

# **ENGLISH HOUSE CONDITION SURVEY**

## **DATA BASE USER GUIDE**

## **Contents**

### **1. The EHCS data base**

- 1.1 Overview of the EHCS Continuous Survey**
- 1.2 Overview of the EHCS data base**
- 1.3 Other sources of information**
- 1.4 Using the data base – first steps**

### **2. How the data bases are produced**

- 2.1 Scope of the EHCS**
- 2.2 Sample structure**
- 2.3 Datasets from the Continuous Survey**
- 2.4 The Core sample**
- 2.5 Levels of data**
- 2.6 Types of data**

### **3. What analysis and tables are published**

### **4. Key Concepts and Definitions<sup>1</sup>**

### **5. Locating information in the database**

- 5.1 General Overview**
- 5.2 Interview (household) survey variables**
- 5.3 Physical survey variables**
- 5.4 Derived Variables**
- 5.5 Market Value Survey Variables**

### **6. How to match files**

### **7. How to apply grossing factors to obtain national estimates**

### **8. How to deal with missing data**

- 8.1 Missing Data Conventions**
- 8.2 Dealing with missing data in analysis**
- 8.3 Checking if results are significant**

---

<sup>1</sup> A full glossary of terms used in the EHCS can be found in the Technical Report on the CLG web site

**Appendix A: Overview of the data base structure**  
**Appendix B: Content and level of Interview Survey Files**  
**Appendix C: Content and level of Physical Survey Files**  
**Appendix D: Content of Derived Files**  
**Appendix E: Content of Detailed Derived Variables**  
**Appendix F: UK Data Archive folder structure**

# 1. The EHCS dataset

## 1.1 Overview of the EHCS Survey

The EHCS comprises three main component surveys<sup>2</sup>:

- Interview (household) Survey,
- Physical Survey,
- Market Value Survey.

From April 2002 the EHCS has been run on a continuous basis with fieldwork conducted in four eight week periods throughout the year. The survey is conducted for around 8000 sampled addresses annually where a household interview (if an occupied dwelling), a visual property inspection and a market valuation are completed. For vacant properties only a physical and market valuation are available.

Results from the continuous survey are reported annually on a two-year rolling basis. The first set of results (EHCS 2003) was based on data collected in the first two years of fieldwork, between April 2002 and March 2004. EHCS 2004 was based on data collected from April 2003 to March 2005 and so forth. The datasets and grossing factors contained on the EHCS database relate to these two-year combined data sets and **not** to single years.

Fieldwork period	EHCS data set
April 02 to March 04	EHCS 2003
April 03 to March 05	EHCS 2004
April 04 to March 06	EHCS 2005
April 05 to March 07	EHCS 2006
April 06 to March 08	EHCS 2007

Prior to 2002 the survey was conducted on a five-yearly basis with the most recent surveys taking place in 2001 and 1996. These followed the same basic methodology but with a larger achieved sample. Some of the key derived variables from these two surveys are also included on the EHCS data base CD provided through CLG and as separate data files on the UK Data Archive site. (See below). The full set of data from these surveys and documentation is available from CLG on request.

Full details on how the survey is managed can be found in the EHCS technical reports – see web link below in section 1.3.

From April 2008 the EHCS was merged with the survey of English Housing to form the English Housing Survey (EHS). Public data sets from the EHS will continue to be

---

<sup>2</sup> data from the Private Landlord Surveys will be available separately on request

made available. Further details will be released via the EHS web site [www.communities.gov.uk/englishhousingsurvey](http://www.communities.gov.uk/englishhousingsurvey)

## 1.2 Overview of the EHCS data set

There are two ways to obtain copies of the EHCS data

- a) **CLG CD** - a complete data base covering 2003 to 2007 plus some 1996 and 2001 data on a single CD from the CLG (contact [ehcs@communities.gsi.gov.uk](mailto:ehcs@communities.gsi.gov.uk)) The database structure for the CD is shown at Appendix A. The CD is provided free of charge.

*or*

- b) **from the UK Data Archive** - data sets are held individually rather than as a complete EHCS data base. So for example to download the 2006 and 2007 data sets you would need to complete two separate downloads. To download or access the data you will need to register with the archive, though it is possible to download descriptions of data and a general overview of the survey for each year without registering (see link below) The data files and documentation are held in separate folder structures, following the archive guidelines. See Appendix F for an overview.

[http://www.data-archive.ac.uk/search/allSearch.asp?ct=xmlAll&q1=ehcs&zoom\\_and=1&zoom\\_cat=-1](http://www.data-archive.ac.uk/search/allSearch.asp?ct=xmlAll&q1=ehcs&zoom_and=1&zoom_cat=-1)

**Please note all the key concepts and technical advice given in this user guide will apply regardless of whether you obtain the data from the UKDA or the CLG direct. However to avoid complexity, this guide assumes the full data base on CD is being used. Users downloading data from the UKDA should refer to the structure at Appendix F in locating individual files.**

### Key concepts

**Primary Data** - The EHCS data set consists of primary data files containing basic survey data collected by the surveyors and interviewers

**Derived Data** - a small number of files containing 'derived' variables are also included on the database holding variables created by calculating with or recoding the basic survey data. These include for example energy efficiency indicators and repair costs. These files will be the most widely used for analysis. (See Appendix D).

### Contents of the database

The information provided on the EHCS data base comprises

- a) The full set of primary and derived files for the continuous survey 2003 forward
- b) Documentation and reports for all years of the continuous survey
- c) The three main derived files and reports only for the 1996 and 2001 surveys.

Since most analysis will make use of variables from the derived files only, users will be able to generate time series analysis from 1996 forward using the data contained on this CD. The primary files for the 1996 and 2001 surveys are available on request from CLG together with more complete documentation.

