



ENHANCING AND ADDING VALUE TO RESEARCH DATA

**LAURENCE
HORTON**

Research Data Management Senior Officer
UK DATA ARCHIVE
UNIVERSITY OF ESSEX

NCRM RESEARCH METHODS FESTIVAL 2010
6 JULY, 2010



KINDS OF DATA UK DATA ARCHIVE DEALS WITH

- **Quantitative**
 - micro data
 - macro data
 - formats: SPSS, Stata and tab delimited formats
- **Qualitative**
 - in-depth interviews, diaries, anthropological field notes
 - complete answers to survey questions
 - formats: Excel, Word and RTF
- **Multimedia**
 - image files, such as photographs, and audio files
 - formats: WAV, Mp3, JPEG, TIFF
- **Non-digital material**
 - photographs, reports, questionnaires and transcriptions
 - formats: paper, analogue audio or audio-visual recordings

WHAT DO USERS DO WITH DATA?

- Use descriptive material
- Comparative research, restudy or follow-up study
- Re-analysis/secondary analysis
- Research design and methodological advancement
- Replication of published statistics
- Teaching and learning

ADDING VALUE: THE INCENTIVE

- “The costs of archiving activities (archival storage and preservation planning and actions) are consistently a very small proportion of the overall costs and significantly lower than the costs of acquisition/ingest or access activities...”

N. Beagrie et.al *Keeping Research Data Safe 2*, p.4

PROVIDING GOOD DOCUMENTATION

- Quality of the information provided by the data creator determines ease of discovery and appropriate re-use
 - comprehensive and comprehensible documentation and metadata
 - complete the deposit form as fully as possible

EXAMPLES

- Quantitative dataset
 - Questionnaire, variable list, codebook
- Qualitative dataset – depends on size and scale
 - data listing, interview schedules

End of Award reports, publications list, methodology

QUANTITATIVE DOCUMENTATION

- smaller-scale study - user guide may just contain survey questionnaire, methodology information

Documentation:

Format	Name	Size in Kilo Bytes	Description
PDF	6112datadocs.pdf	808	Lists of Variables and Derived Variables
PDF	6112interviewingdocs.pdf	1655	Questionnaires, Showcards, Coding Frames and Consent Booklet
PDF	6112supportingdocs.pdf	1573	Interviewer, Nurse, Coding and Editing Instructions
PDF	6112userguide.pdf	288	User Guide
HTML	UKDA Study 6112 Information.htm	25	Study information and citation

Introduction

This document is the most sensible starting point to analysing the HSE data, as it categorises all the variables stored on the dataset to two levels, and it is therefore easier to see the coverage of questions asked at this summary level, rather than ploughing straight into the documentation of the questionnaires and self-completion booklets.

Once you have found the appropriate variables that you want to analyse, you then need to look at the other documentation to see in more detail exactly how the question was asked in the study, or how a derived variable has been defined.

The source of each variable is indicated in the final column of each table of variables with abbreviations as follows:

HHold	Household CAPI Questionnaire
Indiv	Individual CAPI Questionnaire
Nurse	Nurse CAPI Questionnaire
SC ...	Self-Completion Booklet: SC 8-10, SC 11-12, SC 13-15, SC YP, SC Adult, SC 65yrs+, SC PT 4-15, SC Eating Habits or where a question appears in more than one booklet the range is widened (eg SC8-15, SC 16+)
Lab	Results from laboratory, ie from saliva, serum or urine testing
ARF	Address Record Form completed for each issued address
NRF	Nurse Record Form completed for each household where at least one person had agreed to a nurse interview
Derived	A variable derived from other variables, and detailed in the Derived Variable Specification document

The Health Survey for England 2007 - Household Questionnaire

P2727 The Health Survey for England 2007
Program Documentation
Household Questionnaire

Point
SAMPLE POINT NUMBER.
Range: 1..997

Address
ADDRESS NUMBER.
Range: 1..97

Hhold
HOUSEHOLD NUMBER.
Range: 1..9

AdiField
PLEASE ENTER THE FIRST TEN CHARACTERS OF THE FIRST LINE OF THE ADDRESS
TAKEN FROM A.R.F. ADDRESS LABEL. MAKE SURE TO TYPE IT EXACTLY AS IT IS
PRINTED.
Text: Maximum 10 characters

First
INTERVIEWER FOR INFORMATION.....You are in the Questionnaire for
Point no: (Point number)
Address no: (Address number)
Household no: (Household number)

IntDate
PLEASE ENTER THE DATE OF THIS INTERVIEW. ENTER DAY OF MONTH IN NUMBERS,
NAME OF MONTH IN WORDS (FIRST THREE LETTERS), YEAR IN NUMBERS, EG. 2 Jan 98.

