

# NatCen

**Social Research** that works for society

# Health Survey for England 2014

## User Guide

*Joint Health Surveys Unit:*

**NatCen Social Research**

**Department of Epidemiology and Public Health, University College London**

**A survey carried out for the Health and Social Care Information Centre**

**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](http://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.

<b>1</b>	<b>Background</b> .....	<b>4</b>
<b>2</b>	<b>Survey design</b> .....	<b>5</b>
<b>3</b>	<b>Documentation</b> .....	<b>6</b>
<b>4</b>	<b>Using the data</b> .....	<b>7</b>
4.1	Variables on the files.....	7
4.2	Multicoded questions.....	8
4.3	Missing values conventions .....	9
4.4	Valid cases.....	9
4.5	Notes about particular variables .....	10
	4.5.1 HSE 2011 Derived Medications Variables .....	10
	4.5.2 Previous revisions to data.....	10
<b>5</b>	<b>Weighting variables</b> .....	<b>11</b>
5.1	Household weight.....	11
5.2	Interview weight.....	12
5.3	Nurse weight.....	12
5.4	Blood weight .....	12
5.5	Cotinine weight .....	12
5.6	Urine weight .....	12
5.7	LTCP weight .....	13
5.8	Selecting the appropriate weight variable .....	14
<b>6</b>	<b>Combining HSE data</b> .....	<b>15</b>
<b>7</b>	<b>HSE 2014 report</b> .....	<b>16</b>
	<b>Appendix A.</b>	<b>18</b>

---

# 1 Background

The data files contain data from the Health Survey for England 2014 (HSE), the twenty-fourth year of a series of surveys designed to monitor trends in the nation's health. The 2014 Health Survey was commissioned by the Health and Social Care Information Centre and carried out by the Joint Health Surveys Unit of NatGen 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 2014 survey included two new topics for the core survey: hearing and mental health. The survey also provided updates on repeated core topics, including smoking, drinking and general health. Additional non-core modules of questions were also included: average weekly alcohol consumption, well-being, and physical activity.

The report on this survey, including a detailed Methods and Documentation volume, is published at <http://www.hscic.gov.uk/pubs/hse2014>

See also [www.hscic.gov.uk/health-survey-england](http://www.hscic.gov.uk/health-survey-england)

---

## 2 Survey design

The HSE 2014 included a general population sample of adults and children, representative of the whole population at both national and regional level. For the sample, 9,024 addresses were randomly selected in 564 postcode sectors, issued over twelve months from January to December 2014. Where an address was found to have multiple dwelling units, one dwelling unit was selected at random and where there were multiple households at a dwelling unit, one household was selected at random.

In each selected household, all individuals were eligible for inclusion in the survey. Where there were three or more children aged 0-15 in a household, two of the children were selected at random. A nurse visit was arranged for all participants who consented.

A total of 8,077 adults aged 16 and over and 2,003 children aged 0-15 were interviewed. A household response rate of 62% was achieved. Of those where a full interview was achieved, 5,491 adults and 1,249 children also had a nurse visit.

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 (for the analysis of total and HDL cholesterol and glycated haemoglobin) and urine samples were collected from adults aged 16 and over. Saliva samples for cotinine analysis were collected from all participants aged 4 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.

---

## 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 data set have a different name.

---

## 4 Using the data

The HSE 2014 data consists of two files; one at individual level and one at household level:

HSE2014ai.sav	10,080 records	Contains data for all individuals in household who gave a full interview. It contains information from the household questionnaire, main individual schedule, self-completions and the nurse visit (where one occurred).
---------------	----------------	--

HSE2014ah.sav	11,934 records	Contains data on household composition, sex, age and marital status for all individuals in co-operating households (including those not interviewed).
---------------	----------------	---

### 4.1 Variables on the files

Each of the data files contain questionnaire variables (excluding variables used for administrative purposes), demographic information including household composition and derived variables. The variables included in both files 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),
- Source of each variable (e.g. 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 data set. In these cases, the data variable label is shown in red beside the CAPI label in the documentation.

## 4.2 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 2014 data sets 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 2014 Health Survey.

As an example, question ConSubX (in the CAPI)/ConSbX (in the dataset) on the 2014 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 2014, 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 six 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 ConSbX.

## 4.3 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.

## 4.4 Valid cases

In the 2014 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.

---

## 4.5 Notes about particular variables

### 4.5.1 HSE 2011 Derived Medications Variables

Some HSE 2011 medications derived variables have a number of cases coded -9 (to indicate the participant refused to answer) when they should have been coded 0 (to indicate that they were taking some medications, but not that particular medication). The variables affected include: diur, beta, aceinh, calciumb, obpdrug, lipid, iron, bpmedc and bpmedd

Only HSE 2011 data are affected; this problem did not occur in previous years. From 2012 onwards, new derived variables were created which were coded correctly: diur2, beta2, aceinh2, calciumb2, obpdrug2, lipid2, iron2, bpmedc2 and bpmedd2.

### 4.5.2 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
- Cholesterol measurement

---

## 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 2014 sample data. For more detailed information on how the weights were produced see Health Survey for England 2014: Volume 2: Methods and documentation <http://www.hscic.gov.uk/catalogue/PUB19295/HSE2014-Methods-and-docs.pdf>

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 in section 5.1 below. Individual level non-response weights have also been generated for the general population and are described in section 5.2 onwards.

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 2014 the weights are for interview, nurse visit, blood sample, cotinine (from the saliva sample), urine and Itcp (long term care planning module).

