

**Health Survey for England**

**Physical activity  
and fitness**

**'08**

**User Guide**

A survey carried out on behalf of the Information Centre

*Joint Health Surveys Unit*

National Centre for Social Research

Department of Epidemiology and Public Health, University College London

## CONTENTS

<b>1. BACKGROUND .....</b>	<b>3</b>
<b>2. SURVEY DESIGN .....</b>	<b>3</b>
<b>3. DOCUMENTATION .....</b>	<b>5</b>
<b>4. USING THE DATA .....</b>	<b>5</b>
4.1 VARIABLES ON THE FILES .....	5
4.2 MULTICODED QUESTIONS .....	6
4.3 MISSING VALUES CONVENTIONS .....	7
4.4 VALID CASES .....	7
<b>5. WEIGHTING VARIABLES .....</b>	<b>8</b>
5.1 HOUSEHOLD WEIGHT .....	8
5.2 INDIVIDUAL WEIGHT .....	8
5.3 NURSE WEIGHT .....	9
5.4 BLOOD WEIGHT .....	9
5.5 CHILD SAMPLE WEIGHTS .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
<b>6. HSE 2006 REPORT .....</b>	<b>10</b>
<b>7. APPENDIX A .....</b>	<b>11</b>

# 1. Background

The data files contain data from Health Survey for England 2008 (HSE), the eighteenth year of a series of surveys designed to monitor trends in the nation's health. The 2008 Health Survey was commissioned by the Information Centre and carried out by the Joint Health Surveys Unit of the *National Centre for Social Research* and the Department of Epidemiology and Public Health at Royal Free and University College Medical School.

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;
- since 1995, to measure the height of children at different ages, replacing the National Study of Health and Growth;
- since 1995, monitor the prevalence of overweight and obesity in children
- HSE 2008 focused on physical activity and fitness; adults and children were asked to recall their physical activity over recent weeks, and objective measures of physical activity and fitness were also obtained.

## 2. Survey Design

The HSE 2008 included a general population sample of adults and children, representative of the whole population at both national and regional level, and a boost sample of children aged 2-15. A sub-sample was identified in which the main survey was supplemented with objective measures of physical activity and fitness. For the general population sample, 16,056 addresses were randomly selected in 1,176 postcode sectors, issued over twelve months from January to December 2008. Where an address was found to have multiple dwelling units, one was selected at random. Where there were multiple households at a dwelling unit, up to three households were included, and if there were more than three, a random selection was made. At each address, all households, and all persons in them, 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. In addition to the core general population sample, a boost sample of children aged 2-15 was

selected using 19,404 addresses. These were drawn from 996 of the core sampling points. As for the core sample, where there were three or more children in a household, two of the children were selected at random to limit the respondent burden for parents. There was no nurse follow up for this child boost sample.

A sub-sample was identified in which the main survey was supplemented with objective measures of physical activity and fitness. The sub-sample was taken in 384 sampling points, including both core and boost addresses. Up to two individuals in the sub-sample households were selected to wear an accelerometer to measure physical activity; in households where both adults and children of the appropriate age were interviewed, an adult and a child were selected. In these households, eligible adults aged 16-74 were offered the step test in the nurse visit, to measure fitness.

A total of 15,102 adults and 7,521 children were interviewed in 2008, with 3,473 children from the core sample and 4,048 from the boost. A household response rate of 64% was achieved for the core sample, and 73% for the boost sample. Among the general population sample, 10,740 adults and 2,464 children had a nurse visit.

The 2008 survey focused on physical activity and fitness levels. Participants were interviewed, and for those in the core sample this was followed by a visit from a specially trained nurse. Adults and children were asked modules of questions including general health, fruit and vegetable consumption, alcohol consumption and smoking, as well as physical activity.

Height was measured for those aged two and over and weight for all participants. Nurses measured infant length (aged at least six weeks and under two years), 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, and saliva samples for cotinine analysis 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 and showcards)
- Data (contains the list of variables and list of derived variables)
- Other instructions (contains interviewer, nurse and coding & editing instructions).

