Metadata File Name SILC_ESQRS_A_UK_2011_0000 - Version 1

TimeDimension	2011-A0
DataProvider	UK1
DataFlow	SILC_ESQRS_A

1 Contact

1.1 Contact organisation	Office for National Statistics
1.2 Contact organisation unit	Household Income and Expenditure
1.3 Contact name	Richard Tonkin
1.4 Contact person function	EU-SILC UK Project Manager
1.5 Contact mail address	Office for National Statistics,
	Room 1.024
	Government Buildings
	Cardiff Road
	Newport NP10 8XG
	United Kingdom
1.6 Contact email address	EUSILC@ons.gsi.gov.uk
1.7 Contact phone number	+44 (0)1633 456082
1.8 Contact fax number	

2 Introduction

The production of quality reports is part of the implementation of the EU-SILC instrument. In order to assess the quality of data at national level and to make a comparison among countries, the National Statistics Institutes are asked to report detailed information mainly on: the entire statistical process, sampling and non-sampling errors, and potential deviations from standard definition and concepts.

This document follows the ESS standard for quality reports structure (ESQRS), which is the main report structure for reference metadata related to data quality in the European Statistical System. It is a metadata template, based on 13 main concepts, which can be used across several statistical domains with the purpose of a better harmonisation of the quality reporting requirements in the ESS.

For that reason the template of this document differs from that one stated in the Commission Reg. 28/2004.

This report provides quality information for EU-SILC which was collected as part of the General Lifestyle Survey (GLF) in Great Britain and the Living Conditions Survey (LCS) in Northern Ireland in 2011. This version of the 2011 Intermediate Quality Report relates to and is consistent with the indicators and microdata transmitted to Eurostat on the 20th November 2012. Users should be aware that microdata available via Eurostat may not be consistent with the indicators if either have been recently revised and so should contact Eurostat to ensure consistency. See section 13 for further information on version control.

3 Quality management - assessment

Not requested by Reg. 28/2004

4 Relevance

Not requested by Reg. 28/2004

4.1 Relevance - User Needs Not requested by Reg. 28/2004

4.2 Relevance - User Satisfaction Not requested by Reg. 28/2004

4.3 Completeness — Not requested by Reg. 28/2004

4.3.1 Data completeness - rate — Not requested by Reg. 28/2004

5 Accuracy and reliability — The concept of accuracy refers to the precision of estimates computed from a sample rather than from the entire population. design effects and structure of the population under study. In addition to that, sampling errors and non-sampling errors need the variability that occurs at random because of the use of a sample rather than a census and non-sampling errors are errors production process.

5.1 Accuracy - overall — In terms of precision requirements, the EU-SILC framework regulation as well the Commission Regulation on sampling and t sample size to be achieved and to representativeness of the sample. The effective sample size combines sample size and sa design, population structure and non-response rate.

5.2 Sampling error — EU-SILC is a complex survey involving different sampling designs in different countries. In order to harmonise and make sam (with the substantial methodological support of Net-SILC2) has chosen to apply the "linearisation" technique coupled with the estimation. Linearisation is a technique based on the use of linear approximation to reduce non-linear statistics to a linear for estimator. This technique can encompass a wide variety of indicators, including EU-SILC indicators. The "ultimate cluster" ap the variance taking into account only variation among Primary Sampling Unit (PSU) totals. This method requires first stage sa the case. This method allows a great flexibility and simplifies the calculations of variances. It can also be generalised to calcu another.

However, the Eurostat method does not take account of the complete UK design which uses both the UK and Northern Irelan one strata. In order to keep the NI cases the UK have used the households as clusters for NI. In addition to this, the calibratic regression model (with the poverty indicators as the dependent variable) and created standard errors around the residual val values). The AROPE and ARPT60 indicators are complex functions including a median (which depends on the data) - creatir assume that the median is fixed which is not the case, so in order to avoid this, the two measures have been linearised using standard errors.

5.2.1 Sampling error - indicators

	AROPE			At risk of poverty (60%)			Severe Material Deprivation			Very low work intensity		
	Ind. value	Stand. errors	Half CI (95%)	Ind. value	Stand. errors	Half CI (95%)	Ind. value	Stand. errors	Half CI (95%)	Ind. value	Stand. errors	Half CI (95%)
Total	22.7	0.65	1.27	16.2	0.59	1.15	5.1	0.38	0.75	11.5	0.57	1.12
Male	21.4	0.70	1.37	14.8	0.58	1.13	5.0	0.40	0.78	10.7	0.65	1.27
Female	24.1	0.67	1.30	17.6	0.57	1.12	5.1	0.43	0.85	12.2	0.66	1.30
Age0-17	26.9	1.18	2.32	18.0	1.00	1.95	7.1	0.80	1.56	14.0	0.96	1.87
Age18-64	21.5	0.72	1.40	14.0	0.57	1.12	5.5	0.42	0.82	10.5	0.55	1.08
Age 65+	22.2	0.78	1.52	21.3	0.77	1.52	1.3	0.21	0.40			

5.3 Non-sampling error — Non-sampling errors are basically of 4 types:

- Coverage errors: errors due to divergences existing between the target population and the sampling
- Measurement errors: errors that occur at the time of data collection. There are a number of sources t instrument, the information system, the interviewer and the mode of collection
- Processing errors: errors in post-data-collection processes such as data entry, keying, editing and we
- Non-response errors: errors due to an unsuccessful attempt to obtain the desired information from ar non-response errors are considered:

  1. – Unit non-response: refers to absence of information of the whole units (households and/or pe
  2. – Item non-response: refers to the situation where a sample unit has been successfully enumer been obtained

5.3.1 Coverage error — Coverage errors include over-coverage, under-coverage and misclassification:

- Over-coverage: relates either to wrongly classified units that are in fact out of scope, or to units that (
- Under-coverage: refers to units not included in the sampling frame
- Misclassification: refers to incorrect classification of units that belong to the target population

The target population of EU-SILC UK is all private households and their current members at the time of data collection. Perso are excluded from the target population. However, from 2008 students who are living in halls of residence are also included a are not *in situ* at the time of the interview.