**Core Sample** - Where information for a single dwelling is available from more than one of the survey components that dwelling forms part of the 'core sample'. The core sample comprises

*Occupied dwellings with an interview, physical and market value survey  
Vacant dwellings with a physical and market value survey only.*

**Grossing weights** have been produced for all core sample cases in order to produce accurate national estimates. See Section 7 below for advice on applying these weights. The weights are held on the derived file general.sav

**Missing data** - as with all surveys there will be some level of missing data on the primary data files where information was not available or for example, a respondent refused to answer specific questions. For derived variables calculated for the core sample however the aim is to contain no missing data, by imputing missing values based on cases with similar characteristics. Where imputation has taken place there are accompanying flag variables to indicate the level of imputation. Imputation is not always achievable so for some of the key derived variables there will be missing data which you may wish to distribute proportionately using the tool we have provided – see Section 8.2. The method of data imputation is described more fully in the Technical Reports.

### 1.3 OTHER SOURCES OF INFORMATION

A large amount of information about the way in which the survey is organised is available on the EHCS web site. [www.communities.gov.uk/ehcs](http://www.communities.gov.uk/ehcs)

Below are links to specific documents that may be of interest

**EHCS Technical Reports** – these provide details of how the fieldwork is organised, how the sample is drawn and the results grossed each year, estimates of sampling and measurement error together with a full glossary of terms and details of how derived measures such as energy efficiency and income are created. The first report was produced to accompany the reporting from the 2003 data set and updates produced annually. The technical report will be available on the web in due course.

**EHCS Reports** – annual reports from EHCS 1996-2007 are available on the database CD in directory \\EHCS XXXX\Reports. They are also available on the web site from 2001 onwards. (Where XXXX is the survey year).

<http://www.communities.gov.uk/housing/housingresearch/housingsurveys/englishhousecondition/ehcsreports/>

**EHCS Supporting Tables** – an updated set of tables are produced each year and made available on the web site

[EHCS Standard Tables - Housing - Communities and Local Government](#)

**Supporting information provided on the database CD**

A number of documents are included on the CD holding the EHCS database to help users understand how the database is organised and where to find the data they need. These are found in directory \\EHCS200X\Data\Documentation, where X is the Survey year.

**Interview survey documentation** – this document shows all the questions asked of households, the variable names and available response categories. There have been a number of changes to the content of the interview survey over the years so not all variables will be present for every year. Questions, which have been asked in only some years are shown .i.e. it is a cumulative record of the content of the interview since 2002. It also shows on which data base file each of the variables is located. A number of overviews showing routing through each of the main modules can be found in \\EHCS200X\Data\Documentation\Interview.

**Physical survey documentation** – a separate CD is available from CLG complete with a search engine that shows the layout of the physical survey form from 2002/03 to 2006/07 and enables users to search for particular topics or questions, to find the variable name they require and on which file it is located. The content of the physical survey has changed significantly over the years reflecting in particular the shift from the Fitness standard to the new Housing Health and Safety Rating system. Over the two years 2005/06 and 2006/07 information relating to both standards was collected to enable analysis of the impact of the change to be assessed. Full details are in the Technical Reports for these years.

The notes to help guide surveyors and which appear on the facing pages of the paper survey form are also included on this CD. A pdf file showing the survey form and associated variable names and file locations are also provided for each year including 2007/08 in \\EHCS200X\Data\Documentation\Physical

**Market Value survey documentation** - can be found at \\EHCS200X\Data\MarketValue

**Fuel Poverty documentation** - a small number of variables relating to fuel poverty are included on the database in \\EHCS200X\Data\Fuel Poverty. Documentation relating to this data is also located in this directory. The link below provides further information relating to Energy Efficiency and Fuel Poverty

<http://www.berr.gov.uk/energy/fuel-poverty/strategy/index.html>

<http://www.berr.gov.uk/files/file29694.pdf>

## 1.4 USING THE DATA BASE - FIRST STEPS

To use the data set correctly you must first ensure that you familiarise yourself with and be aware of the following:

- How the dataset has been produced
- What information is published
- The key concepts and definitions used in the survey
- How to locate the information you want in the database (find both survey and computed data by variable and file names)
- How to match files
- How to apply grossing factors to obtain national estimates
- How to deal with missing data
- How to deal with sampling and survey error

Each of these is described briefly in the sections below but please note this guide is not intended to be a comprehensive source of information. Reference is made to the more detailed explanations and guidance contained in other files, publications and documents which are either included on the EHCS CD or available from the web site.

The guide assumes knowledge and experience of using SPSS.

## 2. How the dataset has been produced

### 2.1 Scope of the EHCS

The EHCS is a set of interrelated data from household, physical and market value surveys based on a *sample of dwellings*.

The following describes the components of the whole EHCS:

- A questionnaire-based survey of households living in the sampled dwellings which provides information on household characteristics, resources, housing costs and views of respondents about their home and neighbourhood. This is known as the **Interview Survey and is based on households**.
- A physical assessment of the sampled dwellings where permission/access was available or which were vacant at the time the interviewer called. This provides information on dwelling characteristics and a professional assessment of dwelling and neighbourhood conditions. This is known as the **Physical Survey and is based on dwellings**.
- A **Market Value Survey** of all dwellings where a physical survey was carried out. This is a desk exercise conducted by the Valuation Office to provide a professional assessment of the dwelling's market value and a range of indicators of housing demand in the local area around the dwelling. This is dwelling based.



More detail on how the three surveys are conducted is given in Chapter 1 of the Technical Report

## 2.2 Sample structure

The EHCS is a multi-stage clustered sample stratified by tenure (rented tenures are over-represented and owner-occupation under-represented in relation to national estimates). Survey estimates are therefore less reliable than a single-stage, unclustered sample of the same size.

Weighting factors for each sample case have been produced. These **must** be applied to the raw data to ensure accurate national estimates are produced (see Section 7).

The sampling approach has been modified over the course of the continuous survey and full details on sample structure and weighting can be found in Chapter 2 of the annual Technical Report. For example an RSL top-up sample was incorporated in 2002/03 only and from 2005/06 a longitudinal component has been introduced to help improve the accuracy of estimates of change over time compared to comparison of two cross-sectional surveys.

## 2.3 Datasets from the Continuous Survey

The survey has been running annually since 2002. Although the EHCS produces annual data files, these are combined on a two year rolling basis for the purpose of analysis.

Each 2-year data set presents the average position over the two year period. Results may therefore be presented in terms of a mid point survey position, i.e. the 2003 data set refers to a mid point of 01.04.2003.