### 5.1 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 2013, with a small adjustment to exclude the population aged 65 and over living in institutions, based on data from the 2011 census.

---

## 5.2 Interview weight

For analyses at the individual level, the weighting variable to use is (*wt\_int*). These weights are generated separately for adults and children:

- 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 two 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.

## 5.3 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.

## 5.4 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.

## 5.5 Cotinine weight

A cotinine weight (from the saliva sample) has been generated for participants aged 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.

## 5.6 Urine weight

A urine weight (*wt\_urine*) has been generated for participants who had a nurse visit and were eligible (all adults) to have a urine sample taken. This weight should be used on all analysis of questions relating to urine samples.

---

## 5.7 LTCP weight

Adults aged 30 and over were eligible to answer questions about planning for future care. If more than one person in the household was eligible to take part, one participant was randomly select. The LTCP weight (*wt\_ltcp*) was generated to account for possible bias in the selection. The weight should be used on all analysis of questions in the long term care planning module.

---

## 5.8 Selecting the appropriate weight variable

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

- Interview stage
- Nurse visit
- Saliva sample (participants aged 4 and over)
- Blood sample (adults only)
- Urine sample (adults only)
- Long term care planning (adults 30+)

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 2014 HSE data includes stratification (Cluster) and PSU (Primary Sampling Unit) variables with the prefix 14 to represent the survey year. This was included for the first time in 2012 to enable users to differentiate between Strata (Cluster) and PSU variables when combining different HSE years together.

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 2012) **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 2014 report

Further information about the Health Survey for England 2014 is available in the following publications:

- Craig, R., Fuller, E. and Mindell, J. (eds). [\*Health Survey for England 2014: Volume 1. Health, social care and lifestyles\*](#). The Health and Social Care Information Centre, Leeds, 2014.
- Craig, R., Fuller, E. and Mindell, J. (eds). [\*Health Survey for England 2014: Volume 2. Methods and documentation\*](#). The Health and Social Care Information Centre, Leeds, 2014.
- Craig, R. and Mindell, J. (eds). [\*Health Survey for England 2014: Health, social care and lifestyles. Summary of key findings\*](#). The Health and Social Care Information Centre, Leeds, 2014.

These can be found at [www.hscic.gov.uk/pubs/hse2014](http://www.hscic.gov.uk/pubs/hse2014)

See also [www.hscic.gov.uk/health-survey-england](http://www.hscic.gov.uk/health-survey-england)

Trend tables for the Health Survey for England series can be found at <http://www.hscic.gov.uk/pubs/hse2014trend>

- [Adult Trend Tables 2014](#). The Health and Social Care Information Centre, Leeds, 2014.
- [Child Trend Tables 2014](#). The Health and Social Care Information Centre, Leeds, 2014.
- [Health Survey for England 2014: Trend tables \(commentary\)](#). The Health and Social Care Information Centre, Leeds, 2014.
- [Population Number Estimate Tables 2014](#). The Health and Social Care Information Centre, Leeds, 2014.
- [Population Number Estimates - user guide](#). The Health and Social Care Information Centre, Leeds, 2014.



---

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

<http://healthsurvey.hscic.gov.uk/support-guidance/public-health/health-survey-for-england-2014/>

[www.natcen.ac.uk/our-research/research/health-survey-for-england/](http://www.natcen.ac.uk/our-research/research/health-survey-for-england/)

[www.ucl.ac.uk/hssrg/studies/hse](http://www.ucl.ac.uk/hssrg/studies/hse)

# Appendix A.

## HEALTH SURVEY FOR ENGLAND 2014 – CONTENTS

Household questionnaire							
Household size, composition, relationships	Smoking in household						
Accommodation tenure and number of bedrooms	Car ownership						
Economic status / occupation of household reference person	Learning difficulties*						
Household income							
The Interviewer visit							
	Age (years)						
Module	0-1	2-4	5-7	8-9	10-12	13-15	16+
General health, longstanding illness	•	•	•	•	•	•	•
Fruit and vegetable consumption			•	•	•	•	
Self-reported height and weight							•
Personal care plans							•
Doctor diagnosed hypertension, diabetes							•
Planning for future care (30+) *							•
Use of health services*							•
Social care							•
Hearing*							•
Smoking				• <sup>a</sup>	• <sup>a</sup>	• <sup>a</sup>	•
Drinking				• <sup>a</sup>	• <sup>a</sup>	• <sup>a</sup>	•
Economic status, occupation, education							•
Ethnic origin, National identity	•	•	•	•	•	•	•
Self completion							
Learning Difficulties*							•
General health today (EQ5D)*							•
General Health last few weeks (GHQ12)*						•	•
Wellbeing (Warwick – Edinburgh Scale)							•
Mental health attitudes*							•
Physical Activity							•
Perception of weight							•
Sexual orientation, Religion							•
Measurements							
Height		•	•	•	•	•	•
Weight	•	•	•	•	•	•	•

Nurse visit					
	Age (years)				
Module	0-3	4	5-10	11-15	16+
Prescribed medicines	•	•	•	•	•
Nicotine replacement products					•
Blood pressure			•	•	•
<b>Mental health questions*</b>					•
Waist and hip circumference				•	•
Saliva sample (cotinine)		•	•	•	•
<b>Urine sample*</b>					•
Non-fasting blood samples (Total and HDL cholesterol, glycated haemoglobin)					•
<b>Hearing test*</b>					•