Health Survey for England 2018

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

The data file contains data from the Health Survey for England 2018 (HSE), the twenty-eighth year of a series of surveys designed to monitor trends in the nation's health. The 2018 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 UCL (University College London).

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 2018 survey included additional topics on breathing problems for adults and children, and gambling participation among adults. 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, is available here.

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

2 Survey design

The HSE 2018 sample comprised of a core general population sample. There was no boost sample in 2018.

The sample comprised 9,612 addresses selected at random in 534 postcode sectors, issued over twelve months from January to December 2018. Field work was completed in March 2019. 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 2018, 89% of the sample were eligible for a nurse visit. For eligible addresses, everyone who was interviewed was eligible for a 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 5 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 children aged between 4 and 15. 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 8,178 adults aged 16 and over and 2,072 children aged 0-15 were interviewed, including 4,825 adults and 1,103 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, 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 2018 data consists of one individual level file:

HSE18 EUL.sav	10,250 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, 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 for how the variables were derived.

Note that the variable labels 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 archived HSE 2018 dataset in two ways, coded either **by mention** or **by category**. Questions coded by mention are stored as categorical variables where the complete value set is repeated in each of the variables. Questions coded by category are

stored as indicator variables where each value in the set is stored as its own variable. Both approaches have been used in the 2018 Health Survey.

As an example, question ConSubX (in the CAPI)/ConSbX (in the dataset) on the 2018 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

If recorded by mention, four variables would record the (up to) four possible responses to the question, assigning codes 1-5 in the first variable and codes 1-4 in each of the next three variables. In 2018, the variables CONSBX11-15 store the answer to this question by category 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.

Valid cases

In the 2018 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 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 NatCen's Data Release Panel.

4.1.2 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.

4.1.3 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.

4.1.4 English index of multiple deprivation (IMD)

From 2015, HSE data contains the 2015 English Index of Multiple Deprivation (IMD) divided into quintiles.

4.1.5 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. The same strategy as in 2003 has been followed for weighting the HSE 2018 sample data. For more detailed information on how the weights were produced see Health Survey for England 2018: 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 2018 the weights are for: the main interview, self completions and problem gambling, nurse visit, blood sample and cotinine.

Household weight

The household weight (***wt_hhld***) 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 2018, with a small adjustment to exclude the population aged 65 and over living in institutions, based on data from the 2011 census.

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.

Self-completion and gambling weight

Weighting was applied to adjust for non-response to the self-completion booklet (***wt_sc)***, and for whether the problem gambling screen in the self-completion booklet (adults aged 16 and over) was completed (***wt_gambling***). A logistic regression model was fitted for those participants that were eligible to fill in the self-completion booklet. The outcome variable was whether or not the booklet was filled in. The covariates in the model were age group by sex, household type, social class of HRP, smoking status and general health. The weights for not filling in the self-completion booklet (w9) were calculated as the reciprocal of the predicted probability of the self-completion booklet being filled in, estimated from the regression models.

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 15 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)
- Self completion
- Gambling module

- Nurse visit
- Saliva sample (participants aged 4 to 15)
- 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 2018 HSE data includes stratification (Cluster) and PSU (Primary Sampling Unit) variables. In 2018 there are three Cluster variables:

- Cluster195 is the stratification for the whole sample
- Cluster95 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 and children with a valid cotinine sample.

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.

7 HSE 2018 report

Further information about the Health Survey for England 2018 is available in the following publications:

- Health Survey for England 2018 chapters and tables
- Health Survey for England 2018 Methods
- Health Survey for England 2018: 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	
Household income	

Main interview

The Interviewer visit									
	Age (years)								
Module	0-1	2-4	5-7	8-9	10-12	13-15	16-17	18-64	65+
General health, longstanding illness, limiting longstanding illness	•	•	•	•	•	•	•	•	•
Personal care plans							•	•	•
Doctor diagnosed hypertension and diabetes							•	•	•
Breathing problems*	•	•	•	•	•	•	•	•	•
Social care (receipt of) – long module									•
Fruit and vegetable consumption			•	•	•	•	•	•	•
Smoking (revised questions)				•a	•a	•a	•a	•a	•
Drinking				•a	•a	•a	•a	•a	•
Economic status / occupation / shift patterns							•	•	•
Educational attainment							•	•	•
Ethnic origin / National identity	•	•	•	•	•	•	•	•	•

The Interviewer visit									
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	•	•	•	•	•	•	•	•	•

* Additionally funded content

[a] Smoking and drinking modules administered by self-completion for all aged 8-17 and some aged 18-24.

Self completion

Self-completion content	8-12	13-15	Young adults	Adults
Smoking	• [b]	• [b]	• [b]	
Drinking	• [b]	• [b]	• [b]	
GHQ-12 (general health)		•	•	•
EQ-5D* (health today)			•	•
ONS wellbeing question			•	•
Gambling*			•	•
IPAQ (Physical activities)			•	•
Sexual identity / National identity / Religion	• [c]	• [c]	• [c]	• [c]

* Additionally funded content

As usual, all adults aged 16-17 are given the young adult self completion. Interviewers could decide for adults aged 18-24 whether they thought it would be better for respondents to answer smoking and drinking questions in the young adult self completion, or whether to ask these questions in CAPI.

The nurse visit

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)					•	•