WhoHere
INTERVIEWER: COLLECT THE NAMES OF THE PEOPLE IN THIS HOUSEHOLD.
1 Continue

IF First person in household OR More=Yes THEN
Name
What is the name of person number (1-12)?
ENTER PERSON'S FORENAME

More
Is there anyone else in this household?
1 Yes
2 No

ENDIF

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

The Health Survey for England Errata Note



Author: The NHS Information Centre, Lifestyle Statistics

Responsible Statistician: Paul Eastwood, Lifestyles Section Head

Version: 1

Original Date of Publication: 16th December 2008

Date of Publication of Errata Note: 19th November 2009

As a result of a detailed validation carried out during the production of the Health Survey for England (HSE), 2008, the following error has been identified in the Health Survey for England series. The error pertains to childhood obesity data for the years 1995 to 2007 inclusive and will affect all publications during these years.

Between 1995 and 2007 there was an error which meant that small numbers of children that should have been classified as either 'overweight' or 'obese' were omitted from these categories because of rounding of age and Body Mass Index (BMI) thresholds. The revised percentages of those who were overweight in each year differ by less than 0.1 to 0.8 percentage points and for those who were obese in each year they differ by less than 0.1 to 1.1 percentage points from those originally published, and 0.3% to 1.2% of children in each year were misclassified. In no cases were results significantly different from those presented previously.

An updated version of the following latest trend table for 1995 to 2007 has therefore been published:
Children Table 4, Children's overweight and obesity prevalence, by survey year, age-group and sex.

This table gives revised figures for children's overweight and obesity prevalence for 1995 to 2007, as included both in the main reports and separate trend table publications.

References to these figures within the publications have not been corrected, as the changes are small and will be corrected in the next HSE report which is expected to be published in December 2009. Figures within the reports, referring to the children's categories of overweight or obese may be slightly different, but none are significantly different. All figures in the Health Survey for England 2008 (expected to be published on 17th December 2009), main report and trend tables will be calculated using the correct methodology and the time series data for 1995 to 2007 will all be updated using the correct methodology.

This note has been agreed with the HSE contractors, the National Centre for Social Research.

UK DATA ARCHIVE METADATA

- UK Data Archive collects and creates structured metadata for each archived dataset
- created during ingest data processing (Data-level metadata)
 - data dictionaries, format transfer, data listing, ingest processing details and information gathered in 'readme' file for users
- Catalogue record and keyword index
(mix of Study-/Data-level metadata - 'Catalogue metadata'. Also contains 'Administrative metadata', such as access conditions, date of publication, etc.)
 - data deposit form
 - keyword index covers data elements and concepts
 - international standards: DDI, METS, ISAD(G), TEI
 - standardised elements + controlled vocabularies = consistent search and retrieval
 - sufficient information for users to decide if the data suitable
 - information on the provenance of a dataset
 - record of publications



Finding data

Overview

Catalogue search

Help on searching

About catalogue

Glossary of terms

Browse by subject

Major studies

Major depositors

New releases

HASSET thesaurus

Search UKDA-store

Other archives

About data

Login

[Download/Order](#)  | [Explore Online](#) 

[New Data Catalogue search](#)

* Internal User *

[External View](#)

[Glossary of terms in the Data Catalogue](#)

[Documentation](#) [Variable List](#) [Keywords](#) [Publications](#) [Related Studies](#)

Title:

Health Survey for England, 2007
(HSE)

Subject Categories:

Health Survey for England - Major studies
General - Health
Nutrition - Health
Physical fitness and exercise - Health
Health services and medical care - Health

Depositor(s):

National Centre for Social Research

Principal Investigator(s):

National Centre for Social Research
University College London, Department of Epidemiology and Public Health

Sponsor(s):

Information Centre for Health and Social Care

Abstract:

The *Health Survey for England* (HSE) is part of a wider programme of surveys commissioned by the Department of Health, and is designed to monitor trends in the nation's health. The aims of the HSE 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

The survey focuses on different health issues each year, although a number of core questions are included every year. Topics are revisited at appropriate intervals in order to monitor change:

- in 1991, 1992, 1993, 1994, 1998, 2003 and 2006, the survey focused on cardiovascular disease and associated risk factors
- in 1995 and 1996, the major focus of the survey was atopic disease (including asthma and eczema), accidents and disability
- in 1997 the survey report focused on the health of young people aged 2-24 years, and combined data (in report format only) from the 1995-1997 surveys
- in 1999 and again in 2004, the focus was on the health of minority ethnic groups. The boost sample was designed solely to yield additional interviews with members of the most populous minority ethnic groups: Black Caribbean, Black African (covered in 2004 only), Indian, Pakistani, Bangladeshi, Chinese and Irish
- in 2000 and 2005, the survey focused on the health of older people; the HSE 2000 included a sample of care home residents, whereas HSE 2005 included a boost sample of people aged 65 and over resident in private households
- the 2001 survey included for the first time questions on fruit and vegetable consumption
- the 2002 survey focused on children, young people and maternal health, and included a boost sample of young people aged 0 to 24 years and mothers of infants aged under one year
- the 2007 survey focused on knowledge and attitudes towards key aspects of lifestyle, and monitoring any impact from the 2007 smokefree legislation

Further information about the HSE may be found on the [Department of Health](#), the [Information Centre for Health and Social Care](#) and [National Centre for Social Research](#)

Users are advised to visit the [Health Survey for England](#) web pages for support in using these data, additional resources, and news and events.