5.3.1.1 Over-coverage - rate

	Main problems	Size of error
Cross sectional data	·Over-coverage ·Under-coverage ·Misclassification	Over-coverage: 705 ineligible households were selected out of the total of 11,857 (5.9%) for Under-coverage: No estimate available.

Main problems	Size of error

Misclassification: None.

5.3.2 Measurement error

Cross sectional data

Source of measurement errors	Building process of questionnaire	Intervie

Measurement error occurs at the time of data collection. While attempts are made to minimise measurement error, some will occur during data collection from these four main sources:

1. Questionnaire: The design of the questionnaire is particularly important as it can introduce error through ways such as poorly-worded, confusing questions or not allowing for certain legitimate answers. The design of the questionnaire is conducted very carefully in order to minimise any error. The details of this design are covered below.
2. Mode of data collection: For EU-SILC UK, the mode of data collection for the main stage of interviewing is through face-to-face interviewing. This mode of data collection is chosen because it is deemed to be the best mode for obtaining the most accurate data possible. However, this mode of data collection does allow for the interviewer to possibly influence the answers of the respondent. Steps taken to prevent this are outlined below.
3. Interviewer: It may be possible for an interviewer to systematically influence responses in one way or another. To prevent this, UK interviewers undergo rigorous training and monitoring, the details of which are covered below.
4. Respondent: The respondent error is minimised by careful design of the questionnaire (e.g., collecting detailed information about individual income sources in order to construct a total) and interviewer training. See below for details.

Where possible the UK uses questions that have been harmonised across the national social surveys, such as the Labour Force Survey and Household Budget Survey. This is a thorough process whereby representatives from the different surveys and topic experts agree the harmonised standards. Any changes are thoroughly tested via peer review and pilot testing. Cognitive methods are also applied where deemed necessary.

New EU-SILC specific questions are developed by research staff and data collection methodologists and are often tested via a pilot exercise that takes place before the EU-SILC field period. The pilot exercise involves an interviewer briefing and debriefing session, where feedback from the interviewers is used to further develop the questions. Where deemed necessary soft and hard checks are programmed within the Computer Assisted Personal Interviewing (CAPI) data collection instrument to identify invalid responses and data inconsistencies. The existing questionnaire is checked by research and field teams on a monthly basis to ensure that it is functioning correctly. Each year the guidance provided by Eurostat is consulted and compared to the questionnaire to confirm that the correct questions are being asked and the relevant data are being collected.

Interviewers who w are recruited only a careful selection pr take part in an initia this initial training c are trained to ensu interviewing proces themselves in a ma systematically influ the respondents. N interviewers attend sessions on all asp Furthermore new r supervised by bein field by a field man who continue to wc observed regularly continuously provic

5.3.3 Non response error

Non-response errors are errors due to an unsuccessful attempt to obtain the desired information from an eligible unit. Two ma

1) Unit non-response which refers to the absence of information of the whole units (households and/or persons) selected in Regulation 28/2004:

- **Household non-response rates** (*NRh*) is computed as follows:

NRh=(1-(Ra * Rh)) * 100

Where Ra is the address contact rate defined as:

Ra= Number of address successfully contacted/Number of valid addresses selected

and *Rh* is the proportion of complete household interviews accepted for the database

Rh=Number of household interviews completed and accepted for database/Number of eligible households at contac

- **Individual non-response rates** (*NRp*) will be computed as follows:

NRp=(1-(Rp)) * 100

Where *Rp* is the proportion of complete personal interviews within the households accepted for the database

Rp= Number of personal interview completed/Number of eligible individuals in the households whose interviews we

- **Overall individual non-response rates** (*NRp) will be computed as follows:

*NRp=(1-(Ra * Rh * Rp)) * 100

For those Members States where a sample of persons rather than a sample of households (addresses) was selected, the inc 'the selected respondent', for all individuals aged 16 years or older and for the non-selected respondent.

2) Item non-response which refers to the situation where a sample unit has been successfully enumerated, but not all the re

In strictly controlled circumstances, interviewers are allowed to conduct a proxy interview with a close household member to r are only used where it has proved impossible, despite repeated calls, to contact a particular member of a household in perso

For the 2011 survey income components were asked by proxy if it was not possible to obtain an interview with the responden When it has not proved possible to collect income data it has been imputed.

