline and an SHS survey email address.

The web-based surveyor appointment system was central to organising and monitoring the progress of the physical survey fieldwork. The website was used by surveyors, Regional Managers, CA Design Services staff and Ipsos MORI. All website users had their own password and were given access to different parts of the site, depending on their requirements.

Surveyors used the survey website to check the appointments that had been made for them, record outcomes of each appointment, record mileage, and to calculate payments due. The progress of individual cases could be viewed on the website by entering the unique case identification number. Additionally, the website system provided information on the progress of the fieldwork overall. Most appointments resulted in a full survey at the first surveyor visit.

4.4 Surveyor variability

In order to minimise the effect of variability between surveyors in completing the physical survey form, and to minimise the bias that this may have on estimates at local authority level, the physical survey fieldwork was subject to a set of allocation rules. These were developed by Communities Scotland around 2001 and comprised the following rules relating to full surveys:

- Each surveyor must work in at least 2 local authorities in each year of fieldwork and at least 3 LAs over the three-year fieldwork period.
- No surveyor should complete more than 25 per cent of the surveys issued in a local authority per year, with the exception of the local authorities Highland, Na h-Eileanan Siar, Orkney and Shetland. Here the level was set at 33 per cent.
- Each surveyor's allocation should contain a mixture of dwelling types approximate to the profile of the area they are working in, over each year of fieldwork.
- Each surveyor's allocation should contain a balance of urban/rural properties approximate to the profile of the area they are working in, over each year of the fieldwork.
- Each surveyor should conduct no more than a maximum number of surveys over each year of fieldwork. This maximum was set as 1.5 times the average number of full surveys issued each year.

4.4.1 Compliance with surveyor allocation rules

Most of the surveyor allocation rules relating to the physical survey fieldwork during 2019 were met.

Rule 1: Each surveyor must work in at least two local authorities in each year of fieldwork and in at least 3 local authorities over the three-year fieldwork period. There were no breaches of this rule.

Rule 2: No surveyor should do more than 25 per cent of the (full) surveys issued in any local authority in any one year, with the exception of the Highlands and the three island local authorities, where no one surveyor should exceed 33 per cent of all (full) surveys. There were 3 breaches of this rule. In two local authorties, there was one survey over target (Angus and East Renfrewshire) and in one local authority, the target was breached by 4 surveys (Moray).

Rules 3 and 4: Each surveyor's allocation should contain a mixture of dwelling types and a balance of urban/rural properties that approximate the profile of the area in which they are working in over each year of fieldwork. Table 4.1 shows the proportion of full surveys conducted by surveyor and property type. It confirms that each surveyor undertook surveys in a mixture of different dwelling types.

Rule 5: Each surveyor should conduct no more than a maximum number of 1.5 times the average number of full surveys issued to each surveyor each year. For 2019, the maximum was set at 72. There were four breaches of this rule with surveyors exceeding this number by just a few surveys. The maximum number of surveys completed by a single surveyor was 79.

Table 4.1: Full physical surveys by surveyor and dwelling type

Row percentages

Surveyor	Terraced/corner house	Detached	Semi-detached	Tenement flat	Other	Total
1	31%	24%	17%	10%	17%	100%
2	16%	38%	16%	24%	7%	100%
3	22%	47%	16%	10%	6%	100%
4	26%	15%	19%	28%	13%	100%
5	15%	51%	15%	9%	9%	100%
6	17%	23%	13%	33%	13%	100%
7	23%	27%	18%	22%	10%	100%
8	12%	67%	14%	0%	8%	100%
9	10%	32%	32%	20%	7%	100%
10	25%	21%	17%	23%	13%	100%
11	26%	21%	34%	12%	7%	100%
12	24%	24%	12%	24%	18%	100%
13	20%	23%	31%	13%	13%	100%
14	19%	12%	17%	35%	17%	100%
15	19%	28%	28%	9%	16%	100%
16	22%	22%	30%	22%	5%	100%
17	17%	54%	18%	4%	7%	100%
18	21%	31%	31%	5%	13%	100%
19	11%	57%	20%	5%	7%	100%
20	23%	39%	15%	13%	10%	100%
21	14%	31%	21%	14%	21%	100%
22	16%	23%	14%	25%	21%	100%
23	24%	22%	29%	12%	12%	100%
24	11%	35%	32%	8%	14%	100%
25	15%	23%	34%	13%	16%	100%
26	13%	21%	45%	5%	16%	100%
27	26%	32%	19%	15%	8%	100%
28	8%	17%	21%	33%	21%	100%
29	15%	32%	29%	8%	16%	100%
30	23%	30%	21%	16%	9%	100%
31	31%	18%	20%	18%	12%	100%
32	20%	26%	20%	16%	18%	100%
33	19%	3%	19%	32%	26%	100%
34	19%	17%	23%	23%	19%	100%
35	16%	30%	11%	27%	16%	100%
36	18%	18%	27%	30%	6%	100%
37	21%	21%	30%	8%	21%	100%
38	21%	52%	18%	6%	3%	100%
39	12%	22%	27%	24%	14%	100%
40	15%	49%	29%	4%	4%	100%
41	16%	42%	26%	11%	5%	100%
42	23%	14%	26%	14%	23%	100%

43	26%	15%	30%	19%	11%	100%
44	17%	3%	17%	38%	24%	100%
45	21%	18%	15%	27%	18%	100%
46	17%	8%	18%	43%	13%	100%
47	29%	24%	16%	21%	11%	100%
48	16%	14%	21%	33%	16%	100%
49	11%	6%	25%	36%	22%	100%
50	20%	47%	22%	4%	7%	100%
51	25%	29%	18%	4%	25%	100%
52	30%	19%	28%	4%	19%	100%
53	30%	19%	21%	24%	6%	100%
54	27%	39%	18%	3%	12%	100%
55	19%	30%	38%	9%	4%	100%
56	12%	33%	19%	19%	17%	100%
57	33%	28%	33%	3%	5%	100%
58	17%	39%	20%	7%	16%	100%
59	24%	30%	24%	9%	12%	100%
60	19%	31%	31%	8%	12%	100%
61	28%	26%	22%	10%	14%	100%
62	9%	13%	32%	19%	28%	100%
63	9%	41%	18%	32%	0%	100%
64	30%	16%	14%	30%	10%	100%
All	20%	28%	23%	16%	13%	100%

4.5 Physical survey form

The SHS physical survey is a dwelling-based survey of the home and surrounding area and uses a 14 page paper form formatted for use with digital pens. There were no changes to the form used in 2018.

The physical survey form can be found in the technical reports section of the current SHCS website[16]. The survey form included sections relating to:

- type and age of the dwelling;
- types of defects;
- types of amenities;
- heating systems and insulation;
- dwelling measurements;
- external construction and materials used;
- external repairs required; and
- Statutory Action and Tolerable Standards.