The following two-year datasets will become available:

Dataset	Annual data contents	Mid point	Available
2003	2002/03 – 2003/04	April 2003	√
2004	2003/04 – 2004/05	April 2004	√
2005	2004/05 – 2005/06	April 2005	√
2006	2005/06 – 2006/07	April 2006	√
2007	2006/07--2007/08	April 2007	√

## 2.4 The Core sample

The Core Sample is the set of matching cases from the Interview, Physical and Market Value Surveys

The number of (unweighted) cases in the core sample each year is as follows:

Dataset	Core Cases
1996	19438
2001	17628
2003	16648
2004	16502
2005	16670
2006	16269
2007	16217

**It is important to note that on primary files for the 2003 and 2004 data bases only information is also included for some non-core cases.**

These are cases where an interview survey was conducted, but the respondent, for example, refused a physical survey. These cases do not have a grossing weight and need to be excluded from analysis. See section 7. Derived files for all years hold core cases only.

For more information on the sampling strategy, survey management and response rates see the Technical Report.

## 2.5 Levels of data

Most of the data is at household or dwelling level. However, some of the primary data files contain data at sub-household and sub-dwelling level. These are indicated for the Interview Survey at Appendix B and for the Physical Survey at Appendix C. All the files containing derived variables are at dwelling or household level

## 2.6 Types of data

### Primary Survey data

The primary data collected in the field is often very specific and detailed and in many cases is only collected in order to provide the building blocks for computing more useful pieces of information - in particular the core house condition and energy efficiency measures used in the analysis of the survey. There will also be a small amount of missing data in the primary data collected in the field.

In many cases omissions or inconsistencies in the base variables will have been identified in the course of producing derived variables for final analysis. ***Users are recommended wherever possible to make use of the final derived variables rather than the raw data from the primary files as these are likely to contain more complete information.***

### Derived/computed data

While the information that is collected in the field provides all the base variables needed, a considerable amount of further processing and modelling is needed to fully validate the results and produce secondary variables needed for detailed analysis. Producing derived variables often involves reconciliation of conflicting information

from different parts of the survey (e.g. tenure) and/or the imputation of missing data where this has been possible from other data collected in the survey.

Detailed information on how derived variables are produced and how they can be used are given in the Technical Report as follows:

- The decent homes criteria and its application in the EHCS including the updating of the decent home definition arising as a result of the introduction of the new Housing Health and Safety Rating System. A revision within the EHCS was also introduced for 2006 in how the survey implements the thermal comfort requirements for insulating flats.
- Estimated costs to make decent - including changes arising as a result of the 2006 revisions above.
- Using EHCS data to model Decent Homes and Thermal Comfort – including changes arising as a result of the 2006 revisions above.
- Estimating Repair Costs
- Treatment of incomes
- Energy cost rating (SAP) and updating to Sap 2005 methodology.
- Calculating property dimensions and changes introduced to the dimensions model in 2006
- The introduction of the Housing Health and Safety Rating system from 2006 and its impact on Decent Homes
- Changes to the way certain hazards under the HHSRS were modelled from 2007

It is important to note that changes introduced to some of the EHCS models will have a significant impact on time series analysis and users should familiarise themselves with relevant parts of the Technical Report before drawing conclusions about changes over time. Revised variable names have been used wherever there has been a major break in methodology year on year to minimise the risk of data being mis-interpreted.

### **3. What analysis is published by the CLG**

EHCS reporting is undertaken on an annual basis using the two-year rolling sample.

All the EHCS publications including reports and supporting tables can be found on the CLG website. For each year of the continuous survey (from 2003 forward) a Headline Report is available reporting progress on the key indicators of Decent Homes, Liveability and Deprivation. This is followed each year by an Annual Report which builds on these key findings by providing a more detailed account of living conditions and energy efficiency.

The publications available on the website also include reports and supporting tables from the 1996 and 2001 surveys.

Some reports are also included on the database CD to assist users. These include Annual and headline reports and can be found in folder \\EHCS200X\Reports.

A regular EHCS Bulletin is produced to update interested parties on news and developments relating to the survey, including publications. To subscribe to the Bulletin, please contact the EHCS team at **ehcs@communities.gsi.gov.uk**

## **4. Key Concepts and Definitions<sup>3</sup>**

### **Dwelling**

The EHCS is a dwelling based survey. It is therefore important to understand how the term 'dwelling' is defined and applied in the survey.

A dwelling is a self-contained unit of accommodation (normally a house or flat) where all the rooms and amenities (i.e. kitchen, bath/shower room and WC) are for the exclusive use of the household(s) occupying them. In rare cases, amenities may be located outside the front door but provided they are for the exclusive use of the occupants, the accommodation is still classed as a dwelling.

For the most part a dwelling will be occupied by one household but may contain none (vacant dwelling) or may contain more than one (HMO).

A typical example of an HMO is a house divided into a number of bedsits. The complete HMO would be defined as a dwelling. The occupant(s) of an individual bedsit is a single household sharing amenities with other households. Only one interview (selected randomly) is conducted at an HMO.

Where only part of the building is occupied in this way (the rest may be non-residential and/or flats with their own amenities) then the dwelling is the part occupied with shared amenities and will be treated as a single flat.

### **Household**

One person living alone or a group of people, who have the address as their only or main residence and who either share one meal a day or share a living room.

## **5. Locating information in the database**

### **5.1 Data base Overview**

Appendix A provides a diagram showing the overall structure of the EHCS database held on the EHCS CD. Essentially files are held under six main folders comprising

---

<sup>3</sup> A full glossary of terms used in the EHCS can be found in the Technical Report.

Interview  
Physical  
Derived  
Documentation  
Fuel Poverty  
Market Value

For the 1996 and 2001 surveys only two folders are present, Derived and Reports.

## **5.2 Interview (household) survey variables**

The topics covered in the Interview Survey questionnaire are illustrated in Figure 1 below.

