Introductory Guide to the Living Costs and Food Survey

UK Data Service

Author: UK Data Service
Updated: August 2014
Version: 2.0

Contents

1. Introduction

1.1 The guide

This guide is designed to pull together essential information about the Living Costs and Food Survey (LCF) deposited by the Office for National Statistics (ONS) and Department for Environment, Food and Rural Affairs (DEFRA). Although the guide draws heavily on pre-existing documentation it is intended to supplement rather than replace the user guides produced by the Office for National Statistics, which should be considered definitive. Users should therefore consult the user guides before using the data. These user guides are provided with data downloads and can also be accessed from the documentation section of a specific dataset on the Living Costs and Food Survey series page. References to specific locations in the documentation are given where applicable. Links to online material are either provided within the text or as endnotes. Where not made explicit, the LCF 2011 and its documentation are used as a point of reference.

1.2 The Living Cost and Food & Expenditure and Food Surveys

- From 2001-2002, the Expenditure and Food Survey (EFS) brought together the Family Expenditure Survey and the National Food Survey. Both survey series were well established and important sources of information for government and the wider community, charting changes and patterns in Britain's spending and food consumption since the 1950s. Both were completely replaced by the EFS. The merger of the two surveys also brings benefits for users, as a single survey on food expenditure removes the difficulties of reconciling data from two sources.

- From January 2008, the EFS became known as the Living Costs and Food Survey and it also became a module of the Integrated Household Survey (IHS). As a consequence of this change, the EFS questionnaire was altered to accommodate the insertion of a set of questions (the 'core questions'), common to all of the separate surveys which feed into the IHS. From 2015 onwards the LCF will no longer be a module of the IHS.

- In most cases these did not affect the LCF data itself: the new variables added to the survey are not made available in the LCF datasets but instead in the IHS. A few variables names changed, and in a small number of cases, mostly related to summary income indicators, variables were deleted. More information is available in the Volume G of the Documentation to the 2008 data.

- The Office for National Statistics (ONS) has overall project management and financial responsibility for the LCF, whilst the Department for Environment, Food, and Rural Affairs (DEFRA) sponsors the food data. As with the FES and NFS, the LCF continues to be primarily used to provide information for the Retail Prices Index, National Accounts estimates of household expenditure, the analysis of the effect of taxes and benefits, and trends in nutrition. The results, however, are multi-purpose, providing an invaluable supply of economic and social data.

- The LCF is designed on a voluntary sample survey of private households. Although the report based on the survey is called 'Family Spending,' the basic unit of the survey is the household although individual and benefit unit information is also available. Each individual aged 16 or over in the household visited is asked to keep diary records of daily expenditure for two weeks. Information about regular expenditures, such as rent and mortgage payments, is obtained from a household interview along with retrospective

information on certain large, infrequent expenditures such as those on vehicles. The results also include information from simplified diaries kept by children aged between 7 and 15.

- Expenditure captured by the LCF is classified using European Standard Classification of Individual Consumption by Purpose **(COICOP)**, in place of the codes used in the FES and NFS, which were unique to the two surveys. More introductory information about the LCF is contained in the User Guide Volume A.

1.1 Sample design

The LCF sample design for Great Britain is based on a multi-stage stratified random sample with clustering. It is drawn from the Small Users file of the Postcode Address File (PAF) - the Post Office's list of addresses. All Scottish offshore islands and the Isles of Scilly are excluded from the sample because of excessive interview travel costs. Postal sectors are the primary sample unit. Six hundred and seventy two postal sectors are randomly selected during the year after being arranged in strata defined by Government Office Regions (sub-divided into metropolitan and non-metropolitan areas) and two 2001 Census variables – socio-economic group and ownership of cars. These stratifiers were introduced in the 1996/1997 Family Expenditure Survey. The Northern Ireland sample is drawn as a random sample of addresses from the Valuation and Lands Agency list.

The survey is continuous, interviews being spread evenly over the year to ensure that seasonal effects are covered. Further information on sampling can be found in the User Guide Volume A of the LCF documentation.

1.2 Accessing the data

To access UK Data Service supported datasets such as the LCF, all users must register with the UK Data Service. You will need a username and password to register. If you are from a UK higher education institution your usual central username and password will suffice, allowing you to authenticate using the UK Access Management Federation. If in doubt, contact your local IT service.

If you do not have a username and password issued by a UK HE/FE institution, you will need to apply for a UK Data Archive username and password. If you need further advice go to Login help on the UK Data Service website.

Once registered, users can download/order the datasets direct from the UK Data Service website (usually in SPSS, Stata or tab-delimited formats) via our Data Catalogue.

All users requiring data for non-commercial purposes can download data free of charge. Where data is required for commercial purposes fees apply. See Charges on the UK Data Service website for more information.

1.3 Documentation

The LCF is a more complex survey than some others, and reading around the data will lessen the risk of misunderstanding the nature of the datasets.

The UK Data Service website provides the Data Catalogue, from where you can access the data directly. It also provides additional material to help LCF users.

The LCF Series page is where you can find the catalogue record for the date and year you require. Documentation can be obtained either by scrolling down the specific catalogue record or by using the link at the top of that record.

The documentation, created by the data depositors, is also available as part of download. You may find that there is some variation year to year (for example, volume names may change to numbers rather than letters).

The following information relates to the LCF 2011 and can be considered a typical example.

Volume A - Introduction

Introduction
Part 1 Guide to the 2011 User Documentation
Part 2 The Structure of the 2011 Database
Data Flow Chart
Part 3 Database Definitions
Part 4 Descriptions and Response Rates
Part 5 Uses and Definitions
Part 6 Survey Improvements and Weighting

Volume B – The Household Questionnaire User Guide

Explanatory notes
Household questions in questionnaire

Volume C – The Income Questionnaire User Guide

Explanatory note
Income questions in questionnaire

Volume D - Expenditure Codes

Explanatory notes
Part 1 Expenditure codes: Detailed

Volume E - Raw Variable User Guide

Explanatory notes
Part 1 Raw table definitions
Part 2 Raw variables in alphabetical order on variable name
Part 3 Raw variables coding frame in variable name order

Volume F - Derived Variable User Guide

Explanatory note
Anonymisation in the 2011 LCF
Part 1 Raw table definitions
Part 2 Derived variable descriptions
Part 3 Specification of product P-codes
Part 4 Derived Variables Coding Frame

Volume G – Derived Variable Flowcharts

Explanatory note
Flow-charts showing the source and make-up of derived variables in the data

Volume H – Changes Database

Explanatory notes
Part 1 New Variables
Part 2 Deleted Variables
Part 3 Changed Variables
Part 4 New Derived Variables for 2011
Part 5 Deleted Derived Variables for 2011
Part 6 Derived Variables that have changed from 2010
Part 7 Major Changes to the Database for 2011

1.6 Other UK Data Service resources for the LCF

The UK Data Service provides dedicated information pages for each year of LCF/EFS data. Detailed user guides for the data can be found on these pages.