[16] https://www.gov.scot/collections/scottish-house-condition-survey/#technicalreports

Figure 4.2: Physical survey form example

Q. EXTERNAL CONSTRUCTION / MATERIALS

1. Is more than two thirds of the attached flank wall(s) exposed?	unob. 9	detached 4	yes, both flanks 3	yes, one flank 2	no 1						
Principal external wall of original dwelling (c4)											
2. Principal external wall construction (largest wall area)	unob. 9	other 3	cavity 2	solid 1							
3. Principal external wall construction material	unob. 9	other surveyor notes 8	metal 7	concrete 6	timber 5	blockwork 4	brick 3	clay / earth 2	stone 1		
4. Principal external wall finish	other surveyor notes 8	metal 7	concrete panel 6	concrete block 5	timber 4	brick 3	stone 2	render 1			
5. If stone, stone type	unob. 9	n/a not stone 8	other local stone surveyor notes 4	sandstone 3	whin 2	granite 1					
6. Type of stone coursing	n/a not stone 8	random rubble. 4	random squared 3	coursed squared 2	ashlar 1						
7. Primary external wall thickness	unob. 9	>750mm 3	450-750mm 2	<450mm 1							
8. Has additional insulation been added to principal external walls since built?	internal and external 7	cavity and external 6	cavity and internal 5	external 4	internal 3	cavity 2	none 1				
9. Thickness of insulation added to primary external walls (mm)	unob. 999	Specify									
10. Extent of principal external wall **(specify to nearest 10%)**	unob. 99	10	09	08	07	06	05	04	03	02	01
Secondary external wall of original dwelling (c4)											
11. Secondary external wall construction?	unob. 9	n/a 8	other 3	cavity 2	solid 1						
12. Secondary external wall construction material	unob. 9	other surveyor notes 8	metal 7	concrete 6	timber 5	blockwork 4	brick 3	clay / earth 2	stone 1		
13. Secondary external wall finish	other surveyor notes 8	metal 7	concrete panel 6	concrete block 5	timber 4	brick 3	stone 2	render 1			
14. If stone, stone type	unob. 9	n/a not stone 8	other local stone surveyor notes 4	sandstone 3	whin 2	granite 1					
15. Type of stone coursing	n/a not stone 8	random rubble. 4	random squared 3	coursed squared 2	ashlar 1						
16. Secondary external wall thickness	unob. 9	>750mm 3	450-750mm 2	<450mm 1							
17. Has additional insulation been added to secondaryexternal walls since built?	internal and external 7	cavity and external 6	cavity and internal 5	external 4	internal 3	cavity 2	none 1				
18. Thickness of insulation added to secondary external walls (mm)	unob. 999	Specify									
Principal roof type of whole dwelling											
19. Principal roof type	unob. 9	vaulted ceiling 6	half mansard 5	mansard 4	mono 3	flat 2	pitched 1				
20. Extent of principal roof type	unob. 99	10	09	08	07	06	05	04	03	02	01
21. Principal roof cover	unob. 9	other, specify in notes 7	metal 6	asbestos 5	asphalt 4	felt 3	tiles 2	slates 1			
22. If principal roof cover is slates or tiles	unob. 9	n/a 8	other tiles 5	rosemarys 4	pantiles 3	all other slate 2	scots slate 1				

5 Data Processing

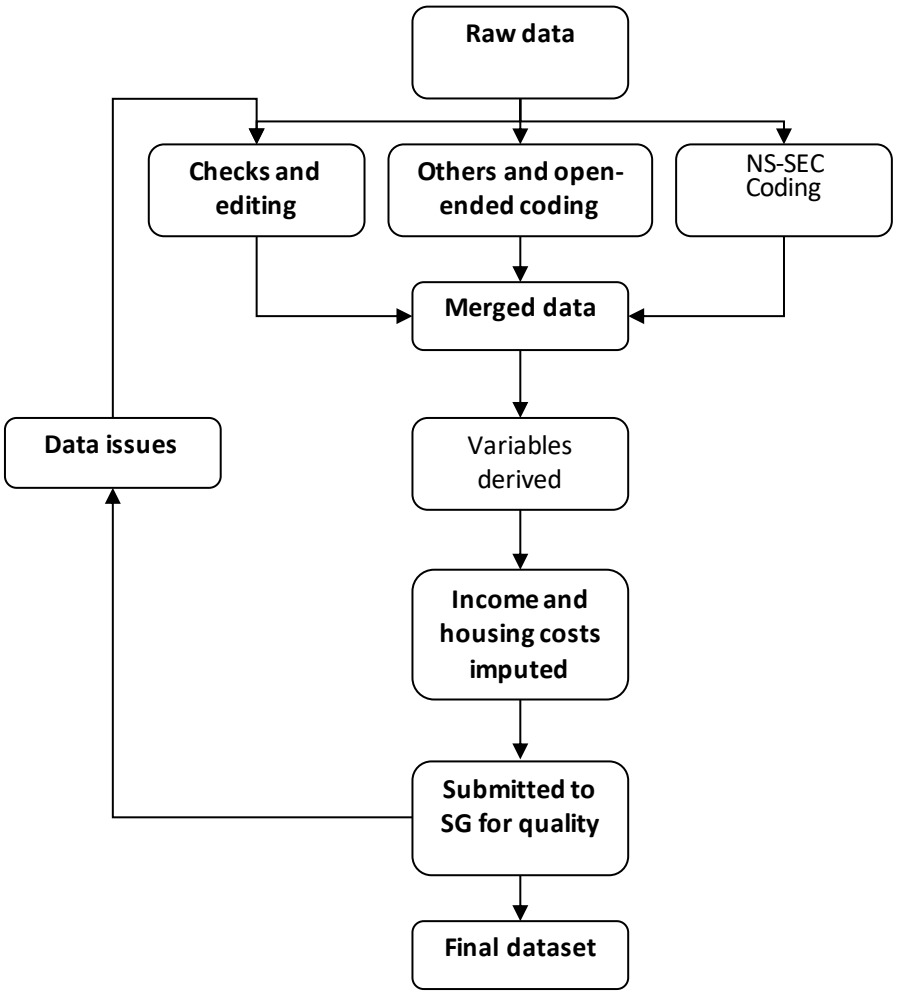
Summary

- Data checks are important to maintain the high quality of the data. The main data file was subject to checks and editing involving range checks, simple logic checks and complex logic checks.
- The data then underwent two additional processes. Firstly, the calculation of derived variables such as the age and sex of the Highest Income Householder and secondly, the imputation of household income.
- Within the SHS, total net annual household income remains the main indicator of household income. A proportion of respondents either did not know how much they received or refused to say how much they received. In order to rectify this non-response, and produce an accurate measure of total net household income, missing values were imputed.
- In 2019, for the first time, missing housing costs were also imputed.
- The edited data was delivered to the Scottish Government, who ran further checks on the data.
- Physical survey data was also subject to a rigorous validation process to ensure the accuracy and validity of each item of data entered. This included range checks on all fields.

5.1 Social data processing

The social data processing routines are summarised in Figure 5.1.

Figure 5.1: SHS social survey data processing procedures



The raw data was initially split into 3 files. Data from the 'other (write in)' variables and open-ended data was extracted for coding separately. Additionally, the variables used to produce NS-SEC variables were extracted into a separate file for coding[17].

The main data file was subject to checks and editing involving:

- Range checks, confirming that all variables were within the acceptable limits established for the question concerned.

[17] NS-SEC can also be automatically assigned to SEG codes, which allow a degree of backward compatibility with Socio-economic Group.

- Simple logic checks ensuring the relationships between questions were logical. For example, that the number of people answering a filtered question is equal to the number of people giving the appropriate response at the filtering question e.g. if 500 people say they smoke then the number of people giving a response to the number of cigarettes they smoke needs to be 500.

- Complex logic checks. These involved examining the relationships between variables and assessing the logic of combinations of responses. Combinations of age and working status, age and relationships to other household members, for example, were checked to assess the logic of someone aged over 60 years and coded as the child of another household member.

The data then underwent two additional processes. Firstly, the calculation of derived variables such as the age and sex of the Highest Income Householder, and secondly, the imputation of household income and, for the first time in 2019, housing costs. The edited data was delivered to the Scottish Government, who ran further checks on the data. Any data issues identified by Scottish Government were discussed and, where necessary, corrected and the data processing routines were amended.

5.1.1 Imputation of income in social data

Within the SHS, total net annual household income remains the main indicator of household income. Prior to 2018 this was defined as the total income from earnings, benefits and a variety of miscellaneous sources for the Highest Income Householder and their spouse, where applicable, with each component of income collected separately. In 2018 the definition was widened to include up to three other adults in the household, where applicable. In the 2019 survey 17 per cent of households had other adults (i.e. adults other than the Highest Income Householder and their spouse) and it was found that more than half (58 per cent) of these adults had earnings.

