Health Survey for England

Cardiovascular Disease

'03

User Guide

A survey carried out on behalf of The Department of Health

1. Background

The data files contain data from Health Survey for England 2003 (HSE), the twelfth year of a series of surveys designed to monitor trends in the nation's health. The 2003 Health Survey was commissioned by the Department of Health 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 2003 was designed to provide data at both national and regional level about the population living in private households in England. The sample for the 2003 was designed to be a cross-section of the population living in private households for which 13,680 addresses were drawn from the Postcode Address File (PAF).

All private households in the general population sample are eligible for inclusion in the survey (up to a maximum of three households per address). Up to two children aged 0-15 are interviewed in each household, as well as up to 10 adults aged 16 and over. Information was obtained directly from persons aged 13 and over. Information about children under 13 was obtained from a parent with the child present.

An interview with each eligible person was followed by a nurse visit both using computer assisted interviewing. The nurse visit was split into two sample types, standard and long. The standard nurse visit collected blood pressure measurements, saliva samples, waist and hip measurements and non-fasting blood samples. At one sixth of the selected addressed the nurse visit was extended to include a fasting blood sample (from those aged 35 and over) and a spot urine sample (from adults aged 16 and over). Blood and saliva samples were sent to a laboratory for analysis.

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 2003 data consists of one individual level file and one household level file:

HSE03ai.sav	18,553 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).

HSE03ah.sav	20,993 records	contains data on household, and 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

In 2003, non-response weighting was introduced to the HSE data. Although the HSE has generally presented a good match to the population, this decision was taken to keep up with the recent changes on many large-scale government sponsored surveys, and with the aim of reducing the possible biases.

Non-response weights have been calculated for both adults and children. Four sets of non-response weights have been generated in total. Firstly a household weight was calculated to adjust for non-contact and for refusals of entire households. In addition, three sets of weights have been calculated to adjust (a) non-response among individuals in responding households (b) non-response to the nurse visit stage and (c) refusal to give a blood sample. The aim of each set of weights is that each of the main datasets (households, individuals, individuals who see a nurse, and individuals who give blood) can be treated as broadly representative of the general household population.

The household weight (**hhld_wt**) is the product of household selection weight (to adjust for addresses with more than three households per address) and the calibration weight. These weights were applied separately within Government Office Region to bring the age and sex distribution of adults and children within responding households into line with each region's population age-sex distribution, but with the constraint that adults and children from the same household are all given the same weight. The rationale behind calibration weighting is that it attaches an estimated probability of response to each household that 'explains' any discrepancy between the survey age-sex distribution and the population age-sex distribution.[1]

The population control totals used for this exercise were the ONS projected population estimates for 2003, but with a small adjustment to exclude (our best estimate of) the population aged 65 and over living in communal establishments.

At the individual level there are three sets of weights, the interview weight (**int_wt**), the nurse weight (**nurse_wt**) and the blood weight (**blood_wt**). The appropriate weight variable should be used for analysis done using data from the relevant sections.

Children aged 0-15: To compensate for limiting the number of children interviewed in a household to two (the sampling fraction therefore being lower in households containing three or more children) it has become necessary to weight the child sample. This 'child weight' is the total number of children aged 0-15 in the household divided by the number of selected children in the household. The weighted sample was then adjusted to ensure that the age/sex distribution matched that of all children in co-operating households.

The variable **child_wt** contains the appropriate selection weights for children aged 0-15.

The variables **int_wt** and **nurse_wt** for children aged 0-15 includes both the child selection weights and non- response weights.

The tables in the published volumes of the HSE2003 have been weighted using the int_wt, nurse_wt or blood_wt variables in the relevant sections.

4.3 Multicoded questions

Multicoded questions are stored in the archived HSE 2003 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 2003 Health Survey.

As an example, question CONSBX1 on the 2003 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 2003, 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.

-6   Schedule not obtained: Used to signify that a particular variable was not answered because the respondent did not complete or agree to a particular schedule (i.e. nurse schedule or self-completions).

