Scottish Health Survey

2020

Telephone Survey

User Guide

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

1. Background

The data file contains data from Scottish Health Survey 2020 telephone survey.

SHeS has been carried out annually since 2008 and prior to this was carried out in 1995, 1998, and 2003. Fieldwork for SHeS 2020 (the sixteenth face-to-face survey in the series) was suspended towards the end of March 2020 as the UK went into a nationwide lockdown at the outset of the COVID-19 pandemic. The decision was taken to instead collect data for some of the key measures from SHeS via a telephone survey.

Commissioned by the Scottish Government Health Directorates, the SHeS series provides regular information on aspects of the public's health and factors related to health which cannot be obtained from other sources. The specific aim of the telephone survey was to provide national level data on health, health conditions and the prevalence of certain risk factors associated with these health conditions for adults over a specified period of time during the COVID-19 pandemic.

This shorter SHeS 2020 telephone survey was undertaken in order to capture data on key survey measures as quickly as possible and to add to the growing evidence base on public health during the pandemic. Of particular interest were the national indicators relevant to health including:

- Wellbeing
- Healthy Weight
- Health Risk Behaviours
- Physical Activity
- Food insecurity

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 Scottish Health Survey 2020 telephone survey main report, which is available online here.

2. Survey Design

The Scottish Health Survey series was designed to yield a representative sample of the general population living in private households in Scotland every year. In line with annual surveys in the series, for the telephone survey, a random sample of addresses (11,000 addresses) was selected from the Postcode Address File (PAF), using a multi-stage stratified design. The number of addresses in the sample was more than would usually be sampled for a survey of this length as it was estimated that the response rate would be lower when using an opt-in method.

Potential participants aged 16 and over were contacted by letter and asked to opt-in to taking part in an interview conducted over the phone. Due to the opt-in approach, achieving a representative sample was harder to control for. See sections 1.6.4 and 1.7 of the technical report for more information on variations in the sample profile and the weighting approach used to attempt to adjust for these differences.

Participating households included in the survey were those from which a participant or

participants contacted ScotCen to opt in to taking part. All adults aged 16 and over within these opt-in households were also given the opportunity to take part once initial telephone contact had been made.

As for the Scottish Health Survey in previous years, those living in institutions were outwith the scope of the survey. This should be borne in mind when interpreting the survey findings as participants living in these settings are more likely to be older and, on average, in poorer health than those in private households.

3. Key changes to the survey

The SHeS 2020 telephone survey was intended to provide a snapshot of the health of Scotland's population, both physical and mental, during a short period within the COVID-19 pandemic. The questionnaire was shorter than the usual SHeS survey and, as such, the scope of the survey was broad rather than permitting a detailed focus on particular topic areas. In addition to interest in general health, long-term conditions and health risk behaviours, mental health (including social capital and loneliness) has also been a topic of particular interest throughout the pandemic and in turn, featured in the SHeS 2020 telephone survey. Cardiovascular disease (CVD) and related risk factors (smoking, poor diet, lack of physical activity, obesity and alcohol use) remained a key focus of the survey.

The accompanying technical report for the SHeS 2020 telephone survey provides further information on the method used for this survey. While every effort was made to retain questions that were consistent with the face-to-face surveys as far as possible, it should be noted that due to a change in the mode used to collect the data (from face-to-face to telephone collection) and the shorter data collection period (August – mid September 2020), this data is not directly comparable with the previous findings from face-to-face SHeS surveys. It is, however, a useful snapshot into the health of the population during the COVID-19 pandemic and a useful exercise in aiding the development of the SHeS approach for data collection in 2021.

4. SHeS 2020 Telephone Survey Dataset

The SHeS 2020 telephone survey dataset contains data from the main interview questions, and additional variables derived from the responses to those questions.

Disclosure control methods applied from 2014 onwards were also applied to the SHeS 2020 telephone survey dataset and are noted in the variable listing.

5. Documentation

The documentation has been organised into the following sections:

- Interview - contains the questionnaire
- Data - contains the list of variables in the file, a list of derived variables with the syntax used to create them and an overview of the variables and DVs used in the main report tables.

