

Health Survey for England

**The health of  
minority ethnic  
groups**

**'04**

**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

# 1. Background

The data files contain data from Health Survey for England 2004 (HSE), the fourteenth year of a series of surveys designed to monitor trends in the nation's health. The 2004 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;
- to assess the frequency with which combinations of risk factors 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.

## 2. Survey Design

The Health Survey for England 2004 was designed to provide data at both national and regional level about the population living in private households in England. The sample design of the 2004 survey had two parts: a general population sample that followed the same pattern as in previous years and a minority ethnic 'boost' sample, designed solely to yield interviews with members of seven largest minority ethnic groups in England: Black Caribbean, Black African, Indian, Pakistani, Bangladeshi, Chinese and Irish.

The general population sample was half the size of the usual sample, and involved selection 6,552 addresses from the Postcode Address File (PAF) in 312 wards, issued over a twelve-month period from January to December 2004. Up to ten adults and up to two children in each household were interviewed, and a nurse visit arranged for those participants in minority ethnic groups who consented.

In the ethnic boost sample, 41,436 addresses were randomly selected from PAF, within another 483 wards, issued over the same 12 month period, January to December 2004. All sampled addresses were fully screened and only informants from the specified minority ethnic groups were eligible for inclusion in the survey. Among those eligible informants at an address, up to four adults and three children were selected for interview, with a random selection of participants if there was more than this number in an eligible household.

In order to increase further the number of Chinese informants in 2004, the sample was supplemented with an extra sample consisting of people with 'Chinese sounding' surnames obtained from the Electoral Register (for further information see HSE 2004 report, Volume 2, Methodology and documentation).

For informants from the specified minority ethnic groups (whether identified in the general population sample or the minority ethnic sample), an interview with each eligible person was followed by a nurse visit both using computer assisted interviewing. The main focus of the 2004 survey for adults from minority ethnic backgrounds was cardiovascular disease (CVD) and related risk factors. At the nurse visit, questions were asked about prescribed medication, vitamin supplements and nicotine replacements. The nurse took the blood pressure of those aged 5 and over, measured lung function of those aged 7-15, and made waist and hip measurements for those aged 11 and over. Saliva samples were collected from 4-15 year olds and blood samples from those aged 11 and over including fasting blood from those aged 16 and over. Blood and saliva samples were sent to a laboratory for analysis.

Informants in the general population sample, unless they were members of the specified minority ethnic groups, were given a shortened version of the questionnaire covering core topics only.

For all informants, information was obtained directly from persons aged 13 and over. Information about children under 13 was obtained from a parent with the child present.

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 2004 data consists of two individual level files and one household level file:

HSE04gpa.sav	8,354 records	contains data for all individuals in the <i>General Population Sample</i> in co-operating households who gave a full interview. It contains information from the household questionnaire, main individual schedule, self-completions and the nurse visit (where one occurred).
HSE04etha.sav	10,114 records	contains data for all individuals in the <i>Ethnic Boost Sample and informants in the General Population Sample who were of the specified ethnic groups</i> in co-operating households who gave a full interview. It contains information from the household questionnaire, main individual schedule, self-completions and the nurse visit (where one occurred).

HSE04ah.sav	21,157 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 Weighting variables

### *General Population Data (HSE04gpa.sav)*

Prior to 2003, the weighting strategy for the core sample in the HSE was to apply selection weights only – no attempt was made to reduce non-response bias through weighting. However, following a review of the weighting, non-response was also included in the weighting strategy for the HSE 2003. We have followed the same approach for weighting the HSE 2004 general population sample data.

Two sets of non-response weights have been generated for the general population sample: household weights which adjust for non-contact and refusal of households, and interview weights which also adjust for the additional non-response among individuals in participating households.

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 2004, but with a small adjustment to exclude (our best estimate of) the population aged 65 and over living in communal establishments.

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 the General Population Sample, taking into account child selection only and not adjusting for non-response, the **child\_wt** variable can be used.

### *Minority Ethnic Group Data (HSE04etha.sav)*

For the HSE 2004, as well as the general population sample, boost samples were selected in areas with (relatively) higher proportions of people in minority ethnic groups. All respondents, whether from the general population sample or the boost sample, are included in the minority ethnic group sample. As well as the main interview, respondents in the minority ethnic group sample were also eligible for nurse visits and to have blood taken. Therefore, three sets of weights were generated: interview weights, plus nurse and blood weights.

The first stage of the weighting process was to generate weights for the probability of selecting an address in the minority ethnic sample. Addresses in areas with different ethnic profiles had different chances of being selected for the HSE - selection weight were generated which corrected for this. These weights were combined with the selection weights for household within addresses and for individuals within household to give the interview weights (**wt\_int**).

All respondents in the minority ethnic group sample were eligible for a nurse visit and were also asked to give a sample of blood (if aged 11 or more). As there was drop-out at both these stages, separate weights were generated for the nurse visit sample (**wt\_nurse**) and blood sample (**wt\_blood**).

The tables in the published volumes of the HSE2004 were weighted using **wt\_int**, **wt\_nurse** and **wt\_blood** in the relevant sections.

### 4.3 Multicoded questions

Multicoded questions are stored in the archived HSE 2004 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 2004 Health Survey.