-7   Refused/ not obtained: Used only for variables on the nurse schedules, this code indicates that a respondent refused a particular measurement or test or the measurement was attempted but not obtained or not attempted.

-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 2003 do not on the whole conform to this scheme. The derived variable specifications should be consulted for details.

4.5 Valid cases

In the 2003 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. The List of Variables document gives details of which variables show only valid codes, and which also include invalid answers.

5. HSE 2003 Report

Further information about the Health Survey for England 2003 is available in:

Sproston K and Primatesta P (eds) *Health Survey for England 2003 Volume 1:Cardiovascular disease.* The Stationery Office, London, 2004.

Sproston K and Primatesta P (eds) *Health Survey for England 2003 Volume 2: Risk factors for cardiovascular disease.* The Stationery Office, London, 2004.

Sproston K and Primatesta P (eds) *Health Survey for England 2003 Volume 3: Methodology and documentation.* The Stationery Office, London, 2004.

or on the Department of Health website:
http://www.dh.gov.uk/PublicationsAndStatistics/PublishedSurvey/HealthSurveyForEngland/fs/en

APPENDIX A
2003 HEALTH SURVEY FOR ENGLAND – CONTENTS

Household level information

Household size, composition and relationships	Smoking in household
Tenure and number of bedrooms	Household income
Economic status/occupation of HRP	Car ownership
Type of dwelling and area	

Individual level information

	Age											
Interviewer visit	<2	2-3	4	5-6	7	8-9	10	11-12	13-15	16-34	35-64	65+
General health, longstanding illness, limiting longstanding illness, acute sickness, fractures	●	●	●	●	●	●	●	●	●	●	●	●
Cardiovascular disease (CVD) (Including diagnoses, treatment and use of services. Rose Angina question, symptoms of stroke)										●	●	●
Physical activity										●	●	●
Smoking						●[a]	●[a]	●[a]	●[a]	●[b]	●	●
Drinking (seven day period)						●[a]	●[a]	●[a]	●[a]	●[b]	●	●
Fruit and vegetable consumption				●	●	●	●	●	●	●	●	●
Economic status/occupation, educational attainment										●	●	●
Ethnic origin	●	●	●	●	●	●	●	●	●	●	●	●
Parental health										●	●	●
Height measurement		●	●	●	●	●	●	●	●	●	●	●
Weight measurement	●	●	●	●	●	●	●	●	●	●	●	●
Cycling safety						●[a]	●[a]	●[a]				
Psychosocial health (GHQ 12)									●[a]	●[a]	●[a]	●[a]
Euroqol general health (EQ-5D)										●[a]	●[a]	●[a]
Social support, social capital										●[a]	●[a]	●[a]
Use of contraceptive pill										●[a]	●[a]	●[a]
Hormone replacement therapy										●[c]	●[a]	●[a]
Nurse visit												
Prescribed medicines and vitamin supplements	●	●	●	●	●	●	●	●	●	●	●	●
Nicotine replacements										●	●	●
Immunisations	●											
Blood pressure				●	●	●	●	●	●	●	●	●
Waist and hip circumference										●	●	●
Saliva sample – cotinine			●	●	●	●	●	●	●			
Blood sample – total & HDL cholesterol, fibrinogen, c-reactive protein, glycated haemoglobin										●	●	●
Infant length	●											
Eating habits (fat, salt)										●[a]	●[a]	●[a]

Additional nurse procedures in the sub-sample (extended nurse visit)											●	●	●
Saliva sample – cotinine											●	●	●
Fasting blood sample – triglycerides, LDL cholesterol, glucose												●	●
Urine sample											●	●	●

[a] These modules were administered by self-completion

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

[c] 18+ only (there are no HRT questions in the young adult self-completion)

[1] In principle, if we had population estimates for age and sex by household composition (for example, the number of households with 2 adults: 1 man aged 70 and 1 woman aged 68) then we could calculate a *direct* estimate of the probability of a household responding in terms of its age-sex composition rather than a calibration estimate. However, because we do not have data to this level of detail, calibration weighting is a means of modelling the probabilities across household compositions whilst controlling to the marginal age-sex distribution.