

Health Survey for England

**Cardiovascular
Disease and risk
factors**

2006

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

1. BACKGROUND	3
2. SURVEY DESIGN	3
3. DOCUMENTATION	5
4. USING THE DATA	5
4.1 VARIABLES ON THE FILES	5
4.2 MULTICODED QUESTIONS	6
4.3 MISSING VALUES CONVENTIONS	7
4.4 VALID CASES	7
5. WEIGHTING VARIABLES	8
5.1 HOUSEHOLD WEIGHT	8
5.2 INDIVIDUAL WEIGHT	9
5.3 NURSE WEIGHT	9
5.4 BLOOD WEIGHT	10
5.5 CHILD SAMPLE WEIGHTS	10
6. HSE 2006 REPORT	11
7. APPENDIX A	12

1. Background

The data files contain data from Health Survey for England 2006 (HSE), the sixteenth year of a series of surveys designed to monitor trends in the nation's health. The 2006 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.

2. Survey Design

The Health Survey for England 2006 was designed to provide data at both national and regional level about the population living in private households in England. The sample for the HSE 2006 comprised of two components: the core (general population) sample and a boost sample of children aged 2-15. The core sample was designed to be representative of the population living in private households in England and should be used for analyses at the national level. The core sample was split in two for some modules of the 2006 survey, further details are shown in Appendix A.

A random sample of 720 PSUs (Primary Sampling Units) was selected for the core sample and 468 PSUs selected for the child boost sample. The PSUs were selected with probability proportional to the total number of addresses within them. Once selected, the PSUs were randomly allocated to the 12 months of the year (60 per month in the core sample, 39 per month in the boost) so that each quarter provided a nationally representative sample.

For the core sample, a sample of 20 addresses was selected within each selected postcode sector, giving a total selected sample of 14,400 (720 x 20) addresses. Because PSUs were sampled with probability proportional to the numbers of addresses, and then a fixed number of addresses was sampled in each PSU, every address had an equal chance of being included in the sample. For the child boost sample, a random sample of 36 addresses was selected in each PSU, giving a total sample of 16,848 addresses (468 x 36). The addresses sampled for the child boost sample also had equal probability of being selected for that sample.

For the HSE core sample, all adults aged 16 years or older at each household were selected for the interview (up to a maximum of ten adults). However, a limit of two was placed on the number of interviews carried out with children aged 0-15. For households with three or more children, interviewers selected two children at random.

At boost addresses interviewers screened for households containing at least one child aged 2-15 years. For households which included eligible children, up to two were selected by the interviewer for inclusion in the survey.

An interview with each eligible person was followed by a nurse visit both using computer assisted interviewing (CAPI). The 2006 survey for adults focused primarily on Cardiovascular Disease (CVD) and its risk factors. All adults were also asked modules of questions on general health, alcohol consumption, smoking, fruit and vegetable consumption and physical activity. Adults aged 65 plus were randomly allocated to one of two questionnaire versions to avoid lengthy interviews. This included either the CVD module and a short physical activity module, or a long physical activity module but not the CVD module (further details are provided in section 5). Adults aged 16-64 completed both the CVD and long physical activity modules.

Children aged 13-15 were interviewed themselves, and parents of children aged 0-12 were asked about their children, with the child interview including questions on physical activity and fruit and vegetable consumption.

Interviewing was conducted throughout the year to take account of seasonal differences.

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 2006 data consists of two files; one at individual level and one at household level:

HSE06ai.sav	21,399 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).
-------------	----------------	--

HSE06ah.sav	30,068 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 2006 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 2006 Health Survey.

As an example, question CONSBX1 on the 2006 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 2006, 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 2006 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 2006 core general population sample data. (For more detailed information on how the weights were produced see Health Survey for England 2006, Volume 3: 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 5.1. Individual level non-response weights have also been generated for the general population and are described in section 5.2 onwards.

These individual weights adjust for the additional non-response among individuals in participating households and additional weights take into account respondents aged 65 or more who were randomly allocated to two groups.

Those in the first group (sample type 1) answered the full CVD module and a short version of the physical activity module, whilst those in the second group (sample type 2) answered the physical activity module but did not answer the CVD module. Therefore it was necessary to produce two additional sets of weights for each interview. If the standard core weights had been used, then people aged 65 or older would have been under-represented in the weighted estimates for these questions.

In order to ensure that people aged 65 or over were correctly represented for these analyses, adjustments were made to the weights to ensure that the weighted age distribution for the two groups matched the weighted age distribution of those aged 65 and over in the full sample. This was done separately for the three sample types: interview, nurse visit and blood samples (i.e. six extra sets of weights were produced).

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 2005, 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 adults (aged 65 or more) who were in sample type 1 and answered the full CVD module, the interview weights are adjusted accordingly and the weighting variable to use is (***wt_int_s1***).
- for adults (aged 65 or more) who were in sample type 2 and answered the short version of the physical activity module, the interview weights are adjusted accordingly and the weighting variable to use is (***wt_int_s2***).
- 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 the Core sample, taking into account child selection only and not adjusting for non-response, the (***child_wt***) 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. Two additional weights have been computed to adjust for;

- any analysis focusing on adults (aged 65 or more) who were in sample type 1 and who answered the full CVD module and a short version of the physical activity module. The nurse weight variable is (***wt_nurse_s1***).