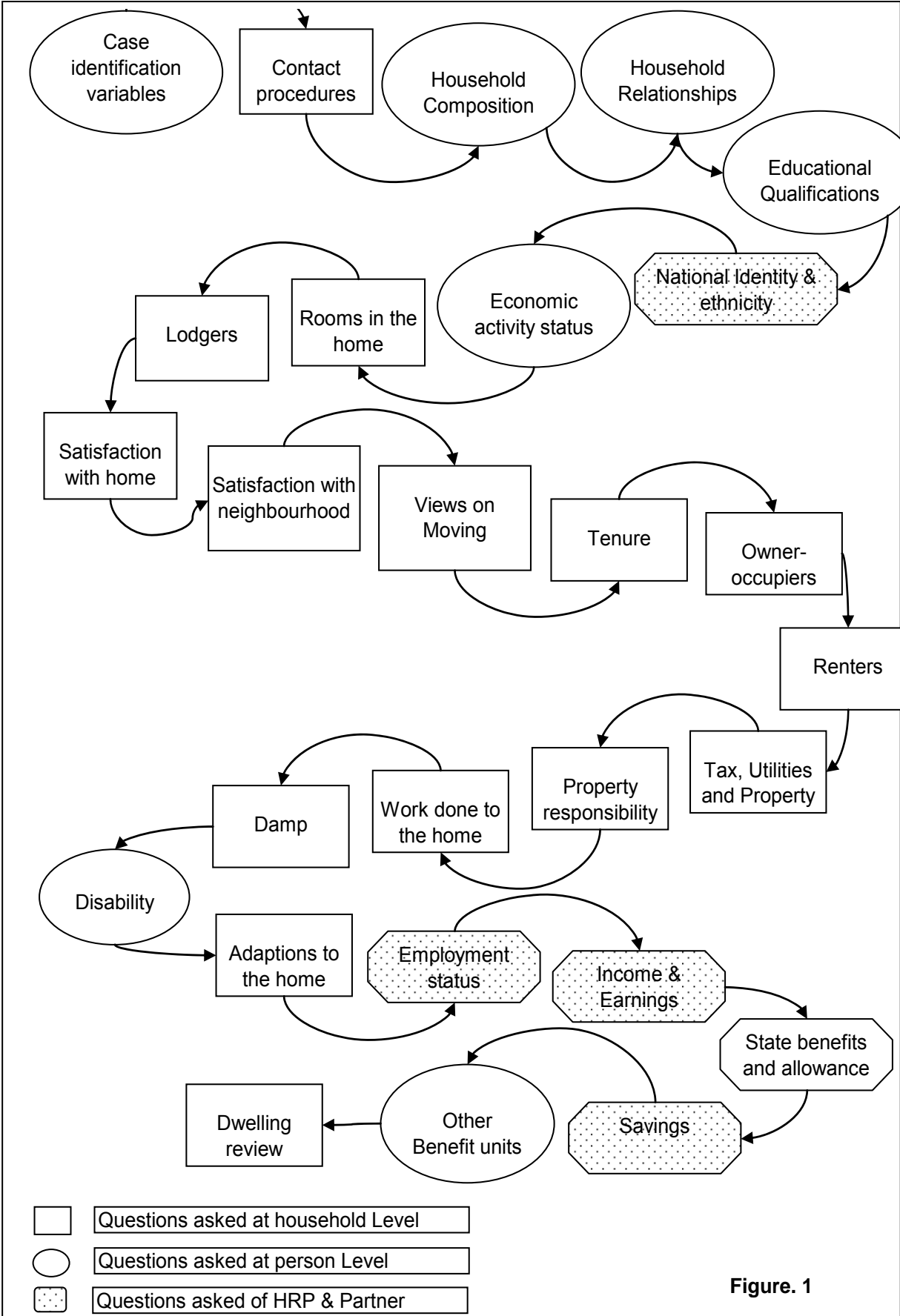


Figure. 1

There are also overviews for each of the following sections of the questionnaire:

Attitudes  
Condensation and damp  
Disability  
Household composition  
Property responsibility  
Rooms  
Tenure  
Workdone

For 2003 forward these can be found in \\EHCS200X\Data\Documentation\ Interview on the database CD.

Detailed documentation for the Interview Survey is provided on the EHCS CD. See \\EHCS200X\Database\Documentation\Interview. This is a cumulative document showing all questions asked from 2002 forward. As well as the questionnaire, the documentation also indicates on which data file the variables can be found. There is an index of variable names to indicate where to locate them on the questionnaire form.

The files containing the Interview Survey data, their key contents and the level of aggregation (household/person) are listed at Appendix B.

### 5.3 Physical survey variables

A separate CD ROM is also available from CLG , which provides full documentation for the Physical Survey from 2002/03 to 2006/07 forward. The CD gives graphic representations of each page of the English House Condition Physical Survey form, annotated with variable names and the files on which the variables are located. There is also a search engine so that users can easily locate questions/variables/topics on the form.

The CD helps you

- identify the location on a form of any variable for which you already know the code (variable name)
- identify the code for any variable on the form
- search by keyword for particular topics and identify codes and page numbers associated with them
- search by page number to identify all the codes on a particular page of a form
- view the **facing pages** for the guidance notes associated with the variables

A separate pdf version of the Physical Survey annotated with variable names form for each year of the continuous survey (including 2007/08) are also provided on the main EHCS CD and can be found in folder \\EHCS200X\Data\Documentation\Physical.

Please note certain variables relating to rats and mice, which are collected in the EHCS are not included on the CD. This information is collected on behalf of Defra who report on this data separately

## 5.4 Derived Variables

There are three files available on each of the 2-yearly EHCS data sets (and for 1996 and 2001), which contain the most commonly used derived variables. These are

- General.sav
- Physical.sav
- Interview.sav

These include the key indicators summarised below:

General.sav	Tenure, nature of area, vacant/non-vacant; grossing factors;
Physical.sav	Dwelling type, decent home criteria, repair costs, heating type, structure; energy efficiency
Interview.sav	Household composition, length of residency/ownership; income; rent, mortgages

There are a number of further derived variables available on the EHCS CD, which includes detailed information on repair costs. New detailed derived files have been added in 2006 and 2007 covering equivalised income and energy performance and potential for improvement. They are located in \\EHCS200X\Data\Derived\Detailed. Further information on content of the three main derived variable files is in Appendix D.

