

Health Survey for England 2019

User Guide

Joint Health Surveys Unit:

NatCen Social Research

Department of Epidemiology and Public Health, University College London

A survey carried out for NHS Digital

At NatCen Social Research we believe that social research has the power to make life better. By really understanding the complexity of people's lives and what they think about the issues that affect them, we give the public a powerful and influential role in shaping decisions and services that can make a difference to everyone. And as an independent, not for profit organisation we're able to put all our time and energy into delivering social research that works for society.

NatCen Social Research
35 Northampton Square
London EC1V 0AX
T 020 7250 1866
www.natcen.ac.uk

A Company Limited by Guarantee
Registered in England No.4392418.
A Charity registered in England and Wales (1091768) and Scotland (SC038454)

Contents.

1	Background	2
2	Survey design	3
3	Documentation	4
4	Using the data	5
	Variables on the file	5
	Multicoded questions	6
	Missing values conventions	7
	Valid cases	7
	Notes about particular variables	8
	4.1.1 Smoking and drinking questions for young adults	8
	4.1.2 Disclosure control review	8
	4.1.3 Cholesterol results (Cholval and HDLval)	9
	4.1.4 Glycated haemoglobin results (glyhbval and iffcval)	9
	4.1.5 English index of multiple deprivation (IMD)	9
	4.1.6 Previous revisions to data	9
5	Weighting variables	10
	Household weight	10
	Interview weight	11
	Nurse weight	11
	Blood weight	11
	Cotinine weight	11
	Selecting the appropriate weight variable	11
6	Combining HSE data	13
7	HSE 2019 report	14
	Appendix A.	15

1 Background

The data file contains data from the Health Survey for England 2019 (HSE), the twenty-ninth year of a series of surveys designed to monitor trends in the nation's health. The 2019 Health Survey was commissioned by NHS Digital and carried out by the Joint Health Surveys Unit of NatCen Social Research and the Department of Epidemiology and Public Health at University College London (UCL).

The aims of the Health Survey series are:

- to provide annual data about the nation's health;
- to estimate the proportion of people in England with specified health conditions;
- to estimate the prevalence of certain risk factors associated with these conditions;
- to examine differences between population subgroups in their likelihood of having specific conditions or risk factors;
- to assess the frequency with which particular combinations of risk factors are found, and which groups these combinations most commonly occur;
- to monitor progress towards selected health targets including the prevalence of overweight and obesity in children;
- to measure the height of children at different ages, replacing the National Study of Health and Growth (since 1995);
- to monitor the prevalence of overweight and obesity in children (since 1995).

The 2019 survey included additional topics on providing unpaid social care, dental health, eating disorders, use of GP and counselling services, and awareness of two mental health resources, Good Thinking and Every Mind Matters. The survey also provided updates on repeated core topics, including general health, long standing illness, smoking and drinking.

The topic reports on this survey, including a detailed Methods and Documentation volume, are available [here](#).

See also: <http://content.digital.nhs.uk/healthsurveyengland>

2 Survey design

The HSE 2019 sample comprised of a core general population sample. There was no boost sample in 2019.

The sample comprised 9,612 addresses selected at random in 534 postcode sectors, issued over twelve months from January to December 2019. Field work was completed in March 2020. Where an address was found to have multiple dwelling units, one dwelling unit was selected at random. Where there were multiple households at a dwelling unit, one household was selected at random.

Adults and children were interviewed at households identified at the selected addresses. Up to four children in each household were selected to take part at random; up to two aged 2 to 12 and up to two aged 13 to 15.

In 2019, not all households were eligible for a nurse visit. In each PSU, 16 addresses were selected at random in advance; in these households all adults and children who were interviewed were eligible for the nurse visit. The nurse visit included measurements and the collection of blood and saliva samples, as well as other questions. Height was measured for those aged two and over, and weight for all participants. Nurses measured blood pressure (aged five and over) and waist and hip circumference (aged 11 and over). Non-fasting blood samples were collected from adults aged 16 and over. Saliva samples for cotinine analysis were collected from adults and children aged four and over. Nurses obtained written consent before taking samples from adults, and parents gave written consent for their children's samples. Consent was also obtained from adults to send results to their GPs, and from parents to send their children's results to their GPs.

A total of 8,205 adults aged 16 and over and 2,095 children aged 0-15 were interviewed, including 4,947 adults and 1,169 children who had a nurse visit.