A Variable and Question Bank for the main UK Data Service supported datasets allows you to search for specific variables within UK Data Service datasets across time.

1.7 Office for National Statistics (ONS) resources

The Office for National Statistics website provides links to the publications based on the LCF as well as some general information.

- Any questions?

Queries regarding accessing or analysing the LCF can be directed towards the UK Data Service Helpdesk:
Email : help@ukdataservice.ac.uk
Tel: +44 (0)1206 872143

UK Data Service website: ukdataservice.ac.uk/

2. Understanding the structure of the LCF

2.1 Introduction

This section outlines the survey instruments used in the collection of LCF data and topics covered by the survey. LCF data is partly collected through household and individual income questionnaires, although the main detail on expenditure is obtained through individual expenditure diaries. How information from these different sources is presented in the LCF datasets is often not completely obvious to new users. Consequently, this section will map out the way the different instruments relate to the various types of data files you will find when you download an LCF data deposit.

The main learning outcomes of this section are:

- becoming familiar with the topics covered by the LCF;
- knowing the different tools used for data collection: the difference between the household questionnaire, income questionnaire, and expenditure diaries (adult and children);
- understanding the different 'levels of hierarchy' in the data: knowing the difference between household level, individual level, and item level data, as well as their uses;
- understanding the difference between the LCF Raw Data Files, Derived Data Files, and Food Diary Database and how these relate to the survey instruments used in data collection;
- knowing who the household reference person is and what a 'primary shopper' is.

2.2 Survey tools used for data collection

Information for the LCF is collected through three main sources. These are: a household questionnaire; an income questionnaire (for each adult household member); and expenditure diaries (for each adult, and for children aged between 7 and 15 years).

2.2.1 The household questionnaire

The household questionnaire collects information on household characteristics and the characteristics of the individuals living there.

The household questionnaire contains information on one-time purchases/irregular payments and regular payments, many of which have relevance when reflecting the expenditure of the household as a whole.

This includes information on expenditure on:

- rent, rates, mortgages, insurance, council tax, pensions;
- consumer durables;
- telephone, electricity and gas, central heating, television and satellite/digital services;
- vehicles;
- season tickets;
- credit cards, loans, hire purchase, club credit, banking;
- capital improvements and house maintenance, house moves, furniture, carpets, second dwellings;

- holidays and flights;
- welfare and free school milk and meals, education;
- separation/maintenance allowances;
- employer refunds;
- items paid from outside the household,
- money given to household;
- Internet purchasing.

Information on irregular purchases is often obtained through retrospective recall (i.e. the question may be 'when did you last purchase a <item x>?'). Expenditure on retrospective recall is converted to a weekly equivalent value. That is, if the recall period is one year, then the weekly equivalent value is calculated by dividing by 52.

Demographic information on household composition (e.g. age, gender of members) and information on accommodation and tenure is also collected in the household questionnaire. The household questionnaire must be completed for EVERY adult in the household for a household to be included in the valid sample. The household questionnaire is typically completed by one household member.

2.2.2 The income questionnaire

The income questionnaire collects information on the income of individuals in households and of the household as a whole. Topics covered include:

- employment status, job description, pay details, income from self-employment and subsidiary employment;
- National insurance contributions;
- redundancy payments;
- concessionary bus passes, social security benefits, pensions, allowances;
- money sent abroad;
- investments and children's income;
- personal information such as age, gender and marital status.

The income questionnaire must also be completed for EVERY adult in the household. See 'Survey Definitions', Volume C of the user guide for income definitions for the LCF 2011.

2.2.3 The expenditure diaries

Each individual aged 16 years and over in the household is asked to keep diary records of daily expenditure for two weeks (14 days). Each expenditure amount collected through the diary and some of those collected through the questionnaire are given a code that represents the type of item purchased. The expenditure coding system used is based on the European standard classification, Classification of Individual Consumption by Purpose (COICOP). This is discussed further in Section 5. The children's diary is shorter and more simplified than the adult diary and is completed for each child aged between 7 and 15 years.

See the explanatory notes of Volume D of the documentation for further information on the collection of data through the diary for the 2011 data. Appendix A in the current guide also gives an example of an adult expenditure diary.

An expenditure diary must be completed by the main shopper to constitute a valid response.

2.3 Who are the household reference person and the main shopper?

Definitions for key terms are to be found in the LCF User Guide. For most years this is volume 1 or A, definitions are to be found in part 5.

The **household reference person** (HRP) is a person who is taken as a representative of the household in lieu of head of household. The concept is one which has been adopted on all government surveys since 2001-2002.

> "The household reference person is the householder, i.e. the person who:
>
> a. owns the household accommodation, or
> b. is legally responsible for the rent of the accommodation, or
> c. has the household accommodation as an emolument or perquisite, or
> d. has the household accommodation by virtue of some relationship to the owner who is not a member of the household.
>
> If there are joint householders the household reference person will be the one with the higher income. If the income is the same, then the eldest householder is taken." (User guide volume A).

2.4 How is the data represented and stored in the datasets?

In the 2011 deposit, there were five main data files forming the LCF database, and an additional food expenditure database stored in Microsoft access format. Multiple files are quite common for large datasets where they are large or where files may represent information at different data levels known as 'levels of hierarchy'.

Information contained in LCF datasets is usually at one of three data levels:

- households level;
- individual level (aka 'person level' or 'personal level');
- (expenditure) item level.

These levels can be nested, so that it is possible to look at individuals within households and items bought by persons and so forth. In household level files the case is a household, by default analyses of this file have the households as units of analysis and are therefore deemed to be 'household level' analyses.