5.3.3.1 Unit non-response - rate

Cross sectional data

	Address contact rate (Ra)*		Complete household interviews (Rh)*		Complete personal interviews (Rp)*		Household Non-response rate (NRh)*		Individ non-respor (NRp
	A*	B*	A*	B*	A*	B*	A*	B*	A*	
	99.7	99.9	74.3	62.2	100	100	27.3	37.9	0	0

* All the formulas are defined in the Commission Regulation 28/2004, Annex II

A* = Total sample; B = * New sub-sample

5.3.3.2 Item non-response - rate

The computation of item non-response is essential to fulfil the precision requirements concerning publication as stated in the non-response rate is provided for the main income variables both at household and personal level.

For EU-SILC UK, non-responses to questions that feed into income components at personal and household levels are impute the imputation strategy for EU-SILC UK. The tables in section 5.3.3.2.1 provide information on the percentages of non-respon personal and household levels.

5.3.3.2.1 Item non-response rate by indicator

	Total hh gross income (HY010)	Total disposable hh income (HY020)	Total disposable hh income before social than old-age and survivors ber (HY022)
% of household having received an amount	99.7	100.0	97.0
% of household with missing values (before imputation)	1.3	0.5	0.9
% of household with partial information (before imputation)	37.1	40.2	26.5

	Imputed rent (HY030)	Income from rental of property or land (HY040)	Family/ Children related allowances (HY050)	Social exclusion payments not elsewhere classified (HY060)	Housing allowances (HY070)	Regu cash re (H
% of households having received an amount	58.0	5.7	28.9	10.0	13.7	3.1

	Imputed rent (HY030)	Income from rental of property or land (HY040)	Family/ Children related allowances (HY050)	Social exclusion payments not elsewhere classified (HY060)	Housing allowances (HY070)	Regu cash re (H
% of household with missing values (before imputation)	100.0	0.7	0.9	2.8	1.6	0.1
% of household with partial information (before imputation)	0.0	0.2	8.1	1.2	0.0	0.0

	Cash or near-cash employee income (PY010)	Other non-cash employee income (PY020)	Income from private use of company car (PY021)	Employers social insurance contributions (PY030)	Cash profits or losses from self-employment (PY050)	Value of goods produced for own consumption (PY070)	Unemployment benefits (PY090)	Old-a benef (PY10
% of **households** having received an amount	55.1	5.6	5.6	51.4	12.8	N/A	3.2	46.3
% of household with missing values (before imputation)	5.0	0.1	0.1	0.0	1.7	N/A	0.1	0.6
% of household with partial information (before imputation)	2.4	0.3	0.3	0.0	0.0	N/A	0.0	7.4

5.3.4 Processing error

Data entry and coding

Editing controls

Data collection is carried out by face-to-face interviewers using Computer Assisted Personal Interviewing (CAPI) on laptop computers. Blaise software (developed by Statistics Netherlands) is used, which is an integrated system for survey processing. The use of Blaise enables a reduction in processing errors as data can be "checked" as it is entered by interviewers. One example of this is that all income data is "checked" at the point of collection to make sure that net values are not greater than gross values for an individual. Another example relates to the receipt of child benefit payments, where if a respondent claims to receive such a benefit and enters an amount which differs to the known amount that someone can receive then this is highlighted for the interviewer to ensure the right answer has been recorded.

Data are converted from Blaise toSP the data are extensively checked for performed. Very high and low values deemed to be erroneous. Period cod consistent with the amounts provided performed to ensure that respondent should not on the basis of their answ edits are made. The editing process of cases, which is due to in part to th built in to the questionnaire and the E

5.3.4.1 Imputation - rate

Not requested by Reg. 28/2004

5.3.4.2 Common units - proportion

Not requested by Reg. 28/2004

5.3.5 Model assumption error

Not requested by Reg. 28/2004

5.3.6 Data revision

Not requested by Reg. 28/2004

5.3.6.1 Data revision - policy

Not requested by Reg. 28/2004

5.3.6.2 Data revision - practice

Not requested by Reg. 28/2004

5.3.6.3 Data revision - average size

Not requested by Reg. 28/2004

5.3.7 Seasonal adjustment	Not requested by Reg. 28/2004
6 Timeliness and punctuality	Not requested by Reg. 28/2004
6.1 Timeliness	Not requested by Reg. 28/2004
6.1.1 Time lag - first result	Not requested by Reg. 28/2004
6.1.2 Time lag - final result	Not requested by Reg. 28/2004
6.2 Punctuality	Not requested by Reg. 28/2004
6.2.1 Punctuality - delivery and publication	Not requested by Reg. 28/2004
7 Accessibility and clarity	Not requested by Reg. 28/2004
7.1 Dissemination format - News release	Not requested by Reg. 28/2004
7.2 Dissemination format - Publications	Not requested by Reg. 28/2004
7.3 Dissemination format - online database	Not requested by Reg. 28/2004
7.3.1 Data tables - consultations	Not requested by Reg. 28/2004
7.4 Dissemination format - microdata access	Not requested by Reg. 28/2004
7.5 Documentation on methodology	Not requested by Reg. 28/2004
7.5.1 Metadata completeness - rate	Not requested by Reg. 28/2004
7.5.2 Metadata - consultations	Not requested by Reg. 28/2004
7.6 Quality management - documentation	Not requested by Reg. 28/2004
7.7 Dissemination format - other	Not requested by Reg. 28/2004
8 Comparability	
8.1 Comparability - geographical	Not requested by Reg. 28/2004
8.1.1 Asymmetry for mirror flow statistics - coefficient	Not requested by Reg. 28/2004

