Scottish Health Survey

2021

User Guide

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

1. Background

The data files contain data from Scottish Health Survey 2021 (SHeS21), the seventeenth 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 2018-2023 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 2021 Scottish Health Survey Report, which is available online here.

2. SHeS 2021 – survey design and fieldwork

Fieldwork for the 2021 survey was significantly affected by the COVID-19 pandemic. Data collection involved a main computer assisted telephone interview (CATI), and online or paper self-completion questionnaire.

As interviews were conducted by telephone, no height and weight measurements or biological measures could be taken. Participants were asked to estimate their own height and weight during the interview. In previous years, the core version B sample completed a biological module, and these addresses were only assigned to trained bio interviewers. For 2021, as no biological measurement could be taken the only real differences between the core version A and version B interviews were a slightly longer telephone interview for version A to cover the rotating modules and a slightly longer self-completion for version B to cover the depression, anxiety, self-harm and attempted suicide questions which are included in the biological module.

Participants aged 16 and above were also invited to complete an online recall using Intake24 (https://intake24.org/). Participants were asked to provide verbal consent, which was recorded in CAPI. Respondents were also sent a leaflet with some key information about Intake24 included. If the respondent indicated to the interviewer that they had not read the Intake24 information leaflet (sent with their advance letter), the interviewer read out key information from this before obtaining consent. Those who agreed were invited to complete two dietary recalls, either independently or via a phone call with Cambridge University. Further information about Intake24 is provided in the 2021 technical report.

The target sample size for the 2021 survey was the same as that for 2018 and 2019. Because of the requirement for telephone interviewing, with no doorstep recruitment allowed until the end of October, the issued sample was much larger than in previous years.

An initial sample of 64,523 addresses was drawn from the Postcode Address File (PAF) in 2021 on the basis of the survey being conducted by opt-in. These addresses comprised three sample types: main (core) sample version A, main (core) sample version B and the child boost screening sample. This sample was split into 9 monthly waves of fieldwork, from April to December. For the core samples, only the first 6 months were issued, as the sample was superseded by a smaller knock-to-nudge sample for the final three months (the sample for knock-to-nudge was smaller due to higher expected levels of response for this approach). The child boost opt-in sample was used for the whole 9 months.

Replacement core sample (3,839 addresses) was drawn for the last three months of 2021 to be issued as knock-to-nudge.

There were two phases of fieldwork for SHeS 2021. During Phase 1, potential participants were contacted by letter and asked to opt-in to taking part in an interview conducted over the phone. This phase began in April, with new invites being sent out each month until September 2021. The survey itself was more comprehensive than 2020, with similar content to earlier survey years, and unlike 2020, interviews with or on behalf of children were also conducted.

Phase 2 began at the end of October, with new core sample issued across three months. Potential respondents were again contacted by letter, but then recruited to participate by interviewers knocking on their door, in what is termed a ‘knock-to-nudge’ methodology. Interviews were still conducted by telephone, as for Phase 1. This second phase only began once COVID-19 restrictions in Scotland had been lifted to the extent that Scottish Government ministers and the Chief Medical Officer gave permission for such doorstep contact to recommence on Scottish Government surveys. The shift to a knock-to-nudge approach significantly increased levels of response to the survey.

Participants from the child boost sample were invited to opt in via letter for the entire fieldwork period.

3. Key changes to the survey

SHeS 2020 and 2021

Fieldwork for the 2020 and 2021 surveys were significantly affected by the COVID-19 pandemic. Fieldwork for SHeS 2020 was suspended in March 2020. Data for some of the key measures from SHeS was collected via a telephone survey in August and September 2020. Due to the testing of a new methodology for the SHeS survey within the context of the COVID-19 pandemic, the survey results for 2020 were presented as experimental statistics.

The telephone survey methods differed from those used prior to 2020 in the SHeS series. While the 2021 survey includes most of the questions and key indicators from the face-to-face surveys, the change in mode of administration, along with the different approach to sampling, is likely to have impacted the responses received and thus comparability with the previous SHeS data.

