



Office for  
National Statistics

# Harmonised Concepts and Questions for Social Data Sources

---

## Introduction to Harmonised Standards

### Introduction

---

## About this document and Update history

### **About this document**

This document forms part of a series that replaces the original *Harmonised Concepts and Questions for Government Social Surveys*. This series will make topics easier to find, with clearer guidance and consistent layouts. By using a series of PDF documents, updates will be simpler and faster, and new topics will be easier to incorporate.

This document provides the introduction to both the primary and secondary harmonised standards, all of which are available from our website.

### **This is version 3.0, published in April 2008.**

This document was updated to Version 3.0 to reflect the changes in the Survey of English Housing (SEH) and changes in contact details for the Harmonisation Team.

Previous versions

### **Version 2.0, published in November 2004.**

This document is part of a series that replaced the Harmonised Concepts and Questions document. We have been working to make our guidance easier to find and clearer to use. Other than the formatting and layout changes, there are no new changes to questions or outputs in this version.

Please note that version control is new to this edition. Please ensure that you always use up-to-date versions of questions and outputs by checking our website for new editions – the version number will be incremented when changes are made.

---

## Contents





About this document and Update history .....	2
Contents .....	3
Introduction .....	4
The need for harmonisation of concepts and questions .....	5
Principles of harmonisation .....	7
Practical assumptions for harmonised questions .....	9
Conventions .....	11
Annex A: Edits.....	12
Annex B: How to deal with missing information when recording dates: Discussion .....	13
Annex C: Uptake of harmonised concepts and questions in major government social surveys.....	14

## Introduction

This document explains the history of harmonisation and its underlying principles. We welcome your questions, so please contact us if you would like to know more about our work or if you would like to join our mailing list.

### Contact us

For more information about Harmonisation or to join our mailing list, please visit our website at [www.statistics.gov.uk/harmonisation](http://www.statistics.gov.uk/harmonisation). If you would like further information or have any questions, please contact:

	Email:	<a href="mailto:harmonisation@ons.gov.uk">harmonisation@ons.gov.uk</a>
	Telephone:	01329 444017
	Fax:	01329 813383
	Mail:	Harmonisation Team Office for National Statistics Room 1400 Segensworth Road Titchfield Fareham PO15 5RR

### The need for harmonisation of concepts and questions

The United Kingdom has a wide range of Government surveys of persons and households which provide sources of social and economic statistics. The decennial Census of Population is the largest and best known but Government departments also commission continuous household surveys on a range of topics. These include:

- economic activity (*Labour Force Survey* - LFS)
- income (*Family Resources Survey* - FRS, *Expenditure and Food Survey* - EFS)
- expenditure (*Expenditure and Food Survey* - EFS)
- food purchase and consumption (*Expenditure and Food Survey* - EFS)
- health (*Health Survey for England* - HSE)
- housing (*Survey of English Housing* - SEH) – English Housing Survey (EHS) from Apr 2008 (**see note 1**)
- transport (*National Travel Survey* - NTS).

The multi-purpose *General Household Survey* (GHS) also links many of these topics and others such as education. There are also several large-scale surveys which are repeated regularly, such as the *British Crime Surveys* - BCS, the *Dental Surveys* and the *House Conditions Surveys*. The Government also commissions single surveys from time to time on subjects of national importance such as the prevalence of disability and psychiatric morbidity.

These surveys were designed at different times, to meet different needs, and were commissioned by a range of departments. Consequently, they were developed to a significant degree in isolation from each other. This lack of cohesion led to differences in:

- concepts and definitions
- design
- fieldwork and processing practices
- outputs

This was a source of frustration for many users. A major factor behind the creation of National Statistics in 1996 was the desire to improve the service to users of statistics, and the availability and accessibility of those statistics. Development plans for the Office for National Statistics (ONS) include improving access across Government by establishing a central database of key economic and social statistics. Producing these using common classifications, definitions and standards will lead to improved comparability of social statistics. The harmonised concepts shown here are a step towards this goal by making the interpretation and analysis of data easier. This will allow users of published sources to see a more coherent picture of British society and trends within it, and to help users plan surveys (at a local level, for example) that can provide data comparable with national surveys.

Account has also been taken of concepts and definitions developed for the 2001 Census of Population, although the different data collection methods and uses of the data mean there are inevitably differences in some cases.