A proportion of respondents either did not know how much they received or refused to say how much they received. In order to rectify this non-response, and produce an accurate measure of total net household income, missing values were imputed.

The process used for the Highest Income Household and their spouse was similar to that which has been used in previous years. A modified version of this process was used for the imputation of the missing income of the other adults in the household.

Missing income data was imputed for each component of income separately:

- earnings from main jobs and other jobs[18];
- 40 different benefit components[19]; and
- ten different components of miscellaneous income.

Before starting the imputation process, the raw data was fully cleaned. For income from benefits, the upper limit of entitlement for each benefit was calculated. Any cases which were above these thresholds were examined, and edited if necessary. It is possible that respondents over-estimate income from one source of benefit and under-estimate income from another. Therefore, in cases where the benefit level was marginally above the threshold, the amount was not edited, but the case was excluded from use as a donor case in the imputation process.

Unlike benefits, clear rules do not exist regarding upper and lower limits of earnings and sources of miscellaneous income. These were examined against key indicators - such as tenure, NS-SEC, and description of employment - and were either edited or excluded from the imputation process.

Imputation of earnings has the largest effect on total net household income because of the proportion of cases with missing earnings data and the fact that earnings are commonly the main source of household income. For earnings from main and other jobs, imputed values were calculated using Hot Deck imputation. In Hot Deck imputation respondents are sorted into imputation groups according to likely determinants. Cases with missing data are donated values from cases with data which are in their imputation groups, according to the characteristics chosen. The determinants were selected from a regression model that related earnings to a set of explanatory variables, such as age and sex, full-time or part-time employment, car ownership, tenure, receipt of means tested benefits, and NS-SEC. When selecting the determinants for Other Adults variables such as the relationship to the Highest Income Householder were also used.

[18] For Other Adults the questionnaire made no distinction between main and other jobs. Furthermore, if the respondent was unable to provide an estimate of the exact amount of earnings they were asked if they could provide a banded estimate.

[19] It is worth noting that the list of benefits asked about was substantially revised in 2018. Overall, 40 benefits and other sources of income such as welfare loans were included in four sets.

Imputation of income from benefits was undertaken for each benefit separately. For benefits which were received by only a few people, no modelling could be undertaken and the median value of receipt for these benefits was imputed. For other benefits which are received by a significant number of respondents (e.g. State Retirement Pension), Hot Deck imputation was used, with the imputation classes reflecting the entitlement rules as closely as possible. Imputation of income from benefits was undertaken after imputation of earnings and other sources of income, as income from benefits can be dependent on the income of the household.

Imputation of miscellaneous income was undertaken separately. Most miscellaneous sources of income were received by a small number of respondents and no modelling could be undertaken. The median value of receipt was imputed for these components. For components where modelling could be undertaken (e.g. Investment income, non-State pensions and Student loans) - Hot Deck imputation was used, with the imputation classes based on the variables in the models that had the most explanatory power.

Following imputation, income from all components were summed to create a total net household income variable. All households with a net total household income were set to ‘missing’ if the computed figure was less than £25 a week. Although a small proportion of households will have had a lower income than this – and be living off savings or loans – it is likely that some households will have either under-reported receipt of benefits or earnings, or the imputation process has resulted in a low value being given.

Overall, imputation was undertaken for one or more component in 50% of households. After imputation, household income was missing for 4% of households.

With imputation, there is a danger that the donor groups may differ from those with missing information. While this factor can be minimised with careful specification, it can never be totally excluded. In order to guard against analyses that might be sensitive to the imputation procedures, a set of flag variables were created in order that analysts could identify cases and components where income had been imputed.

For full details of the processing routings users are encouraged to read the memorandum on income imputation in the SHS that is a supporting file accompanying this publication.

5.1.2 Imputation of housing costs in social data

The Scottish Household Survey collects data on how much households pay in housing costs (mortgage and rent). Prior to 2019, missing values were not imputed and unlikely outliers were not edited. As these costs are used in the calculation of total net household income after housing costs, for the first time in 2019, processing routines have been implemented to impute missing values and edit outliers.

For full details of the processing routings users are encouraged to read the memorandum on SHS housing costs imputation that is a supporting file accompanying this publication.

It is worth noting that the extent of missing values for housing costs is less than for income. In 2019 mortgage costs were imputed for 17% of respondents and rent costs were imputed for 23% of respondents. For comparison, imputation was undertaken for one or more component of income in 50% of households.

5.2 Physical survey data validation

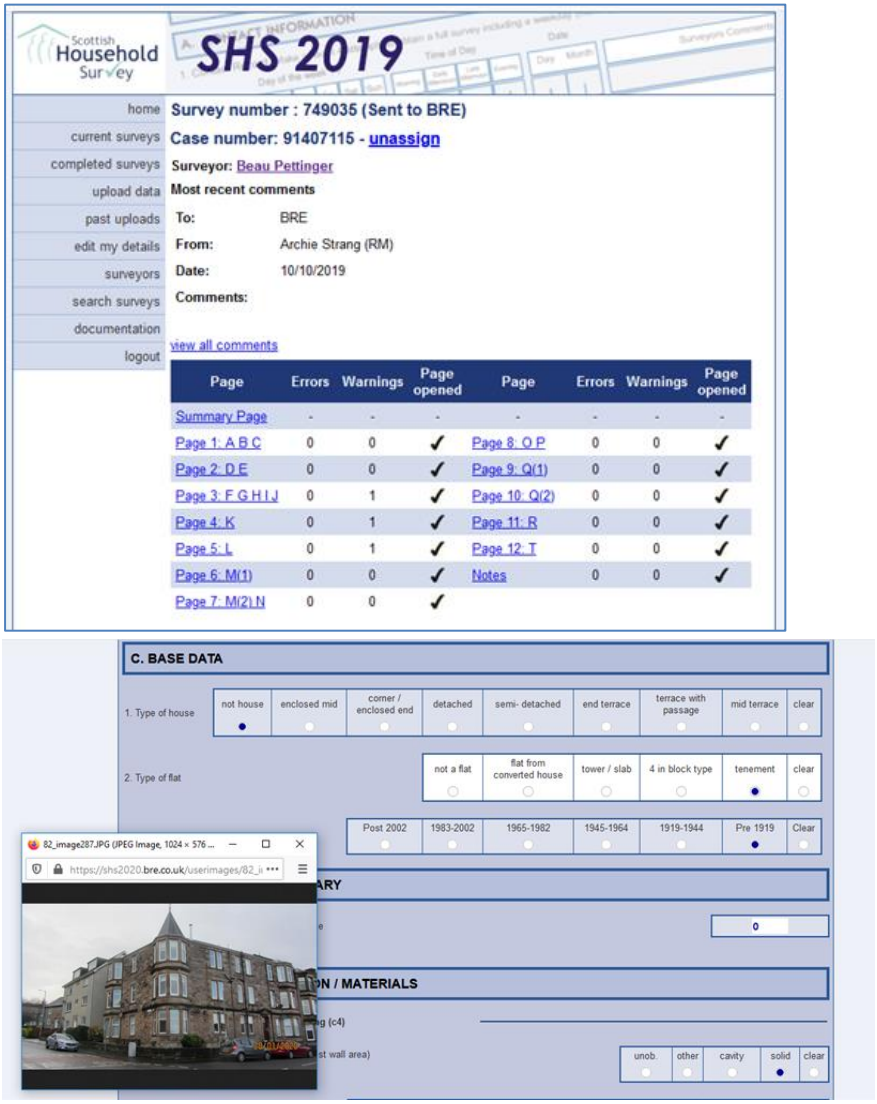
The data from the physical survey forms were uploaded into the physical survey validation system together with the photographs of each dwelling.