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.

Changes to survey methodology in 2018

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. SHeS21 Dataset

The SHeS21 dataset contains data from the main (core) interview questions, the self-completion questionnaires and additional variables derived from the responses to those questions. As interviews were conducted by telephone, no height and weight measurements or biological measures could be taken so the 2021 dataset does not include these variables. Self-reported height and weight variables and derived variables, along with the adjusted versions, are included instead (see section 6.7 below).

Combined datasets for the following years are also provided through UKDS: 2019/2021 and 2017/2018/2019/2021.

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 2021 data consists of one individual level file and one household level file:

SHeS2021 EUL.sav	6157 records	contains data for all individuals who gave an interview. It contains information from the household questionnaire, main individual schedule and self-completions.

SHeS2021h EUL.sav	8290 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 2021 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 because of changes to wording or response categories, have been given new variable names. These variables are usually suffixed either with a letter such as 'a' (e.g. Disc10a) or with the year of the change (e.g. PTRUST19).

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 2020 mid-year household population estimates for Scotland.

Weight name	**When it should be used**
int21wt	Analysis of items in core questionnaire (adults)
vera21wt	Analysis of items in version A of questionnaire (variables with labels starting "VERA") (adults)
cint21wt	Analysis of child data
cvera21wt	Analysis of version A and child boost data
cmint21wt	Analysis of main interview child data (not boost)
bio21wt	Analysis of bio data
SHeS_Intake24_wt_sc	Analysis of Intake24 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, the question COPDOth in the main interview is a "CODE ALL THAT APPLY" question which asks “What treatment or advice are you currently receiving because of your COPD, chronic bronchitis or emphysema?”.

The code frame consists of eight values:

1 Regular check-up with GP / hospital / clinic
2 Taking medication (tablets / inhalers) 3
Advice or treatment to stop smoking
4 Using oxygen
5 Immunisations against flu / pneumococcus
6 Exercise or physical activity
7 Advice or treatment to lose weight
8 Other

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

COPDOth1 – code 1: Regular check-up with GP / hospital / clinic, code 0: no
COPDOth2 – code 1: Taking medication (tablets / inhalers), code 0: no
COPDOth3 – code 1: Advice or treatment to stop smoking, code 0: no
COPDOth4 – code 1: Using oxygen, code 0: no
COPDOth5 – code 1: Immunisations against flu / pneumococcus, code 0: no
COPDOth6 – code 1: Exercise or physical activity, code 0: no
COPDOth7 – code 1: Advice or treatment to lose weight, code 0: no
COPDOth8 – code 1: Other, code 0: no

Because a respondent could have replied with more than one answer, that respondent could have a value 1 for a number of these variables. The missing values are the same across all eight 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: VitTake, Vacnot1 to Vacnot15, Supprt1R_19, Supprt2R_19, RelStrnR_19, ChangeR_19. The affected cases have been assigned the code of -4 for these variables.

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 Adjusted height and weight derived variables

BMI has been used as a measure of obesity in SHeS since its inception in 1995. In 2021, it was not possible to obtain the standardised height and weight measurements using a telephone approach. Therefore, where they were happy to do so, participants were asked to estimate their height and weight, with these self-reported measures used to calculate estimated BMI. So, while these estimates have been calculated, this data should be used with caution given the self-reported nature of these measures.

Work undertaken to analyse data from the Health Survey for England (HSE) and Active Lives Surveys found that, on average, participants overestimated their height and underestimated their weight. In SHeS 2021, adjustments for adults have been made to account for this and while these corrections appear to remove some bias from the estimates, caution is advised in interpreting this data. The syntax for the adjustments can be found in the derived variable specification.

For more information on the adjustments used, see the 2021 Technical report.

6.8 Intake24

Participants aged 16 and over were invited to complete an online recall using Intake24 (https://intake24.org/). Those who consented to take part were invited to complete two dietary recalls, either independently or via a phone call with Cambridge University. The SHeS 2021 EUL dataset includes the person level data from Intake24 and the Intake24 weight (SHeS_Intake24_wt_sc). Separate food level and day level datasets are also available under End User Licence and contain CPSerialA so that the datasets can be linked.