8.1.2 Reference population	Reference population	Private household definition	
	All private households and their current members residing in the UK at the time of data collection. This is comparable with the EU-SILC definition.	In 2011, the definition for what constitutes a household was changed from: "one person living alone or a group of people (not necessarily related) living at the same address with common housekeeping – that is sharing either a living room or sitting room or at least one meal a day" to: "one person living alone or a group of people (not necessarily related) living at the same address who share cooking facilities and share a living room or sitting room or dining area" This reflected a change in the household definition used for the 2011 UK Census. The national definition differs slightly from the Eurostat definition. The UK definition does not state that a group of people living together in a dwelling must share expenditures to be classified as a single household.	A person is in general reg informant) considers the a however, certain rules wh Children of any age away under 16 at boarding sch From 2008 students who residents of the househol the interview. However, c home for the purposes of holidays are not included Anyone who has been aw longer is excluded. Anyone who has been liv longer is included even if Addresses used only as s residence.

8.1.3 Reference Period	Period for taxes on income and social insurance contributions	Income reference periods used	R(

| | See section on income reference periods used. | EU-SILC UK, like other official income surveys in the UK, uses continuous interviewing with interviews spread evenly throughout the year. The survey measures current income. For example, with regards to income from earnings and benefits, respondents will provide figures which relate most commonly to the last week, two weeks or month. With earnings in particular, respondents are asked for usual earnings. These figures, which represent current (and usual) incomes, are then annualised (weekly estimates multiplied by 52, monthly by 12 etc). Income from self-employment can be reported for a variety of periods, but it is always up-rated (using the UK's average earnings index) to the interview date. For income from investment and employee non-cash income respondents are most likely to provide their most recent annual or half-yearly income that they received from this source. This income would be annualised, although there is no up-rating. | |

This approach is adopted in the UK because it is much easier for respondents to provide estimates of current income, rather than income for a specific reference period, say the most recent financial year. In the UK, only a relatively small proportion of the adult population fill in tax returns, and the rest of the population probably never actually calculate what their annual income is. For this reason, it would be very difficult to collect an estimate of annual income corresponding to a fixed reference year.

So the estimates of income do not correspond strictly to an income reference year. However, we can regard each household's estimate of annualised current income as corresponding to a 12 month period centred around the interview date. So for a household interviewed in early January 2011, we can regard their income as being measured for the period July 2010 to June 2011, and similarly for a household interviewed in December 2011, the income estimate can be regarded as referring to the period July 2011 to June 2012. Since interviews are spread evenly throughout the year, for any one survey year, the interview reference periods collectively are centred around the calendar year. Therefore it is reasonable to regard aggregate statistics produced from the full annual datasets as measuring annual income in the current survey year. So the EU-SILC UK 2011 survey measures current annual income in 2011.

In the UK, household income statistics, and especially aggregate statistics such as those that are produced from EU-SILC, are generally used and interpreted on the assumption that this distinction between annualised current income, and what might be called a 'true' annual income, is small (Bönheim and Jenkins, 2006).

Bönheim, R. and Jenkins, S.P. 2006. A comparison of current and annual measures of income in the British Household Panel Survey. *Journal of Official Statistics* **22** (4), 733-758.

8.1.4 Statistical concepts and definitions	Total hh gross income (HY010)	Total disposable hh income (HY020)	Total disposable hh income before social transfers other than old-age and survivors' benefits (HY022)	
	F	F	F	

Imputed rent (HY030)	Income from rental of property or land (HY040)	Family/ Children related allowances (HY050)	Social exclusion payments not elsewhere classified (HY060)	Housing allowances (HY070)	Regular inter-hh cash transfers received (HY080)	Interest, dividends, profit from capital investments in incorporated businesses (HY090)	Int pa mo (H
F	F	F	F	F	F	F	

Cash or near-cash employee income (PY010)	Other non-cash employee income (PY020)	Income from private use of company car (PY021)	Employers social insurance contributions (PY030)	Cash profits or losses from self-employment (PY050)	Value of goods produced for own consumption (PY070)	Unemployment benefits (PY090)	Old-age benefits (PY100)	Surviv benefi (PY11
F	L – for EU-SILC UK, income from private use of company cars is used to calculate	F	F	F	N/A	L – the only minor difference relates to redundancy payments. For EU-SILC UK, only information on regular redundancy payments	F	F

Cash or near-cash employee income (PY010)	Other non-cash employee income (PY020)	Income from private use of company car (PY021)	Employers social insurance contributions (PY030)	Cash profits or losses from self-employment (PY050)	Value of goods produced for own consumption (PY070)	Unemployment benefits (PY090)	Old-age benefits (PY100)	Survivo benefi (PY11
	PY020. However, none of the other goods and services highlighted in the EU-SILC definition are used. To collect data on these other goods and services would impart a high burden on the respondent and the accuracy of the collected data would likely be poor.				is collected. One-off lump sum redundancy payments are excluded.			

The source or procedure used for the collection of income variables	The form in which income variables at component le have been obtained
All income variables are collected at the point of interview. It is not mandatory for respondents to provide any documentation to support their answers. However, interviewers are being encouraged to ask respondents whether it is possible to consult their payslip (if they are working). No information is collected from registers.	For most income components which are subject to taxation and/ social security contributions, respondents are asked to provide n gross amounts. The only exception to this is income from interes dividends and capital investments, which is collected either gros net, and for which tax paid is then estimated. Total income for an individual/household refers to income at the the interview. If the last pay packet/cheque was unusual, for exa included holiday pay in advance or a tax refund, the respondent asked for usual pay. No account is taken of whether a job is tem or permanent.