The validation system worked by applying a set of rules (the same rules as used in previous years) provided by the Scottish Government, to the raw data, to ensure the accuracy and validity of each item of data entered. This included range checks on all fields, detailed consistency checks making use of the redundancy built into the survey schedule and plausibility checks on all appropriate items. Rules cross-reference different parts of the survey form (e.g. if the dwelling is a house, then aspects of common dwelling section should not be completed; if the house is a flat, then details for common parts should be present).

Surveyors were shown a list of all the errors picked up by the validation program. Additionally, they were shown a list of all the entered data, with a description of the variable next to each bit of data, and with the data split into representations of each page of the form. The validation system showed the data and the failed edits as well as showing the photographs of the property.

Corrections were then made and each form rechecked until it passed all edits. Changes to the data were made simply by overtyping the incorrect data where it was displayed. Once a surveyor had completed validation, the data was forwarded to their Regional Manager for sign-off. Validation of each form was completed when all errors had been eliminated or a supervisor had determined that the dwelling genuinely falls outside the validation criteria. An audit trail of changes made to the data was kept.

Figure 5.2: Physical data validation screens



6 Survey Response

Summary

- The final number of social survey interviews achieved was 10,577. This represents a response rate of 63 per cent and exceeds the target of 10,450 interviews.
- The SHS response rate has fallen in recent years, having fallen by three percentage points since 1999/2000. The response rate of 63% in 2019 was one percentage point lower than the previous four years and four percentage points lower than the 2014 response rate of 67%.
- The long-term average response rate for 1999-2011 was 67.9%. However, it should be noted that the calculation has changed slightly for 2012 onwards as a portion of the addresses of unknown eligibility are considered to be eligible whereas previously they would all have been classed as ineligible. This calculation change would have led to a lower response rate in years prior to 2012, if it had been calculated on the same basis.
- There was a target of at least 80 completed physical surveys for each local authority along with a target of 3,004 surveys for Scotland.

6.1 Introduction

This section presents the fieldwork outcomes for the sampled addresses.

The final number of social survey interviews achieved was 10,577. This represents a **response rate of 63 per cent** and exceeds the target of 10,450 interviews. Since the design changes to the SHS were introduced in 2012 the target number of interviews has been met in recent years, namely 2017 to 2019. However, the SHS response rate has fallen by three percentage points since 1999/2000 and the impact of this drop in response rate is discussed in section 6.3. The performance of the physical survey is described in section 6.5.

The SHS response rate has fallen in recent years, having fallen by three percentage points since 1999. The response rate of 63% in 2019 was one percentage point lower than the previous four years and four percentage points lower than the 2014 response rate of 67%. The maintenance of the response rate in a climate of declining response rates on other population surveys is commendable.

Previously the contractual required response rate was 67%. However, in 2017, the SHS steering group agreed to lower the target in the new contract. The group agreed on a 65% rate for the contract specification 2018-2021. The 2018-2021 contract states that "the target response rate for the survey will be at least 65%".

The SHS 2018-2021 contract states that "The household response rate assumption for each Local Authority will be updated annually and set as the mean response rate for each Local Authority over the last three sweeps of the survey for which response rate data are available. Response rate here is the successful completion of the household interview." This is subject to the following conditions: (i) the response rate assumption for any Local Authority will not be below 55%; (ii) the response rate assumption for any Local Authority will not be above 80%; and (iii) the Scotland level response rate will not be below 65%. If the third condition is at risk of being breached then each Local Authority's response rate assumption will be uniformly increased to ensure the Scotland level response is 65% or above for each survey year.

Survey response is an important indicator of survey quality as non-response can introduce bias into survey estimates. Standardised outcome codes (based on an updated version of those published in Lynn et al (2001)[20]) for survey fieldwork were applied across the SHS, SHeS and SCJS. The outcome codes paper includes guidance on the appropriate categorisation of interview outcomes. This will allow consistent reporting of fieldwork performance and effective comparison between the performance of the surveys.

[20] Lynn, Peter, Beerten, Roeland, Laiho, Johanna and Martin, Jean (October 2001) 'Recommended Standard Final Outcome Categories and Standard Definitions of Response Rate for Social Surveys', Working Papers of the Institute for Social and Economic Research, paper 2001-23. Colchester: University of Essex.

6.2 Scotland-Level Summary

Table 6.1 shows a detailed breakdown of the SHS response for all sampled addresses for Scotland. The addresses of unknown eligibility have been allocated as eligible and ineligible proportional to the levels of eligibility for the remainder of the sample. This approach provides a conservative estimate of the response rate as it estimates a high proportion of eligible cases amongst the unknown eligibility addresses.

The table shows that the overall household response rate was 62.6 per cent. This is below the long-term (1999/2000 to 2011) average response rate for the SHS of 67.9 per cent, below the 2012 to 2014 rates (67%) and below the 2015 to 2018 rates (64%). The effects of the drop in response rate are discussed in section 6.3. It should be noted that from 2012 the calculation was changed slightly and a portion of the addresses of unknown eligibility are now considered to be eligible, whereas previously they would all have been classed as ineligible.

Table 6.1: Fieldwork outcomes, Scotland

Fieldwork Outcome	Sample	% of issued	% of eligible
Responding households	**10,577**	**57.4%**	**62.6%**
Random adult interview complete	**9,776**	**53.0%**	**57.9%**
Refused			
Office refusal	733	4.0%	4.3%
Refusal at introduction/before interview	3,806	20.7%	22.5%
Broken appointment - no re-contact	237	1.3%	1.4%
Total refused	**4,776**	**25.9%**	**28.3%**
Non-contact			
No contact with anyone at the address	488	2.6%	2.9%
Contact made at address, but not with target respondent	366	2.0%	2.2%
Total non-contact	**854**	**4.6%**	**5.1%**
Other non-response			
Ill at home during field period	31	0.2%	0.2%
Away or in hospital throughout field period	124	0.7%	0.7%
Physically or mentally unable/incompetent	209	1.1%	1.2%
Language barrier	63	0.3%	0.4%
Lost interview	0	0.0%	0.0%
Other non-response (not covered by categories above)	0	0.0%	0.0%
Total other non-response	**427**	**2.3%**	**2.5%**
Unknown eligibility			
Inaccessible	0	0.0%	
Unable to locate address	57	0.3%	
Unknown if occupied, due to non-contact	209	1.1%	
Other unknown eligibility	4	0.0%	
Total unknown eligibility	**270**	**1.5%**	
Estimated eligible addresses in set of unknown eligibility addresses	253	1.4%	1.5%
Total eligible addresses	**16,887**	**91.6%**	**100.0%**
Not eligible			
Not yet built / under construction	17	0.1%	
Demolished/derelict	51	0.3%	
Vacant/empty	887	4.8%	
Non-residential	241	1.3%	
Address occupied but not resident household	266	1.4%	
Communal establishment / institution	26	0.1%	
Other ineligible	37	0.2%	
Estimated ineligible addresses in set of unknown eligibility addresses	17	0.1%	
Total not eligible	**1,542**	**8.4%**	
All issued addresses	**18,429**	**100.0%**	

6.3 Drop in response rate

The response rate of 63% in 2019 was one percentage point lower than the previous four years and four percentage points lower than the 2014 response rate of 67%. Response rates are shown in the Table 6.2 below.

Table 6.2: Response rates over time

	1999/00	2001/02	2003/04	2005/06	2007/08	2009/10	2011	2012	2013	2014	2015	2016	2017	2018	2019
Response rate	66%	67%	68%	69%	66%	69%	69%	67%	67%	67%	64%	64%	64%	64%	63%
Achieved sample	30,227	30,639	30,822	31,013	27,238	28,404	14,358	10,644	10,652	10,633	10,325	10,470	10,683	10,532	10,577

The long-term average response rate for 1999-2011 was 67.9%. However, it should be noted that the calculation has changed slightly for 2012 as a portion of the addresses of unknown eligibility are considered to be eligible whereas previously they would all have been classed as ineligible. This calculation change would have led to a lower response rate in years prior to 2012, if it had been calculated on the same basis.