## 5.5 Market Value Survey Variables

Documentation on the Market Value Survey from 2003 forward is provided in folder \\EHCS200X\Data\Market Value.

This provides a list of the questions that valuers were required to answer for each core case and the associated variable name.

## 6. How to match files

Key points when matching files:

- Primary survey data files (from the physical and interview) for years 2003 and 2004 hold non-core cases (some of which contain survey data e.g. where an interview was obtained but the respondent refused a physical survey). These do not have a grossing factor and must be excluded from analysis. The primary files from year 2005 onwards and derived files for all years only contain core cases .i.e. those that have a grossing factor.
- Some statistical procedures in SPSS ignore weightings and are therefore likely to include any cases found on the active file whether or not they have a grossed



value. It is therefore best practice to include only core cases in the active file (the single or merged file being analysed). This is achieved either

- a) through merging files using general.sav as the table file; or
  - b) through selecting for cases where a relevant grossing variable (household or dwelling depending on the type of analysis) is greater than zero
  - c) and preferably deleting all other cases
- All the derived files and most of the primary data files are at household or dwelling level i.e. one case per household or dwelling but some files are at lower levels of aggregation e.g. one case per person, one case per window type etc.

### **Matching files that exist at the dwelling or household level**

The key variable for matching at the dwelling/household level within or across any of the survey data is **aacode** (an eight character string variable). You must specify this as the key field when matching. All files on the database have been sorted by this variable so should match. 'Sort file' or 'split file' commands, however, re-sort the file. Matching with standard database files will fail if the data is not first re-sorted by aacode.

### **Matching lower level files (i.e. below dwelling or household level) up to the grossing file**

Files at a lower level of organisation will have more than one case per household or dwelling. Some of them have exactly the same number of cases per address e.g. the windows.sav file has 7 cases for each address each representing a type of window, the chimneys.sav has 2 cases per address. With other files, the number of cases varies e.g. for people.sav the number of cases is the number of people in the household so one person households will just have one case, 6 person households will have 6 cases etc. As there is always more than one case per address, these cannot simply be matched into the dwelling or household level files to give results at the dwelling or household level. There are 2 ways to approach these files:

- Match the household or dwelling level files in as 'tables'. Here the information at household or dwelling level will be copied for each person etc. within that aacode. The results will then represent the number of people, number of types of window etc. rather than the number of dwellings or number of households. This can be useful for very specific analyses, e.g. on numbers of **people** in unfit dwellings.
- Aggregate the lower level file up to address code level and then match it in with the household or dwelling level file. (Most key variables for analysis have been derived at the household level so this will not normally be necessary).

## 7. How to apply grossing factors to obtain national estimates

The EHCS is based on a stratified sample with over-sampling of the rented tenures to achieve large enough sample sizes to produce reliable results for the social and private rented sectors.

Grossing factors have been calculated to

1. Compensate for the design of the sample i.e. the over sampling of some dwellings and under sampling of others; and
2. Take account of non-response bias: the survey response rates achieved for different groups of households and dwellings; and
3. Take account of the relative weight attached to each of the sample sources that were used to design the EHCS sample.

Some cases will have no grossing value as they do not form part of the core sample. These are interview addresses that were occupied at interview, but no physical survey took place. These cases should be excluded from analysis (see above)

**Within SPSS, results must be based on data weighted by the relevant grossing factor to produce national estimates.**

Each dataset from the EHCS comes with its own grossing factors covering the core sample of cases (see above). The grossing factors can be found in the file **general.sav** in the 'derived' folder. The relevant variables are:

- Dwelling grossing      aagcdxx
- Household grossing    aagdhxx

Where xx indicates the two years across which fieldwork was conducted. For example, Aagcd23 is the dwelling grossing factor for the 2-year data set covering 2002 and 2003 i.e. for the EHCS 2003 data set. Aagdh34 is the household grossing factor for the data set covering 2003 and 2004 i.e. for the EHCS 2004 data set

**Aagcdxx** should be used for any analysis in which the aim is to provide estimates of dwellings and that includes physical survey data (e.g. percentage of dwellings unfit) and market value survey data;

**Aagchxx** should be used for any analysis in which the aim is to provide estimates of households and that includes interview survey data (e.g. percentage of households in unfit dwellings).

The grossing factors can only be used on full 2-year data sets. They cannot be used on the data split into separate years.

Using the combined sample, the following numbers of cases and national estimates should be produced for the two current grossing factors:

### Basic control totals for grossing dwellings and households in the combined samples

Year	Dwellings - number of real cases	Dwellings - grossed counts (thousands)	Households - number of real cases	Households - grossed count (thousands)
2003	16,648	21,484	15,950	20,724
2004	16,502	21,613	15,874	20,931
2005	16,670	21,781	16,059	21,134
2006	16269	21,988	15,648	21,221
2007	16217	22189	15604	21380

It is vital that these figures are used as a basic check on analysis i.e. that these grossed counts are being achieved before embarking on any more detailed analysis on sub-sections of households or dwellings.

Grossing factors for the 1996 and 2001 surveys are also included on the general.sav files for those years.

## 8. How to deal with missing data

### 8.1 Missing Data Conventions

The files contain no system missing values. Where values are unknown or the question was not asked, specific codes are used and the values are set to user defined missing. The conventions adopted are:

Unknown/don't know      9 for single digit fields, 99 for 2-digit fields etc.  
 Not applicable, not asked   8 for single digit fields, 88 for 2-digit fields etc.

(and for some variables 7,77 etc for additional 'not applicable' situations.)

Users need to be aware of the implications of having missing values switched on or off within SPSS. It is useful for example to leave missing values switched on (the SPSS standard) if users wish to establish the level of missing data in a frequency or cross tab. The SPSS output Case Processing Summary will show the level of missing data. This can be used to decide on whether the level of missing data is significant and whether it is appropriate to re-distribute missing cases.

It is particularly important to leave missing values switched on when dealing with a continuous variable e.g. costs to make decent. Including 99, 88 etc as a normal value (i.e. not as a missing value) would distort any outputs.

When you create any new variables, you should declare missing values as below:

```
Missing values var1 (9).
```

This code sets up '9' as a user defined missing so will be excluded from any crosstabs or statistics of var1.