ONS is increasingly positioning itself to make greater use of administrative data, for example, to support Neighbourhood Statistics. In order to exploit such data it is important to extend the principles of harmonisation beyond surveys and the Census. This is the focus of new harmonisation activities over the next few years; more information can be found by looking under the 'Other Harmonisation' tab.

Other factors influencing harmonisation activities may be summarised as follows:

- updates to reflect changes to existing harmonised concepts (for example, reflecting changes to the benefit options)
- Eurostat's requirements - the desirability for the UK to be working with Eurostat in developing harmonised Key Social Indicators (KSIs, which are at the heart of Europe-wide social reporting.)
- ever-increasing emphasis on coherence and priority given to the need to develop harmonised concepts for topics such as households, income and urban/rural, sub-groups within households.
- emerging social topics, such as social capital, e-society and cultural identity.

The material shown here sets out the outputs, concepts and questions which it has been agreed with the sponsoring Departments should be harmonised wherever possible among the key group of Government surveys of persons and households listed above. However, it will be for the commissioners of individual surveys to decide which questions are appropriate for their surveys and how outputs should be presented, bearing in mind the needs of their users.

**Notes:**

1 From April 2008 the English House Condition Survey and the Survey of English Housing will be integrated to form the English Housing Survey (EHS). The EHS will bring together the key components of both surveys into a single fieldwork operation.

### Principles of harmonisation

Harmonisation concerns concepts which are **inputs** (interview questions and answer categories) or **outputs** (analysis variables derived from the inputs) or both (the question on sex, for example). These are explained in more depth below.

#### Inputs

Different surveys have different purposes and so cover topics in different depth. Harmonised questions are designed to provide the recommended minimum information to allow common classifications and help data analysis from different surveys in combination. Not all surveys will include questions on all topics or in every year but the recommendation is that where a topic is covered, harmonised questions should be included wherever possible.

Some surveys will require more detail on topics than can be obtained from the harmonised questions alone. It will normally be the case that such surveys already ask for that detail. The harmonised questions have been designed so that the surveys which ask for more detail can either derive them, without asking them directly, or combine them with the extra detail, without adding to the length of interview.

The idea that a survey for which it is appropriate to derive an equivalent variable to a harmonised question should additionally ask the harmonised question directly has been considered and rejected. To do so would duplicate effort and seem inappropriate to respondents.

Harmonisation involves some compromises, since surveys' prime concerns vary so widely. For example, surveys vary in the extent to which they allow information to be given by one respondent on behalf of another who is absent at the time of interview. It would be unrealistic to expect Departments which currently accept proxy data as adequate to their needs on certain of their surveys to find the resources to harmonise on data given in person.

The harmonised questions are intended to fit flexibly into the designs of different surveys. There is no intention that they should form a unified sequence within a questionnaire. Questions and groups of questions are intended to be placed in existing questionnaires in the most appropriate places. This will often mean substituting a harmonised question for an existing one on the same topic.

Avoiding an increase in respondent burden has been a major consideration in designing harmonised questions. Some of the harmonised questions have more detailed sets of answer categories than some of the current surveys use for these topics. However, classifying respondents' answers to a more detailed set does not necessarily increase the time needed to answer a question. More detailed categories have been included only where they will not add to interview length. There is no intention to probe for detail which is not volunteered, unless a question specifically demands it. The aim has been to save time by providing clear categories for the rarer answers, where these are of interest for analysis. However, account has also been taken of the need to make it easy for interviewers and respondents to find the major answer categories and not to lose them in a host of details. Finding the right balance in such compromises has an important bearing on survey quality.

The harmonised questions have been built on the current surveys' experience, in particular that of the surveys which are sponsored by Departments with the lead interest in a topic. For example, harmonised questions on economic activity have been based on those developed for the ONS's Labour Force Survey to obtain, in particular, the International Labour Office's measure of unemployment.

#### Outputs

While it will be neither feasible nor desirable to produce outputs from different surveys in a completely harmonised way, it is hoped that unless there are strong reasons for doing otherwise, Government surveys will as far as possible adopt the harmonised outputs and questions. This will not only allow users to interpret data more easily but also help those outside Government better to plan their own data collection and analysis. However, as

noted above, it will be for the commissioners of individual surveys to decide how outputs should be presented, bearing in mind the needs of their users.