Due to a drop in the response rate in 2015, an analysis was conducted to investigate the likely impact of this drop in response rates. This was undertaken by modeling the 2014 sample results to examine what the impact of a 3% drop in response rates would have been had a lower number of interviews been achieved, and comparing this against the 67% full sample results.

Analysis showed that the demographic composition of the sample was largely unchanged (age, gender), with only the most deprived SIMD quintile and other urban areas showing a 1 percentage point drop in their respective shares of the total sample.

Analysis of mean differences in the population estimates from the two samples for a basket of full sample questions from the SHS, and full one third sample questions from the SHCS module, were also undertaken.

This showed that the absolute mean differences for the total population estimates across the different basket of questions within the household and random adult parts of the survey, including the Scotland Performs National Indicators, were very small, at around 0.1 percentage points.

A few sub-group categories had one or two ‘maximum’ differences in estimates of around 1 percentage points, including age and social and private-rented sector sub-groups. However, these differences are unlikely to be statistically significant due to small subgroup sample sizes.

Sub-national analysis was not considered. It is expected that there would be a greater impact of this lower response rate for Local Authorities and other sub-national geographies.

6.4 Local authority performance

Table 6.3 shows levels of ineligible addresses, response rate and random adult conversion.

Table 6.3: Response rate and eligibility by local authority

		Ineligible addresses		Responding households		Random adult interviews	
Local Authority	Sampled addresses	n	%	n	% of eligible	n	% of HH ints
Aberdeen City	677	122	18%	314	55%	304	97%
Aberdeenshire	620	63	10%	363	65%	342	94%
Angus	472	39	8%	251	58%	234	93%
Argyll and Bute	457	89	19%	249	68%	243	98%
Clackmannanshire	372	24	6%	222	64%	213	96%
Dumfries and Galloway	431	43	10%	277	71%	256	92%
Dundee City	411	38	9%	244	65%	219	90%
East Ayrshire	488	31	6%	275	60%	259	94%
East Dunbartonshire	468	13	3%	276	61%	258	93%
East Lothian	431	30	7%	297	74%	273	92%
East Renfrewshire	457	17	4%	259	59%	235	91%
Edinburgh City	1,489	110	7%	871	63%	839	96%
Falkirk	442	19	4%	237	56%	223	94%
Fife	911	68	7%	526	62%	477	91%
Glasgow City	1,987	163	8%	977	54%	853	87%
Highland	648	88	14%	345	62%	329	95%
Inverclyde	435	32	7%	247	61%	236	96%
Midlothian	457	15	3%	291	66%	279	96%
Moray	431	42	10%	256	66%	225	88%
Na h-Eileanan Siar	383	49	13%	278	83%	275	99%
North Ayrshire	481	41	9%	251	57%	233	93%
North Lanarkshire	885	49	6%	522	62%	462	89%
Orkney	367	50	14%	258	81%	246	95%
Perth and Kinross	428	22	5%	237	58%	203	86%
Renfrewshire	478	14	3%	263	57%	249	95%
Scottish Borders	432	42	10%	267	68%	239	90%
Shetland	376	40	11%	244	73%	217	89%
South Ayrshire	449	43	10%	263	65%	256	97%
South Lanarkshire	862	60	7%	480	60%	423	88%
Stirling	376	27	7%	240	69%	212	88%
West Dunbartonshire	426	22	5%	243	60%	236	97%
West Lothian	402	20	5%	254	66%	228	90%
Scotland	18,429	1,525	8%	10,577	63%	9,776	92%

Argyll and Bute, Aberdeen City, Orkney, Highland, Na h-Eileanan Siar, Shetland and Aberdeenshire[21] were the areas where the highest levels of ineligible addresses were recorded. For Argyll and Bute and Na h-Eileanan Siar, high levels of ineligible addresses were expected as both areas contain a high number of holiday and second homes which are not eligible for the survey. Expected levels of ineligible addresses for Highland, Orkney and Shetland were all above 10%.

The two lowest household response rates in 2019 were in Glasgow City (54%) and Aberdeen City (55%). In addition, seven other local authorities had a response rate under 60 per cent in 2019.

The conversion from household interview to random adult completion was 92% in 2019 (this was 92% in 2016 to 2018, 91% in 2015, 92% in 2014 and 93% in 2012 and 2013).

6.5 Monitoring and reducing the respondent burden

The Code of Practice for Statistics[22] states that “Statistics producers should be transparent in their approach to monitoring and reducing the burden on those providing their information, and on those involved in collecting, recording and supplying data.” Furthermore it states that “The burden imposed should be proportionate to the benefits arising from the use of the statistics.”

The following steps are/have been taken to reduce respondent burden.

[21] These are highlighted in red in Table 6.3.

[22] https://code.statisticsauthority.gov.uk

- Addresses selected for any of the surveys (SHS, SHeS, SCJS) are removed from the sample frame for a minimum of 4 years so that they cannot be re-sampled for another survey.

- Samples sizes are no greater than required to obtain robust local authority estimates.

- The size of the SHS questionnaire was not allowed to increase following the recent consultation[23] on its contents. Where there was a strong policy need for new questions, the space/time required for these new questions was created by dropping existing questions or moving them to a biennial basis.

- The SHS social interview is carried out using Computer Aided Personal Interviewing (CAPI). Routing is built into the CAPI script which ensures that respondents are only asked questions that are relevant to them, e.g. respondents are only asked questions on mortgage repayments if they have a mortgage.

- Interviewer briefings are held annually. These events provide interviewers with an opportunity to feedback on improvements that could be made to the questionnaire to assist them conducting interviews and reduce the burden on respondents.

- Interviewers were asked to choose the logo on the tote bag which was given to respondents as an incentive to complete the survey in 2019.

- The Scottish Household Survey team engages its users in the design of new products and in the steering and evaluation of the dissemination of the survey results. The Scottish Household Survey run an annual user day and we normally have around 70-80 attendees every year. We advertise this on the Scotstat email distribution list and through twitter. In addition, we send out annual report evaluation questionnaires to our users to find out views of our publications.

- For large-scale new developments such as the change away from our Excel tables to our new web-based data explorer, we have collected views user through quantitative questionnaires – but also through face-to-face testing and meetings with local authority staff.

[23] https://www2.gov.scot/Topics/Statistics/16002/Publications/QuestReviewFullReport

6.6 Physical survey

As described in section 2, just under half of the SHS sample was assigned to the physical survey module. For completion of the physical survey, respondents had to agree to make an appointment for a surveyor to make a follow-up visit and to complete the appointment. Table 6.4 shows the number of households assigned to the physical module which responded to the main SHS and the conversion to completion of the physical survey.

There was a target of at least 80 completed physical surveys for each local authority along with a target of 3,004 surveys for Scotland. The result of continuing lower than estimated levels of conversion from household survey to physical survey, meant that there were fewer than 80 physical survey responses for ten of the local authorities in 2019[24]. However, the decision to 'open up' an additional stream to the physical survey gave a lower shortfall on the minimum target of 80 per local authority than in previous years (11 local authorities in 2018, 9 in 2017, 16 in 2016, 17 in 2015, 24 in 2014, 21 in 2013 and 20 in 2012). The 2,997 surveys achieved for Scotland in 2019 was 7 fewer than the target of 3,004.

[24] These are highlighted in red in Table 6.4.