## 8.2 Dealing with missing data in analysis

### Analysis that produces national estimates

Analysis that uses derived variables with no missing cases for the core sample will have no problem with missing data. However, when undertaking analysis using other variables that have missing or unknown codes, these cases can be redistributed to produce national estimates.

Some procedures like 'Cross tabulation' exclude missing cases from the table. To include these as the basis for any redistribution it is necessary to switch off the user defined missing values for each variable used. This is done by

***Missing values Var1 ().*** (missing data is included in outputs)

Missing values can then be switched back on using:

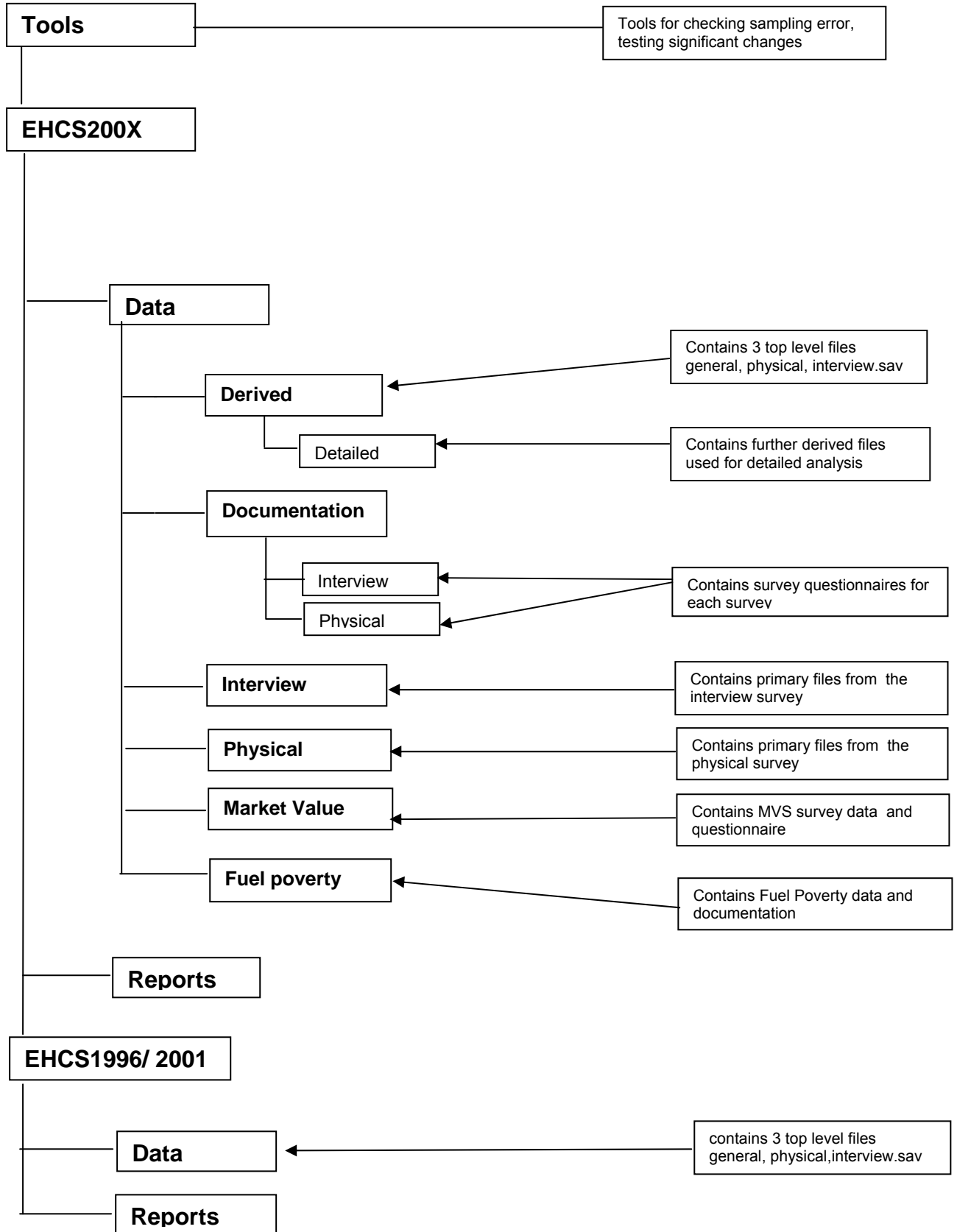
***Missing values var1 (9).*** etc

Where you have missing values within the table, unless there is a clear reason to do otherwise, these need to be allocated pro rata. An Excel spreadsheet ***distributor.xls*** is included with this database to help to ensure that this is done in a consistent fashion. (\\Data\\Tools).

## 8.3 Checking if results are significant

An additional Excel spreadsheet, ***Significant change tester.xls***, is included on the EHCS CD in folder \\Data\\Tools that provides a quick method of checking whether changes in a particular indicator over time are statistically significant over time. There is an accompanying word document in the same location, which explains how the Excel macro works.

# Appendix A: Overview of Database Structure – for CLG CD



## Appendix B: Content and level of Interview Survey Files

File name	level	Key identifier	Contents
<b><i>Adapt.sav</i></b>	household	aacode	Information on disability aids used by most disabled person and degree of difficulty experienced due to disability
<b><i>Adaptation.sav</i></b>	adaptation	aacode	Information on internal and external adaptations for those with disability (not available for 2006 or 2007 , see note in Interview folder for 2006 and 2007)
<b><i>Attitudes.sav</i></b>	household	aacode	Information on attitudes about the home, neighbourhood and moving
<b><i>Benefits.sav</i></b>	benefit unit	aacode	Information on benefits received by 'other benefit units' in the household and their income
<b><i>Contact.sav</i></b>	household	aacode	Key information from the doorstep form
<b><i>Damp.sav</i></b>	room	aacode	Information on the types of damp problems
<b><i>Disability.sav</i></b>	person	persno	Information on types of disability
<b><i>Employment.sav</i></b>	household	aacode	Employment information and ethnicity
<b><i>FirstImp.sav</i></b>	household	aacode	Key information on the first impressions of the dwelling and neighbourhood as recorded by interviewer
<b><i>Income.sav</i></b>	household	aacode	Information on amounts and sources of income including benefits
<b><i>Intenure.sav</i></b>	household	aacode	Tenure information and details of length of residence and housing costs
<b><i>People.sav</i></b>	person	persno	Key info for each person (age, sex etc.) and relationships with other people in the household
<b><i>Repairs.sav</i></b>	household	aacode	Information on repairs, builders and part of the set of questions on damp
<b><i>Rooms.sav</i></b>	household	aacode	Key information about the number of rooms and whether shared
<b><i>Vacant.sav</i></b>	household	aacode	Information relating to vacant properties from the doorstep form
<b><i>Workdone.sav</i></b>	item	aacode	Key information about work done to the home
<b><i>Workdone3.sav</i></b>	item	aacode	Key information about work done to the home