### 4. Using the data

The 2008 data consists of two files; one at individual level and one at household level:

HSE08ai.sav	22,623 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).
-------------	----------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

HSE08ah.sav	31,927 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) and derived variables. The variables included in the individual file are detailed in the “**List of Variables**” document in the data section of the documentation. This document is the best place to look at in order to plan your analysis. It includes:

- Major categories of variables (eg Accidents, Anthropometric measurements)
- Sub categories of variables (eg Attitudes to cycling, Major accidents within the Accidents category)
- Source of each variable (eg Individual questionnaire, Nurse visit, 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 derived variables.

## 4.2 Multicoded questions

Multicoded questions are stored in the archived HSE 2008 data sets in two ways. 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, were stored as array variables in the QUANTUM DBMS system which was used to read and edit the data. However, in SPSS (which was used for analysis and archiving the data) multicoded variables must be stored as ‘flat’ variables, 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 2008 Health Survey.

As an example, question CONSBX1 on the 2008 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 2007, the variables CONSBX11-15 store the answer to this question by category as follows:

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

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

CONSBX13 - coded 1 for those who drank alcohol in the last half hour and 0 for those

that didn't.

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

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

Because a respondent could have replied with more than one answer, that respondent 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.

In most instances **by category variables** are denoted by a C after the original variable name, **by mention variables** are denoted by an M. Documentation for the CAPI questionnaires (household and individual) shows only the name of the first variable (which stores the number of mentions).

### 4.3 Missing values conventions

- 1 Not applicable: Used to signify that a particular variable did not apply to a given respondent usually because of internal routing. For example, men in women only questions.
- 2 Schedule not applicable: Used when a whole module is missed i.e. all nurse variables when a nurse visit was not achieved or self completion variables when the respondent is not of the given age range.
- 8 Don't know, Can't say.
- 9 No answer/ Refused.

These conventions have also been applied to most of the derived variables. The derived variable specifications should be consulted for details.

### 4.4 Valid cases

In the 2008 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.

## 5. Weighting variables

Prior to 2003, the weighting strategy for the core sample in the HSE was to apply selection weights only, and no attempt was made to reduce non-response bias through weighting. However, following a review of the weighting for the HSE, 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 2008 core general population sample data. (For more detailed information on how the weights were produced see Health Survey for England 2008: Volume 2: Methodology and documentation).

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 7.1 Individual level non-response weights have also been generated for the general population and are described in section 7.2 onwards.

The individual weights adjust for the additional non-response among individuals in participating households and additional weights take into account respondents participation in five phases: interview, nurse visit, saliva sample, blood sample and accelerometer sample.

### 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 where the limit of three households are selected at addresses with more than three.) Note that the population control totals used for the calibration weighting were the ONS projected mid-year population estimates for 2007, but with a small adjustment to exclude (our best estimate of) the population aged 65 and over living in communal establishments.

### 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 in both the Core and the Boost sample, taking into account child selection only and not adjusting for non-response, the (***wt\_child***) variable can be used. For analysis of children aged 2-15 in the only Boost sample the (***wt\_childb***) variable can

### **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 and agreed or 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 Saliva weight**

- A saliva weight has been generated for all adults and children that are 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 Accelerometer weights**

- Only respondents in a sub-sample of the selected core addresses were eligible to be selected to wear an accelerometer. This generated an additional set of calibration weights (***wt\_hhld\_acc***) which adjust for dwelling unit and household selection, and for the age/sex and region profiles of participating households selected for the accelerometer sample.

Within the selected households, selection weights were calculated to correct for the selection of two respondents (either two adults in households with no children, or one adult and one child in households with children).

Further weights have been generated that could be used to analyse the interview data (**wt\_int\_acc**), nurse data (**wt\_nurse\_acc**) and blood (**wt\_blood\_acc**) and cotinine data (**wt\_cot\_acc**) for the accelerometer sample.