Table 6.4: Conversion to full physical survey

Local Authority	Conversion rate from household interview to physical survey	Physical survey complete
Aberdeen City	63.1%	77
Aberdeenshire	59.0%	82
Angus	73.2%	90
Argyll and Bute	60.3%	70
Clackmannanshire	71.0%	71
Dumfries and Galloway	72.0%	90
Dundee City	61.6%	77
East Ayrshire	52.6%	70
East Dunbartonshire	59.4%	82
East Lothian	57.5%	77
East Renfrewshire	62.7%	84
Edinburgh City	57.3%	208
Falkirk	70.3%	90
Fife	59.1%	140
Glasgow City	55.1%	225
Highland	64.7%	86
Inverclyde	53.4%	71
Midlothian	59.2%	90
Moray	66.1%	82
Na h-Eileanan Siar	61.8%	89
North Ayrshire	71.7%	86
North Lanarkshire	51.8%	103
Orkney	82.6%	100
Perth and Kinross	72.1%	88
Renfrewshire	48.2%	54
Scottish Borders	54.2%	77
Shetland	76.6%	85
South Ayrshire	68.9%	82
South Lanarkshire	52.6%	110
Stirling	64.9%	74
West Dunbartonshire	65.1%	82
West Lothian	76.6%	105
Scotland	61.9%	2,997

7 Survey Weighting

Summary

- Ideally, a selected sample is a miniature of the population it came from. This should be reflected in the sample being representative with respect to all variables measured in the survey. Unfortunately, this is often not the case. One of the problems is non-response. It may cause some groups to be over- or under-represented within the responding sample. If such problems occur, no reliable conclusions can be drawn from the observed survey data, unless a correction is made to ensure the sample is representative of the population. A commonly applied correction technique is weighting adjustment. It assigns an adjustment weight to each survey respondent.
- Weighting ensures that the sample in the Scottish Household Survey represents the population of Scotland as a whole.
- The procedures for the implementation of the weighting methodology were developed by the Scottish Government working with the Methodology Advisory Service at the Office for National Statistics
- The weighting procedures for the SHS incorporate a selection weighting stage to address the unequal selection probabilities and calibration weighting to correct for non-response bias. Calibration weighting derives weights such that the weighted survey totals match known population totals.
- Three weights were derived for the main section of the 2019 SHS; a household weight; random adult weight; and a random schoolchild weight. Further weights were required for analysis of the travel diary and physical survey sections.
- In the SHS, there are seven weights that can be used but two of these - LA_WT (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) and IND_WT (used for analysis of data in derived variables about the random adult or collected from the random adult) - are used for most analyses.

7.1 Introduction

This section presents information on the weighting procedures applied to the survey data. From the 2012 SHS onwards, the weighting has been undertaken by the Scottish Government rather than the survey contractor (as had previously been the case), but the methodology applied has been largely consistent with that from previous sweeps of the survey. The procedures for the implementation of the weighting methodology were developed by the Scottish Government working with the Methodology Advisory Service at the Office for National Statistics.[25]

Weighting procedures for survey data are required to correct for unequal probabilities of selection and variations in response rates from different groups. The weighting procedures for the SHS incorporate a selection weighting stage to address the unequal selection probabilities and calibration weighting to correct for non-response bias. Calibration weighting derives weights such that the weighted survey totals match known population totals. For the 2019 SHS the population totals used were the National Records of Scotland's (NRS) "Mid-2018 population estimates Scotland"[26] and for households the NRS "Estimates of Households and Dwellings in Scotland, 2018"[27] were used. For the physical survey weighting household totals were used from the NRS "Estimates of Households and Dwellings in Scotland, 2018"[28]. To undertake the calibration weighting the ReGenesees Package for R was used and, within this to execute the calibration, a linear distance function was implemented.

Three weights were derived for the main section of the 2019 SHS; a household weight; random adult weight; and a random schoolchild weight. Further weights were required for analysis of the travel diary and physical survey sections. The procedures to calculate the weights are described in the following subsections.

[25] A report on the development of the weighting procedures is available here: http://www.scotland.gov.uk/Topics/Statistics/About/Surveys/WeightingProjectReport

[26] NRS, Mid-2018 population estimates Scotland: https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-estimates/mid-year-population-estimates/mid-2018

[27] NRS, Estimates of Households and Dwellings in Scotland, 2018: https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/households/household-estimates/2018

[28] NRS, Estimates of Households and Dwellings in Scotland, 2018: https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/households/household-estimates/2018

7.2 Household weights

This weight is for use with variables that relate to the household. There were three steps to the household weight:

1) Selection weights

The address selection weights were calculated to compensate for unequal probabilities of selection of addresses in different survey strata. For the SHS there were 32 strata – one for each local authority. The address selection weight for each stratum was calculated as the proportion of Scottish households (from NRS estimates) in the stratum divided by the proportion of the sample in the stratum:

$$\text{Stratum selection weight} = \frac{\text{per cent of Scottish households in the stratum}}{\text{per cent of all responding addresses in the stratum}}$$

2) Calibrated household weight

The stratum selection weight was applied to the survey data to act as entry weights for the calibration. The execution of the calibration step then modified the entry weights so that the weighted total of all members of responding households matched NRS population totals for age bands and sex within each local authority.

3) Households adjustment

The calibration step ensured that survey totals matched the population totals for local authorities but not the household totals. To make the sample representative of households at local authority level a scaling factor was applied so that the weighted number of households from the sample matched the NRS local authority household estimates.

7.3 Random Adult weights

Within responding households a random adult was selected to answer individual questions. There were three stages to the random adult weights:

1) Stratum selection weight

A new stratum selection weight is required for the random adults as the reference population is all adults within the stratum. Also, not all random adults in households that completed the household section responded to the survey. The stratum selection weight was calculated as:

$$\text{Stratum selection weight} = \frac{\text{per cent of Scottish adults in the stratum}}{\text{per cent of all responding adults in the stratum}}$$

NRS mid-year population estimates were used to calculate the population percentages for each stratum.

2) Adult selection weight

The probability of an adult within a household being selected for the random adult interview was inversely proportional to the number of adults within a household – i.e. in a single adult household the only adult resident must be sampled but in a three adult household each adult only has a one in three chance of being selected. To correct for this unequal probability of selection an adult selection weight equal to the number of adults in the household was applied.

3) Calibrated weight

The stratum selection weight and adult selection weight were multiplied together and applied to the survey data. The execution of the calibration step then modified these combined entry weights so that the weighted total of responding random adults matched NRS adult population totals for age bands and sex within each local authority.

7.4 Random Schoolchild weights

A separate weight was required for information collected about a random schoolchild within responding households. The weighting procedures for the random schoolchild are similar to those for the random adult:

1) Stratum selection weights

Stratum selection weights were calculated as:

$$= \frac{\text{Stratum selection weight}\ \text{per cent of Scottish school age children in the stratum}}{\text{per cent of all responding school children in the stratum}}$$

NRS mid-year population estimates were used to calculate the population percentages for each stratum.

2) Random schoolchild selection weight

As with the random adult weight, only one child was selected within each household so a selection weight equal to the number of eligible children in the household was required.

3) Calibration weight

Population estimates for the number of schoolchildren resident in each local authority are not available. Population estimates only give estimates in terms of ages and the school census gives the local authority totals for place of schooling rather than residence. Therefore, the population of schoolchildren was estimated using the survey data by applying the household grossing weight to calculate the total number of pupils in each local authority by age group. The selection weights were then combined and applied to the data before the calibration was run to match the random schoolchild totals to the target populations by age group and local authority.

7.5 Travel Diary weight

The travel diary questions were asked as part of the random adult interview. The travel diary collects information on all travel undertaken on the day prior to interview. Over the fieldwork period significantly fewer interviews took place on Fridays, Saturdays and Sundays when compared to other days.

The working status of respondents was also found to vary across day of response, with disproportionately more adults in full-time employment interviewed at the weekend.