- Any analysis focusing on adults (aged 65 or more) who were in sample type 2 and who answered the full physical activity module but did not answer the CVD module, the nurse weight variable is (***wt_nurse_s2***).

5.4 Blood weight

A blood weight has been generated for individuals aged 16 and over who had a nurse visit and were eligible for a blood sample. This weight (***wt_blood***) should be used on all analysis of questions asked relating to blood samples. Two additional weights have been computed to adjust for;

- Any analysis focusing on adults who were in sample type 1 and had answered the CVD module against relevant blood information should use the weight variable (***wt_blood_s1***).
- Any analysis focusing on adults who were in sample type 2 and had answered the Physical Activity module against relevant blood information should use the weight variable (***wt_blood_s2***).

5.5 Child sample weights

The child sample is defined as all children aged 0-15 from the core sample and all children aged 2-15 from the boost sample addresses. Child weights were calculated to take into account, household and child selection, plus seasonal adjustment to compensate for any bias when the boost sample was only being carried out during the winter months. The final calibrated weight is called (***wt_intch***).

6. HSE 2006 Report

Further information about the Health Survey for England 2006 is available in:

Craig R and Mindell J (eds). *Health Survey for England 2006. Volume1: Cardiovascular disease and risk factors in adults*. The Information Centre, Leeds, 2007.

Craig R and Mindell J (eds). *Health Survey for England 2006. Volume2: Obesity and other risk factors in children*. The Information Centre, Leeds, 2007.

Craig R and Mindell J (eds). *Health Survey for England 2006. Volume3: Methodology and Documentation*. The Information Centre, Leeds, 2007.

Or on the Information Centre website:

<http://www.ic.nhs.uk/pubs/hseupdate06>

For the general population, tables showing selected trends from 1993 to 2006 can be found on The Information Centre web page: <http://www.ic.nhs.uk/statistics-and-data-collections/health-and-lifestyles/health-survey-for-england/health-survey-for-england--updating-of-trend-tables-to-include-2006-data>

7. APPENDIX A

HEALTH SURVEY FOR ENGLAND 2006 – CONTENTS

Household data

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
Household income	

Individual level information

	0-1	2-3	4	5-7	8-10	11-12	13-15	16-64	65+
Interviewer visit									
General health, longstanding illness, limiting longstanding illness, acute sickness, fractures	●	●	●	●	●	●	●	●	●
Use of social care services									●
Carers' responsibilities								●	●
CVD								●	● ^a
Child Physical activity		●	●	●	●	●	●		
Adult Physical activity (short version)									● ^a
Adult Physical activity (long version)								●	● ^a
Smoking					● ^b	● ^b	● ^b	● ^c	●
Drinking (seven day period)					● ^b	● ^b	● ^b	● ^c	●
Fruit and vegetable consumption (and salt)				●	●	●	●	●	●
Eating habits (fat, sugar)		●	●	●	●	●	●		
Economic status/occupation, educational achievement								●	●
Ethnic origin	●	●	●	●	●	●	●	●	●
Height measurement		●	●	●	●	●	●	●	●
Weight measurement	●	●	●	●	●	●	●	●	●
Reported birth weight	●	●	●	●	●	●	●		
Cycling safety					● ^b	● ^b			
Psychosocial health (GHQ 12)							● ^b	● ^b	● ^b
Euroqol general health (EQ5D)								● ^b	● ^b
Social support, social capital								● ^b	● ^b
Strengths and difficulties			● ^d	● ^d	● ^d	● ^d	● ^d		
Perception of weight					● ^b	● ^b	● ^b		
Use of contraceptive pill								● ^{b, e}	● ^{b, e}
Hormone replacement therapy								● ^{b, e}	● ^{b, e}

^a To avoid an overlong interview for informants aged 65, they were randomly allocated to one of two groups; one group answered CVD questions and a short version of the physical activity questions, and the second group completed the full version of the physical activity questions but not the CVD questions.

^b These modules were administered by self-completion.

^c This module was administered by self-completion for those aged 16-17 and some aged 18-24.

^d This module was administered by self-completion to parents of 4-15 year olds.

^e This is asked of women aged 18 and over only (there are no HRT questions in the young adult self-completion).

	0-1	2-3	4	5-7	8-10	11-12	13-15	16-64	65+
Nurse visit									
Prescribed medicines and vitamin supplements	●	●	●	●	●	●	●	●	●
Nicotine replacements								●	●
Immunisations	●								
Blood pressure				●	●	●	●	●	●
Eating Habits								● ^b	● ^b
Infant length	●								
Waist and hip circumference						●	●	●	●
Demi-span									●
Blood sample – total & HDL cholesterol, ferritin, haemoglobin, glycated haemoglobin, fibrinogen, c-reactive protein								●	●
Saliva sample (cotinine)			●	●	●	●	●		
Urine sample								●	●

^b This module was administered by self-completion.