

Scottish Health Survey

18

# User Guide

# 1. Background

The data files contain data from Scottish Health Survey 2018 (SHeS18), the fourteenth of a series of surveys designed to monitor trends in the nation's health. Commissioned by the Scottish Government Health Directorates, the series provides regular information on aspects of the public's health and factors related to health which cannot be obtained from other sources. The SHeS series was designed to:

- estimate the prevalence of particular health conditions in Scotland
- estimate the prevalence of certain risk factors associated with these health conditions and to document the pattern of related health behaviours
- look at differences between regions and between subgroups of the population in the extent of their having these particular health conditions or risk factors, and to make comparisons with other national statistics for Scotland and England
- monitor trends in the population's health and health related behaviour over time
- make a major contribution to monitoring progress towards health targets.

Each survey in the series includes a set of core questions and measurements (height and weight and, if applicable, blood pressure, waist circumference, and saliva samples), plus modules of questions on specific health conditions that vary from year to year. Each year the core sample has also been augmented by an additional boosted sample for children. Since 2008 NHS Health Boards have also had the opportunity to boost the number of adult interviews carried out in their area.

The 2018-2021 surveys are being conducted by ScotCen Social Research in collaboration with the Office for National Statistics (ONS), the Social and Public Health Sciences Unit (MRC/CSO SPHSU) at the University of Glasgow, the Centre for Population Health Sciences at the University of Edinburgh and the Public Health Nutrition Research Group at the University of Aberdeen.

This user guide provides an overview of the study and the key elements of the dataset that all users need to know before conducting any analysis. It is designed to be read in conjunction with the documentation supplied with the dataset and the Technical Report to the 2018 Scottish Health Survey Report, which is available online [here](#).

## 2. Survey Design

The 2018-2021 surveys are designed to yield a representative sample of the general population living in private households in Scotland every year.

Those living in institutions, who are likely to be older and, on average, in poorer health than those in private households, were beyond the scope of this survey. This should be borne in mind when interpreting the survey findings.

A random sample of 6,080 addresses were selected from the Postcode Address File (PAF), using a multi-stage stratified design. Where an address was found to have multiple dwelling units, one was selected at random. Where there were multiple households at a dwelling unit, a single household was selected at random. Each individual within a selected household was eligible for inclusion. Where there were more than two children in a household, two were randomly selected for inclusion, to limit the burden on households.

Two further samples were selected for the survey in 2018: a child boost sample (5,448 addresses) in which up to two children in a household were eligible to be interviewed but adults were not, and a Health Board boost sample (273 addresses) for those Health Boards which opted to boost the number of adults interviewed in their area. Information was obtained directly from persons aged 13 and over. Information about children under 13 was obtained from a parent or guardian with the child present.

Data collection involved a main (core) interview, and if applicable, adults also completed the biological module. Of the main addresses issued, 2,027 were flagged as eligible for the 'biological module sample'. At these addresses all adults (16+) that participated in the main interview were eligible to take part in the module. Only interviewers that were specially trained in administering biological measures and samples were allocated these addresses to work on. There was no biological module at the remaining main (core) sample addresses or at child boost or health board boost addresses.

The contents of the Stage 1 interview and the 'biological module' visit are listed in Appendix A.

Interviewing was conducted throughout the year to take account of seasonal differences.

### 3. Key changes to the survey

#### **SHeS 2012-2017**

A number of changes to the survey methodology were introduced following the 2011 Scottish Government review of Scotland's major household surveys. The key changes to SHeS introduced in 2012 for the 2012-2017 surveys were:

- Sample of addresses drawn by the Scottish Government.
- Inclusion of a set of harmonised core questions asked across the three major Scottish Government household surveys.
- Reduction in the achieved sample size.
- Discontinuation of a module of questions on Knowledge, Attitudes and Motivations (KAM) to health.
- Introduction of interviewer administered biological samples and measurements to replace the nurse interview.

These changes are discussed in greater detail in the Scottish Health Survey 2012: Volume 2 Technical Report and in the Scottish Health Survey Questionnaire Review Report 2012-2015.