2.4.1 Household level data

The definition of a household used by ONS is:

"One person or a group of people who have the accommodation as their only or main residence AND (for a group) either share at least one meal a day or share the living accommodation, that is, a living room or sitting room"

Household level data gives information on the characteristics of entire households, and so this will be identical for every person in a given household. For example, information on utility bills will be the same for each person as this relates to the household.

Household level data provide details on the entire expenditure on a certain type of item for households (e.g. how much the entire household spends on a type of milk or on satellite television). This information is obtained through aggregating information contained in

individual level diary files, or from the household or income questionnaires. Such data can be used for household level analysis or be linked to individual level data.

The LCF is predominantly designed to be used as a household level survey both in its sample design and in the topics it seeks to address. Household level analysis can often be more suitable for analysing expenditure, particularly where items are shared between household members. For example, households often have a 'primary shopper' who shops for food on a regular basis. Although expenditure does not directly inform us about consumption, looking at household level expenditure on groceries can give a clearer picture of what people actually consume than individual level expenditure. If we alternatively examined individual expenditure, those who were not primary shoppers would look like they consumed very little whereas the primary shopper would look like they consumed a lot. A more sensible approach would be to share the overall household level of expenditure on food between household members in some way. Admittedly, this approach is not perfect – some households will contain multiple families or unrelated individuals where the assumption that households share goods might be less persuasive.

2.4.2 Individual level data

Individual level data gives information on the characteristics and expenditure of specific respondents within households. This information can be used for individual level analysis.

2.4.3 Item level data

Item level data contains information on specific expenditure items so each row of data represents an item of expenditure for a household or an individual. Consequently, there are likely to be multiple rows for a given individual or household reflecting different items of expenditures. In some files, item level data is aggregated to reflect total level of individual or household expenditure on a given type of item rather than each specific instance.

2.4.4 Other variations

While most files adhere to the above levels, you may find that files may only apply to a particular type of item or individual. Individual level files may be applicable to only adults or children for example.

One file is at an intermediate level between individual and household; benefit unit. A **benefit unit** is an adult plus their spouse or cohabitee (if applicable) plus any dependent children they are living with[1]. The benefit unit has analytic value in that the number relationships between benefit unit members are more readily understandable than the more diffuse relationships involved in, say, a multi-family household or student house. The benefit unit file, however only contains information about the benefit unit type.

The next section describes each of the files, and the above variations will be explained in full. In the documentation produced by the Office for National Statistics, you may find that 'household level' data contain information stored at the person or item level. This occurs, for example, where the data have been collected from the household form or for household analysis, but where multiple records per household are held.

[1] Benefit units therefore resemble 'family units'. However, the treatment of adult children differs somewhat.

2.5 The raw, derived, and food expenditure datasets

The different datasets can further be classified into one of three groups:

- raw datasets;
- derived datasets;
- food diary database.

Together the raw and derived datasets form the LCF general expenditure deposi

Although these groups of datasets contain information obtained from across the different levels of hierarchy and survey tools, they use or compile the information in different ways.

2.5.1 The raw database

The raw datasets contain questionnaire responses as recorded. The structure of the raw datasets most closely relates to that of the actual questionnaires. The raw database thus contains data 'as received'. Expenditure is stored in weekly equivalent values (wevs) to allow comparison and some imputation has been carried out. Care should be taken when using raw data. In particular, a new feature of the LCF is the inclusion of 'looping' variables – where a question is repeated, a variable may be repeated (see Volume E, Explanatory Notes).

The names of raw data files begin with the prefix 'raw' (e.g. 'rawhh' contains raw household level information whereas 'rawper' contains raw person level variables.

Much of the data contained in the raw datasets is also represented in the derived datasets. Consequently, there is a lot of redundancy in the raw data. Most users will therefore only need to use the derived datasets.

For more information on the derived variables, see the derived database volume of the documentation.

2.5.2 The derived datasets

The derived database consists of variables that have been adjusted in some way in relation to the state in which they were received. This adjustment might involve only a change in name, or it might involve some calculation or aggregation. The variables are grouped in files that have DV in the name. Examples are DVHH or DV_SET89. To the majority of users, this data is likely to be the most relevant and user-friendly.

2.5.3 The food diary database

The food diary database contains calculated information on the nutritional values of food and drink in households. Unlike the other data files, this is presented in Microsoft Access (single file) or Tab format (multiple files) The LCF records information on the weight/volume and amount paid for all food and drink brought home (including free food), as well as amounts paid for takeaway meals and snacks eaten at home. Information on the amount paid for all food and drink consumed away from home and type of meal is also recorded. Together, this information is used by Department for Environment, Food, and Rural Affairs (Defra) to estimate portion size and nutrition. The results are published in the annual report 'Family Food' at United Kingdom level. The calculation of nutritional values and structure of the food diary database is discussed in further detail in Section 6.

3. What files will you find in the LCF microdata?

3.1 Introduction

In this section, we describe the different data files that make up the LCF data composed of six main data files. The files are called:

Dv_set89_ukanon
Dvhh_ukanon
Dvper_ukanon_v2
Rawper_ukanon
Rawhh_ukanon

The food database is described separately and described in part 5 of this document. Although the above list gives the key names as referred to in the user guides, the actual names of the files you find in a data download are likely to differ slightly depending on the version you download, although the names will reflect those given above.

3.2 The LCF data files

Dv_set89_ukanon

This file contains information on expenditure. The variables contained within are:-

- case number (household level);
- person number;
- COICOP expenditure codes;
- wxpenditure amount;
- Whether person is an adult or child.

Dvhh_ukanon

The main derived household level file. This file is likely to be central for creating datasets at the household level. It contains information on household characteristics such as housing tenure, and items that are contained in a household (e.g. whether there is a home computer or a microwave oven). It provides information on household composition, such as the number and age of children in a household.

Dvhh importantly contains the majority of the 'C code' variables (prefixed with a letter c) that indicate items of expenditure, representing aggregate household spending. These variables are discussed further in Section 4.

Dvper_ukanon_v2

The main derived person level file. This file is an important file for person level analysis. Dvper gives information on the characteristics or demographics of individual household members, such as their age, gender, individual income, and social security benefits claimed. It also indicates whether a respondent is the household reference person or not.

Rawper_ukanon

This file is the main raw person level file. Most users are likely to require the derived files rather than the raw files.

Rawhh_ukanon

This is the equivalent raw household level file.

Matching cases across files

It is possible to link different LCF files and also link them across years. In order to do so you need only identify the correct linking variable and then link according to the analysis package you are using. More information on this can be found in the hierarchy and linking guide, also available from the UK Data Service website.