Of course, not all surveys will be able to provide information in full detail or for all tabulations, owing to the limitations of sample size. But it is hoped that so far as possible analyses will use a standard hierarchy, so that results can be compared. For some outputs, where the categories are likely to vary considerably in size (such as ethnic groups), the commentary suggests how groups might be further amalgamated on a harmonised basis. It is suggested that information included in the notes attached to the harmonised output categories be included in each publication where the categories are used (though not necessarily in every table); and that those publications also refer to this website entry for further clarification. Where output categories differ from the harmonised ones, it is also hoped that this will be made clear.

The output categories may be applied to households or individuals as appropriate, or to households containing one (or more) individuals with the characteristic specified. The unit of analysis should be made clear in the table heading or documentation.

Harmonised outputs are not intended to restrict published outputs. Where more detail is currently published from a particular survey, it is expected that this will continue, but including some aggregate data on a harmonised basis. Users who have access to micro data (for example, through the Data Archive) should not generally be affected by the harmonisation of survey outputs, except where it impacts on classification systems themselves, or where it is suggested that electronic databases should include harmonised categories.

### Practical Assumptions for harmonised questions

When specific harmonised questions are proposed, practical issues of how they should be asked are inevitably raised. The questions are framed within specific methods (such as interview surveys; and at a more detailed level, that certain questions should be asked within special grids). All the surveys considered are conducted by face-to-face interview, except that second and subsequent interviews in the LFS (which is a panel survey) may be carried out by telephone. Most of the surveys are carried out with computer assisted interviewing (CAI), which allows for quality checks during the interview (for example on completeness of information in complex grids) and ensures that routing is correctly followed in every case. Nevertheless, all the harmonised questions can be asked in the proposed forms, using paper and pencil methods, by suitably trained interviewers.

While it might be desirable that harmonised questions should be asked in exactly the same ways on different surveys, it is recognised that this may not be achievable. The emphasis in the harmonisation project is on harmonising question wording, answer categories and the sub-samples to whom the questions are addressed. Matters such as question sequence (for the factual type of question involved), the use of proxy respondents, and (at a rather more minor methodological level) use of specific kinds of grids were regarded as secondary in the sense that achievement of the main type of harmonisation would be worthwhile even if the second were not practicable. Documentation of the questions used, and their routing, can be found in the published reports of surveys and may also be available from commissioning Departments.

These differences in the way in which data are collected, even using harmonised survey questions, are extremely important. ONS research has pointed to a variety of factors which lead to different estimates for harmonised questions – factors including:

- question wording and context effects
- definitive differences
- non-response bias
- geographical coverage
- sampled population
- mode effect
- acceptance of proxy information
- treatment of multi-households
- unit of analysis
- field procedures
- item non-response
- time period
- organisational effects.

More information is available in the White/McCreith article in *Survey Methodology Bulletin* 43, 7/98, p9-22. For a copy of this bulletin, contact SVS Project Support Branch (Tel: 0845 601 3034) or email [info@ons.gov.uk](mailto:info@ons.gov.uk). In addition, ONS is currently undertaking research to update these findings using data for further survey years. The results will be included when they are available.

### The scope of harmonised concepts and questions

Harmonisation which extends to all or nearly all major Government household surveys can be thought of as covering a **primary set** of concepts and questions. Concepts and questions which apply only for a selected group of surveys can be thought of as belonging to a **secondary set**. There may be a number of secondary sets (one might involve a set of questions on the EFS, FRS and GHS, and another different set of questions on the EFS, GHS and SEH).

#### **The primary set of harmonised concepts and questions**

Common definitions of person and household response units are vital steps towards harmonisation. For Government surveys, there is already a standard definition of adults as persons aged 16 years or more. Definition of the household response unit has differed between surveys. Most use the household definition which was adopted in the 1981, 1991 and 2001 Censuses of Population, which focuses on shared living accommodation.

#### **The secondary set of harmonised concepts and questions**

The secondary sets of concepts and questions have been based on the shared interests in particular topics of different groups of surveys.

### Conventions

The questions shown in this document apply to the whole sample of persons or households (as appropriate), and only one answer is expected at each question, unless otherwise specified.

Instructions to interviewers which appear with the questions are shown in *italics*.