## Appendix C: Content and level of Physical Survey Files

<b>File Name</b>	<b>Section on survey form</b>	<b>Level</b>	<b>Notes</b>
<b>Amenity.sav</b>	5. Interior - Amenities	Dwelling	
<b>around.sav</b>	19. Around the house/module 23. Local area	Dwelling	
<b>chimney.sav</b>	18. Exterior of house/module (front and rear views)	Type of chimney	
<b>Commac.sav</b>	9. Common parts of flat surveyed: accessways (common parts by accessway, 3 cases per address)	Access	
<b>Common.sav</b>	9. Common parts of flat surveyed	Address	
<b>Damppc.sav</b>	18. Exterior of house/module (front and rear views)	Type of damp course	
<b>doors.sav</b>	18. Exterior of house/module (front and rear views)	Type of door	
<b>Dormers.sav</b>	18. Exterior of house/module (front and rear views)	Type of dormer	
<b>elevate.sav</b>	16. Elevation features 17. Specification of views	Address	
<b>Firstimp.sav</b>	1. Survey Record 2. Address Information 3. Dwelling description and occupancy 4. Module associated with address surveyed	Address	
<b>fitness.sav</b>	22. Summary of fitness	Address	HHSRS.sav file replaces fitness as criterion for measure of decency (from 2006)
<b>flatdets.sav</b>	8. Details of flat	Address	
<b>Hhsrs.sav</b>	23. Health and safety rating	Address	Available from 2006
<b>hmodata.sav</b>	Variables from separate HMO form		
<b>hq.sav</b>	7. Household questionnaire conducted by surveyor	Address	
<b>interior.sav</b>	5. Interior	Address	
<b>introoms.sav</b>	5. Interior	Rooms	
<b>numflats.sav</b>	10. Number of flats in the module	Address	
<b>plotlvl.sav</b>	18. Exterior of house/module (front and rear views)	Address	
<b>plotwall.sav</b>	18. Exterior of house/module (front and rear views)	Address	
<b>roofcov.sav</b>	18. Exterior of house/module (front and rear views)	Address	
<b>rooffeat.sav</b>	18. Exterior of house/module (front and rear views)	Address	
<b>roofstru.sav</b>	18. Exterior of house/module (front and rear views)	Address	
<b>services.sav</b>	5. Interior	Address	
<b>shape.sav</b>	12. House/module shape 13. External dimension of house/module 14. Material and construction of house/module	Address	

	15. Improvements/alterations and when done		
<b>shared.sav</b>	11. Shared facilities and services	Address	
<b>structure.sav</b>	21. Structural defects	Address	
<b>wallfin.sav</b>	18. Exterior of house/module (front and rear views)	Address	
<b>wallstru.sav</b>	18. Exterior of house/module (front and rear views)	Address	
<b>windows.sav</b>	18. Exterior of house/module (front and rear views)	Address	

## Appendix D Content of derived variables

Filename	Variable	Label
<b>General.sav</b>	aacode	Key field
	tenure8x	Tenure
	tenure4x	Tenure
	tenure2x	Tenure
	vacantx	Type of vacancy
	vacprob	Type of vacancy
	goregx	Government office region
	arnatx	Nature of area
	neivisx	Appearance of area
	region3x	Overall region of England
	area3x	Type of area
	lv1upkpx	Poor quality environment - upkeep problems
	lv2trafx	Poor quality environment - traffic problems
	lv3utilx	Poor quality environment - utilisation problems
	lvnumx	Number of liveability problems present
	lvanyx	Poor quality environment
	valout	Valuation of home
	demand1	Level of demand in the area
	mktdemfs	Market area factor score for demand for housing
	mktvolfs	Market area factor score for volume of properties on market
	mktdem10	Market areas ranked by factor score for demand for housing
	mktvol10	Market areas ranked by factor score for volume of properties on market
	mktdem5	Market areas ranked by factor score for demand for housing
	mktvol5	Market areas ranked by factor score for volume of properties on market
	rumorph	Rural urban morphology (COA)
	pthfind	Market Renewal Pathfinder areas



	nrf88	NRF districts (original funding)
	nrf86	NRF districts (2006 revised funding)
	nrf91	NRF districts (original or revised)
	nrf91stat	Status of 91 NRF reporting districts
	imd1004	2004 decile ranking of areas (lower layer SOA)
	imd1007	IMD07 decile ranking of areas
	aagcdXX	Dwelling weight (core cases )
	aagchXX	Household weight (core cases )
<b>Interview.sav</b>	aacode	Key field
	hhcomp	Household composition
	hrpage2x	Age of household reference person
	hrpage4x	Age of household reference person
	hrpage6x	Age of household reference person
	hrpagex	Age of household reference person
	hrpemp6x	Employment status (primary) of HRP
	hrpemp3x	Employment status (primary) of HRP
	hhemp	Employment status of HRP and partner combined
	famnum	Number of families in household
	depchild	Number of dependent children in primary family
	otherfam	Additional families present in household
	othfamlp	Type of additional families in household
	hhtype11	Household type
	hhtype7	Household type
	hhtype6	Household type
	pyngx	Age of youngest person in household
	pyngbx	Age of youngest person in household
	poldx	Age of oldest person in household
	poldbx	Age of oldest person in household
	hhsizex	Number of persons in the household
	olderx	No of people aged 60 plus who are HRP or partner
	loncoup	Single householder or with partner
	hrpeth8x	Ethnic group of HRP
	hrpeth4x	Ethnic group of HRP
	hrpeth2x	Ethnic group of HRP
	psagex	Age of partner
	hrpsex	Sex of household reference person
	lenres	Length of residence (years)
	lenresb	Length of residence
	lenown	Length of ownership (years)
	lenownb	Length of ownership
	bedstd	Bedroom standard
	nbeds	Number of bedrooms available
	bedrq	Number of bedrooms required
	hhitsick	Household includes anyone with long term illness or disability
	hpregdis	HRP or partner registered disabled
	rentwk	Weekly rent