## **SHeS 2018- 2021**

A number of changes were introduced in light of the 2017 Scottish Government review of the Scottish Surveys Core Questions, and following the 2016 Scottish Government consultation on the Scottish Health Survey questionnaire content which was published in Spring 2017. The key changes implemented in 2018 for the 2018-2021 surveys include:

- Increased sample size allowing for analysis at Local Authority level by 2021.
- Removal of local police force, contraception and cosmetic procedures questions.
- Removal of urine sample from the biological module.
- A number of modules will no longer appear in the questionnaire each year, but will appear approximately every 2 years: gambling, problem drinking, dental health services, parental history, respiratory health, CPR training and use of health services.
- New questions introduced asked about satisfaction with key public services, Nicotine Replacement Therapy (NRT), asthma, type of diabetes and gender identity.

These changes are discussed in greater detail in The Scottish Health Survey 2018 Technical Report.

## **4. SHeS18 Dataset**

The SHeS18 dataset contains data from the main (core) interview questions, the self-completion questionnaires and additional variables derived from the responses to those questions. The biological module data is designed to be analysed when data from two years can be combined. Combined datasets for the following years are also provided through UKDS: 2017/2018; and 2015/2016/2017/2018.

Disclosure control methods applied in the survey were reviewed in 2015. Changes took effect in the 2014 survey and are noted in the variable listing.

## **5. Documentation**

The documentation has been organised into the following sections:

- Interview - contains the CAPI documentation for household and individual questionnaires, self-completion booklets and showcards.
- Data - contains the list of variables in the file, and a list of derived variables with the syntax used to create them.
- Other instructions - contains interviewer and coding & editing instructions

## 6. Using the data

The 2018 data consists of one individual level file and one household level file:

SHeS2018i.sav	6790 records	contains data for all individuals who gave an interview. It contains information from the household questionnaire, main individual schedule, self-completions
SHeS2018h.sav	9355 records	contains data on household, and sex and age of all individuals in co-operating households.

### 6.1 Variables on the files

Each of the data files contain questionnaire variables (excluding variables used for administrative purposes and any variables that are potentially disclosive) and derived variables. The variables included in the individual file are detailed in the “**Variable List**” document in the data section of the documentation. This document is the best place to look at in order to plan your analysis. It includes:

- Major categories of variables (e.g. Drinking, Anthropometric measurements)
- Sub categories of variables (e.g. Drinking in the last 7 days within the Drinking category)
- Source of each variable (e.g. Individual questionnaire, Self-completion, Derived variable etc.)

Once you have decided which variables to include in your analysis, you should look up details of the question wording using the interview section documentation (all variables on the data file are given by name in the copy of the interview schedules provided), or use the “**Derived Variables**” document in the data section of the documentation for the syntax which produced the derived variables. You cannot rely on the individual variable and value labels to always capture the detail of the question asked, or the answer categories used, so reading the interview documentation is essential.

To assist users, particularly those unfamiliar with the survey series, we have produced a guide to the variables used in the tables in the main 2018 report. In most cases these files identify the key variables for each of the main topics covered in the survey.

Existing questions that changed notably between survey years, for example via changes to their wording or response categories, have been given new variable names. These variables are usually suffixed either with a letter such as ‘a’ (e.g. BarSprt1a-BarSpMaia) or with the year of the change (e.g. TvWeek18).

## 6.2 Weighting variables

Weighting has been used to correct for different selection probabilities and for non-response. The non-response weights were designed to adjust for non-contact, refusals of entire households and the non-response of individuals within responding households. Separate weights exist for adults and children. The aim of each set of weights is that the data can be treated as broadly representative of the general household population. The weights were designed so that the weighted age/sex profile of the sample matched the NRS 2017 mid-year household population estimates for Scotland.