See [Exploring differences in health SPSS workbook for students](#)

Main Topics:

For adult respondents, the HSE 2007 focused on knowledge, attitudes and behaviour on key aspects of lifestyle - smoking, drinking, eating and physical activity. Both adults and children were asked about their views on what constitutes healthy behaviour, their knowledge of government recommendations and the factors that may encourage or discourage healthy behaviour.

A secondary focus was the impact of the smokefree legislation. The 2007 survey allowed an initial examination of the effect of the legislation by looking at adults' and children's smoking behaviour and their exposure to other people's smoke, pre and post 1 July 2007 (the date the legislation came into effect). As well as questions being asked, saliva samples were taken and tested for cotinine, an indicator of recent nicotine exposure.

Questionnaire interviews were followed by a nurse visit, both using computer assisted interviewing (CAPI) and self completion booklets. Parents of children aged 0-12 years were asked about their children, with the child interview including questions on physical activity and fruit and vegetable consumption. Children aged 13-15 were interviewed themselves.

Standard Measures:

- Strengths and Difficulties Questionnaire (SDQ)

Coverage:

Time Period Covered: January 2007-December 2007

Dates of Fieldwork: January 2007-April 2008

Index date: 2007

Country: England

Spatial Units: Government Office Regions (GORs); Health Authorities (England)

Observation Units: Individuals

Kind of Data: Numeric data; Individual (micro) level

Universe Sampled:

Location of Units of Observation: National

Population: Adults (aged 16 and over) and children (aged 0-15 years) in the general population, living in private households in England during 2007. A boost sample of children aged 2-15 was also included.

Methodology:

Time Dimensions: Repeated cross-sectional study

The survey is conducted annually.

Sampling Procedures: Multi-stage stratified random sample

Number of Units: 14,386 cases (individual file), 24,910 cases (household file).

Method of Data Collection: Face-to-face interview; Self-completion; Clinical measurements; Physical measurements; CAPI

Weighting: Several types of weighting variable have been used. See User Guide for details.

Language(s) of Written Materials:

Study Description: English

Study Documentation: English

Documentation:

Permission for commercial orders from Patrick Tucker email Patrick.Tucker@doh.gsi.gov.uk. No royalty requests so far.

(28/3/2007) permission from Andy Sutherland -andy.sutherland@ic.nhs.uk

Operations Directorate

The information centre for health and social care

2nd Floor South

1 Trevelyan square

Boar lane

Leeds LS1 6AE

0113 254 7200

Andy Sutherland would like to receive a copy of the signed commercial agreement.

Andy Sutherland is no longer responsible for the HSE. Permission for commercial requests now needs to be obtained by emailing surveys.queries@ic.nhs.uk . A copy of the signed commercial agreement should be attached. BA June 2009

[Print-friendly page](#)

SN 6112 - Health Survey for England, 2007

Variables

[New Data Catalogue search](#)

GROUP 1: hse07ah

- Variable 0(Final outcome code)
- Variable 1(Individual outcome codes)
- Variable 2(Sex)
- Variable 3(Age last birthday)
- Variable 4(Number of adults aged 16+)
- Variable 5(Number of children)
- Variable 6(Number of infants)
- Variable 7(Enter number of households covered by this address)
- Variable 8(Address number)
- Variable 9(Household)
- Variable 10(Relationship to other persons in household)
- Variable 11(Relationship to other persons in household)
- Variable 12(Relationship to other persons in household)
- Variable 13(Relationship to other persons in household)

[Show variable](#)

GROUP 2: [hse07aj](#)

Recent searches

- [61412](#) (All of Study Description)
- [6112](#) (All of Study Description)

[TOP](#)



Finding data

- Overview
- Catalogue search**
 - Help on searching
 - About catalogue
 - Glossary of terms
- Browse by subject
- Major studies
- Major depositories
- New releases
- HASSET thesaurus
- Search UKDA-store
- Other archives

About data

Login



- Finding data**
- Overview
 - Catalogue search**
 - Help on searching
 - About catalogue
 - Glossary of terms
 - Browse by subject
 - Major studies
 - Major depositors
 - New releases
 - HASSET thesaurus
 - Search UKDA-store
 - Other archives

- About data
- Login

SN 6112 - Health Survey for England, 2007

[Print-friendly page](#)

Variable detail

GROUP 1 : hse07ah

[Explore Online](#) 

Variable Id	2
Variable Name	sex
Variable Text	Sex

Value Labels	Unweighted Frequency	Percentage
Value 1 Men	12084	48.5106
Value 2 Women	12826	51.4893

[Full list of variables](#)

Catalogue Frequencies

Where frequencies are displayed here they are unweighted and users should consult the associated documentation for information on whether, and how, to correctly weight the data. Where weighting is required, registered users can download the dataset for offline analysis or, for those datasets displaying the Nesstar icon, produce weighted analyses online.