Originally, the person level data (daily intakes for foods and nutrients) were calculated using the ‘day average’ method. To enable comparisons with the 2024 data, the person level data has been recalculated using the ‘usual intakes’ method. 'Usual intakes' is the accepted method to estimate population habitual nutrient and food intakes and has been used for the 2019 to 2023 National Diet and Nutrition Survey (NDNS). This is in place of calculating ‘day average’ intakes. The usual intakes method reduces the impact of day-to-day variation in individual intakes and so reduces the extremes of the distribution while having a minimal impact on estimates of average consumption. This gives a better estimation of proportions above or below a recommended level. However, it is not possible to calculate usual intakes for foods with insufficient numbers of consumers. In these cases, the day average method is still used. Further details can be found in Appendix D of the NDNS Stage 3 evaluation report.

The recalculated variables using the ‘usual intakes’ method have been matched onto the SHeS 2021 EUL dataset. This includes additional person level variables that were not reported in the main SHeS 2021 report and were previously in a separate dataset, so that all person level data can be accessed in one file. Where the ‘day average’ method is used, variable names are suffixed with ‘_DA’. Where the ‘usual intakes’ method is used, they are suffixed with ‘_UI’. The methodology is also referenced in the variable label.

6.9 Equivalised income

The OECD equivalisation scale used in the Household Below Average Income poverty estimates was used to equivalise incomes in the 2021 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 2021 data only contains the OECD method.

6.10 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).

6.11 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.

6.12 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 2021 adult respondents were asked what their mother and father did for a living when the respondent was 14 so there are additional NS-SEC codes for both parents.

Classification level	Variable name prefix
Household reference person	HPNSSEC
Individual respondent	NSSEC
Respondent's father	FANSSEC
Respondent's mother	MANSSEC

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

6.13 Scottish Index of Multiple Deprivation (SIMD)

The 2021 data uses the 2020 Scottish Index of Multiple Deprivation (SIMD). The new SIMD quintiles are indicated by a '20' in the variable name, for example SIMD20_RPa. More information on SIMD, including how it is calculated, can be found on the Scottish Government website.

7. SHeS 2021 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/collections/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 2021 – CONTENTS

Points to note:

- There are three versions of the questionnaire in the mainstage: Core Version A; Core Version B (biological module); and Child Boost.
- Children are not eligible for the biological module in Core Version B.
- The biological module content only includes anxiety, depression and self-harm (no physical measurements or sample are being collected in the telephone version of the questionnaire.
- 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]	
Accidents 0+		
	General CVD (16+) and use of services [CVD] 0+	
	Asthma core [Asthma] 0+	
	COVID-19 module	
	Physical activity adults - including Qs on activity at work, [AdPhysic] 16+	
	Physical activity kids – [ChPhysic] 2+	
	Eating habits kids [Eating] 2 - 15	
	Fruit and Veg [Fruitveg] 2 - 15	
	Vitamin Supplements [Vitamin] 0+	
	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). Including where drank and who with.	
	Dental health [Dental] 16+	
Dental services (16+)		
	CPR training [CPRTrn]	
Discrimination and harassment (16+)		
Stress at work (16+)		
	Education and employment details 16+	
	Ethnicity (0+) place of birth (0+)and religion (16+) [Ethnic]	
	Parental history [Parent] 16+	
	Family health [Parent] 16+	
	Self-completions [Selfcomp] 4+	
	Height and weight [Measure] – self-report 2+	
	Consents [Consents] 0+	

		Biological module (16+)includes: • Anxiety • Depression • Self harm

Child Boost
Household questionnaire [HHgrid]+[GenHHold]
General health including caring [GenHlth]
Use of services [CVD] 0+
Asthma core [Asthma] 0+
Accidents [Accid] 0+
Physical activity kids – [ChPhysic] 2+
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+