6. Using the data

The 2020 Telephone data consists of one individual level file:

SHeS20_Tel W1.sav	1920 records	contains data for all individuals who gave an interview. It contains information from the telephone questionnaire.

6.1 Variables on the file

The data file contains 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 in order to plan your analysis. It includes:

- Major categories of variables (e.g. Drinking, Smoking)
- Source of each variable (e.g. Individual questionnaire, 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 questionnaire (all variables on the data file are given by name in the copy of the questionnaire), 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 for the 2020 telephone survey report tables.

Some variables and derived variables in the telephone survey dataset have been given new names. These variables are suffixed with a '_t20'. Variables were renamed:

- If the routing or response categories for questions were different to the face-to-face survey
- Where the mode in the usual SHeS survey was self-completion or CASI. Aside from this, variables were not renamed to indicate the overall change in mode (from face-to-face to telephone)
- Where the derivations of variables were different to the face-to-face survey.

6.2 Weighting variables

Weighting has been used to correct for different selection probabilities and for non-response. The methodology for the 2020 telephone survey weighting was adapted from the 2019 approach to account for the smaller number of completed interviews which took place and to account for some changes in patterns of participants via telephone which are not seen in the face-to-face survey.

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

Weight name	When it should be used
Int20wt	Analysis of questionnaire data

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 Passive is a "CODE ALL THAT APPLY" question which asks “Are you regularly exposed to other people’s tobacco smoke in any of these places?”.

The code frame consists of seven values:

1 At own home
2 At work
3 In other people's homes
4 In cars, vans etc
5 Outside of buildings (e.g. pubs, shops, hospitals)
6 In other public places
7 No, none of these

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

Passive1 - code 1: those who are regularly exposed at own home; code 0: those who aren’t.

Passive2 - code 1: those who are regularly exposed at work; code 0: those who aren’t.

Passive3 - code 1: those who are regularly exposed in other people’s homes; code 0: those who aren’t.

Passive4a - code 1: those who are regularly exposed in cars, vans etc; code 0: those who aren’t.

Passive5a - code 1: those who are regularly exposed outside of buildings; code 0: those who aren’t.

Passive6a - code 1: those who are regularly exposed in other public places; code 0: those who aren’t.

Passive7a - code 1: not regularly exposed in any of the places listed; code 0: those who are.

Because a participant could have replied with more than one answer, that participant could have a value 1 for a number of these variables (however, the nature of the question dictates that having a code 1 at Passive7a precludes having a code 1 at any of the other variables (Passive1 – Passive6a). The missing values are the same across all seven variables.

6.4 Missing values conventions

-1 Not applicable: Used to signify that a particular variable did not apply to a given participant usually because of internal routing

-8 Don't know

-9 Refused

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

Additional missing value conventions which have been used in previous SHeS surveys (-2 and -6) were not used in the 2020 telephone survey.

6.5 Derived variables

In addition to the questions and measurements collected directly in the survey, 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.

6.6 Adjusted height, weight and BMI derived variables

In the usual format of the Scottish Health Survey, interviewers undertake height and weight measurements. As this was not possible in the telephone survey, participants were asked to provide these measurements themselves.

Other studies have shown that self-reported measurements tend to overestimate height and underestimate weight on average. To help address this, adjustment factors based on a comparison study undertaken for the Health Survey for England were used to adjust the self-reported measurements. Adjusted height, weight and BMI DVs are indicated by '_adj'. The syntax for these adjustments is available in the derived variable specification.

6.7 Scottish Index of Multiple Deprivation (SIMD)

The 2020 telephone survey 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 2020 telephone survey report

The full report is available online 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 2020 TELEPHONE SURVEY – CONTENTS

Main interview outline
Household questionnaire including household composition
General health
CVD and diabetes
Asthma and COVID-19
Physical activity
Diet
Smoking
Alcohol
Dental health
Food insecurity
Social capital and loneliness
Mental wellbeing
Self-reported height and weight
Ethnic background and religion
Data linkage and follow-up research consent