NatCen
Social Research that works for society

Health Survey for England 2011

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

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 files contain data from the Health Survey for England 2011 (HSE), the twenty-first year of a series of surveys designed to monitor trends in the nation's health. The 2011 Health Survey was commissioned by the Health and Social Care Information Centre 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.

The main focus of the HSE in 2011 was cardiovascular disease. The survey also provided updates on core topics including smoking, drinking and fruit and vegetable consumption. Additional modules of questions were also included covering social care, chronic pain and well-being. In 2011 there was also a drinking diary designed to measure weekly consumption of alcohol.

The report on this survey, including a detailed Methods and Documentation volume, is published at www.ic.nhs.uk/pubs/hse11report

2 Survey design.

The HSE 2011 included a general population sample of adults and children, representative of the whole population at both national and regional level. For the sample, 8,992 addresses were randomly selected in 562 postcode sectors, issued over twelve months from January to December 2011. 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 8,610 adults aged 16 and over and 2,007 children aged 0-15 were interviewed. A household response rate of 66% was achieved for the core sample. Among the general population sample, 5,715 adults and 1,257 children 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) were collected from adults aged 16 and over. Saliva samples for cotinine analysis were collected from adults aged 16 and over and children aged 4-15. 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.

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, drinking diary and showcards)
- Data (contains the list of variables and list of derived variables including SPSS syntax specification)
- Other instructions (contains interviewer, nurse and coding & editing instructions).

4 Using the data.

The HSE 2011 data consists of two files; one at individual level and one at household level:

HSE2011ai.sav	10,617 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, drinking diary and the nurse visit (where one occurred).

HSE2011ah.sav	12,313 records	Contains data on household composition, sex, age and marital status for all individuals in co-operating households.

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.

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 2011 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 2011 Health Survey.

As an example, question CONSBX1 on the 2011 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 2011, 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 ConSubX.

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 2011 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.

4.5 Notes about particular variables

4.5.1 Revisions to some Cardiovascular and Blood pressure variables

Revisions have been made to some of the cardiovascular and blood pressure variables in light of comments from users and changes to definitions since last using these modules in the HSE series. Variables that have been changed are specified in the variable listing with a footnote and the variables have been renamed and relabelled so that comparisons can be made between years.

One example is the variable diabete2. In previous years this had been derived so that those participants who had not answered whether or not they had diabetes, had been assumed not to have the condition. The question 'DocInfo1' (Were you told by a doctor that you had diabetes?) is only asked if the participant has answered yes at 'Everdi' (Do you now have, or have you ever had diabetes?). Therefore, it was assumed that all those who were not asked DocInfo1, had replied 'No' at Everdi; and were therefore assumed (diagnosed) diabetes free. However, some participants either refused to answer, or did not know whether they had diabetes when asked at Everdi. In the 2011 data, these participants have been coded to their original refused/not answered value. For comparison purposes the original derivation of Diabete2 has been kept on the data along with the revised version – Diabete2r. Women who had diabetes only during pregnancy were correctly classified as not diabetic at diabete2 both before and after the change.

Up to 20 people in each survey year either refused, or didn't know the answer to the original cardiovascular condition questions, and were therefore miscoded as not having those conditions. In addition, in 1998, 1999, 2003, 2004, 2005, and 2006, under-16s (who were not asked the question) were miscoded as not having cvd conditions; as were adults aged under 65 in the 2005 sample, and white participants in both 1999 and 2004.

As these groups were never intended to be asked the cardiovascular questions, it is hoped that researchers restricted their analysis to the correct sample groups, and therefore were using derived cvd variables correctly.

4.5.2 Cholesterol measurement

During 2010 (from the 12th of April, 2010), the laboratory that carries out the analyses on the blood and urine samples taken during the HSE interview introduced new analytical equipment. This had no affect on most analytes, but resulted in a slight change in the reference range for total and HDL cholesterol after that date. A flag variable was computed for 2010 data (called CHOLFLAG) which shows whether a sample was tested before or after the equipment change during that year. If analysts are making comparisons between 2010 and 2011 results and those from previous years, they may make allowance for this difference if they wish to adjust the cholesterol results. Note that the difference (an average

of 0.1mmol/L) was very small and in most analyses this difference is unlikely to be statistically significant.

5 Weighting variables.

Prior to 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 2011 sample data. (For more detailed information on how the weights were produced see Health Survey for England 2011: Volume 2: Methods and documentation http://www.ic.nhs.uk/ pubs/hse11report).

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: interview, nurse visit, blood sample, cotinine (from the saliva sample) and the drinking diary.

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 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 2010, but with a small adjustment to exclude the population aged 65 and over living in institutions, based on data from the 2001 census.

5.2 Individual 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 or more)**,** 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.

For analysis of children aged 0-15 the ***wt_child*** variable can be used.