8.2 Comparability - over time

DB100 Degree of urbanisation

There are differences in the classification of degree of urbanisation used by EU-SILC UK pre- and post-2010. In 2010, the me updated to match that used by the Labour Force Survey. This new classification method increased the proportion of thinly pop populated areas. In 2010 under the new method, densely populated areas covered 62.7% of the population, intermediate area was not classifiable. Under the old method, densely populated areas covered 74.6% of the population, intermediate areas 16. classifiable.

Highest ISCED level attained (PE040)

Prior to 2010 respondents who replied that they had 'other' qualifications were coded as having post-secondary non-tertiary le so that the 'other' category could not be used. Longitudinal data were used where possible to code cases who answered 'othe and 2013 questionnaires to assign a more accurate ISCED level to respondents answering 'other'.

Variables not asked in error or not asked by proxy

There was an error on the UK questionnaire whereby a number of labour variables were not asked through 2008 and January PL190 and PL200. Additionally, the update to ask these variables by proxy was not implemented on a number of labour variab PL140, PL160, PL190, PL200 and PL210A-K. Also in 2009, the variables HS020 and HS030 were inadvertently omitted from August.

The UK imputation method uses a donor method and therefore the extent of the missing data would make it difficult to predict methodology team aim to impute the arrears variables using two years data either side of the gap (2007-08 and 2010-11). An possible.

PY021 Company car

The method for calculating income earned from having a company car changed in 2010. Prior to 2010, PY021 was calculated capacity obtained by looking up vehicle models on the UK Vehicle Certification Agency. In 2010 the method changed to use the various tax rates applied to different fuel types.

8.2.1 Length of comparable time series

Not requested by Reg. 28/2004

8.3 Comparability - domain

Not requested by Reg. 28/2004

9 Coherence

The coherence of two or more statistical outputs refers to the degree to which the statistical processes, by which they were generated, used the same concepts and harmonised methods. A comparison with external sources for all income target variables and the number of persons who receive income from each 'income component' will be provided, where the Member States concerned consider such external data to be sufficiently reliable.

9.1 Coherence - cross domain

Results from two other survey sources have been used to validate EU-SILC results – the Family Resources Survey and the Living Costs and Food Survey.

Family Resources Survey

The Family Resources Survey (FRS) collects information on the incomes and circumstances of private households in the United Kingdom (or Great Britain before 2002-03). The survey is sponsored by the Department for Work and Pensions.

The FRS is used primarily to validate the indicators of poverty and social exclusion. Before the introduction of EU-SILC, the Laeken and Pensions indicators were produced using data from theFRS. Comparisons between EU-SILC andFRS-based indicators continue so that any apparent differences between national poverty estimates and EU-SILC estimates can be explained. This work will be ongoing, and in the first seven years of EU-SILC has served as a useful way of validating the new EU-SILC data, and highlighting any possible problems that there might be with the EU-SILC data.

For 2011 the figures from EU-SILC and FRS are largely in strong agreement. The overall levels of poverty produced from both sources are within one percentage point of each other. Where differences in the figures from both sources occur they are consistent with the slight differences in the methodologies between EU-SILC and FRS.

Living Costs and Food Survey

The Living Costs and Food Survey (LCF), the UK's Household Budget Survey, is a comprehensive overview of all aspects of household expenditure and income for the year 2011 derived from a survey of around 5,000 households in the UK. It contains analyses of household expenditure on goods and services by household income, composition, size, type and location. The results are widely seen as providing one of the most accurate pictures available of what households in the UK spend their money on today.

EU-SILC income variables have been compared with the detailed income information collected through the LCF particularly that which is published in the ONS report 'The Effects of Taxes and Benefits on Household Income'. The LCF is an extremely useful source for comparing figures from EU-SILC on income and household expenditure, especially relating to mortgages, council tax, rent and utility bills.

9.1.1 Coherence - sub annual and annual statistics

Not requested by Reg. 28/2004

9.1.2 Coherence - National Accounts

Not currently available.

9.2 Coherence - internal

Not requested by Reg. 28/2004

10 Cost and Burden

Not requested by Reg. 28/2004

11 Confidentiality

Not requested by Reg. 28/2004

11.1 Confidentiality - policy

Not requested by Reg. 28/2004

11.2 Confidentiality - data treatment

Not requested by Reg. 28/2004

12 Statistical processing

Detailed information concerning sampling frame, sampling design, sampling units, sampling size, weightings and mode of data collection can be found in this section. Such information is mainly used for the computation of the accuracy measures.

12.1 Source data

Data for EU-SILC UK 2011 are collected from two sources. First, data for Great Britain are collected by the ONS, using the GLF. Second, to ensure that EU-SILC is representative of the UK, a sample of approximately 200 households is selected by NISRA (Northern Ireland Statistics and Research Agency) using the LCS. This small additional sample represents the (approximately) 2% of the UK population that live in Northern Ireland. All of the data analysis and processing is undertaken by ONS. The sampling frame used to select households is the Postcode Address File which is maintained by the UK Post Office. This sample frame contains all the known postal addresses in the UK. The Postcode Address File is updated four times a year by the Post Office and updated twice a year on the sampling system of the ONS. Addresses from the Isles of Scilly and north of the Caledonian Canal in Scotland are excluded from the sample frame because of the excessive costs that would be involved in sending interviews to these areas. These parts of the country comprise less than 2% of the UK population, which complies with the limit of population exclusion allowed for in EU-SILC.