Weight name	When it should be used
int18wt	Analysis of items in core questionnaire (adults)
vera18wt	Analysis of items in version A of questionnaire (variables with labels starting "VERA") (adults)
cint18wt	Analysis of child data
cvera18wt	Analysis of version A and child boost data
cmint18wt	Analysis of main interview child data (not boost)
bio18wt	Analysis of bio data

The different aspects of the survey are obviously all linked. The weighting variable selected should always match the dependent variable in the analysis.

Full details of the weighting are provided in the main Technical Report.

## 6.3 Multicoded questions

Some questions in the survey enabled participants to give more than one answer. In the final dataset each of the answer options has been converted into a binary variable with the people who selected that option coded 1 and the rest coded 0.

As an example, question CONSUBX in the adult nurse interview is a "CODE ALL THAT APPLY" question which asks "Have you eaten, smoked, drunk alcohol or done any vigorous exercise in the past 30 minutes?".

The code frame consists of five values:

- 1 - eaten
- 2 - smoked
- 3 - drunk alcohol
- 4 - done vigorous exercise
- 5 - none of these

The five answer options have been converted into five separate binary variables as follows:

CONSUBX1 - code 1: those who ate in the last half hour; code 0: those who didn't.

CONSUBX2 - code 1: those who smoked in the last half hour; code 0: those who didn't.

CONSUBX3 - code 1: those who drank alcohol in the last half hour; code 0: those who didn't.

CONSUBX4 - code 1: those who did vigorous exercise in the last half hour; code 0: those who didn't.

CONSUBX5 - code 1: those who did none of the above in the last half hour; code 0: everyone else.

Because a respondent could have replied with more than one answer, that respondent could have a value 1 for a number of these variables (however, the nature of the question dictates that having a code 1 at CONSUBX5 precludes having a code 1 at any of the variables CONSUBX1 – CONSUBX4). The missing values are the same across all five variables.

## 6.4 Missing values conventions

- 1 Not applicable: Used to signify that a particular variable did not apply to a given respondent usually because of internal routing. For example, men in women only questions.
- 2 Schedule not applicable: Used for variables when the respondent was not of the given age range or sample type.
- 6 Schedule not obtained: Used only for variables on the self-completion schedules this code indicates that a self-completion booklet was not completed for this respondent
- 8 Don't know, Can't say.
- 9 No answer/ Refused

These conventions have also been applied to most of the derived variables, but the derived variable specifications should be consulted for full details.

## 6.5 CAPI routing errors

There were some instances where there was an error within the CAPI questionnaire routing. This affects a limited number of cases within the following CAPI variables: HNotAsk, TrtWze18, SchAb, Wlk5Ch, and UseNRT1c-9d. The affected cases have been assigned the code of -4 for these variables.

## 6.6 Valid cases

In the 2018 Scottish Health Survey report, as in previous reports, cases were excluded from the analysis of anthropometric and other physiological measurements if their measurement was invalid. For example, those who had smoked, drunk or eaten within 30 minutes of having their blood pressure taken were excluded from the main blood pressure variables as this can affect blood pressure.

## 6.7 Derived variables

In addition to the questions and measurements collected directly in the survey, a large number of derived variables have been created for use in the analysis. These variables are sometimes just straightforward recodes of existing variables, for example a summary variable that collapses some categories to make the data more succinct, or a conversion of continuous data (e.g. age) into categories (e.g. age groups). In most cases the derived variables make use of the underlying data in a number of variables to create new variables. For example, the height and weight data is used in combination to derive the Body Mass Index variable. It can sometimes appear to users that there are multiple measures of the same item within the dataset, especially in more complex parts of the questionnaire (e.g. the smoking and drinking sections). In these instances, it is advisable to use the derived variable listing provided, or the listing of variables used in the report, to identify variables for potential analysis, and to refer back to the questionnaires to confirm your selection, rather than to look at the questionnaire documentation first. All derived variable labels start with "(D)" to help distinguish them from other types of variable. Some of the more complex derived variables require the use of look-up tables (e.g. children's BMI groups) and the syntax has not been included. Further information on these variables can be obtained from research team.