Not all datasets or variables will have frequencies displayed. See [help on searching](#) for more information.

Neither the original data collectors, depositors or copyright holders, the funders of the Data Collections nor the UK Data Archive bear any responsibility for the accuracy, comprehensiveness or interpretation of the data supplied.

[TOP](#)



Keywords

- [ENGLAND](#)
- [ACCIDENTS](#)
- [ACUPUNCTURE](#)
- [AGE](#)
- [ALCOHOL CONSUMPTION](#)
- [ALCOHOLIC DRINKS](#)
- [ANTHROPOMETRIC DATA](#)
- [ANXIETY](#)
- [ATTITUDES](#)
- [BACTERIAL AND VIRUS DISEASES](#)
- [BEDROOMS](#)
- [CARDIOVASCULAR DISEASES](#)
- [CHILD BEHAVIOUR](#)
- [CHILD CARE](#)
- [CHILD NUTRITION](#)
- [CHILD PSYCHOLOGY](#)
- [CHILDREN](#)
- [CHRONIC ILLNESS](#)
- [CLINICAL TESTS AND MEASUREMENTS](#)
- [CLUBS](#)
- [COHABITATION](#)
- [CONCENTRATION](#)
- [CONFECTIONERY](#)
- [CONFUSION](#)
- [COOKING](#)
- [CULTURAL IDENTITY](#)
- [CYCLING](#)
- [DAIRY PRODUCTS](#)
- [DAY CENTRES](#)
- [DEBILITATIVE ILLNESS](#)
- [DENTAL CARE](#)
- [DIABETES](#)
- [DIET AND NUTRITION](#)
- [DIGESTIVE SYSTEM DISORDERS](#)
- [ECONOMIC ACTIVITY](#)
- [EDIBLE FATS](#)
- [EDUCATIONAL BACKGROUND](#)
- [EMOTIONAL DEVELOPMENT](#)
- [EMOTIONAL STATES](#)
- [EMPLOYEES](#)
- [EMPLOYMENT](#)

[New Data Catalogue search](#)

Recent searches

- [61412](#) (All of Study Description)
- [6112](#) (All of Study Description)

[Clear search history](#)

Finding data

- [Overview](#)
- Catalogue search**
 - [Help on searching](#)
 - [About catalogue](#)
 - [Glossary of terms](#)
 - [Browse by subject](#)
 - [Major studies](#)
 - [Major depositors](#)
 - [New releases](#)
 - [HASSET thesaurus](#)
 - [Search UKDA-store](#)
 - [Other archives](#)

About data

Login

Publications

[New Data Catalogue search](#)

By Principal Investigator(s):

Breeze, E., et al. (1994) *Health Survey for England 1992*, London: HMSO.

Bennett, N., et al. (1995) *Health Survey for England 1993*, London: HMSO.

Colhoun, H. and Prescott-Clarke, P. (eds.) (1996) *Health Survey for England 1994, 2v.* London: HMSO. ISBN 011 321 8958.

Prescott-Clarke, P. and Primatesta, P. (eds.) (1997) *Health Survey for England 1995, 2v.* London: HMSO. ISBN 011 3220219.

Prescott-Clarke, P. and Primatesta, P. (eds.) (1998) *Health Survey for England 1996, 2v.* London: HMSO. ISBN 011322091X.

Prescott-Clarke, P. and Primatesta, P. (eds.) (1998) *Health Survey for England: the health of young people '95-97 2 vols.*, London: The Stationery Office. ISBN 0113222661.

Erens, B. and Primatesta, P. (eds.) (1999) *Health Survey for England: cardiovascular disease '98*, volumes I and II, London: The Stationery Office. ISBN 0113223072.

Erens, B., Primatesta, P. and Prior, G. (eds.) (2000) *Health Survey for England: the health of minority ethnic groups 1999*, volumes I and II, London: The Stationery Office. ISBN 0113224486.

Falascetti, E., Malbut, K. and Primatesta P. (2002) *Health Survey for England 2000: the general health of older people and their use of health services*, London: The Stationery Office.

Hirani, V. and Malbut, K. (2002) *Health Survey for England 2000: disability among older people*, London: The Stationery Office.