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 all adults and children that were aged 4-15yrs 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 Drinking diary weight

The drinking diary was given to all participants aged 18 and over who completed the main HSE interview and had had an alcoholic drink in the previous 12 months. A drinking diary weight has been generated for all adults eligible for the drinking diary. This weight (***wt_drink***) should be used on all analysis of drinking diary questions.

5.7 Priority of using weight variables

When carrying out analysis using variables from different modules that consequently have different survey weights (i.e. a variable from the individual questionnaire and one from a nurse module) please use the weights in the following order:

Blood weight (for blood or other analyte variables)

Nurse weight

Interview weight.

For example when analysing blood pressure by a cvd condition, the nurse weight should be used.

Where weights have been generated for specific modules, i.e for 2011 the drinking diary, or when anlaysing cotinine, please ensure you use the specific weights rather than the generic interview or nurse weights.

6 Combining HSE data

The 2011 HSE data comprises of stratification (Cluster) and PSU (Primary Sampling Unit) variables with the prefix 11 for the survey year. This has been included for the first time to enable users to differentiate between strata and psu variables when combining HSE years together. It is the intention that the data for previous years of HSE will be updated to include new PSU and strata variables with the survey year prefix.

If you are intending to carry out analysis on multiple years of HSE by combining data, before the datasets can be updated retrospectively, it would be worthwhile to prefix the survey year to both of these variables to avoid duplication of these variables when merging the datasets.

7 HSE 2011 report.

Further information about the Health Survey for England 2011 is available in the following publications:

- Craig R and Mindell J (eds). *Health Survey for England 2011: Volume 1. Health, social care and lifestyles*. The Health and Social Care Information Centre, Leeds, 2012.

- Craig R and Mindell J (eds). *Health Survey for England 2011: Volume 2. Methods and documentation*. The Health and Social Care Information Centre, Leeds, 2012.

- Craig R and Mindell J (eds). *Health Survey for England 2011: Health, social care and lifestyles. Summary of key findings*. The Health and Social Care Information Centre, Leeds, 2012.

These can be found at t http://www.ic.nhs.uk/pubs/hse11report

Trend tables for the Health Survey for England series can be found at www.ic.nhs.uk/pubs/HSE11trends :

- Adult Trend Tables 2011. The NHS Information Centre, Leeds, 2011.

- Child Trend Tables 2011. The NHS Information Centre, Leeds, 2011.

- Health Survey for England 2011: Trend tables (commentary). The Health and Social Care Information Centre, Leeds, 2012.

- Population Number Estimate Tables 2011. The Health and Social Care Information Centre, Leeds, 2012.

- Population Number Estimates - user guide. The Health and Social Care Information Centre, Leeds, 2012.

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):

www.ic.nhs.uk/statistics-and-data-collections/health-and-lifestyles-related-surveys/health-survey-for-england

www.natcen.ac.uk/series/health-survey-for-england

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

Appendix A.

HEALTH SURVEY FOR ENGLAND 2011 – CONTENTS

Household data

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

Individual level information

	Age							
	0-1	**2-3**	**4**	**5-7**	**8-10**	**11-12**	**13-15**	**16+**
Interviewer visit								
General health, longstanding illness, limiting longstanding illness, acute sickness	●	●	●	●	●	●	●	●
Self care								●
Self-reported height and weight								●
Cardiovascular disease, including doctor-diagnosed hypertension and diabetes								●
Chronic pain								●
Dental health								●
Social care								●
Fruit and vegetable consumption				●	●	●	●	●
Smoking					● [a]	● [a]	● [a]	● [b]
Drinking (heaviest drinking day last week, regular drinking)					● [a]	● [a]	● [a]	● [b]
Economic status/occupation, educational achievement								●
Ethnic origin	●	●	●	●	●	●	●	●
Height measurement		●	●	●	●	●	●	●
Weight measurement	●	●	●	●	●	●	●	●
Reported birth weight	●	●	●	●	●	●	●	

Consent to linkage to NHS Central Register/Hospital Episodes Statistics								●
Self completion								
Attitudes to personal health and lifestyle								●
Warwick Edinburgh mental wellbeing scale								●
EQ5D								●
Happiness							●	●
Perception of weight					●	●	●	
Sexual orientation, religion								● [a]
Strengths and difficulties, including parent perception of child's weight			● [c]	● [c]	● [c]	● [c]	● [c]	
Drinking diary								
Drinking diary								● [d]
Nurse visit								
Immunisations	●							
Prescribed medicines and vitamin supplements	●	●	●	●	●	●	●	●
Nicotine replacement treatments								●
Waist and hip circumference						●	●	●
Blood pressure				●	●	●	●	●
Saliva sample (cotinine)			●	●	●	●	●	●
Blood sample								●

[a] This module was administered by self-completion.

[b] This module was administered by self-completion for those aged 16-17 and some aged 18-24.

[c] This questionnaire was administered by self-completion to parents of children aged 4-15.

[d] The drinking diary was left with participants aged 18 and over to complete in the week following the interview; young people aged 16-17 were asked to complete the diary retrospectively during the nurse visit.