## 6.8 Equivalised income

The OECD equivalisation scale used in the Household Below Average Income poverty estimates was used to equivalise incomes in the 2018 survey. This change was introduced in the 2015 data; previous survey years used the McClements scoring system. In the 2015 data, the McClements method was retained alongside the OECD method for continuity, however, the 2018 data only contains the OECD method.

## 6.9 Long-term conditions

The 2015 report presented experimental statistics on multiple conditions, which vary the way in which long-term conditions were defined. Long-term and limiting long-term conditions chapter IV, diabetes and other endocrine and metabolic illnesses were counted separately (compm2a and compm2b), and in chapter IX, stroke, angina, hypertension, other heart problems, and other circulatory system problems were all counted separately (compm7a, compm7b, compm7c, compm7d, compm7e). Thus from 2015, up to 20 different conditions were counted. The number of physical conditions was counted in the same way, but with conditions coded under chapter V of the ICD (mental and behavioural disorders) excluded (derived variable: condphy15).

## 6.10 Drinking guidelines

From 2015, the data contains derived variables for drinking that reflect the revised weekly limits for men (up to 14 units). The new derived variables are denoted by a '15' in the variable name. Some of the DVs reflecting the old guidelines for men have been retained for trend reporting.

## 6.11 Socio-economic classification and social class measures

As detailed in the main technical report, the survey uses the National Statistics Socio-Economic Classification (NS-SEC) introduced in 2001. Information about all adult respondents' employment history is collected and where applicable two NS-SEC codes are derived: one for the Household Reference Person (HRP) and one for the individual respondent. The HRP is the householder with the highest income within the household. In 2018 adult respondents were not asked what their mother and father did for a living when the respondent was 14 so there are no additional NS-SEC codes for both parents

Classification level	Variable name prefix
Household reference person	HPNSSEC
Individual respondent	NSSEC

More information about NS-SEC and RG social class is available from the ONS website [here](#).

## 6.12 Scottish Index of Multiple Deprivation (SIMD)

The 2018 data uses the 2016 Scottish Index of Multiple Deprivation (SIMD). The new SIMD quintiles are indicated by a '16' in the variable name, for example SIMD16\_RPa. More information on SIMD, including how it is calculated, can be found on the [Scottish Government website](#).



## 7. SHeS 2018 Report

The full report, and a set of further tables with selected results for topics not covered in the report, is available on the web at:

<https://www.gov.scot/Topics/Statistics/Browse/Health/scottish-health-survey>

The Scottish Health Survey website also contains a large amount of useful information including the background to the study and plans for future dissemination.

## APPENDIX A

### SCOTTISH HEALTH SURVEY 2018 – CONTENTS

Points to note:

- There are four versions of the questionnaire in the mainstage: Core Version A; Core Version B (biological module); Child Boost; and Health Board Boost.
- Children are not eligible for the biological module in Core Version B or at Health Board Boost sampled addresses.
- The below table indicates what should be in each version and the order of the interview. The associated CAPI block names are in [] after the topic.

Topic	Adults (Core A)	Adults (HB boost)	Children (Core A)	Child Boost
Household questionnaire	•	•	•	•
General health including caring (age 0+)	•	•	•	•
Respiratory symptoms [CVD] 16+	•			
General CVD (16+)	•	•	•	•
Asthma core (0+)			•	
Asthma additional (0+)	•		•	•
Physical & sedentary activity adults (16+) and children (2-15)	•	•	•	•
Additional physical activity questions adults (16+)	•			
Eating habits adults (16+)	•			
Eating habits (2-15)			•	•
Fruit and veg consumption (2+)	•	•	•	•
Smoking and Drinking (16+) [16-19 in a self-completion]	•	•		
Dental health (16+)	•	•		
Economic activity (16+)	•	•		
Education (16+)	•	•		
Ethnicity (0+) and religion (16+)	•	•	•	•
Self-completions (13+ & parents of 4-12 year-olds)	•	•	•	•
Height (2+) and Weight (2+)	•	•	•	•
Consents	•	•	•	•