Linking for the LCF is undertaken using a number of identifier ('Key') variables. The data levels and corresponding linking variables are:

households **(case)**
individuals **(person)**
expenditure Items **(itemnum)**

Using these key variables will enable you to link and merge files for broader or more detailed analysis.

4. Understanding LCF variable names and value codes

4.1 Introduction: suffixes and prefixes

The naming of variables and coding of expenditure in the LCF is systematic. Consequently, information on what variables represent can be inferred from their prefixes and suffixes. The main variable prefixes and suffixes are briefly described below. More information on the derived database is given in Volume F of the User Documentation (explanatory notes for what the prefixes and suffixes mean and part two for the list of variables).

The main learning outcomes of this section are:

- understanding what the different prefixes and suffixes used in the naming of LCF variables mean;
- knowing where information for the different types of variables is obtained from.

The following information is given in the LCF 2011 user guides.

4.2 Prefixes

4.2.1 Variables prefixed 'a' and 'b'

A-codes and B-codes are aggregate variable codes, some at **household level** and some at **individual (personal) level**. These variables are identified through an 'a' or 'b' prefix respectively and are derived from the questionnaires (not the diaries).

A-Codes contain information on individual and household characteristics such as children in the household, the type of heating in a household or age. B-codes contain income and expenditure information, such as on mortgages, holidays or benefits paid.

A and B codes are listed in The Derived Database volume of the User Documentation.

4.2.2 Variables pre-fixed 'c'

Variables representing household aggregates of the expenditure codes are called C-codes (or C-variables): The variable name is the COICOP expenditure code (Classification of Individual Consumption by Purpose, discussed in further detail in Section 5) pre-fixed with 'C'. C–codes are used for variables that give a variable for each type of item expenditure. Each expenditure amount collected through the diary and some collected through the questionnaire is given a code that represents the type of item that was purchased. C-codes generally reflect diary information, although some may also be derived from the household questionnaire.
A complete listing of the expenditure codes and more details on their use can be found in Expenditure codes volume of the User Documentation.

4.2.3 Variables Prefixed 'p': Product codes

Product codes (or p-codes) are variables which have been derived from the questionnaire data. These need not be 'products' in the conventional sense and include such variables as gross wage. The derived database volume of the User Documentation gives descriptions of these codes as an aid to understanding their structure, along with more detailed explanatory notes on their use.

As a rule of thumb, codes P001 to P199 are calculated for ***each person*** (unless otherwise indicated). Codes P200 and above are calculated for ***each household***.

Unless stated otherwise, all monetary values appear as weekly equivalents. For Northern Ireland only, rateable value (code B070U) appears on the household schedule as an annual value but should be recorded on the database as a weekly equivalent; rate poundage (code B080U) should appear at the rate in the pound payable.

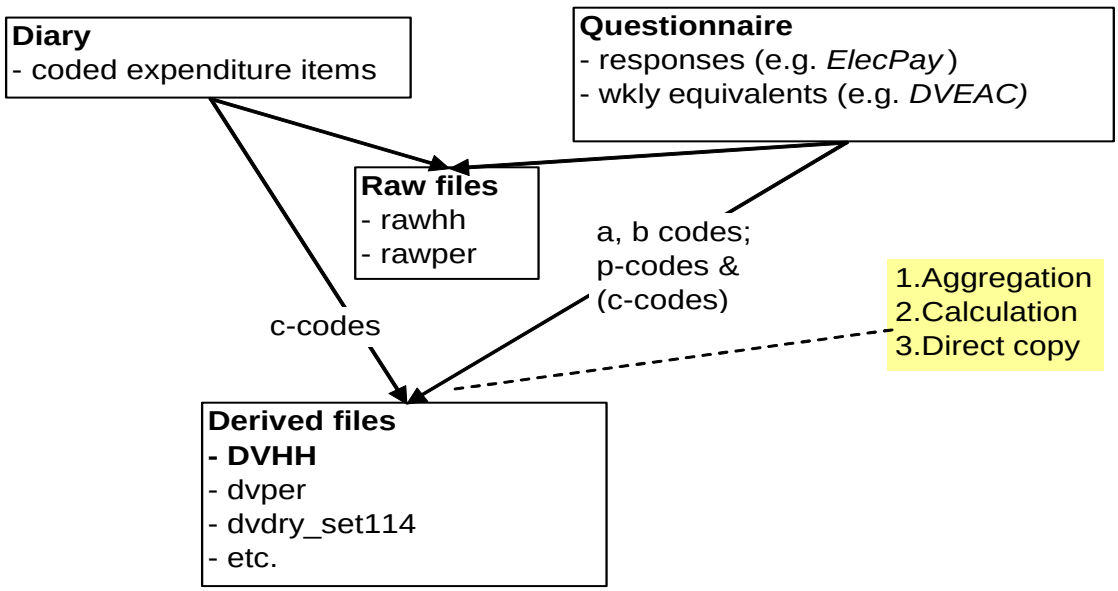
4.2.4 Weekly equivalent values (wev) - codes prefixed 'DV'

So that expenditure from different sources can be aggregated, it is necessary for all expenditure amounts to be on an equivalent basis. In order to achieve this, all expenditure amounts collected through the questionnaire are converted to WEEKLY EQUIVALENT VALUES. In the dataset, weekly equivalent values can be identified because they are pre-fixed 'DV'. Information on weekly equivalent values is included as part of the questionnaire. Weekly equivalent values are RAW variables and so are listed in Raw Database volume of the Documentation.

4.2.5 How do the prefix codes relate to the survey instruments?

The below diagram (Fig. 1) summarises the process for creating the raw and derived files. The diary information is used to create the c-code variables in the derived datasets whereas the 'a', 'b', and 'p'-code, (and some c-code) variables are created from the information collected on the questionnaires.

Fig 1. *Household Summary of how the Raw and Derived Data files are created*



Note that:

Household derived variable dataset (dvhh)

This database contains the following types of variables:

- A- Codes
- B- Codes
- C- Codes
- P-Codes
- anonymised variables, weighting variables, and administrative variables

Person level derived dataset (dvper)

The person level derived file contains:

- A- Codes
- B- Codes
- P-Codes
- anonymised variables, weighting variables, and administrative variables

4.3 Suffixes

4.3.1 Variables suffixed 'c'

These codes represent the expenditure of children, aged between 7 and 15, who completed a diary. The codes are aggregated to household level. Variables suffixed 'c' are only available as expenditure or product codes.