Interviewers must ask the question as worded, reading as far as the question mark. The only exceptions to this rule are:

- if a question carries the instruction *ASK OR RECORD*: at such questions the interviewer can record the answer if it has already been clearly volunteered by the respondent. If in any doubt, the interviewer must ask the question. The layout of some questions means that the answer categories are included in the wording, and the question mark is found at the end of the final answer category to be read out;
- if a phrase is in brackets, the interviewer may decide if it is appropriate to read it out; the interviewer also chooses which words or phrases separated by / to read.

Questions refer to 'you', as is appropriate when talking with the person whose information is being collected or asking about the collective 'you' which refers to the household as a whole. Interviewers will substitute the appropriate name when asking the respondent about someone other than themselves.

Questions which ask for monetary amounts expect answers to be given in pounds and pence, unless specified as to the nearest pound.

### Edits

In this document, the term *edits* refers to checks on respondents' answers.

Under the heading *Edits*, this booklet sets out harmonised consistency checks which have been agreed as a minimum set associated with particular harmonised questions. For their own purposes, different surveys may add further consistency checks.

**Hard checks** in edit programs are written to look for inconsistencies which clearly require at least one of the answers involved in the check to be corrected before the data can be accepted (if, for example, the respondent is said to receive Guardian's Allowance but not child benefit when it is not possible to receive Guardian's Allowance without also receiving child benefit).

**Soft checks** are written to warn of possible inconsistencies needing further inspection, like hard checks, but which might turn out to be valid (for example if the respondent said to be a parent and said to be aged less than 15) and could then be accepted. Soft checks are particularly useful in checking numeric variables, like the cost of a house, where the valid range must cover all possibilities but is worth checking that answers with very high or low values are valid and not miss-recorded.

Computer assisted interviewing (CAI) allows checks to be applied during the interview, so that apparent inconsistencies can be resolved or corrected by the respondent. The government surveys covered by this booklet use CAI, so the edits are shown here in the form of the message which an interviewer sees on screen when a check is triggered by an apparent inconsistency.

### How to deal with missing information when recording dates: Discussion

Missing information in dates may be noted in a variety of ways, depending on the recording medium and, in the case of Computer Assisted Interviewing (CAI), the software that is used. For example:

IF AGE <20

- **(As you are under 20, may I just check) What is your date of birth?**

ENTER ?? FOR ANY PART OF THE DATE WHICH IS MISSING

The instruction on how to deal with missing parts of the date applies for CAI only, and its specific details of how to record missing values will vary with the software package used. On paper, it would be necessary to record the date elements in separate fields.

We now propose that, instead of recording one element of the date is missing (day, month), an imputation (*if appropriate*, see below) should be made according to the following rules:

If day or month is missing (not known, refused) - enter 15 for day, and 6 for month, respectively.

This proposal is embodied in the new proposed question on Date of Birth to replace Age, but is intended to apply generally for dates. Surveys which wish to record imputed day or month can do so by a subsequent question for the interviewer if the day or month element of the date of birth is a possibly imputed value.

To reiterate, the survey must decide when it is appropriate to impute in this fashion, or indeed in some other way, depending on the particular needs of the survey. It is assumed here that it is never appropriate (at the data collection stage) to impute the year. (Just to be clear: we are not talking here about an interviewer obtaining a respondent's best guess, which may or may not be a legitimate procedure, depending on a survey's specification, but about purely automatic imputation to a specific value as proposed for day and, in appropriate circumstances, month). If year is missing, the answer should be recorded as Don't Know or Refusal or other appropriate missing value. In some cases, if the information is sufficiently important, it may be possible to route missing values to a further question which tries another approach (asks age, for example, if year of birth is refused or not known). A missing day or month should also lead to the whole answer being treated as missing if the imputation rule above is not appropriate (for example, if recording a date for comparison with another which is a very short period apart).