These factors resulted in two stages of rescaling the random adult weights for travel diary analysis:

1) To ensure the travel diary was representative of travel patterns for the week as a whole, the random adult weights were rescaled so that the weighted number of interviews was equal for each day of the week. The scaling factor was given by:

$$\text{Day i scaling factor} = \frac{\left(\text{Total interviews}/7\right)}{\text{Weighted (using rand ad weight) total of interviews on day i}}$$

To ensure the travel diary was representative of working status across each day a second scaling factor was derived such that the working status breakdown for each day was equal.

$$\text{Status j and day i scaling factor} = \frac{\left(\text{Total responses of status j}/7\right)}{\text{Weighted (using scaled rand ad weight) total of status j, day i interviews}}$$

The final travel diary weight was then calculated as:

$$\text{Travel diary weight} = (\text{Random adult weight}) \times (\text{Day scale factor}) \times (\text{Day and working status factor})$$

7.6 Physical survey weight

A subsample of the total SHS sample was allocated to the physical survey. This subsample completed a specific module of the SHS in the main interview and received a visit from a fieldworker to conduct a physical survey of the property. Not all of the properties which completed the interview resulted in a completed physical survey. Therefore, two weights are required for the physical survey, one for the interview and one for the completed physical survey paired with the interview. The derivation for both weights followed exactly the same steps:

1) Selection weights

The address selection weights were calculated to compensate for unequal probabilities of selection of addresses in different survey strata. For the physical survey there were 32 strata – one for each local authority. The address selection weight for each stratum was calculated as the proportion of Scottish households (from NRS estimates) in the stratum divided by the proportion of eligible selected addresses in the stratum:

$$\text{Stratum selection weight} = \frac{\text{per cent of Scottish households in the stratum}}{\text{per cent of selected eligible addresses in the stratum}}$$

2) Calibration

The stratum selection weight was applied to the survey data to act as entry weights for the calibration. The execution of the calibration step then modified the entry weights so that the weighted total of responding households matched:

a) The number of households in each local authority

b) Dwelling age at Scotland level

c) Dwelling type at Scotland level

d) Urban-rural classification at Scotland level

The totals for targets (b) and (c) were generated from the sample itself. For almost all of the addresses in the physical survey sample, even where an interview or physical survey were not completed, a visual inspection of the selected address was conducted to record information on dwelling type and age. The Scotland-level targets were then generated from the frequencies for dwelling age and type from the sample weighted with the selection weight. The targets are shown in Table 7.1.

Table 7.1: Physical Survey calibration targets

Category	Calibration target
Dwelling type	
Detached	571,640
Semi	490,135
Terrace/corner	520,746
Tenement	582,465
Other flat	312,289
Total	2,477,275
Dwelling age	
Pre-1919	475,795
1919-1944	271,162
1945-1964	514,611
1965-1982	544,235
1983-2002	378,202
Post 2002	293,270
Total	2,477,275
Urban-Rural	
Large urban	884,363
Other urban	859,148
Accessible small towns	222,891
Accessible rural	272,580
Small remote towns	89,195
Remote rural	149,099
Total	2,477,275

7.7 Summary of weighting variables to apply in analysis

In the SHS, there are seven weights that can be used but two of these – LA_WT and IND_WT – are used for most analyses, with the others used for smaller specific subsets of the sample. The table below shows the different types of weights available for the survey.

Table 7.2: Survey weighting variables

	Calibrated weights	Grossing to population estimates
Main sample		
Household	LA_WT	LA_GRWT
Random Adult	IND_WT	IND_GRWT
Random Schoolchild	KID_WT	KID_GRWT
Travel Diary	TRAV_WT	-
Physical Survey		
Social survey	SWGHT12_N	SWGHT12
Paired social and physical survey	PWGTH12_N	PWGHT12
Household scaling weight	SHSWGHT12_N	

- LA_WT 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.
- KID_WT which is used for analysis of questions related to the random schoolchild (HE1 to HE13N inclusive) (see separate Questionnaire publication).
- TRAV_WT, contained in the Travel Diary data, which is used for analysing that data.

8 Limitations of the Data

Summary

- 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. The factors are sample coverage; sampling variability, the number of cases that analysis is based on and the bias in the achieved sample
- 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).
- 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.
- Although the SHS has a large sample that covers the whole of Scotland, it has some geographical limitations. Users should not use it to undertake geographical analysis below local authority level. Instead, the Scottish Surveys Core Questions should be used for this.
- Users need to be mindful of the sampling errors for analysis but especially when this is based on breakdowns within a single local authority.

Introduction

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 2019, the unweighted sample of adults aged 16 to 34 is 20 per cent and after basic weighting 29 per cent of the sample are adults aged 16 to 34 which is in line with the 2018 population estimate[29] of adults aged 16 to 34. This is an example of bias caused by young adults being difficult to contact or refusing to take part in the survey.

Although the use of calibration weighting addresses the disparity between the age/sex composition of the sample and the known composition of the population, it does so on the assumption that respondents do not differ in terms of survey measures that do not form part of the weighting. The review of the weighting strategy generally found that calibration brought the survey estimates closer to census estimates but like all surveys, the potential for bias remains a limitation that should be considered.

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).

[29] NRS, Mid-2018 population estimates Scotland: https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-estimates/mid-year-population-estimates/mid-2018

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 completely accurate.

- The SHS only collects information from, or about, the Highest Income Householder and, if there is one, their spouse or partner. From 2018 information was also collected for up to three other adults in the household.
- 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 half[30] of households in the sample have missing data for at least one component of income, values have to be imputed.

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).

[30] However, note that for each individual components, the level of item non-response is generally considerably lower.

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.

9 Bias and Data Quality

Summary

- Bias arises in every sample survey.
- Some bias comes from survey design (such as the sampling frame or who is deemed eligible for interview). Bias also reflects aspects of fieldwork outcomes e.g. whether potential respondents can be found at home at times when interviewers call and whether they are able to participate i.e. not restricted by ill health, disability or communication barriers; and the willingness of members of the public to participate in the survey.
- Traditionally, response rates have been used as a proxy measure of survey quality, with a high response rate and a large sample ensuring accurate estimates.
- However, the response rate is not a measure of survey error or bias and its use as such (although widespread) is inappropriate. If 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 weighting strategy employed by the survey is intended to minimise the extent of bias.
- Comparisons with Scotland's Census 2011 and the mid-year population estimates show that the weighted SHS sample appears to be generally robust.
- The survey weighting reduces the difference between the unweighted SHS survey results and the Census 2011 estimates, though differences do still remain.

Introduction

The issue of bias arises in every sample survey. 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 previously:

- 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 i.e. not restricted by ill health, disability or communication barriers; and
- 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, if 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 weighting strategy employed by the survey (described in section 7) is intended to minimise the extent of bias. The issue of residual bias is considered by comparing key results from the SHS with comparator data. The publication of the 2011 Census is the most accurate source of population data which is used by National Records of Scotland (with other sources of data on migration) to produce mid-year population estimates. While the 2011 Census figures are eight years behind the 2019 SHS data, they ought to be comparable as changes in the distribution of age and household types are relatively small year to year.

9.1 Comparisons with Scotland's Census 2011 and mid-2018 Population Estimates

Comparisons with Scotland's Census 2011[31] and the mid-year population estimates show that the weighted SHS sample appears to be generally robust in terms of variables associated with accommodation/property characteristics. Table 9.1 shows that outright ownership appears to be over-represented whilst private rented accommodation is under-represented. The survey weighting reduces the difference between the unweighted SHS survey results and the Census 2011 estimates, though differences do still remain.

This may reflect the difficulties in obtaining interviews with particular sub-groups of the population.