12.1.1 Sampling design and procedure

Type of sampling design

The EU-SILC Great Britain 2011 survey is based on a probability stratified two-stage sampling design. The sample design in Northern Ireland is a simple random sample. Stratification of the postcode sectors is done by geographical criteria for Great Britain.

In 2011, 11,857 addresses were sampled. Each year approximately 70% of the sample is rolled forward from previous years and the remaining 30% is a new "Wave 1" sample. EU-SILC UK aims to interview all adults aged 16 or over at every household at the sampled address.

Stratification and sub stratification criteria

Stratification involves the division of the population into sub-groups, or strata, from which independent samples are taken. This ensures that a representative sample is drawn with respect to the stratifiers. Stratification of a sample can lead to substantial improvements in the precision of the survey estimators provided that the strata are chosen such that members of the same strata are as similar as possible in respect to the characteristics of interest. The bigger the differences between strata, the greater the gain in the precision of the survey estimates.

Initially, postcode sectors for Great Britain were allocated to 30 major strata. These were based on the 10 regions in England (subdivided between the former Metropolitan and non-Metropolitan counties), five subdivisions in Scotland, two in Wales and one in Northern Ireland (Annex 2). In addition, London was subdivided into quadrants (Northwest, Northeast, Southwest and Southeast) with each quadrant being divided into inner and outer areas (Annex 2). Using a finer division of London significantly improves the precision of estimates.

It should be noted that regions and strata do not exactly map onto each other. There are 30 strata in Great Britain but 37 regions. Some strata contain cases from two or more regions and some regions contribute cases to more than one stratum.

Within each major stratum, postcode sectors were then stratified according to selected indicators taken from the 2001 Census. Sectors were initially ranked according to the proportion of households with no car, then divided into three bands containing approximately the same number of households. Within each band, sectors were re-ranked according to the proportion of households with a household reference person in socio-economic groups 1 to 5 and 13 (Annex 3), and these bands were then subdivided into three further bands of approximately equal size. Finally, within each of these bands, sectors were re-ranked according to the proportion of people who were pensioners. The ranking by pensioners and socio-economic group is carried out in reverse order so as to maximise similarity between one band and the next. A systematic sample of postcode sectors (PSUs) is selected from the ordered frame resulting in an implicit stratification of the sample. PSUs were then paired up to form pseudo-minor strata. The implicit stratification of the sample makes it possible to increase the precision of the survey estimates while ensuring good geographical coverage. It is just the major strata that are provided in the microdata D file.

Major strata were then divided into minor strata with equal numbers of addresses, the number of minor strata per major strata being proportionate to the size of the major stratum, so larger PSUs have more chance of being selected. In 2011 there were 912 minor strata. These consisted of 684 strata of waves 1, 2 and 3 from 2010 and 228 new wave 1 strata which replaced the 228 wave 4 strata from 2010.

Each PSU formed a quota of work for an interviewer. Within each of the 228 new PSUs, 23 addresses were randomly selected.

Sample selection schemes

EU-SILC Great Britain uses a two-stage sampling scheme:

1. Selection of Primary Sampling Units (PSUs) utilising a probability proportional to size sampling scheme.
2. Systematic random sampling of 23 addresses within aPSU.

The sample design in Northern Ireland is a simple random sample.

Sample distribution over time

See section 12.2.

12.1.2 Sampling unit

The sample frame, the Postcode Address File, is ordered by postcode sectors. The postcode sectors are the Primary Sampling Units (PSU-1) for EU-SILC and the Secondary Sampling Units (PSU-2) are addresses within those sectors. Further information on the sampling unit is given in section 12.1.1.

12.1.3 Sampling rate and sampling size

Concerning the SILC instrument, three different sample size definitions can be applied:

- the actual sample size which is the number of sampling units selected in the sample

- the achieved sample size which is the number of observed sampling units (household or individual) with an accepted interview

- the effective sample size which is defined as the achieved sample size divided by the design effect with regards to the at-risk-of poverty rate indicator

Given that the effective sample size has been already treated in the section dealing with sampling errors, in this section the attention focuses mainly on the achieved sample size.

Achieved sample size for the cross-sectional data		
	No of households	No of persons 16+
1st rotational group	8,058	15,134

12.2 Frequency of data collection

Household interviews for EU-SILC UK are spread evenly throughout the calendar year. Typically a small number of interviews will be completed in January and February of the following year. The distribution of interviews for 2011 EU-SILC UK is shown below.

Distribution of the EU-SILC UK sample over time with information based on data presented in the household data file.

Date of interview	Number of households
01/01/11 – 31/01/11	547
01/02/11 – 28/02/11	637
01/03/11 – 31/03/11	709
01/04/11 – 30/04/11	634
01/05/11 – 31/05/11	737
01/06/11 – 30/06/11	704
01/07/11 – 31/07/11	713
01/08/11 – 31/08/11	672
01/09/11 – 30/09/11	654
01/10/11 – 31/10/11	706
01/11/11 – 30/11/11	788
01/12/11 – 31/12/11	499
01/01/12 – 31/01/12	56
01/02/12 – 28/02/12	2
Total	8,058

Renewal of sample: rotational groups

In the UK, 2005 was the initial year for the EU-SILC survey. In 2005, the EU-SILC survey adopted a new sample design in line with EU-SILC requirements, changing from a cross-sectional to a longitudinal design.