## Annex C

# Uptake of harmonised concepts and questions in major government social surveys

### HARMONISATION DATES

**Key:** a = harmonised pre 1996-97; 'year' = harmonised for that 'year'; n = no current plan to introduce; x topic not in survey

	EFS	FRS	GHS	HSE	LFS	NTS	SEH*	SHCS**	EHCS	OMNI	SHS	CHS	WHS <sup>1</sup>
<b>PRIMARY SET</b>													
Household response unit	n	a	a	a	a	a	x	a	02/03	a	99	a	n
Household composition -													
Gender	a	a	a	a	a	a	a	02	02/03	a	99	a	03
age last birthday	96/7	96/7	a	a	a	98	a	a	02/03	a	99	a	03
date of birth - if aged under 20 years	96/7	96/7	a	a	a	00	x	a	02/03	a	99	a	x
Marital status (legal)	96/7	96/7	96/7	98	a	02	97/8	02	02/03	97/8	n	a	x
living arrangements	96/7	97/8	96/7	98	a	99	97/8	02	02/03	97/8	99	96/7	x
who owns or rents accommodation	96/7	a	96/7	98	96	02	96/7	a	02/03	97/8	n	96/7	?
Household reference person (HOH)	a	a	a	a	a	00	a	02	02/03	a	99	a	03
Relationship to HOH	96/7	96/7	a	98	a	98	a	02	02/03	a	n	a	x
Ethnic origin	96/7	96/7	96/7	a	a	01	96/7	a	02/03	00/01	n	n	03
National Identity	00/1		01/2		01			03	02/03	01/02			
Tenure	96/7	96/7	96/7	96/7	96	98	a	n	n	97/98	n	96/7	2003 <sup>2</sup>
Economic Activity (ILO)	96/7	96/7	96/7	96/7	a	98	n	a	n	97/98	n	a	?
Industry, occupation, socio-economic classifications	96/7	03/4	96/7	98	a	98	96/7	a	n	97/98	03	a	2003
Full-time and part-time work	96/7	96/7	96/7	96/7	a	98	x	96/02	02/03	97/98	99	a	x <sup>3</sup>

	EFS	FRS	GHS	HSE	LFS	NTS	SEH	SCHS	EHCS	OMNI	SHS	CHS	WHS <sup>1</sup>
<b>SECONDARY SET</b>													
Benefits / Tax Credits	97/8	a	97/8	x	x	x	x	02	02/03	x	n	97/8	x
Consumer durables	96/7	a	a***	x	x	x	x	n	x	x	n	a	x
Income as a classificatory variable							n	02	n	n			x
Length of time since last paid work (new outputs)	96/7	96/7	98/9	x	a	98	x	02	02/03	n	n	98/9	?
Time in present job (new outputs)	96/7	96/7	98/9	x	a	x	x	02	n	x	n	98/9	x
Usual hours in main job (new outputs)	96/7	97/8	98/9	x	a	x	x	02	n	x	n	98/9	x
Accommodation type	96/7	96/7	96/7	a	x	00	a	03	n	x	n	a	x
Housing costs and benefits	96/7	97	05/6	x	x	X	96/7	a	n	x	n	x	n
Length of residence	96/7	96/7	96/7	x	96	98	96/7	a	n	x	n	96/7	03
Vehicle ownership/ continuous use	96/7	97	97/8	98	x	98	96/7	n	x	n	n	97/8	x
Health	x	03/4	a	a	x	X	x	x	n	03/04	n	a	2003 <sup>4</sup>
Carers	x			n	n****	X		x	x	x		x	2003 <sup>5</sup>
Educational Attainment	n		as LFS	a	As GHS	X		x	n	n		x	2003 <sup>5</sup>
Qualifications	n		as LFS	a	As GHS	X		x	n	n		x	n <sup>6</sup>
Crime and fear of crime	x		x	00	x	n	n	x	n	x		x	x
Social Capital	x		04/5	00	x	x	n	x	n	x		x	x

**Last Updated: March 2008**

**Notes:**

\* Survey of English Housing (SEH) will become English Housing Survey (EHS) w.e.f. Apr 2008

\*\* Scottish House Condition Survey being undertaken as a continuous survey by ONS in consortium with MMBL since October 2003

\*\*\* GHS Consumer durables - from 2005/6 the harmonised questions on consumer durables will be reduced to a few key items on GHS to reflect requirements for statistics on income and living conditions for the European Union.

\*\*\*\* Non-harmonised question on the 2005 Eurostat ad hoc module. No current plan to implement more widely.

1 Welsh Health Survey, October 2003 – September 2005;

2 Harmonised output only

3 Only in survey for HRP

4 Mixture – Census LLTI question, which gives limiting long-standing illness only (but not subdivision); plus 3-point general health scale

5 Harmonised output (census question used as input)

6 Qualifications asked only for the purpose of deriving educational attainment (census question and groups used)