4.3.2 Variables suffixed 't'

These codes represent total expenditure by children and adults. They are aggregated to household level by adding the original (adult) variable to those suffixed 'c', (i.e. p522 + p522c = p522t). Variables suffixed 't' are only available as expenditure or product codes.

4.3.3 Variables suffixed 'L'

These represent expenditure in nominated large supermarkets.

4.3.4 Variables suffixed 'x, y or z'

These suffixes signify expenditure on clothing the suffixes are 'x' for selected clothing chains, 'y' for large supermarkets and 'z' for charity shops.

4.3.5 Variables suffixed 'w': Internet purchases

These codes represent expenditure for goods and services ordered via the internet, i.e. **w**eb purchases. Variables suffixed 'w' are only available as expenditure codes for certain goods and services ordered via the internet.

4.3.6 Variables suffixed 'p' or 'u': Anonymisation

Variables which could lead to the identification of a household or an individual are not released to users. An 'anonymised' version of sensitive raw or derived variables is provided in the user dataset. Variables which have been anonymised can be identified because they are suffixed 'p' (which stands for **p**roxy), those which are not have been detailed in the appropriate User Documentation.

Unanonymised versions are, on the main, suffixed 'u' and are not accessible to external users of the LCF. Raw diary information is not provided to any external users.

The principles of anonymisation are explained in the Derived Database guide in the User Documentation. This information is also contained in Appendix B of this guide.
FS codes

FS codes, where present, are only used for the preparation of the publication 'FAMILY SPENDING'.

4.4 Period codes

Where expenditure amounts are not given on a weekly equivalent basis, a weekly equivalent is calculated by dividing the amount given by the appropriate period. The standard period codes can be found at the end of User Guide Volume A.

5. COICOP: Classification of expenditure

5.1 Introduction

The current expenditure coding system used in the LCF is based on a European classification, 'The Classification of Individual Consumption by Purpose' (COICOP). This is an internationally standardised system used by many countries. In this section, we will consider:

- the difference between what are referred to as 'EFS codes' and COICOP;
- how the COICOP classifies expenditure into 12 main categories;
- its hierarchical structure- as well as individual expenditure codes, COICOP breaks types of expenditure into 'divisions', 'groups', and 'classes';
- what COICOP-plus is;
- how information coded by COICOP is stored in the datasets either as a coding frame for single variables or as separate variables for each COICOP code.

The following information is largely obtained from Expenditure codes volume the user guides. The user guides are central to understanding the COICOP codes and should constantly be used as a reference when conducting your analysis.

5.2 What is the difference between EFS codes and COICOP?

- Items of expenditure in the LCF are initially assigned a five digit code, referred to as an 'EFS code,' which represents a type of expenditure (the code might represent 'bread', 'mortgage payments' or 'visits to the cinema', for example). These five digit codes are listed in column 'EFS Code' of the sheet 'Expenditure Codes - Detailed in volume D of the user guide.
- In the course of processing LCF data, the EFS Codes are mapped to the COICOP coding frame.
- For most purposes, the derived variables coded using the COICOP will be suitable for your analysis and you will not require the original EFS codes.

5.3 How does COICOP classify expenditure?

In COICOP, there are 12 main categories of expenditure plus an extra category for 'Non-consumption' expenditure. This extra 'Non-consumption' category is not covered (i.e. not broken down) by COICOP. The 12 main categories (plus non-consumption category) are:

- Food and Non-Alcoholic Beverages
- Alcoholic Beverages, Tobacco and Narcotics
- Clothing and Footwear
- Housing, Water, Electricity, Gas and Other Fuels
- Furnishings, Household Equipment and Routine Maintenance of the House
- Health
- Transport
- Communication
- Recreation
- Education
- Restaurant and Hotels
- Miscellaneous Goods and Services
- Non-Consumption Expenditure

5.4 What are 'divisions', 'groups', and 'classes'?

COICOP codes are hierarchical, meaning that there are major and minor sub-groups of different types of expenditure items. More precisely, COICOP has four levels of detail:

01	Division
01.1	Group
01.1.1	Class
01.1.1.1	COICOP expenditure code

Consider the following example:

1 FOOD AND NON-ALCOHOLIC BEVERAGES

1.1	**Food**
01.1.1	**Bread and cereals**
01.1.1.1	Rice
01.1.1.2	Bread
01.1.1.3	Pasta products
01.1.1.4	Pastry-cook products
01.1.1.5	Other products

Here: 01.1 is the group (food) 01.1.1 is the class (bread and cereals) 0.1.1.1.1 is the COICOP expenditure code (rice)

The level relevant to you will depend on what question you are trying to answer. For example, you might wish to know how much households spend on fruit. If so, the class code for fruit is 01.1.6. However if you were specifically interested in bananas a further digit is needed; the COICOP expenditure code for bananas is 01.1.6.2.

5.5 What is COICOP-plus?

For the LCF, expenditure coding has an extra level, called COICOP-plus, which allows for more detailed recording of expenditure items. This allows coding down to more specific brands of products, which may to an extent be specific to the UK context.

5.6 How is information using COICOP stored in the datasets?

COICOP codes are either used as:

- the coding frames for the values of variables;
- to create specific variables for different types of expenditure, where the names of the variables are based on the COICOP system (C-codes).

Where COICOP codes are used as coding frames the code numbers are stored as variable values. For example, set114 is the item level table which details items bought for business. In it, the variable coi_plus describes the item bought. The coding frame is very extensive so it is impractical to list it in full here. The meanings of these values can be looked up by looking in the expenditure codes volume of the documentation.

6. The Food Diary Database

6.1 Introduction

The food diary database forms the basis of the annual Defra publication, 'Family Food'. The report presents trends in purchases by type of food and converts these into energy and nutrient intakes based on a database of nutrient values provided by the Food Standards Agency.

Methodological guides which accompany this report are available from the Defra website.

6.2 Family Food Report: Calculations

A number of calculations are undertaken to derive the figures contained within the Family Food report.

6.2.1 Weighting

Households are first weighted. Each household is given a weight calculated by ONS to adjust for non-response and to ensure a match with population totals. Weights range from 0.001 to 12.33318. The average is around 3.7. A survey household with a weight of 3.7 essentially represents 3700 UK households.