The child accelerometer sample is defined as all children aged 4-15 from core and boost sample addresses who provided usable data from an accelerometer (**wt\_acc\_ch**).

## 6. HSE 2008 Report

Further information about the Health Survey for England 2008 is available in:

- Craig R, Mindell J and Hirani V (eds). Health Survey for England 2008: Physical Activity and fitness. Volume 1. Physical Activity and fitness. The Information Centre, Leeds, 2009.
- Craig R, Mindell J and Hirani V (eds). Health Survey for England 2008: Physical Activity and fitness. Volume 2. Methods and documentation. The Information Centre, Leeds, 2009.
- Craig R, Mindell J and Hirani V (eds). Health Survey for England 2008: Summary of key findings. The Information Centre, Leeds, 2009.
- Health Survey for England 2008: Latest trends. The Information Centre, Leeds, 2009.
- Population Number Estimates - user guide. The Information Centre, Leeds, 2009.
- Adult Trend Tables 2008. The Information Centre, Leeds, 2009.
- Child Trend Tables 2008. The Information Centre, Leeds, 2009.
- Population Number Estimate Tables 2008. The Information Centre, Leeds, 2009.

Or on the Information Centre website:

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

For the general population, tables showing selected trends from 1993 to 2008 can be found on The Information Centre web page: <http://www.ic.nhs.uk/statistics-and-data-collections/health-and-lifestyles-related-surveys/health-survey-for-england/health-survey-for-england--2008-trend-tables>

## 7. APPENDIX A

### HEALTH SURVEY FOR ENGLAND 2008 – CONTENTS

<b>Household data</b>	Household income
Household size, composition and relationships	Smoking in household
Accommodation tenure and number of bedrooms	Type of dwelling and area
Economic status/occupation of Household Reference Person	Car ownership

#### Individual level information

Table 1

#### Topics covered in 2008 Health Survey for England

	Age								
	0-1	2-3	4	5-7	8-10	11-12	13-15	16-74	75+
<b>Interviewer visit</b>									
General health, longstanding illness, limiting longstanding illness, acute sickness	●	●	●	●	●	●	●	●	●
Fruit and vegetable consumption				●	●	●	●	●	●
Children's eating habits (fat, sugar)		●	●	●	●	●	●		
Child physical activity		●	●	●	●	●	●		
Adult physical activity								●	●
Smoking					● <sup>a</sup>	● <sup>a</sup>	● <sup>a</sup>	● <sup>b</sup>	●
Drinking (seven day period)					● <sup>a</sup>	● <sup>a</sup>	● <sup>a</sup>	● <sup>b</sup>	●
Economic status/occupation, educational achievement								●	●
Ethnic origin	●	●	●	●	●	●	●	●	●
Height measurement		●	●	●	●	●	●	●	●
Weight measurement	●	●	●	●	●	●	●	●	●
Reported birth weight	●	●	●	●	●	●	●		
Consent to linkage to NHS Central Register/Hospital Episodes Statistics								●	●
GHQ12							● <sup>a</sup>	● <sup>a</sup>	● <sup>a</sup>
EQ5D								● <sup>a</sup>	● <sup>a</sup>
Cycling safety					● <sup>a</sup>	● <sup>a</sup>			
Perception of weight					● <sup>a</sup>	● <sup>a</sup>	● <sup>a</sup>		
Accelerometer			●	●	●	●	●	●	●
<b>Nurse visit</b>									
Immunisations	●								
Infant length	●								
Prescribed medicines and vitamin supplements	●	●	●	●	●	●	●	●	●
Nicotine replacements								●	●
Waist and hip circumference						●	●	●	●
Blood pressure				●	●	●	●	●	●
Step test								●	
Saliva sample (cotinine)			●	●	●	●	●	●	●
Blood sample								●	●
Adult eating habits								● <sup>a</sup>	● <sup>a</sup>

<sup>a</sup> This module was administered by self-completion.

<sup>b</sup> This module was administered by self-completion for those aged 16-17 and some aged 18-24.