

**Scottish Health Survey**

**2017**

**2018**

**2019**

**2021**

**COMBINED DATASET  
User Guide**

# 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 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 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.

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. 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. This should be borne in mind when conducting any analysis using the 17/18/19/21 combined dataset.

As interviews were conducted by telephone, no height and weight measurements or biological measures could be taken for SHeS 2021. 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.

This user guide provides an overview of the key elements of the 17/18/19/21 combined dataset<sup>1</sup> that users need to know before conducting any analysis. It is designed to be read in conjunction with the documentation supplied with the 2021 datasets and the 2021 technical report.

---

<sup>1</sup> Fieldwork for SHeS 2020 was suspended in March 2020 due to the COVID-19 pandemic. Data for some of the key measures from SHeS was collected via a telephone survey in August and September 2020. Due to disruption to the survey at the onset of the pandemic, the survey data collected in 2020 was published as experimental statistics and was not comparable with the time series. The 2020 data has not been included in the four year combined dataset.

## 2. Documentation

Full documentation for each of the individual survey years can be found on the UKDS website.

## 3. Dataset

The combined dataset consists of one individual level file. It contains data for all individuals who gave an interview in 2017, 2018, 2019 or 2021. As this is a repeated cross-sectional study, different households and individuals were interviewed each year.

It contains information from the household questionnaire, main individual schedule and self-completions.

This combined dataset has been provided to give a larger base for analysis of variables. The *individual year* datasets should be used for analysis of individual years, including comparison between years.

SPSS file name	Number of records	Number of variables
SHeS17181921.sav	25128	1780

## 4. Weighting variables for combined dataset

New weights were produced for the combined dataset

Weight name	When it should be used
int17181921wt	Analysis of core adult data
cint17181921wt	Analysis of child data
bio17181921wt	Analysis of bio sample data <sup>2</sup>
cmint17181921wt	Analysis of child core data

Full details of the weighting are provided in the main Technical Reports published for each year.

---

<sup>2</sup> There was no biological module in 2021. Participants have a bio17181921wt if they had a productive bio interview in 2017, 2018 and 2019 (bioout = 81) or are in the core bio sample in 2021 (styp12 = 2).