Scottish Health Survey '16

User Guide

A survey carried out on behalf of The Scottish Government Health Directorates and NHS Health Scotland

ScotCen Social Research

1. Background

The data files contain data from Scottish Health Survey 2016 (SHeS16), the twelfth 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, urine 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 2012-2017 surveys are being conducted by ScotCen Social Research and Medical Research Council Social and Public Health Sciences Unit (MRC SPHSU) in collaboration with 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 2016 Scottish Health Survey Report which is available online here.

2. Survey Design

The 2012 - 2017 surveys were 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 outwith the scope of the survey. This should be borne in mind when interpreting the survey findings.

A random sample of 9,623 addresses was drawn 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. A maximum of 10 adults within a selected household were 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 2016: a child boost sample (4,181 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 (946 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, 1,689 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 Appendix A.

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

3. **Key changes to the survey in 2012-2017**

A number of changes to the survey methodology were proposed during a review by the Scottish Government of all the major household surveys in Scotland. These changes were adopted for SHeS in 2012-2017. The key methodological changes introduced in 2012 were:

Sample drawn by the Scottish Government
Prior to 2012, the contractor for SHeS drew the sample for each survey year. For the 2012-2017 surveys, the sample is being drawn by the Scottish Government in conjunction with the samples for two of the other large population surveys commissioned by the Scottish Government (the Scottish Household Survey and the Scottish Crime and Justice Survey). This approach reduces the burden on households in Scotland as they can only be selected once to take part in one of the three surveys in the 2012-2017 period.

Harmonised core questions
From 2012, there are around 20 core questions that will appear in all three of the Scottish Government population surveys each year. As a result it will be possible to conduct detailed analysis, by small geographical areas and subgroups for these key indicators. Further information about the harmonised core questions can be found here: http://www.scotland.gov.uk/Topics/Statistics/About/SurveyHarm

Reduced sample size
The sample size on SHeS has varied over the years. While in 2008-2011 the aim was to achieve 6,400 interviews with adults and 2,000 interviews with children each year, this has been reduced to a target of 4,000 adult and 1,800 child interviews annually for the 2012-2017 surveys.

Discontinuation of the Knowledge, Attitudes and Motivations to Health module
The NHS Health Scotland funded module of questions on knowledge, motivations and attitudes to health which was included in the 2008-2011 surveys was discontinued in 2012.

Interviewer administered biological samples and measurements
Since its inception in 1995, the SHeS interview included a second stage follow-up visit from a survey nurse to collect biological samples and measures. Prior to 2008 all participants were eligible to take part in this visit and between 2008 and 2011, a sub-sample of adult

participants were invited to take part. The stage 2 nurse visit was discontinued in 2012. Instead, between 2012 and 2017, specially trained interviewers are taking many of the measurements and samples previously collected by survey nurses. As in 2008-2011, only a sub-sample of adult participants are eligible for the additional biological measures and samples. If the eligible participant agrees, they are offered the option of completing the module either at the end of the main interview or at another more convenient time.

Lung function, demispan and hip measurements are not included from 2012 onwards. In 2012 the biological module included:

- blood pressure
- questions on depression, anxiety and self-harm
- waist measurement
- saliva sample
- urine sample

The protocol for taking blood pressure and waist measurement changed in 2012, when the nurse interview was discontinued and trained interviewers started taking the measurement instead. Consequently, the dataset includes two sets of blood pressure and waist circumference data; adjusted and unadjusted figures. The adjusted figures have the same names variable names used in previous years, the unadjusted figures have the suffix _int. A full description of the calibration equation used to make adjustments is available in the Technical report. Some other variable names in this section have also changed to reflect the change in data collection methods. For example the questions on depression, anxiety and self-harm are now asked via computer assisted self-completion when previously they were asked by the nurse.

4. SHeS16 Dataset

The SHeS16 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: 2014/2016; 2015/2016 and 2013/2014/2015/2016.

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, a list of derived variables with the syntax used to create them, and a list of variables used in the main 2016 report tables
- Other instructions - contains interviewer and coding & editing instructions

6. Using the data

The 2016 data consists of one individual level file and one household level file:

SHeS16i.sav	5884 records	contains data for all individuals who gave an interview. It contains information from the household questionnaire, main individual schedule, self-completions

SHeS16h.sav	8027 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 2016 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 changes to their wording or response categories, have been given new variable names including the a figure for the year changed (e.g. “08” or “16”).

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 2015 mid-year household population estimates for Scotland.