Tait, C. and Fuller, E. (2002) *Health Survey for England 2000: psychosocial well-being among older people*, London: The Stationery Office.

Bajekal, M. (2002) *Health Survey for England 2000: care homes and their residents*, London: The Stationery Office.

Boreham, R., Stafford, M. and Taylor R. (2002) *Health Survey for England 2000: social capital and health*, London: The Stationery Office.

Prior, G., et al. (2002) *Health Survey for England 2000: methodology and documentation*, London: The Stationery Office.

Bajekal, M. and Purdon, S. (2002) *Social capital and social exclusion: development of a condensed module for the Health Survey for England*, London: The Stationery Office.

Elam, G., McMunn, A. and Nazroo J. (2002) *Feasibility study for health surveys among black African people living in England: final report - implications for the Health Survey for England 2003*, London: The Stationery Office.

Malbut K. and Falascetti E. (2003) *Health Survey for England 2001: non-fatal accidents*, London: The Stationery Office.

Bajekal M. and Prescott, A. (2003) *Health Survey for England 2001: disability*, London: The Stationery Office.

Primatesta P. and Stamakatakis (2003) *Health Survey for England 2001: respiratory symptoms, atopic conditions and lung function*, London: The Stationery Office.

Doyle M. and Hosfield N. (2003) *Health Survey for England 2001: fruit and vegetable consumption*, London: The Stationery Office.

Prior G., et al. (2003) *Health Survey for England 2001: methodology and documentation*, London: The Stationery Office.

Sproston, K. and Primatesta, P. (eds.) (2003) *Health Survey for England 2002: volume 1: the health of children and young people*, London: The Stationery Office.

Sproston, K. and Primatesta, P. (eds.) (2003) *Health Survey for England 2002: volume 2: maternal and infant health*, London: The Stationery Office.

Finding data

[Overview](#)

Catalogue search

[Help on searching](#)[About catalogue](#)[Glossary of terms](#)[Browse by subject](#)[Major studies](#)[Major depositors](#)[New releases](#)[HASSET thesaurus](#)[Search UKDA-store](#)[Other archives](#)

About data

Login



[SN 6112 - Health Survey for England, 2007](#)

Related studies

[New Data Catalogue search](#)

[Group](#) [Group Constituents](#) [Other Related](#)

Group: [33261](#)

Group Constituents:

[6397 - Health Survey for England, 2008](#)

[6112 - Health Survey for England, 2007](#)

[5809 - Health Survey for England, 2006](#)

[5675 - Health Survey for England, 2005](#)

[5439 - Health Survey for England, 2004](#)

[5098 - Health Survey for England, 2003](#)

[4912 - Health Survey for England, 2002](#)

[4628 - Health Survey for England, 2001](#)

[4487 - Health Survey for England, 2000](#)

[4365 - Health Survey for England, 1999](#)

[4150 - Health Survey for England, 1998](#)

[3979 - Health Survey for England, 1997](#)

[3886 - Health Survey for England, 1996](#)

[3796 - Health Survey for England, 1995](#)

[3640 - Health Survey for England, 1994](#)

[3316 - Health Survey for England, 1993](#)

[3238 - Health Survey for England, 1991-1992 : Combined Data File](#)

Other Related:

[5172 - English Longitudinal Study of Ageing \(ELSA\); Wave 1, 2002-2003: Teaching Dataset](#)

[5050 - English Longitudinal Study of Ageing: Wave 0 \(1998, 1999 and 2001\) and Waves 1-4 \(2002-2009\)](#)

Finding data

[Overview](#)

Catalogue search

[Help on searching](#)

[About catalogue](#)

[Glossary of terms](#)

[Browse by subject](#)

[Major studies](#)

[Major depositories](#)

[New releases](#)

[HASSET thesaurus](#)

[Search UKDA-store](#)

[Other archives](#)

About data

Login

Andy Sutherland would like to receive a copy of the signed commercial agreement.

Andy Sutherland is no longer responsible for the HSE. Permission for commercial requests now needs to be obtained by emailing surveys.queries@ic.nhs.uk. A copy of the signed commercial agreement should be attached. BA June 2009

If user requires additional variables,(e.g. strategic health authorities), user needs to complete a request form from NatCen. Contact Rachel Craig - r.craig@natcen.ac.uk

In response to a request from the Information Centre for Health and Social Care, who note that they took over sponsorship of the HSE from 2004, 'Department of Health' has been removed from the generic catalogue record and added to the chicks from 1991-2003. See red folder correspondence for details. SB 12/9/07.

Access:

Access Conditions: The depositor has specified that registration is required and standard conditions of use apply. The depositor may be informed about usage. See [terms and conditions](#) for further information.