Measures are given in weekly form. These are therefore based on the fortnightly purchases which are then halved to produce the weekly purchase measure.

Household calculation (eating at home)=

$$\frac{\text{sum of quantity x weight}}{\text{sum of household people x weight}} \times \frac{1}{2}$$

The numerator (top line) of the calculation refers to the sum of quantities x household weights for households with a positive consumption for a given item.

i.e. (quantity X household weight)+ (quantity X household weight) .... + (quantity X household weight)

The denominator (bottom line) is basically...

(number of household members x household weight) + (number of household members x household weight) .....+ (number of household members x household weight)

...for all households, so including those who do not make a positive purchase. The calculation therefore provides an estimate of average household consumption for the population, so includes non-purchasing households. For the eating out calculation, this is:

$$\frac{\text{sum of number of purchases x portion size x weight}}{\text{sum of household people x weight}} \times \frac{1}{2}$$

6.2.2 Nutrition Calculations

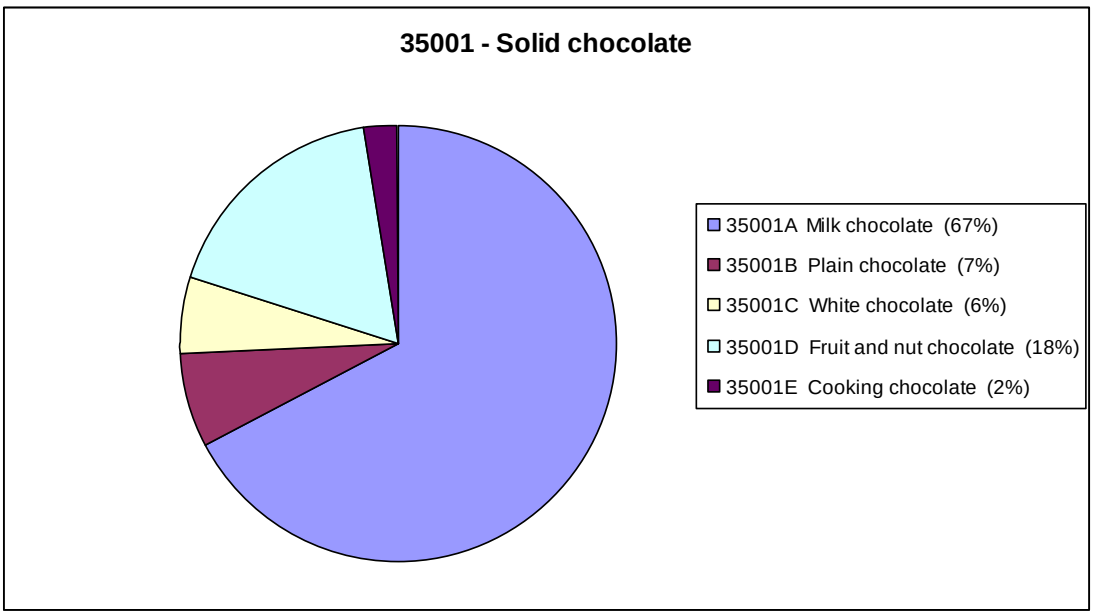
Nutrient intake is calculated as the average per person per day. The nutrition of items is calculated using nutrient conversion factors:

- firstly, Defra uses the food and drink coding structure used by the LCF to identify foods (referred to as **Defra codes**);
- the Food Standard Agency (FSA) provide nutrient profiles for subgroups of the Defra codes- the **FSA nutrient codes**.
- the Defra codes from the LCF may be of a higher level of aggregation than the FSA nutrient codes. Defra consequently undertakes a market share analysis of the FSA codes to determine how to calculate an overall nutrient profile per code from its sub-components;
- the below example shows the relationship between the Defra code and FSA code subgroup for chocolate:

Defra code	Description	FSA Subcode	Name
35001	Soild chocolate	35001A	Milk chocolate
		35001B	Plain chocolate
		35001C	White chocolate
		35001D	Fruit and nut chocolate
		35001E	Cooking chocolate

- As there are different types of chocolate, the overall nutrition figures are based on the market share of the different types:

Market share:

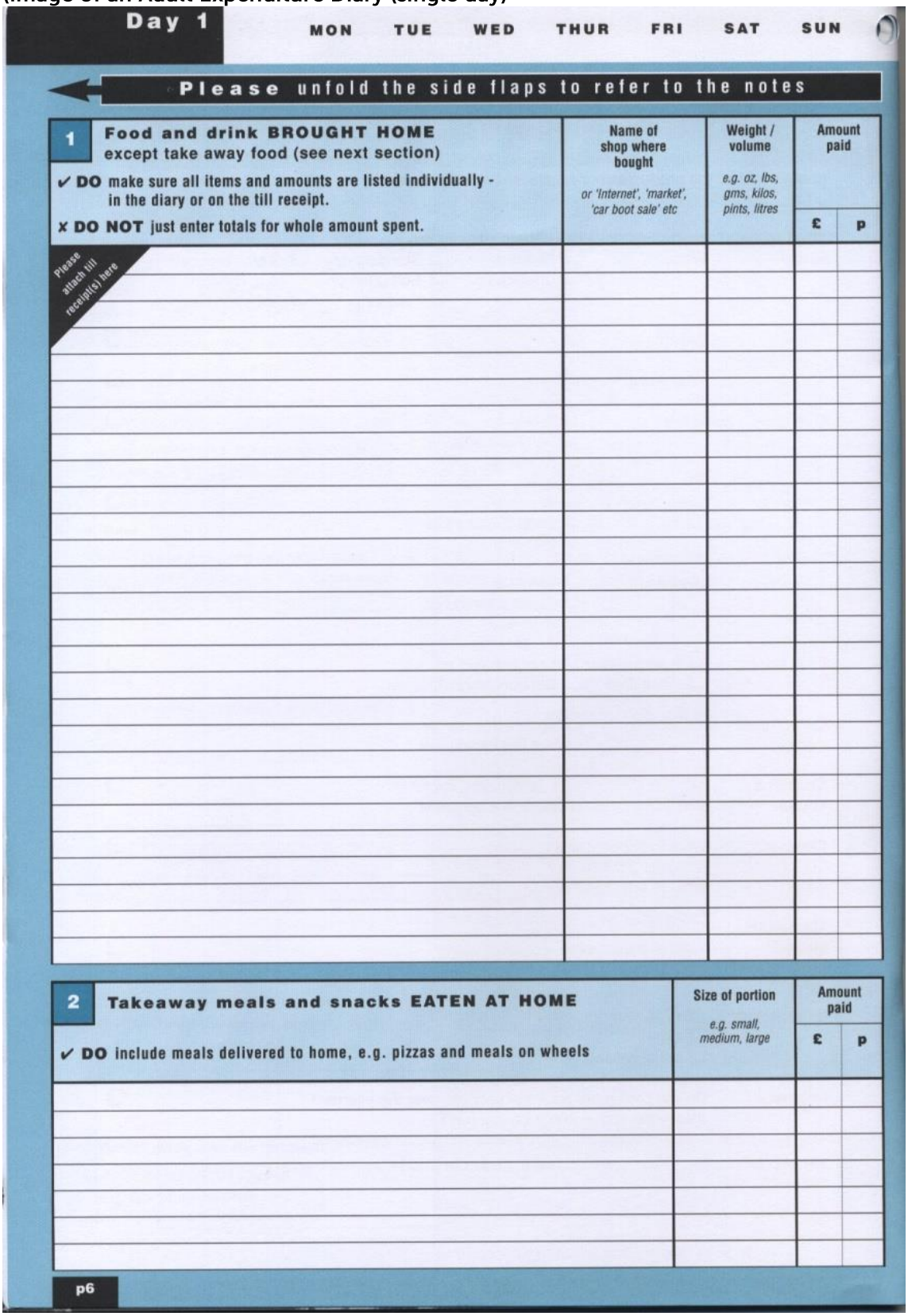


Nutrient intakes are calculated by multiplying purchased quantities by nutrient co-efficient to reach final figure per nutrient:

$$\frac{\text{sum of quantity x weight x nutrient coefficient}}{\text{sum of household people x weight}} \times \frac{1}{2} \times \frac{1}{7}$$

Appendix A. Example Diary Page

(Image of an Adult Expenditure Diary (single day)



Day 1 MON TUE WED THUR FRI SAT SUN

Please unfold the side flaps to refer to the notes

1 Food and drink BROUGHT HOME except take away food (see next section)

✓ **DO** make sure all items and amounts are listed individually - in the diary or on the till receipt.

✗ **DO NOT** just enter totals for whole amount spent.

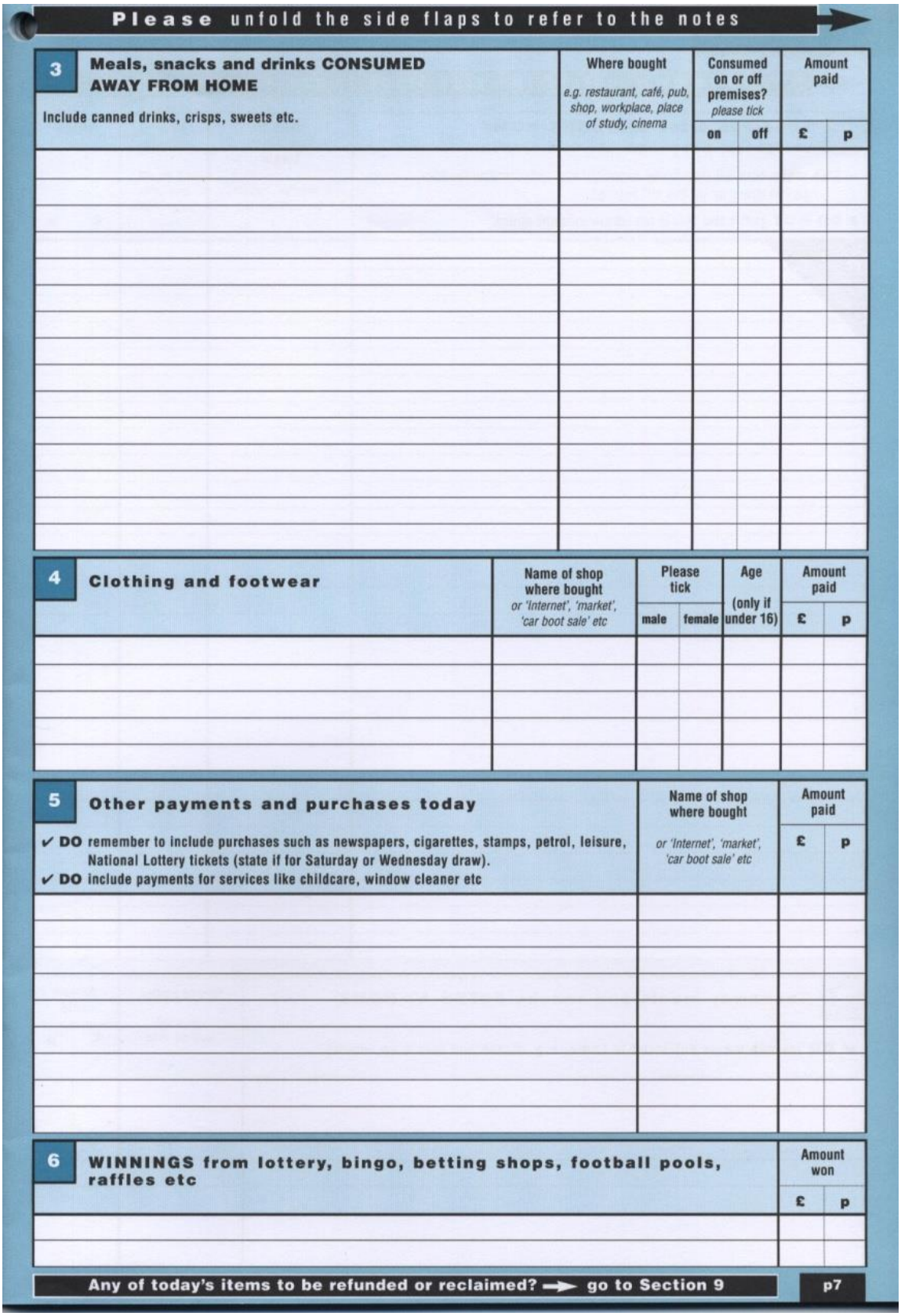
Please attach till receipt(s) here	Name of shop where bought *or 'Internet', 'market', 'car boot sale' etc*	Weight / volume *e.g. oz, lbs, gms, kilos, pints, litres*	Amount paid £	p

2 Takeaway meals and snacks EATEN AT HOME

✓ **DO** include meals delivered to home, e.g. pizzas and meals on wheels

	Size of portion *e.g. small, medium, large*	Amount paid £	p

p6



Please unfold the side flaps to refer to the notes

3 Meals, snacks and drinks CONSUMED AWAY FROM HOME

Include canned drinks, crisps, sweets etc.