Weight name	**When it should be used**
int16wt (adults)	Analysis of items in core questionnaire
vera16wt (adults)	Analysis of items in version A of questionnaire (variables with labels starting “VERA”)
cint16wt (children)	Analysis of child data
cvera16wt (children)	Analysis of version A and child boost data
cmint16wt (children)	Analysis of main interview child data (not boost)
bio16wt	Analysis of bio data
uri16wt	Analysis of urine 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 Valid cases

In the 2016 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.

6.6 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.

6.7 Equivalised income

The OECD equivalisation scale used in the Household Below Average Income poverty estimates was used to equivalise incomes in the 2016 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 2016 data only contains the OECD method.

6.8 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 in 2015 and 2016, 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).

6.9 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.

6.10 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 addition, adult respondents were asked what their mother and father did for a living when the respondent was 14 so there are also additional NS-SEC codes for one or both parents (if sufficient information was collected).

Classification level	**Variable name prefix**
Household reference person	HPNSSEC
Individual respondent	NSSEC
Respondent's father (when R aged 14)	FANSSEC
Respondent's mother (when R aged 14)	MANSSEC
Respondent's parents (when R aged 14) – highest of father/mother	PNSSEC

More information about NS-SEC and RG social class is available from the ONS website here.

6.11 Scottish Index of Multiple Deprivation (SIMD)

The 2016 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 2016 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:

http://www.scotland.gov.uk/Topics/Statistics/Browse/Health/scottish-health-survey/Publications

Users might also be interested in viewing the rest of the Scottish Health Survey website:

http://www.scotland.gov.uk/Topics/Statistics/Browse/Health/scottish-health-survey

This contains a large amount of useful information including the background to the study, the consultation that informed the changes made to the survey in 2008, and plans for future dissemination.

APPENDIX A
SCOTTISH HEALTH SURVEY 2016 – 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 / mop-up.
- 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.

Core Version A	Both A&B	Core Version B
	Household questionnaire [HHgrid]+[GenHHold]	
	General health including caring [GenHlth] 0+	
Respiratory symptoms [CVD] 16+		
	General CVD (16+) and use of services [CVD] 0+	
	Asthma core [Asthma] 0+	
Asthma additional [Asthma] 0+		
	Physical activity adults - including Qs on activity at work, time spent at screens and other sedentary activity [AdPhysic] 16+	
	Physical activity kids – inc Qs on time spent at screens and other sedentary activity [ChPhysic] 2+	
Additional physical activity questions [AdPhysic] [ChPhysic] 2+		
Eating habits adults [Eating] 16+		
	Eating habits kids [Eating] 2 – 15	
	Fruit and Veg [Fruitveg] 2+	
	Smoking [Smoking] 18+ (16-17 year olds do self-comp/18-19 yr olds optional)	
	Passive Smoking [Smoking] 0+	
	Drinking [Drinking] 18+ (16-17 year olds do self-comp/18-19 yr olds optional)	
	Dental health [Dental] 16+	
	Education and employment details 16+	
	Ethnicity (0+) place of birth (0+)and religion (16+) [Ethnic]	
	Family health [Parent] 16+	
	Self-completions [Selfcomp] 4+	
	Height and weight [Measure] 2+	
	Consents [Consents] 0+	

		Biological module (16+)includes: • Prescription drugs • Blood Pressure • Waist • Saliva • Urine • Anxiety • Depression • Self harm

Child Boost
Household questionnaire [HHgrid]+[GenHHold]
General health including caring [GenHlth]
Use of services [CVD] 0+
Asthma core [Asthma] 0+
Asthma additional [asthma] 0+
Physical activity kids – inc Qs on time spent at screens and other sedentary activity [ChPhysic] 2+
Physical activity additional
Eating habits kids [Eating] 2+
Fruit and Veg [Fruitveg] 2+
Passive Smoking [Smoking] 0+
Ethnicity [Ethnic] 0+
Self-completions [Selfcomp] 4+
Height and weight [Measure] 2+
Consents [Consents] 0+

Health Board Boost
Household questionnaire [HHgrid]+[GenHHold]
General health including caring [GenHlth]
General CVD and use of services [CVD] 16+
Asthma core [Asthma] 16
Physical activity adults - including Qs on activity at work, time spent at screens and other sedentary activity [AdPhysic] 16+
Fruit and Veg [Fruitveg] 16+
Smoking [Smoking] 16+ 18/20+ in CAPI
Passive Smoking [Smoking] 16+
Drinking [Drinking] 16+ 18/20+ in CAPI
Dental health [Dental] 16+
Ethnicity and religion [Ethnic] 0+
Family health [Parent] 16+
Self-completions [Selfcomp] 16+
Height and weight [Measure] 16+
Consents [Consents] 16+