# Health Survey for England

# Knowledge and Attitudes

# '07

# 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

The data files contain data from Health Survey for England 2007 (HSE), the seventeenth year of a series of surveys designed to monitor trends in the nation's health. The 2007 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
- to monitor the impact of the national smoking ban 2007.

# 2. Survey Design

The Health Survey for England 2007 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 2007 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 2007 survey, further details are shown in Appendix A.

A random sample of 720 PSUs (Primary Sampling Units) was selected for the core and the boost sample, an additional 180 PSUs were used to supplement 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, 15 per month in the additional child boost) so that each

quarter provided a nationally representative sample.

Within each of the 720 core PSUs a sample of 36 addresses was selected. The selected addresses were randomly allocated to either the core or child boost sample: 10 addresses to the core sample and 26 to the child boost sample. In total therefore, there were 10 core addresses allocated within each PSU, giving a total sample of 7,200 (720 x 10) core addresses, and 18,720 child boost addresses (720 x 26).

For the 180 additional child boost PSUs, a random sample of 41 addresses was selected in each PSU, giving a total sample of 7,380 addresses (180 x 41) for the additional child boost sample. The total child boost sample was thus 26,100 addresses (18,720 from the child boost sample in core points and 7,380 from the additional child boost 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 2007 survey for adults focused on lifestyle behaviour, knowledge and attitudes. Adults were asked modules of questions on general health, alcohol consumption, smoking, and fruit and vegetable consumption. Knowledge and attitudes were covered in self-completion questionnaires.

Children aged 13-15 were interviewed themselves, and parents of children aged 0-12 were asked about their children, with the interview including questions on eating habits (fat and sugar consumption) and fruit and vegetable consumption. Children in the boost sample only were asked about physical activity.

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 2007 data consists of two files; one at individual level and one at household level:

| | | |
|---|---|---|
| HSE07ai.sav | 14,386 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). |

| | | |
|---|---|---|
| HSE07ah.sav | 24,910 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 2007 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 2007 Health Survey.

As an example, question CONSBX1 on the 2007 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 if 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 2007 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 2007 core general population sample data. (For more detailed information on how the weights were produced see Health Survey for England 2007: 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 three phases: interview, nurse visit and saliva samples.

## 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 2006, 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 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.

# 6. HSE 2007 Report

Further information about the Health Survey for England 2007 is available in:

- Craig R and Shelton N (eds). Health Survey for England 2007: Volume 1 – Healthy lifestyles: knowledge, attitudes and behaviours. The Information Centre, Leeds, 2008.
- Craig R and Shelton N (eds). Health Survey for England 2007: Volume 2 - Methodology and documentation. The Information Centre, Leeds, 2008.
- Craig R and Shelton N (eds). Health Survey for England 2007: Summary of key findings. The Information Centre, Leeds, 2008.
- Health Survey for England 2007: Latest trends. The Information Centre, Leeds, 2008.
- Population Number Estimates - user guide. The Information Centre, Leeds, 2008.
- Adult Trend Tables 2007. The Information Centre, Leeds, 2008.
- Children Trend Tables 2007. The Information Centre, Leeds, 2008.
- Population Number Estimate Tables 2007. The Information Centre, Leeds, 2008.

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 2007 can be found on The Information Centre web page: www.ic.nhs.uk/pubs/hse07trends

# 7. APPENDIX A

**HEALTH SURVEY FOR ENGLAND 2007 – CONTENTS**

**Household data**

| | |
|---|---|
| Household size | Smoking in household |
| Age/sex of household members | Household income |
| Marital status and living arrangements aged 16+ | Car ownership |
| Relationship between household members | Accommodation tenure and number of bedrooms |
| Economic status/occupation of Household Reference Person | Type of dwelling and area |

**Individual level information**

Table 1

**Topics covered in 2007 Health Survey for England**

| ***Individual questionnaire*** | **Age** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **0-1** | **2-4** | **5-7** | **8-10** | **11-12** | **13-15** | **16-17** | **18-64** | **65+** |
| **CAPI interview** | | | | | | | | | |
| General health, longstanding illness, limiting longstanding illness, acute sickness, fractures, childhood diabetes, estimated height at age 25 (for those aged 30+) | • | • | • | • | • | • | • | • | • |
| Fruit and vegetables (and salt questions) | | | • | • | • | • | • | • | • |
| Smoking (includes new questions on places of exposure to second hand smoke) | | | | •[a] | •[a] | •[a] | •[a] | •[a] | • |
| Drinking (heaviest day in the last week) | | | | •[a] | •[a] | •[a] | •[a] | • | • |
| Child physical activity (including sitting questions) | •[b] | •[b] | •[b] | •[b] | •[b] | •[b] | | | |
| Eating Habits (fat & sugar) | | • | • | • | • | • | | | |
| Economic status / occupation | | | | | | | • | • | • |
| Educational attainment | | | | | | | • | • | • |
| Ethnic origin | • | • | • | • | • | • | • | • | • |
| Reported birth weight | • | • | • | • | • | • | | | |
| **Self completion** | | | | | | | | | |
| Perception of weight | | | | • | | | | | |
| Cycle helmets | | | | • | | | | | |
| Smoking and drinking | | | | • | • | • | • | | |
| Strengths and difficulties for parents of children aged 4-15 | | •[b, c] | •[b] | •[b] | •[b] | •[b] | | | |
| Knowledge of and attitudes to smoking, drinking, healthy eating | | | | | • | • | • | • | • |
| Knowledge and attitudes to physical activity | | | | | • | • | • | • | |
| **Physical measurements** | | | | | | | | | |
| Height measurement | • | • | • | • | • | • | • | • | • |
| Weight measurement | • | • | • | • | • | • | • | • | • |

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

[b] These questions are only asked in the boost sample, that is at screening addresses within core points as

well as in boost only points.

[c] Starts at children aged 4 only

| ***Nurse visit*** | **Age** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **0-1** | **2-3** | **4** | **5-10** | **11-12** | **13-15** | **16-64** | **65+** |
| | | | | | | | | |
| **Nurse visit** | | | | | | | | |
| Infant length measurements | • | | | | | | | |
| Immunisations | • | | | | | | | |
| Prescribed medicines, vitamin supplements, nicotine replacement (16+) | • | • | • | • | • | • | • | • |
| Cigarette brand choice | | | | | | | • | • |
| Blood pressure | | | | • | • | • | • | • |
| Waist and hip circumference | | | | | • | • | • | • |
| Demi-span | | | | | | | •[d] | • |
| Saliva sample (cotinine) | | | • | • | • | • | • | • |
| Urine | | | | | | | • | • |
| **Self completion** | | | | | | | | |
| Eating Habits | | | | | | | • | • |

[d] Demi-span measurements will be taken of those aged 25-45 for the first 1000 individuals (500 men and 500 women).