Availability: ESDS Government, UK Data Archive

Contact: Help desk: govsurveys@esds.ac.uk

Access Code : 2003A

Date of Release:

First Edition: 26 February 2009

Latest Edition: 27 April 2010 (2nd Edition)

Copyright:

Crown copyright material is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland

Documentation:

Format	Name	Size in Kilobytes	Description
PDF	6112datadocs.pdf	808	Lists of Variables and Derived Variables
PDF	6112interviewingdocs.pdf	1655	Questionnaires, Showcards, Coding Frames and Consent Booklet
PDF	6112supportingdocs.pdf	1573	Interviewer, Nurse, Coding and Editing Instructions
PDF	6112userguide.pdf	288	User Guide
PDF	bmi_errata_note.pdf	143	BMI Errata Note
HTML	UKDA_Study_6112_Information.htm	26	Study information and citation

 [NOTE File](#)

 [READ File](#)



[Download Acrobat PDF Reader](#)

Catalogue record last updated:

27 April 2010

[Download/Order](#)  | [Explore Online](#)  | [nesstor](#) |

[\[TOP\]](#)

UK DATA ARCHIVE: IMPORTANT STUDY INFORMATION

Study Number 6112 - Health Survey for England, 2007

Details of Study Received

First edition received January 2009 and second edition received March 2010.

Anna Strauss, Manager National Centre for Social Research 35 Northampton Square London EC1V 0AX Email: a.strauss@natcen.ac.uk

Details of data files received

SPSS format:

1st edition:

HSE07ai.sav HSE07ah.sav

2nd edition:

bmi07ai.sav

Details of doc files received

1st edition:

MS Word format:

DERVAR07.DOC Derived variable specification VAR07.DOC List of variables CONSENT 07.DOC Consent booklet DRUG2007.DOC Drug coding EDIT07.DOC Edit instructions HSE2007 HQ.DOC Household questionnaire HSE2007 IQ.DOC Individual questionnaire HSE2007 NQ.DOC Nurse questionnaire INT INSTRUCTIONS 07.DOC Interviewer instructions NURSE INSTRUCTIONS 07.DOC Nurse instructions NURSE SHOWCARDS 07.DOC Nurse show cards and protocols SC 8-10 07.DOC Self-completion booklet 8-10s SC 11-12 07.DOC Self-completion booklet 11-12s SC 13-15 07.DOC Self-completion booklet 13-15s SC YAd pre ban 07.DOC Self-completion booklet Young adults pre smoking ban SC YAd post ban 07.DOC Self-completion booklet Young adults post smoking ban SC Adults pre ban 07.DOC Self-completion booklet Adults 16-64 pre smoking ban SC Adults post ban 07.DOC Self-completion booklet Adults 16-64 post smoking ban SC Adults 65+ pre ban 07.DOC Self-completion booklet Adults 65+ pre smoking ban SC Adults 65+ post ban 07.DOC Self-completion booklet Adults 65+ post smoking ban SC Eating Habits 07.DOC Self-completion booklet Adults SC SDQ Boost 07.DOC Self completion booklet Parents of boost 4-15 year olds HSE07 Showcards.DOC Show cards & code frames Individual H07 Nurse Showcard.DOC Show cards & code frames Nurse USER07.DOC User documentation 2nd edition:

Adobe PDF format:

bmi_errata_note.pdf

Details of hard copy received

All editions:

None

Filename changes: files for issue

All editions:

Spaces were removed and capitals changed to lower case.

Filename changes: files in /noissue

All editions:

Spaces were removed and capitals changed to lower case.

Level of processing

A*

Data file conversion: methods, formats created and validation

All editions:

Spss files were converted to Stata and Tab via scripts.

During conversion of these data files within the Nesstar Publisher (V3.56) it was found that the variable measure definitions were set incorrectly i.e. scale, nominal etc. The definitions were fixed using the Publisher software for the Nesstar version of the study but the original SPSS file remains unchanged.

Details of hard copy received

All editions:

None

Filename changes: files for issue

All editions:

Spaces were removed and capitals changed to lower case.

Filename changes: files in /noissue

All editions:

Spaces were removed and capitals changed to lower case.

Level of processing

A*

Data file conversion: methods, formats created and validation

All editions:

Spss files were converted to Stata and Tab via scripts.

During conversion of these data files within the Nesstar Publisher (V3.56) it was found that the variable measure definitions were set incorrectly i.e. scale, nominal etc. The definitions were fixed using the Publisher software for the Nesstar version of the study but the original SPSS file remains unchanged.

Doc file conversion: methods, formats created and validation

1st edition:

Word files were converted to PDF, amalgamated and bookmarks added in line with HSE conventions.

2nd edition:

PDF added as a separate note. Bookmarks and standard UKDA header added.