The sample design follows a four-yearly sample rotation in which households remain in the sample for four years (waves) and one quarter of the sample is replaced each year. Each quarter of the sample is known as a replication. Rotational groups and the renewal of the sample for EU-SILC UK are displayed below.

Renewal of sample: rotational groups

Sample replication	Year 1 (2005)	Year 2 (2006)	Year 3 (2007)	Year 4 (2008)	Year 5 (2009)	Year 6 (2010)	Year 7 (2011)
1	1st						
2	1st	2nd					
3	1st	2nd	3rd				
4	1st	2nd	3rd	4th			

5			1st	2nd	3rd	4th		
6				1st	2nd	3rd	4th	
7					1st	2nd	3rd	4th
8						1st	2nd	3rd
9							1st	2nd
10								1st

12.3 Data collection

Mode of data collection

The survey was carried out using Computer Assisted Personal Interviewing (CAPI) on laptop computers by face-to-face interviewers. In addition, some telephone interviewers were used to convert EU-SILC UK proxy interviews to full interviews. However, all household interviews are completed by CAPI.

1-PAPI (% of total)	2-CAPI (% of total)	3-CATI (% of total)	4-Self administrated (% of total)
0	100	0	0

The mean interview duration

The mean interview duration per household is calculated as the sum of the duration of all household interviews plus the sum of the duration of all personal interviews, divided by the number of household questionnaires completed. Only households accepted for the database have to be considered.

The variables PB120 (length of personal interview) and HB100 (length of household interview) cannot be used to calculate the mean interview duration per household. The GLF is the survey used to collect data for EU-SILC UK. However, the GLF is not used for the sole purpose of EU-SILC and involves a number of questions unrelated to EU-SILC. The lengths of interviews recorded refer to the whole GLF and not just the EU-SILC questions. Nonetheless, it is estimated that the EU-SILC component of the GLF takes, on average, 44 minutes for the household to complete.

Average interview duration =44 mins.

12.4 Data validation

Not requested by Reg.

12.5 Data compilation

Not requested by Reg.

12.5.1 Weighting procedure

Design factor	Non-response adjustments	Adjustment to external data	Final cross sectional weights
Addresses are selected for the first wave of each panel using a random probability design, the detail of which is outlined in the preceding sections of this report. The design weight for a household is calculated as the inverse of the inclusion probability for the samples address (e.g., a standard Horvitz-Thompson (HT) estimator). The HT estimator is then adjusted by a two-step procedure to produce the wave 1 cross-sectional weight.	Non-response to the surveys (GLF and LCS) used to produce the EU-SILC data can introduce bias into the estimator. For the UK data, an attempt is made to correct for this bias through weighting households based on their estimated propensity to respond. For EU-SILC, non-response can occur at any given wave.

A non-response model exists for the GLF which comprises a number of adjustment classes. These classes were constructed by linking households selected for the 2001 General Household Survey (the earlier version of the GLF) to the 2001 Census. The Census is mandatory in the UK and so both responders and non-responders to the GLF can be matched to Census records. Response classes were formed based on households' propensity to respond to the survey, condition on certain combinations of characteristics available in both | None. | Calibration is used in the weighting procedure both to improve precision and to ensure consistency with known population totals. The EU-SILC sample is based on the population of private households, which means that the population totals used in the weighting need to be those created from counts of people living in private households.

At the time the weights were being constructed the most appropriate version of the population totals available for weighting were those produced for the Labour Force Survey (LFS). The LFS derives household population estimates by excluding residents of institutions from population projections based on mid-year estimates. However, certain groups in institutions are included in the population totals (e.g., nurses in nursing homes).

The population information and EU-SILC UK data were grouped into twelve age by sex categories and into six regional categories to form weighting classes. The initial non-response adjusted HT weight is adjusted, using Stats Canada's Generalized Estimation System (GES), so that the final weights ensure that the weighted totals for the above demographic categories match the population totals. |

Design factor	Non-response adjustments	Adjustment to external data	Final cross sectional weights

Age-group by sex

0-4 Males and Females	0-4
5-15	5-15 Males and Females
16-24 Males 16-24	Females
25-44 Males 25-44	Females
45-64 Males 45-64	Females
65-74 Males 65-74	Females
75+ Males 75+	Females

the Census and the survey. The reciprocal of the response propensity is used as the non-response weight.

Regions

Metropolitan

Non-metropolitan

London

South East

Wales

Scotland

Northern Ireland

12.5.2 Estimation and imputation	Imputation procedure used	Imputed rent	Company car

The strategy used to impute UK EU-SILC was consistent with the options proposed in the following Eurostat task-force documents associated with donor-based imputation methodology:

EU-SILC 74/02

EU-SILC 136/04

EU-SILC 154/05

The UK EU-SILC Imputation Strategy was developed with the primary

A UK EU-SILC imputed rent variable was supplied for 2011. Estimates of imputed rent were generated through the use of hedonic regression modelling, using the Heckman Two-Step method. The explanatory variables used in the regression were region, type of dwelling (flat, semidetached/terraced house, detached house), ownership of a car, value of dwelling (council tax band, except Northern

In the UK, company cars are taxed based on their carbon dioxide (CO_2) emissions. Therefore, UK EU-SILC assigns the benefit of having access to a company car as being equal to the level of tax. However, it is difficult to estimate the level of tax and therefore the following method is used.