	amthbenx	Weekly housing benefit
	housbenx	Household receives any housing benefit
	rentflg	Rent changed/imputed
	hhvulx	Vulnerable household (in receipt of means tested or disability)
	hhbensx	Household in receipt of any means tested benefits
	hhincx	Annual net household income (including income from savings)
	hhinc5x	Annual net household income (including income from saving)
	hhincoox	Annual net household income - all households
	hhincprx	Annual net household income - owner occupiers
	hhinclax	Annual net household income - private tenants
	hhinchax	Annual net household income - RSL tenants
	hhincflg	Income imputation flag
	freehold	Freehold/leasehold ownership
	bdgfreeh	Leasehold flat but respondent owns freehold for whole building
	tenex	Tenure of household
	workless	Workless households (no one of working age employed) - ILO defn
	mortwkx	Weekly mortgage payments
	equity	Equity in home
	equity5	Equity in home
	hhincpsx	Annual net household income - private sector households
	hhincscx	Annual net household income - social sector households
	BHCinceq	BHC equivalised weekly income (modified OECD scale)
	BHCinceqv5	BHC equivalised income quintiles (weighted by peoplegross)
	BHCinceqv60h	below 60% of median income (weighted by people group)
	AHCinceq	AHC equivalised weekly income (modified OECD scale)
	AHCinceqv5	AHC equivalised income quintiles (weighted by peoplewt)
	AHCinceqv60h	below 60% of median income (weighted by peoplewt)
<b>Physical Sav</b>	aacode	Key field
	dwtypenx	Dwelling type
	dwtipe8x	Dwelling type
	dwtipe7x	Dwelling type
	dwtipe3x	Dwelling type
	housex	Dwelling type
	dwage9x	Dwelling age
	dwage6x	Dwelling age

	dwage5x	Dwelling age
	dwage4x	Dwelling age
	alltypex	Dwelling age and type
	floorx	Usable floor area (sqm)
	floor5x	Useable floor area    Scale
	storeyx	Number of floors above ground
	typercov	Predominant type of roof covering
	typerstr	Predominant type of roof stucture
	typewstr	Predominant type of wall stucture
	typewfin	Predominant type of wall finish
	typewin	Predominant type of window
	dbiglaz2	Extent of double glazing
	dbiglaz4	Extent of double glazing
	secure	Secure windows and doors
	parking	Parking provision of survey dwelling
	attic	Attic present in dwelling
	basement	Basement present in dwelling
	heat4x	Main heating system
	heat7x	Main heating system
	sysage	Age of heating system
	fuelx	Main fuel type
	mainfuel	Main heating fuel
	loftinsx	Loft insulation thickness
	loftins4	Loft insulation thickness
	loftins6	Loft insulation thickness
	wallinsx	Type of wall and insulation
	wallcavx	Type of wall
	sap05	Energy efficiency (SAP05) rating
	sap405	Energy efficiency (SAP05) rating
	EPceeb05e	Energy efficiency rating band (EHCS SAP 2005)
	EPceir05e	Environmental impact rating (EHCS SAP 2005)
	EPceib05e	Environmental impact rating band (EHCS SAP 2005)
	dhomesx	<a href="#">Decent homes - 2001 original definition (incorporating Fitness) (upto 2006)</a>
	dhomesy	<a href="#">Decent homes - 2006 updated definition (incorporating HHSRS) (2006 forward)</a>
	dhunfitx	Decent homes fitness criterion
	dhhhsrsx	Decent homes HHSRS criterion
	dhthermx	Decent homes thermal comfort criterion
	dhthermy	Decent homes thermal comfort criterion (2006 revised model)
	dhdisrx	Decent homes repair criterion
	dhmodx	Decent homes modern facilities criterion
	dhnumx	Decent homes: number of criteria failed (2001 original definition)
	dhnumy	Decent homes: number of criteria failed (2006 updated definition)

	dhreasnx	Decent homes criteria not met (2001 original definition)
	dhreasny	Decent homes criteria not met (2006 updated definition)
	dhtcreasx	Reason for failing decent homes on thermal comfort (2001 original definition)
	dhtcreasy	Reason for failing decent homes on thermal comfort (2006 updated definition)
	dhtcactx	Requirement to pass decent homes thermal comfort criterion (2001 original definition)
	dhtcacty	Requirement to pass decent homes thermal comfort criterion (2006 updated definition)
	cstactux	Urgent repair costs (actual)
	cstactbx	Basic repair costs (actual)
	cstactcx	Comprehensive repair costs (actual)
	cststdux	Urgent repair costs (per square metre)
	cststdbx	Basic repair costs (per square metre)
	cststdcx	Comprehensive repair costs (per square metre)
	cstunfx	Cost to make fit (actual)
	dhcost	<a href="#">Cost to make decent (2001 original definition)</a>
	dhcosty	<a href="#">Cost to make decent (2006 updated definition)</a>

## Appendix E: Content of Detailed Derived Variables

Filename	Variables	notes
Actual costs.sav	Variables for external and internal works	
De-regionalised costs.sav	Comprehensive repair costs,	
Dimensions.sav	Number of rooms, wall and floor areas	
Energy performance.sav	SAP ratings, energy efficiency ratings, light, water and space heating costs, energy upgrade costs	
Energydims.sav	Area of ground and highest floor, ceiling heights, external wall and window area	
Standardised costs.sav	All costs per square metre of floor area	
HHSRS.sav	Variables based on 15 hazards	

Appendix F: UK Data Archive folder structure