3 Documentation

The documentation has been organised into the following sections:

- Interview: contains the CAPI documentation for household and individual questionnaires, CASI questionnaire, nurse visit questionnaires, self-completion booklets and showcards;
- Data: contains the list of variables and list of derived variables, including SPSS syntax specification;
- Other instructions: contains interviewer, nurse and coding and editing instructions;

Note that the questionnaires show the variable names used in the CAPI programme. In some cases the variables in the dataset have a different name or have been renamed due to disclosure control.

4 Using the data

The HSE 2019 data consists of one individual level file:

HSE19 EUL.sav	10,299 records	Contains data for all individuals in households who gave a full interview. It contains information from the main individual schedule, self-completions and the nurse visit (where one occurred).
---------------	----------------	--

Variables on the file

The data file contains questionnaire variables (excluding variables used for administrative purposes), demographic information and derived variables. The variables included in the dataset are detailed in the “**List of Variables**” 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. General Health, Blood Sample)
- Sub categories of variables (e.g. Longstanding illness (within General Health), Measurements from laboratory analysis (within Blood Sample)),
- The source of each variable (e.g. Household questionnaire, Individual questionnaire, CASI, Nurse Visit, Self-completion booklet, Derived variable etc.)

Once you have decided which variables to include in your analysis, you can 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 Specification**” document in the data section of the documentation to see how the variables were derived.

Note that the variable names used in the interview/CAPI documentation are sometimes different from the variable names used in the dataset.

Multicoded questions

Multicoded questions, where for example the interviewer or nurse is instructed to “CODE ALL THAT APPLY” or where an open ended question has elicited more than one answer, are stored in the HSE dataset as binary variables, with the people who selected each option coded 1 and the rest coded 0.

As an example, the question ConSubX (in the CAPI)/ConSbX (in the dataset) in the adult nurse schedule 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 (CONSBX11-15) as follows:

CONSBX11 - coded 1 for those who ate in the last 30 minutes and 0 for those that didn't.

CONSBX12 - coded 1 for those who smoked in the last 30 minutes and 0 for those that didn't.

CONSBX13 - coded 1 for those who drank alcohol in the last 30 minutes and 0 for those that didn't.

CONSBX14 - coded 1 for those who did vigorous exercise in the last 30 minutes and 0 for those that didn't.

CONSBX15 - coded 1 for those who did none of the above in the last 30 minutes and 0 for everyone else.

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 CONSBX15 precludes having a code 1 at any of the variables CONSBX11 – CONSBX14). The missing values are the same across all five variables.

Documentation for the CAPI questionnaires (household and individual) shows only the name of the first variable (which stores the number of mentions). So, for the example given above, this variable name is ConSubX.

Missing values conventions

These missing value conventions have also been applied to most of the derived variables as well as the original questionnaire variables. The derived variable specifications should be consulted for details.

- 1 Not applicable: Used to signify that a particular variable did not apply to a given participant usually because of internal routing. For example, men in women only questions or self completion variables when the participant is not of the given age range to answer that particular self-completion booklet.
- 8 Don't know, Can't say.
- 9 No answer/ Refused.
- 98 Don't know: CASI smoking and drinking questionnaire variables.

Valid cases

In the 2019 Health Survey report, as in previous reports, cases were excluded from the analysis of anthropometric and blood pressure measurements if their measurement was invalid. For example, those who had smoked, drunk, eaten, or exercised within 30 minutes of having their blood pressure taken were excluded from analysis as this can affect blood pressure. Individual report chapters will specify any exclusions.

Notes about particular variables

4.1.1 Smoking and drinking questions for young adults

In HSE 2019, there is no longer a young adult self completion, everyone aged 16+ was given an adult self completion instead. Smoking and drinking questions for all 16-17 year olds (and 18-24 year olds if this option was chosen) were answered in CASI rather than paper self-completion.

For the majority of questions, the CASI smoking and drinking questions are now the same as the CAPI modules and follow the same routing. As in previous years where CAPI and self completion variables were combined into one derived variable, the CAPI and CASI variables have been combined, unless there is a difference in routing between the two modes.

To reflect the change in mode, the smoking and drinking derived variables have all been renamed, with ‘_19’ appended to the end of the variable name. Some of the CAPI smoking and drinking questionnaire variables have also been renamed to reflect a change to the question wording or routing. The variable listing and questionnaire documentation should be consulted for further detail.

4.1.2 Disclosure control review

A review of the archive data was undertaken for HSE 2015. Variables and topics were assessed for their risk of disclosure in conjunction with guidance on the external release of survey data. As a result of the review, in the HSE 2015 data and onwards, some variables or groups of variables have been adjusted via top coding or re-grouping. Other variables or sets of questions have been removed entirely, such as the individual medication codes.