[31] Scotland's Census 2011: Census Data Explorer (Data warehouse) http://www.scotlandscensus.gov.uk/ods-web/home.html

Table 9.1: Comparison of tenure of household between Census 2011 and SHS 2019 estimates

Households	Census 2011	SHS 2019 unweighted	SHS 2019 weighted
Owned	**62.0**	**64.1**	**61.6**
Owned outright	27.8	37.0	32.9
Buying with help of loan/mortgage[1]	34.2	27.1	28.7
Social rented	**24.3**	**23.0**	**23.5**
Council (Local Authority)	13.2	13.5	13.9
Other social rented	11.1	9.5	9.6
Private rented	**11.1**	**11.7**	**13.8**
Other [2]	**2.6**	**1.1**	**1.1**
Base	*2,372,777*	*10,580*	*10,580*

Notes:
1 includes shared ownership (part owned and part rented);
2 includes living rent free and 2011 Census category 'Rented: Other'

When a single adult is randomly selected within households, the unweighted sample of adults always under-represents those living in multi-adult households, since each has a smaller chance of selection for interview. Table 9.2 shows the differences in the unweighted sample and how the weighting has reduced the differences from other estimates. For instance, the unweighted SHS sample contained only 19.8 per cent of adults aged 16 to 34 and the weighting increases this proportion to 29.3 per cent - much closer to both the 2011 Census and the mid-2018 population estimates. The result is that the age/sex distribution of the weighted sample is much closer aligned to the 2011 Census and the mid-2018 population estimates.

Table 9.2: Comparison of age of adults between Census 2011, mid-2018 population estimates and SHS 2019 estimates

Adults	Census 2011	Mid-2018 population estimates	SHS 2019 unweighted	SHS 2019 weighted
All	**100**	**100**	**100**	**100**
16-34	29.7	29.3	19.8	29.3
35-64	50.0	48.0	47.6	48.0
65 plus	20.3	22.7	32.6	22.7
Male	**47.9**	**48.2**	**44.6**	**48.2**
16-34	14.7	14.7	8.9	14.7
35-64	24.4	23.3	22.0	23.3
65 plus	8.8	10.2	13.7	10.2
Female	**52.1**	**51.8**	**55.4**	**51.8**
16-34	14.9	14.6	10.9	14.6
35-64	25.6	24.6	25.6	24.6
65 plus	11.6	12.5	18.9	12.5
Base	*4,379,072*	*4,518,598*	*9,780*	*9,780*

Notes:
1 Scotland's Census 2011: Census Data Explorer (Data warehouse) http://www.scotlandscensus.gov.uk/ods-web/home.html;
2 National Records of Scotland, Mid-2018 population estimates Scotland, table 1 https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-estimates/mid-year-population-estimates/mid-2018

9.2 Total survey error

Traditionally, response rates have been used as a proxy measure of survey quality – with a high response rate indicating good quality. However, the response rate is not a measure of survey error or bias and its use as such (although widespread) is inappropriate. The response rate in a survey indicates the risk of non-response bias in one or more of the variables measured, not that it is actually present in any of them.

Contrary to previous belief, a high response rate does not necessarily create a quality, unbiased survey sample. Indeed, there is growing, strong evidence that the use of reissues to maintain response rates has a very marginal impact on improving data quality. This has been seen in several academic and industry studies over the past decade across a range of surveys. For example, D'Souza et al (2017)[32] state much of the literature finds a weak link between response rate and non-response bias.

Analysis has been conducted, in partnership with Ipsos MORI, to assess the impact of the reissues on SHS data and showed that, after weighting, only a very small number of measures were changed by reissuing and that the scale of the change was small[33]. However, a first issue response rate would have more of an impact on SHCS estimates, with some sub-group differences greater than two percentage points: It concluded that reissuing has a minimal impact on reducing non-response bias in the Scottish Household Survey. While there are differences between households and people who respond at the first issue, these differences do not make a substantial impact on the results of the survey. Sub-national analysis was not considered in this paper. It is expected that there would be a greater impact of lower response rates at the Local Authority and other sub-national geographies.

It should be noted, however, that the SHS covers a wide range of issues and patterns of non-response bias are likely to differ across different measures.

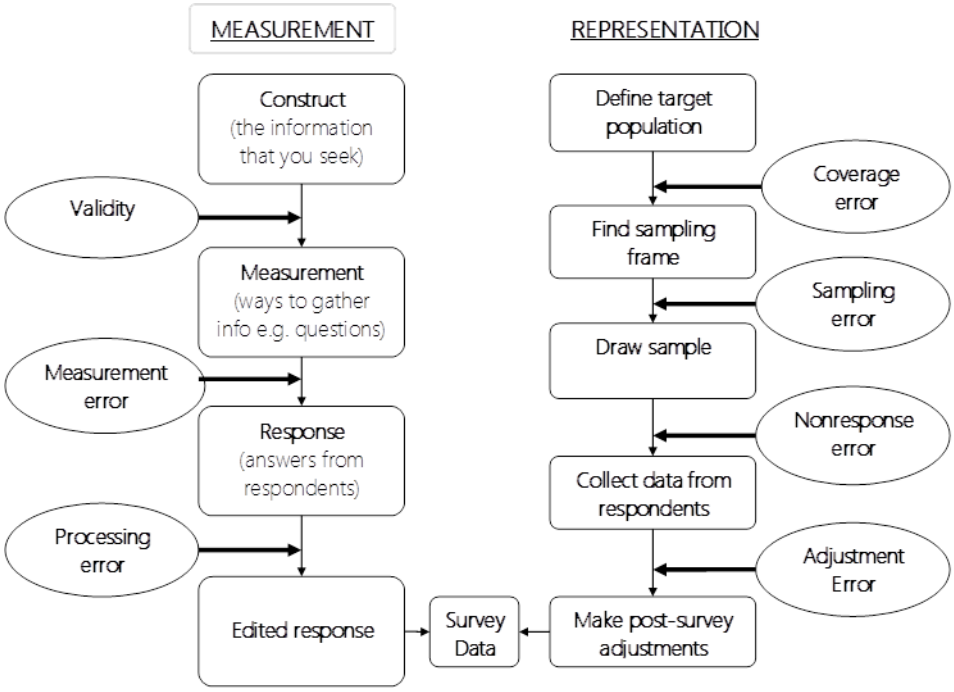
When assessing survey quality, the Total Survey Error framework should not be overlooked. There is a tendency for some sources of error to be overlooked more than other types. The TSE approach methodically identifies all possible errors which can arise at each stage of the survey process. In so doing the survey process is divided into two main strands, a representation strand and a measurement strand. The relationship between survey process and error type is shown in Figure 9.1.

[32] John D'Souza, HM Treasury; Patten Smith and Kevin Pickering, Ipsos MORI; Kathryn Gallop, KSG Research; and Angela Thompson, Royal College of Occupational Therapists Does reissuing unproductive cases in a face-to-face survey reduce non-response bias? Evidence from the Citizenship Survey

'Social Research Practice' - the SRA journal Volume 4, Summer 2017

[33] https://www.gov.scot/publications/scottish-household-survey-response-rates-reissuing-survey-quality

Figure 9.1: The lifestyle of a survey from a quality perspective[1]



The less quantifiable the type of error, the more likely it is to be overlooked. Groves[34], one of the authors of the Total Survey Error framework described this as "the tyranny of the easily measurable". This has tended to mean an over emphasis on errors around representation in Figure 9.1 at the expense of the consideration of validity and measurement error.

[34] Total Survey Error: Past, Present, and Future Robert M. Groves, Lars Lyberg. Public Opinion Quarterly, Volume 74, Issue 5, 1 January 2010, Pages 849–879, https://doi.org/10.1093/poq/nfq065

Published: 01 January 2010

Scottish Household Survey
Help Shape Scotland

Scottish Government
Riaghaltas na h-Alba
gov.scot

This publication is available at **www.gov.scot**

Any enquiries regarding this publication should be sent to us at
The Scottish Government
St Andrew's House
Edinburgh
EH1 3DG

ISBN: 978-1-80004-354-1

Published by The Scottish Government, December 2020

Produced for The Scottish Government by APS Group Scotland, 21 Tennant Street, Edinburgh EH6 5NA
PPDAS799786 (12/20)