Dataset validation and confidentiality checks and solutions to problems raised

1st edition:

All frequencies were checked. Variable labels were tidied and value labels added where necessary. Resolved queries (see emails for full details). Variable XBPDRUG was removed after consultation with the depositor. It had no labelling and turned out to be a scratch variable. Appendix D was not supplied and not needed so this page was removed.

2nd edition:

Both data files were sorted by pserial and the new variables were matched onto the individual file. Scripts were run again.

Data and documentation problems

None.

New Edition Information

For the second edition (April 2010), three new children's Body Mass Index (BMI) variables have been added to the individual data file (bmicat1, bmicat2, bmicat3). The original variables (bmicut1, bmicut2, bmicut3) are unreliable and should not be used. Further information is available in the documentation and on the Information Centre for Health and Social Care [Health Survey for England](#) web page.

Notes to data delivery

Notes from data delivery and post order corrections

NEW EDITION INFORMATION

For the second edition (April 2010), three new children's Body Mass Index (BMI) variables have been added to the individual data file (bmicat1, bmicat2, bmicat3). The original variables (bmicut1, bmicut2, bmicut3) are unreliable and should not be used. Further information is available in the documentation and on the Information Centre for Health and Social Care [Health Survey for England](#) web page.

DATA PROCESSING NOTES

Data Archive Processing Standards

The data were processed to the UK Data Archive's A* standard. This is the Archive's highest standard, and means that an extremely rigorous and comprehensive series of checks was carried out to ensure the quality of the data and documentation. Firstly, checks were made that the number of cases and variables matched the depositor's records. Secondly, checks were made that all variables had comprehensible variable labels and all nominal (categorical) variables had comprehensible value labels. Where possible, either with reference to the documentation and/or in communication with the depositor, labels were accordingly edited or created. Thirdly, logical checks were performed to ensure that nominal (categorical) variables had values within the range defined (either by value labels or in the depositor's documentation). Lastly, any data or documentation that breached confidentiality rules were altered or suppressed to preserve anonymity.

All notable and/or outstanding problems discovered are detailed under the 'Data and documentation problems' heading below.

Data and documentation problems

None.

Data conversion information

From January 2003 onwards, almost all data conversions have been performed using software developed by the UKDA. This enables standardisation of the conversion methods and ensures optimal data quality. In addition to its own data processing/conversion code, this software uses the SPSS and Stat/Transfer command processors to perform certain format translations. Although data conversion is automated, all data files are also subject to visual inspection by a UKDA data processing officer.

With some format conversions data, and more especially internal metadata (i.e. variable labels, value labels, missing value definitions, data type information), will inevitably be lost or truncated owing to the differential limits of the proprietary formats. A UKDA Data Dictionary file (in rich text format), corresponding to each data file, is usually provided for viewing and searching the internal metadata as it existed in the originating format. These files are called: [data file name]_UKDA_Data_Dictionary.rtf

Important information about the data format supplied

The links below provide important information about the format in which you have been supplied the data. Some of this information is specific to the *ingest format* of the data, that is the format in which the UKDA was supplied the data in. The ingest format for this study was SPSS

Please click below to find out information about the *format that you have been supplied the data in*.

[SPSS \(*.por\)](#)

SPSS portable (*.por files)

If SPSS portable was not the ingest format, this format will generally either have been created via the SPSS command processor (e.g. if the ingest format is SPSS .sav, SAS, Excel, or dBase), or if the ingest format was STATA, the SPSS version will be created via the Stat/Transfer command processor. If the ingest format was undelimited text, the data will have been read into SPSS using an SPSS command file.

Issues: There is very seldom any loss of data or internal metadata when importing data files into SPSS. Any problems will have been listed above in the Data and Documentation Problems section of this file.

[STATA \(*.dta\)](#)

STATA (*.dta files)

If STATA was not the ingest format, all STATA files will have been created from SPSS .sav format via the Stat/Transfer command processor. Importantly, Stat/Transfer's optimisation routine is run so that variables with SPSS write formats narrower than the data (e.g. numeric variables with 10 decimal places of data formatted to F.X.2) are not rounded upon conversion to STATA because they are converted to 'doubles' rather than floats. User missing values are copied across into STATA (as opposed to being collapsed into a single system missing code).

Issues: There are a number of data and metadata handling mismatches between SPSS and STATA. Where any data or internal metadata has been lost or truncated, this will have been automatically logged in this file: 6112_SPSS_to_STATA_conversion.rtf Note that the complete internal metadata has been supplied in the UKDA Data Dictionary file(s): [data file name]_UKDA_Data_Dictionary.rtf

[Tab-delimited text \(*.tab\)](#)

If tab-delimited text was not the ingest format, tab-delimited files will have been created from SPSS portable files via the SPSS command processor, and also from Excel and MS Access files. When exporting from Access data tables to tab-delimited text, potentially problematic special characters (tabs, carriage returns, line feeds, etc.) allowed by Access memo and text fields are stripped out by the UKDA.