The majority of the household data, including the household serial number and household relationship variables, have been removed from the dataset. To aid intra-household analysis though, a selection of derived variables for parents have been appended to the records of their children. The HSE household file is no longer archived with UKDS.

Additional data which is not available on the End User Licence dataset can be requested via NatGen’s Data Release Panel.

4.1.3 Cholesterol results (Cholval and HDLval)

New equipment introduced in April 2010 and in June 2015 meant a slight change in the reference range for total and HDL cholesterol. These changes and the impact on the data are detailed in the 2015 User Guide and Section 9.2.2 in [Health Survey for England 2015: Methods](#).

4.1.4 Glycated haemoglobin results (glyhbval and iffcval)

From 19th September 2013, the laboratory that carries out the analyses on the blood and urine samples taken during the HSE interview used a new calibration lot for the processing of glycated haemoglobin. These changes are also detailed in the 2015 User Guide and the HSE 2015 Methods chapter.

4.1.5 English index of multiple deprivation (IMD)

From 2019, the HSE data contains the 2019 English Index of Multiple Deprivation (IMD) divided into quintiles.

4.1.6 Previous revisions to data

There have been revisions to the data in previous years. For information please see the 2012 user guide which has details of changes to:

- Longstanding illness questions
- Some Cardiovascular and Blood pressure variables
- HSE 2011 medications derived variables
- Cholesterol measurement (referred to above)

5 Weighting variables

Before 2003, the weighting strategy for the core sample in the HSE was to apply selection weights only (used for instance when a single household was selected from multiple households at an address, or where there were more than two children in a household), and no attempt was made to reduce non-response bias through weighting. However, following a review of the weighting for the HSE and other government funded surveys, non-response weighting has been incorporated in the weighting strategy since 2003.

In 2018, another review of the weighting strategy recommended changes to the age groups used in calibration weighting for children and regional level adjustments to population of residents over 65 living in communal establishments. These changes have been implemented in the 2019 weighting. For more detailed information on how the weights were produced see [Health Survey for England 2019: Methods](#).

A household weight has been generated for the general population sample which adjusts for non-contact and refusal of households; this is described in more detail below. Individual level non-response weights have also been generated for the general population.

The individual weights adjust for the additional non-response among individuals in participating households and additional weights take into account participants' participation in different elements of the survey. In 2019 the weights are for: the main interview, nurse visit, blood sample and cotinine.

Household weight

The household weight (*wt_hhld*¹) is a household level weight that corrects the distribution of household members to match population estimates for sex/age groups and GOR. These weights were generated using calibration weighting, with the household selection weights as starting values. The household selection weights also correct for the selection of a single household at addresses with more than one. Note that the population control totals used for the calibration weighting were the ONS projected mid-year population estimates for 2019, with a small adjustment to exclude the population aged 65 and over living in institutions, based on data from the 2011 census.

¹ Not included in the EUL dataset.

Interview weight

For analyses at the individual level, the weighting variable to use is (*wt_int*). These are calculated separately for adults and children in the core sample.

- For adults (aged 16 and over), the interview weights are a combination of the household weight and a component which adjusts the sample to reduce bias from individual non-response within households;
- For children (aged 0 to 15), the weights are generated from the household weights and the child selection weights – the selection weights correct for only including a maximum of four children in a household. The combined household and child selection weight were adjusted to ensure that the weighted age/sex distribution matched that of all children in co-operating households.

Nurse weight

To take into account non-response to the nurse section of the survey, a nurse weight has been generated (*wt_nurse*) and should be used on all analysis of questions asked during the nurse visit.

Blood weight

A blood weight has been generated for all adults who had a nurse visit, were eligible for, agreed, and were able to give a blood sample. This weight (*wt_blood*) should be used on all analysis of questions asked relating to blood samples.

Cotinine weight

A cotinine weight (from the saliva sample) has been generated for participants aged between 4 and over who had a nurse visit and were eligible for a saliva sample. This weight (*wt_cotinine*) should be used on all analysis of questions asked relating to saliva samples.

Selecting the appropriate weight variable

Different weights have been provided, for data from different stages of the survey:

- Interview stage (core sample)
- Nurse visit
- Saliva sample (participants aged 4 and over)
- Blood sample (adults only)

If questions from different stages of the survey are combined in analysis, the weights for the latest stage of the survey should be used (that is, the latest in the list above). For instance, if blood sample results are being cross-tabulated with questions from the interview stage, the blood sample weight should be used; or if waist circumference results (from the nurse visit) are cross-tabulated with BMI data from the interview, the nurse visit weight should be used. Where weights have been generated for specific modules, i.e. when analysing cotinine, please ensure you use the specific weights rather than the generic interview or nurse weights.