EU-SILC UK asks several questions about company cars. First, the survey establishes whether the household has any company cars. Second, it establishes what the manufacturer's list price for the vehicle was when it was new. If the respondent is unable to provide an answer, they are asked which price band they think the company car sits in. If the respondent gives a band price the answer is translated into a mid-point price. For example, a Mazda saloon with a band price between £10,001-£13,000 would be given a list price of £11,500.

The estimation of the value of using a company car for private purposes (excluding payment of fuel) is done using the following elements:

    1. Type of fuel used;

    2. Data from VCA (Vehicle Certification Agency, UK);

Imputation procedure used	Imputed rent	Company car

aims of imputing for all item level missingness, resolving inconsistencies, and preserving both cross-sectional and longitudinal relationships in the responses for the households and persons affected. The strategy was also designed to preserve the maximum amount of observed data.

Meeting the aims of the strategy was not trivial as the cross-sectional and longitudinal correlations were both nested and complex. In any one year, the UK EU-SILC dataset contains over 400 routing and income variables; routing variables indicate whether or not the respondent receives an amount whilst the amount itself follows on in one or more consecutive variables. Missing values may be present in both the routing and the amounts collected.

Further complications include:

- legal constraints which make some combinations of the routing variables invalid;
- highly correlated relationships amongst subsets of the variables, for example, earnings before and after taxation followed by an associated time period for which the payment relates;
- the panel aspect of the survey introduces further correlations between years in addition to those within year.

To meet the aims of the imputation strategy the ONS implemented an iterative, two-stage imputation process: Stage 1 focused on the imputation of missing routing; Stage 2 focused on the imputation

Imputed rent column:

Ireland), thermal comfort (ability to keep home adequately warm) and seniority (year of contract). The Heckman Two-Step procedure requires the dependent variable, in this case rent, to be converted to a log linear variable. Hence, predicted imputed rent was estimated as log linear variable. A back-log transformation was done to produce imputed rent in its proper form.

Company car column:

3. Price of the car.

Once the price of the car is known (using one of the methods described above) a factor based on fuel type and emissions of the engine is applied to that list price. However, this is problematic. Although data on the make and model of each car is collected, the quality of answers given by respondents is extremely variable, for instance, answers such as 'a red Ford' offer little value to a calculation. Therefore the estimates are based on average tax bands for cars of certain price bands.

The factors used for 2011 are shown in below. Diesel cars have a 3% uplift from petrol cars, which reflect the extra tax charged for these vehicles. Electric vehicles have no tax charged on them whilst hybrid vehicles and vehicles which run on fuel consisting of up to 85% denatured ethanol (E85) incur a tax rate of 15% regardless of the list price of the vehicle.

Tax rate based on CO_2 emission rates (%)

Fuel	Car price (£)	CO_2 tax emission rate (%)
Petrol	0 – 18,999	15
Petrol	19,000 – 39,999	26
Petrol	40,000 – 99,999	35
Diesel	0 – 18,999	18
Diesel	19,000 – 39,999	29
Diesel	40,000 – 99,999	35
Electric	Any	0
Hybrid/E85/Other	Any	15

These percentage rates are the factors that are applied to the car prices to produce a monetary benefit for each company car in a household.

Car benefit = Car price $\times$ CO_2 tax emission rate

Imputation procedure used	Imputed rent	Company car

of missing amounts and time periods.

The imputation process was supported by statistical tools and used standard statistical techniques for panel data, including:

- SAS (Statistical Analysis System) – to facilitate deductive imputation. This was applied to correct for missing values by implementing propositional relationships in the data based on logical rules and legal constraints. For example, using gross values with auxiliary variables to derive missing net values;
- SPSSAnswerTree - to identify key predictors to partition the data into homogeneous classes for subsequent imputation;
- CANCEIS (CANadian Census Edit and Imputation System) - for stochastic imputation. CANCEIS implements a highly efficient nearest neighbour imputation method that preserves the shape of the distribution whilst also estimates and maintains observed relationships and distributional parameters. Stochastic imputation ensures less

Imputation procedure used	Imputed rent	Company car
distortion in the estimates of variance. Asymmetric trimming was also applied as a refinement to exclude outlying values which might have otherwise caused excessive influence.		

The quality of the final data was validated in two ways: by calculating expected values and observing the pre-and post-imputation distributions.

12.6 Adjustment   Not requested by Reg.

(This item can be filled in on voluntary basis)

13 Comment   There have been two versions of the 2011 cross-sectional EU-SILC UK data sent to Eurostat as listed below.

Version control		
Version	**Date delivered to Eurostat**	**Revision summary**
V1	9/11/2012	Provisional data sent to be included in EU aggregate data.
V2	20/11/2012	First full sent after final quality assurance with some minor adjustments made to V1.

National questionnaire is available at:

http://epp.eurostat.ec.europa.eu/portal/page/portal /income_social_inclusion_living_conditions/quality/national_quality_reports

Annexes	**Description**
Annex 1.docx	Annex 1-Ad-hoc Module variables
Annex 2.docx	Annex 2-Regional stratifier
Annex 3.docx	Annex 3-Socio-economic groups