Issues: Date formats in SPSS are always exported to mm/dd/yyyy in tab-delimited text format - so there be a mismatch with the documentation on such variables. Variables that include both date and time such as dd-mm-yyyy hh:mm:ss (e.g. 18-JUN-2001 13:28:00), will lose the time information and become mm/dd/yyyy. If the time information is critical, a new variable will have been created in the tab-delimited data file by the UKDA. All users of the data in tab-delimited format should consult the UKDA Data Dictionary file(s): [data file name]_UKDA_Data_Dictionary.rtf

If the data was exported from MS Access, more limited 'data documenter' information is supplied in the file(s): [data table name]_variableinformation.rtf These files may also contain SQL setup information.

[MS Excel \(*.xls files\)](#)

If MS Excel was not the ingest format, Excel files will have been created via the SPSS command processor. The date and time issues noted under tab-delimited format apply to SPSS to Excel conversion via the SPSS command processor.

[SAS \(supplied as *.dat and *.sas\)](#)

If SAS was not the ingest format, all SAS files will have been created from SPSS .sav format via the Stat/Transfer command processor. The data files are provided as a fixed-width text file (*.dat) and a SAS command file (*.sas), which when run will create a SAS dataset. This enables the user to recreate the SAS dataset and formats library in almost all versions of SAS and all operating systems.

Issues: The main loss of information when converting from SPSS to SAS is user-missing value definitions. By editing the .sas file, the user can choose whether to collapse all user-missing values into system missing or preserve their value and lose the user-missing definition. To achieve the latter the following section of the .sas file should be removed before running it:

```
/* User Missing Value Specifications */
```

Note that the complete internal metadata has been supplied in the UKDA Data Dictionary file(s): [data file name]_UKDA_Data_Dictionary.rtf

[MS Access \(*.mdb files\)](#)

Due to the substantial incompatibilities between versions of MS Access, the UKDA only make data available in MS Access format if this is the ingest format and the database contains important information in addition to the data tables (coding information, forms, queries, etc.).

Conversion of documentation formats

HOW?

- Researchers can embed certain annotations in data files:
 - quantitative data: variable/value labels; worksheet information; table relationships and queries in relational database; GIS data layers/tables

Examples:

SPSS: variable attributes documented in Variable View (label, code, data type, missing values)

MS-Access: variable descriptions and attributes documented in Design View; relationships

ArcGIS: shapefiles (layers) and tables in geodatabase; metadata created in ArcCatalog.

Excel: base worksheet data-related documentation.

HOW?

- qualitative data/text documents: interview transcript speech demarcation (speaker tags); document header with brief details of interview date, place, interviewer name, interviewee details, context
- Additional documents: final report, published reports, user guide, working paper, publications, lab books

RURAL MARKERS IN DATASETS AT UK DATA ARCHIVE

“to identify national scale survey data for England that have potential to be used for rural and urban analyses by the Commission for Rural Communities.”

ONS RURAL AND URBAN DEFINITION, 2004

Settlement type		Context
Urban	Urban >10k	Less sparse
		Sparse
Rural	Town and fringe	Less sparse
		Sparse
	Village	Less sparse
		Sparse
	Hamlet and isolated dwellings	Less sparse
		Sparse

TWO REVIEWS

1. Datasets with rural variables

- ONS 2004 rural/urban area classification or contain Output Areas as spatial units
- Post 2000 national datasets for England with representative population sample.

Rural marker type	Number of datasets where variable type used
ONS 2004 rural/urban definition, distinguishing either 8 or 6 categories	23
Binary rural/urban variable	6
Variable distinguishing rural/urban/other category	2
Other variable indicating 'rural areas'	8
Respondent evaluation of rural/urban area	7
Indirect indication through settlement size or population density	3

TWO REVIEWS

2. National survey datasets for England (post 2000), with representative population sample that do not contain a rural/urban variable. Data owners contacted to investigate

- whether a rural/urban variable may exist for the dataset and can be made available
- dataset can be retrospectively coded to include a rural/urban variable, based on the presence of postcodes in the dataset
- nineteen datasets identified
- retrospective coding dependent on willingness of depositor

WHY ENHANCEMENT MATTERS

- Good metadata enables easy discovery of data provide meaningful information (titles, descriptions, abstract, keywords)
- Submitting good documentation and metadata enables data archives/data centres to create rich and detailed metadata
- Good data documentation leads to good data re-use
- Detailed yet comprehensive documentation
- Where possible integrate within dataset

CONTACT

- Contact the UK Data Archive if not sure what to produce or provide:
 - Advice on our Managing and Sharing web pages:
<http://www.data-archive.ac.uk/sharing/metadata.asp>
 - The UK Data Archive Data Management Team:
datashar@essex.ac.uk