6 Combining HSE data

The 2019 HSE data includes stratification (Cluster) and PSU² (Primary Sampling Unit) variables. In 2019 there are three Cluster variables:

- Cluster194 is the stratification for the whole sample
- Cluster94 can be used for analysis of subgroups, for example children aged 0-15, adults aged 65 and over, and those who took part in the nurse interview
- Cluster48 can be used for the analysis of smaller subgroups, for example children aged 8-15.

If you are intending to carry out analysis combining multiple years of HSE, **it is recommended that you add a survey year prefix** to the PSU and Cluster variables for each year **before combining** the datasets. This is because the same numbers are used for PSU and Cluster each year, although they do not represent the same geographical area from year to year. In more recent HSE years, the cluster and strata variables in the dataset are already pre-fixed with the survey year.

² Called 'PSU_SCR' in the EUL dataset.

7 HSE 2019 report

Further information about the Health Survey for England 2019 is available in the following publications:

- [Health Survey for England 2019 chapters and tables](#)
- [Health Survey for England 2019 Methods](#)
- [Health Survey for England 2019: Summary of key findings.](#)

Further information about the Health Survey for England in general can be found on the respective websites of NHS Digital, NatCen Social Research and UCL (University College London):

<http://www.content.digital.nhs.uk/healthsurveyengland>

www.natcen.ac.uk/our-research/research/health-survey-for-england/

www.ucl.ac.uk/hssrg/studies/hse

Appendix A.

Household interview

Household questionnaire	
Household size, composition, relationships	Smoking in household
Accommodation tenure and number of bedrooms	Car ownership
Economic status / occupation of household reference person	Smoking in household
Household size, composition, relationships	Household income

Main interview

The Interviewer visit									
Module	Age (years)								
	0-1	2-4	5-7	8-9	10-12	13-15	16-17	18-64	65+
General health, longstanding illness, limiting longstanding illness	•	•	•	•	•	•	•	•	•
Doctor diagnosed hypertension and diabetes							•	•	•
Dental health*	•	•	•	•	•	•	•	•	•
Social care (receipt of) – short module									•
Social care (provision)							•	•	•
Carers*							•	•	•
Good Thinking campaign*							•	•	•
Every Mind Matters campaign*							•	•	•
Smoking				•a	•a	•a	•a	•a	•a
Drinking				•a	•a	•a	•a	•a	•a
Economic status / occupation / shift patterns							•	•	•

The Interviewer visit

Educational attainment							•	•	•
Ethnic origin / National identity	•	•	•	•	•	•	•	•	•
Self-reported height and weight							•	•	•
Attitude towards NHS / Any comments							•	•	•
Self-completion placement				•	•	•	•	•	•
Height measurement		•	•	•	•	•	•	•	•
Weight measurement	•	•	•	•	•	•	•	•	•
Arranging nurse appointments (if address eligible)	•	•	•	•	•	•	•	•	•
Re-contact information	•	•	•	•	•	•	•	•	•

a Smoking and drinking modules administered by self-completion for all aged 8-15, by CASI for those aged 16-17, CASI or CAPI for those aged 18-24 and CAPI for those aged 25+

*Additionally funded content

Self completions

Self completion content			
	8-12	13-15	Adults (16+)
Smoking	• ^b	• ^b	
Drinking	• ^b	• ^b	
General wellbeing - <i>Warwick-Edinburgh Mental Well-Being Scale (WEMWBS)</i>			•
Eating habits*			•
ONS wellbeing question			•
Sexual identity / National identity / Religion	• ^c	• ^c	• ^c

*Additionally funded content

b Smoking and drinking modules administered by self-completion for all aged 8-15

c Sexual orientation questions asked of adults.

There is no longer a young adult self completion in HSE 2019, everyone aged 16+ was given an adult self completion instead. Smoking and drinking questions for all 16-17 year olds (and 18-24 year olds if this option was chosen) are now answered in CASI rather than paper self-completion.

The nurse visit

This is what the nurse visit includes in 2019.

The nurse visit						
	Age (years)					
Module	0-3	4	5-10	11-15	16-17	18+
Prescribed medicines, folic acid supplements	•	•	•	•	•	•
Blood pressure			•	•	•	•
Waist and hip circumference				•	•	•
Saliva sample (cotinine)		•	•	•	•	•
Non-fasting blood samples (Total and HDL cholesterol, glycated haemoglobin)					•	•