As an example, question CONSBX1 on the 2004 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 2004, 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).

### 4.4 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 mainly for variables on the self-completions when the respondent was not of the given age range, also used for children without legal guardians in the home who could not participate in the nurse schedule.
- 8 Don't know, Can't say.
- 9 No answer/ Refused

These conventions have also been applied to most of the derived variables. Those variables created in earlier years of the Health Survey and used again in 2004 do not on the whole conform to this scheme.

The derived variable specifications should be consulted for details.

#### **4.5 Valid cases**

In the 2004 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 or eaten within 30 minutes of having their blood pressure taken were excluded from analysis as this can affect blood pressure.

## **5. HSE 2004 Report**

Further information about the Health Survey for England 2004 is available in:

Sproston K and Mindell J (eds) *Health Survey for England 2004. Volume 1: The Health of Minority Ethnic Groups*. National Centre for Social Research 2006

Sproston K and Mindell J (eds) *Health Survey for England 2004. Volume 2: Methodology and documentation*. National Centre for Social Research 2006

or on the Information Centre website:

<http://www.ic.nhs.uk/pubs/healthsurvey2004ethnicfull>

For the general population, tables showing selected trends from 1993 to 2004 can be found on The Information Centre web page:

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

## APPENDIX A

### 2004 HEALTH SURVEY FOR ENGLAND – CONTENTS

<b>Household level information</b>												
Household size, composition and relationships							Smoking in household					
Accommodation tenure and number of bedrooms							Car ownership					
Economic status/occupation of HRP							Heating/cooking appliances					
Type of dwelling and area							Mould and dampness					
Household income							Household pets					
<b>Individual level information</b>												
	<b>Age</b>											
	<2	2-3	4	5-6	7	8-9	10	11-12	13-15	16-64	65+	
<b>Interviewer visit</b>												
General health, longstanding illness, limiting longstanding illness, acute sickness, fractures	●	●	●	●	●	●	●	●	●	●	●	
Cardiovascular disease (CVD) (Including use of services and Rose Angina questionnaire)										●●	●●	
Physical activity		●	●	●	●	●	●	●	●	●	●	
Smoking						● <sup>a</sup>	● <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>a</sup>	● <sup>b</sup>	●	
Fruit and vegetable consumption				●	●	●	●	●	●	●	●	
Respiratory problems	●●	●●	●●	●●	●●	●●	●●	●●	●●			
Complimentary and alternative medicine										● <sup>d</sup>	● <sup>d</sup>	
Economic status/occupation, educational attainment										●	●	
Ethnic origin	●	●	●	●	●	●	●	●	●	●	●	
Country of birth										●●	●●	
Parental health										●●	●●	
Height measurement		●	●	●	●	●	●	●	●	●	●	
Weight measurement	●	●	●	●	●	●	●	●	●	●	●	
Reported birth weight	●	●	●	●	●	●	●	●	●			
Cycling safety						● <sup>a</sup>	● <sup>a</sup>	● <sup>a</sup>				
Psychosocial health (GHQ 12)									● <sup>a</sup>	● <sup>a</sup>	● <sup>a</sup>	
Euroqol general health (EQ-5D)										● <sup>a</sup>	● <sup>a</sup>	
Social support, social capital										● <sup>a</sup>	● <sup>a</sup>	
Use of contraceptive pill										● <sup>a</sup>	● <sup>a</sup>	
Hormone replacement therapy										● <sup>c</sup>	● <sup>a</sup>	
<b>Nurse visit</b>												
Prescribed medicines and vitamin supplements	●●	●●	●●	●●	●●	●●	●●	●●	●●	●●	●●	
Nicotine replacements										●●	●●	
Immunisations	●●											
Blood pressure				●●	●●	●●	●●	●●	●●	●●	●●	
Waist and hip circumference									●●	●●	●●	
Infant length	●●											
Lung function					●●	●●	●●	●●	●●			
Saliva sample – cotinine			●●	●●	●●	●●	●●	●●	●●	●●	●●	
Urine sample										●●	●●	

	<2	2-3	4	5-6	7	8-9	10	11-12	13-15	16-64	65+
Blood sample - total & HDL cholesterol, ferritin, haemoglobin, glycated haemoglobin								●●	●●	●●	●●
Blood sample - fibrinogen, c-reactive protein										●●	●●
Blood sample – IgE, HDM IgE								●●	●●		
Fasting blood sample - triglycerides, LDL cholesterol, glucose										●●	●●
Eating habits (fat, salt)										●● <sup>a</sup>	●● <sup>a</sup>
Strength and difficulties			●● <sup>e</sup>	●● <sup>e</sup>	●● <sup>e</sup>	●● <sup>e</sup>	●● <sup>e</sup>	●● <sup>e</sup>	●● <sup>e</sup>	●● <sup>e</sup>	

● Asked of general population and minority ethnic informants

●● Asked minority ethnic informants only

<sup>a</sup> These modules were administered by self-completion

<sup>b</sup> These modules were administered by self-completion for those aged 16-17 and some aged 18-24

<sup>c</sup> 18+ only (there are no HRT questions in the young adult self-completion)

<sup>d</sup> This module was introduced in April to December points only

<sup>e</sup> This module was asked by proxy and administered by self-completion for parents of 4-15 year olds.