Where bought *e.g. restaurant, café, pub, shop, workplace, place of study, cinema*	Consumed on or off premises? *please tick* on	off	Amount paid £	p

4 Clothing and footwear

Name of shop where bought *or 'Internet', 'market', 'car boot sale' etc*	Please tick male	female	Age (only if under 16)	Amount paid £	p

5 Other payments and purchases today

✔ **DO** remember to include purchases such as newspapers, cigarettes, stamps, petrol, leisure, National Lottery tickets (state if for Saturday or Wednesday draw).
✔ **DO** include payments for services like childcare, window cleaner etc

Name of shop where bought *or 'Internet', 'market', 'car boot sale' etc*	Amount paid £	p

6 WINNINGS from lottery, bingo, betting shops, football pools, raffles etc

Amount won £	p

Any of today's items to be refunded or reclaimed? → go to Section 9 p7

Appendix B. Anonymisation in the LCF

A common question asked of the UK Data Service Helpdesk is how the data is anonymised. Below is an extract from volume f of the 2011 user guide documentation which explains the process in detail.

The anonymisation principle

The undertaking given to LCF respondents specifies that the information they provide will be treated in confidence and that no identifiable information about individuals or households will be made available to data users. Therefore a precaution taken in releasing microdata is to 'anonymise' certain variables which could lead to the disclosure of the identity of a household. This means that the values stored for these variables must be adjusted in some way. In practice, sensitive variables are 'anonymised' by either top-coding, re-coding (for example, aggregating cases by combinations of variables), entirely dropping disclosive variables, or rounding variables. The most complex anonymisation is that of the council tax group of variables. This method is described in more detail below.

Anonymising the council tax variables

One of the general precautions taken to preserve confidentiality is to provide no information to users about the local/unitary authority of the household. In Great Britain, each local authority sets its own council tax level (and water and sewerage charge levels in Scotland). Therefore, in general, the amount of council tax payable in a certain local authority is different from that payable in others. This means that if the amount of council tax paid is known, the local authority can probably be identified. This in turn increases the risk of a particular household or individual being identified. Therefore no information on the exact amount of council tax (or water or sewerage charge) paid is released to users.

Instead of making available information on the exact amounts to users, 'anonymised' variables are calculated. These indicate the level of council tax (and water and sewerage charges in Scotland) for each household. The approximation is calculated by pooling several authorities within a region, from which a pool average is calculated. These averages are then applied to each household such that the value of council tax shown is close to the real value but such that the local authority cannot be identified.

Identifying anonymised variables in the dataset

Any variable suffixed with the letter **p** has been anonymised. Any variable suffixed with the letter **u** is unanonymised and access to it is restricted to staff in the LCF branch at the ONS.

All variables which use an unanonymised variable in their derivations are recalculated to exclude the unanonymised variable. For example, variable P550u is total expenditure and is available only to authorised ONS staff; variable P550p is the anonymised total expenditure variable which is released to other users and has been recalculated to include anonymised council tax, water and sewerage charges.

Top-coded variables in the dataset

The following section details variables that have been top-coded or re-coded during the anonymisation process for the 4 main data files – dvhh, dvper, rawper and rawhh. Note that when these variables have been anonymised they will have the suffix **p**.

Dvhh

The anonymised age variables - A065p (age of HRP by range - anonymised), A070p (age of oldest person in hhold - anonymised) & P396p (age of HRP - anonymised) - are top-coded so that anyone over 80 is anonymised.

P200p (number of rooms occupied - anonymised), A111p (rooms used solely by household - anonymised) and A114p (rooms in accommodation - total - anonymised) are also top-coded to 6, so that any household with more than 6 rooms is anonymised.

Dvper

The age variable (a005) is top-coded so that anyone over 80 is anonymised, creating variable a005p.

The ethnicity variables A012 (Ethnic origin of HRP) and A013 (Ethnic origin of partner of HRP) are re-coded into anonymised variables A012p and A013p with the following format:

Not applicable	0
White	1
Mixed race	2
Asian or Asian British	3
Black or Black British	4
Other ethnic group	10

The civil partnership codes for A006 (marital status) are combined so that the anonymised variable A006p has the following format applied:

Married and living with your husband/wife	1
Married and not living with your husband/wife	2
Cohabitee	3
Single	4
Widowed	5
Divorced	6
Separated	7
Civil Partner or Former Civil Partner.	8

Rawhh

The variables indicating the number of rooms (or bedrooms occupied) are top-coded to 6 rooms. These are: Nrmsp, Nrms2p, Nrms3p, Nrms4p, Nrms5p, Nrms6p, DVrmsp and DVbedp. Also the total number of shared rooms (DVsharep) is top coded to 2.

Rawper

The anonymised age variable (Dvage_p) is top-coded for anyone over 80 to 80.

The ethnicity variables for countries in the UK (EthE, EthW, EthNI, EthSQ1 and EthSQ24 for England, Wales, Northern Ireland and Scotland respectively) are re-coded into the anonymised variables EthEp, EthWp, EthNIp and EthSp with the following format:

Not applicable	0
White	1
Mixed race	2
Asian	3
Black	4
Other	5

Marital status variables (Marsta and xMarsta) are recoded to combine the civil partnership codes i.e. MarSta_p has the format below:

Single, that is never married	1
Married and living with your husband/wife	2
Married and separated from your husband/wife	3
Divorced	4
Widowed	5
Civil Partner or Former Civil Partner	6

and xMarSta has the following format:

Single, that is never married	1
Married and living with your husband/wife	2
Civil Partner or Former Civil Partner	3
Married and separated from your husband/wife	4
Divorced	5
Widowed	6

16 August 2014

T +44 (0) 1206 872143
E help@ukdataservice.ac.uk
W ukdataservice.ac.uk

The UK Data Service delivers quality social and economic data resources for researchers, teachers and policymakers.

UK Data Service

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL