EU statistics on income and living conditions (EU-SILC) methodology – sampling

Statistics Explained

This article has been archived.

This article is part of the Eurostat online publication EU statistics on income and living conditions (EU-SILC) methodology .

EU-SILC is a sample survey. The legislation specifies that data shall be based on nationally representative probability samples and prescribes minimum effective sample sizes, but leaves to the country the choice of a specific sampling design. This article describes the main characteristics of the sampling: sampling frame, sample design and sample size. It contains also information on the sampling errors. Finally tracing rules are explained.

Sampling frame

The big strength of EU-SILC is the usage of the best sampling frames available in each National Statistical Institute (NSI). According to the EU-SILC Framework Regulation, data are to be based on a nationally representative probability sample of the population residing in private households within the country, irrespective of language, nationality or legal residence status. All private households and all persons aged 16 and over within the household are eligible for the operation. Persons living in collective households and in institutions are generally excluded from the target population. The sampling frame as well as methods of sample selection should ensure that every individual and household in the target population is assigned a known probability of selection that is not zero. As shown in Table 1, the vast majority of countries used
for the 2013 EU-SILC operation population registers, or national census or a master sample derived from the census.

	Source name	Update
Belgium	Central Population Register	Twice a month
Bulgaria	2011 Population Census and National Civil Registration System	Regularly
Czech Republic	Census Enumeration Districts (CEUs)	Continuously
Denmark	Central Population Register	Continuously
Germany	Subsample of the German microcensus (DSP)	Continuously
Estonia	2011 Population and Housing Census	:
Ireland		
Greece	Population Census	Every ten years
Spain	Municipal Register	Continuously
France	Population Census and Survey of new dwellings (Base de sondage de logements neufs - BSLN) plus OCTOPUSSE (from 2010)	Last update 2005 (BSLN) and 1999 (Census)
Italy	Registers of the municipalities	Continuously
Croatia	2001 Population Census	2001
Cyprus	2011 Census of Population	Last update 1.10.2011 (Census)
Latvia	Population Census 2000 and Population register	Regularly
Lithuania	Residents register	Regularly
Luxembourg	Luxembourg Social Security database (IGSS)	Continuously
Hungary	Updated dataset of addresses used in the 2001 population and housing census	Last update 2004
Malta	Census of Population & Housing 2011 database	Annually
Netherlands	Population register	Monthly
Austria	Population register (ZMR)	Continuously
Poland	The Domestic Territorial Division Register (TERYT)	Annually
Portugal	Master Sample based on 2001 Census of Population and Housing and Natonal Dwellings Register (NDR) based on 2011 Census	Non periodic
Romania	Multifunctional Sample of Territorial Areas (EMZOT)	Last update 2007
Slovenia	Central Register of Population (CRP) and enumeration areas	:
Slovakia**	2001 Population and Housing Census	:
Finland	The Population Information System (PIS)	Continuously
Sweden**	Total Population Register of Sweden (TRP)	Daily
United Kingdom	Postcode address file (PAF)	Regularly
Iceland	The Population register	Continuously
Norway*	Copy of the central population register (BEREG)	Daily
Switzerland	Register of phone number (CASTEM)	Quarterly
Serbia		

* Information from 2012; 2013 not available
** Information from 2011; 2013 and 2012 not available

Table 1: Source of the sampling frame (2013) Source: National Quality Reports 2013

Coverage errors

Coverage errors are caused by the imperfections of a sampling frame for the target population of the survey. The target population is the set of elements for which estimates are desired while the frame population is composed of the units which are eligible for inclusion through a given sampling procedure. Ideally, there must be a one-to-one relation between target and frame population elements. If not, the following frame imperfections can be encountered:

- over-coverage which relates either to wrongly classified units that are in fact out of the scope, or to units that do not exist in practice;
- under-coverage which refers to units not included in the sampling frame;
- misclassification which refers to incorrect classification of units that belong to the target population.

Detailed information about coverage problems and errors for each country when such kind of information is available can be found in the national quality reports.

Sampling design

The sample design describes all the steps to be carried out when selecting a sample of households or persons. It aims to improve the quality of the estimates produced and to control costs. Various strategies are in place in different countries to achieve this objective. The table 2 summarizes the sampling design used in each country for the 2013 operation. Countries choose a specific sampling design according to the structure of the country and the population, according to existing information and taking into account budgetary constraints. The most used sampling design is stratified multistage sampling. Only five countries do not use stratification criteria to draw their sample. In details: Malta, Denmark, Island and Norway use a simple random sample design and Sweden uses a systematic sample. Concerning all the remaining countries, they apply one or more stratification criteria, mainly a geographical stratification. Among them, the majority uses a multi-stage sampling with the exception of Luxembourg, Germany, Cyprus, Slovakia, Switzerland, Austria and Lithuania which use a stratified simple random sample. Estonia uses a systematic stratified sample and Hungary is the only country to apply a different sampling design for drawing each rotational group. Countries send every year to Eurostat general information on the sampling design used and detailed information at the level of micro-data on the strata and PSU from which each household is drawn. The efficiency of the sampling design has a big impact on standard error and should be monitored over time. On the other side, changing it is extremely costly.

Sampling design	Country
Without stratification	
Simple random sampling	DK, MT, IS, NO
Systematic sampling	SE
With stratification	
Stratified sampling according to different design by rotational group	HU
Stratified simple random sampling	DE*, CY, LT, LU, AT, SK, CH
Stratified and systematic sampling	EE
Stratified two-stage sampling	IT, HR, LV, NL, PT, SI
Stratified multi-stage sampling	BE, BG, CZ, IE, EL, ES, FR, IT, PL, RO, UK
Stratified two-phase sampling	FI

* from former participants of micro census

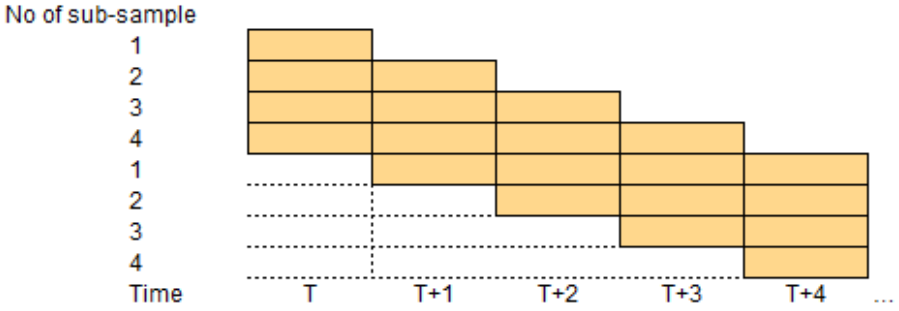
Table 2: Main characteristics of countries' sampling designs Source: National Quality Reports 2013

Integrated design

Although one characteristic of EU-SILC is flexibility in terms of sampling design, Eurostat recommends a rotational design with four sub-samples or replications. All countries adopted for their 2013 operation the four-year rotational design recommended by Eurostat,
with the exception of France and Norway where a longer panel duration (eight and nine years, respectively) was used.

Illustration of the 4-year rotational design

The 4-year rotational design refers to a sample selection based on a number of 4 sub-samples or replications, each of them similar in size and design and representative of the whole population. From one year to the next, three out of four replications are retained, while the first selected one (the "oldest' one) is dropped and replaced by a new replication.



According to this design, the cross-sectional component of EU-SILC operation in year T is composed of four sub-samples each of them drawn in different years: T-3, T-2, T-1 and T.

Sampling unit

Concerning the sampling unit, it can be the address/dwelling, the household or the individual accordingly to the design chosen by the country. In the case of a sample of dwellings /addresses, if more than one household share the same dwelling, dwellings must be regarded as clusters of households. Households are clusters of individuals and all members of the household aged 16 and over at the end of the income reference period are eligible for inclusion in the sample. Countries that carry out a sampling of individuals, instead, only select persons of age 16 and over and the household is defined as the household of which the selected person is a member at the beginning of the survey. As showed in the Table 3, Nordic countries as well as Slovenia and the Netherlands select a sample of individuals while thirteen other countries select a sample of dwellings or addresses and only eleven countries select a sample of households.

Sampling unit	Country
Dwellings/addresses	CZ, DE, ES, FR, HR, LV, LU, MT, AT, PL, PT, RO, UK
Households	BE, BG, EE, IE, EL, IT, CY, LT, HU, SK, CH
Individuals	DK, NL, SI, FI, SE, IS, NO

Table 3: Sampling units

Sample size

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

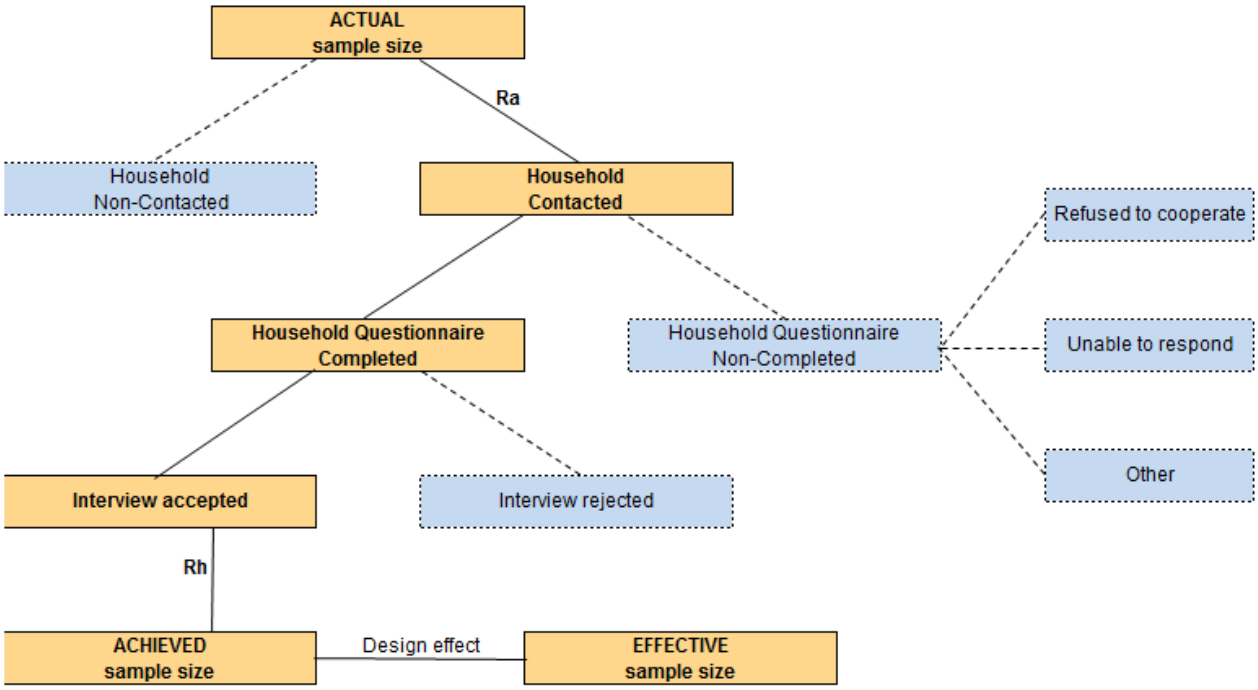
- the actual sample size, that is the number of sampling units selected in the sample;
- the achieved sample size which is the number of observed sampling units (household or individual) with an accepted interview;
- and finally, the effective sample size which is defined as the achieved sample size divided by the design effect.

The Framework Regulation of EU-SILC and its updates define the minimum effective sample size, which is the size that would be required if the survey was based on a simple random sampling (see Table 4). The actual sample sizes have to be larger to the extent that the design effect exceeds 1.0 in order to compensate the loss of effectiveness namely by the use of complex sampling design. The design effect is basically the ratio of the actual variance, under the sampling method actually used, to the variance computed under the assumption of simple random sampling. Below different concepts used when defining the sample size as well as the relation between them are presented.

Countries	Households		Persons aged 16 and over to be interviewed	
	Cross-sectional	Longitudinal	Cross-sectional	Longitudinal
Belgium	4 750	3 500	8 750	6 500
Bulgaria	4 500	3 500	10 000	7 500
Czech Republic	4 750	3 500	10 000	7 500
Denmark	4 250	3 250	7 250	5 500
Germany	8 250	6 000	14 500	10 500
Estonia	3 500	2 750	7 750	5 750
Ireland	3 750	2 750	8 000	6 000
Greece	4 750	3 500	10 000	7 250
Spain	6 500	5 000	16 000	12 250
France	7 250	5 500	13 500	10 250
Croatia	4 250	3 250	9 250	7 000
Italy	7 250	5 500	15 500	11 750
Cyprus	3 250	2 500	7 500	5 500
Latvia	3 750	2 750	7 650	5 600
Lithuania	4 000	3 000	9 000	6 750
Luxembourg	3 250	2 500	6 500	5 000
Hungary	4 750	3 500	10 250	7 750
Malta	3 000	2 250	7 000	5 250
Netherlands	5 000	3 750	8 750	6 500
Austria	4 500	3 250	8 750	6 250
Poland	6 000	4 500	15 000	11 250
Portugal	4 500	3 250	10 500	7 500
Romania	5 250	4 000	12 750	9 500
Slovenia	3 750	2 750	9 000	6 750
Slovakia	4 250	3 250	11 000	8 250
Finland	4 000	3 000	6 750	5 000
Sweden	4 500	3 500	7 500	5 750
United Kingdom	7 500	5 750	13 750	10 500
Total for EU Member States	135 000	101 500	282 150	210 850
Iceland	2 250	1 700	3 750	2 800
Norway	3 750	2 750	6 250	4 650
Switzerland				
Serbia				
Total including non-EU countries	141 000	105 950	292 150	218 300

Table 4: Minimum effective sample sizes for countries

Illustration of different concepts used in the definition of the sample size



Sampling errors

Given the high policy relevance of EU-SILC there is increasing demand from the stakeholders for accuracy measures of the published indicators and for measures of the significance of net change of indicators over time for correct monitoring of the evolution of social exclusion phenomena. As seen, EU-SILC is a complex survey involving different sampling design in different countries. For this reason, "to the book" standard methods for calculating accuracy measures are not directly applicable. Eurostat with the substantial contribution of Net-SILC2 has put in place a simple method for standard error estimation based on linearization and coupled with the ultimate cluster approach. Linearization is a technique based on the use of linear approximation to reduce non-linear statistics to a linear form, justified by asymptotic properties of the estimator. This technique can encompass a wide variety of indicators, including EU-SILC indicators. The "ultimate cluster" approach is a simplification consisting in calculating the variance taking into account only variation among Primary Sampling Unit (PSU) totals. This method requires first stage sampling fractions to be small which is nearly always the case. This method allows a great flexibility and simplifies the calculations of variances. It can also be generalized to calculate variance of the differences of one year to another.

For further details on this method for standard error estimation, please consult the working paper Standard error estimation for the EU-SILC indicators .

Sampling error calculations for main EU-SILC indicators

The method for estimating the standard error described above has been already applied for many EU-SILC indicators. In order to present how it can be implemented in practise, the example of its application for the indicator AROPE (At-risk-of poverty or social exclusion) is described. This indicator is the proportion of persons being in one or more of the three following situations: at-risk-of poverty, i.e. below the national poverty threshold (60% of median national equivalized income), severely materially deprived, living in a household with very low work intensity. This indicator has been considered as a proportion making the assumption that the poverty threshold is a fixed amount and equal to the point estimate. According to the characteristics and availability of data for different countries different variables have been used to specify strata and cluster information. In particular, countries have been split into three groups:

1. Belgium, Bulgaria, Czech Republic, Ireland, Greece, Spain, France, Croatia, Italy, Latvia, Hungary, the Netherlands, Poland, Portugal, Romania, Slovenia and the United Kingdom whose sampling design could be

assimilated to a two stage stratified type we used DB050 (primary strata) for strata specification and DB060 (Primary Sampling Unit) for cluster specification;

2. Germany, Estonia, Cyprus, Lithuania, Luxembourg, Austria, Slovakia, Finland, Switzerland whose sampling design could be assimilated to a one stage stratified type we used DB050 for strata specification and DB030 (household ID) for cluster specification;

3. Denmark, Malta, Sweden, Iceland, Norway, whose sampling design could be assimilated to a simple random sampling, we used DB030 for cluster specification and no strata.

The approach used can take account of stratification, multi-stage selection, unequal probabilities of inclusion for the sample units and re-weighting for unit non-response. However it does not reflect the gain in accuracy caused by calibration weighting. The effect of calibration on variance could be significant especially in the countries where powerful auxiliary information from income registers has been used to adjust the sampling weights. This in some cases may lead to overestimation of sampling errors. Results are shown in Table 5.

	Percent	StdErr	CI95%LB	CI95%UB
EU-28	24.5	0.2	24.2	24.9
Belgium	20.8	0.9	19.2	22.5
Bulgaria	48.0	1.1	45.8	50.2
Czech Rep.	14.6	0.6	13.4	15.8
Denmark	18.9	1.0	17.0	20.9
Germany	20.3	0.4	19.7	21.0
Estonia	23.5	0.7	22.2	24.9
Ireland	29.5	0.9	27.8	31.3
Greece	35.7	0.9	34.0	37.4
Spain	27.3	0.6	26.0	28.6
France	18.1	0.5	17.1	19.1
Croatia	29.9	1.0	27.9	31.9
Italy	28.4	0.5	27.4	29.5
Cyprus	27.8	0.8	26.2	29.4
Latvia	35.1	1.1	32.8	37.5
Lithuania	30.8	1.1	28.8	32.9
Luxembourg	19.0	0.9	17.2	20.8
Hungary	33.5	0.9	31.8	35.3
Malta	24.0	0.8	22.4	25.6
Netherlands	15.9	0.9	14.1	17.7
Austria	18.8	0.7	17.5	20.1
Poland	25.8	0.6	24.7	26.9
Portugal	27.4	0.9	25.6	29.3
Romania	40.4	1.2	38.0	42.8
Slovenia	20.4	0.5	19.4	21.4
Slovakia	19.8	0.7	18.3	21.2
Finland	16.0	0.4	15.2	16.8
Sweden	16.4	0.5	15.4	17.4
United Kingdom	24.8	0.7	23.5	26.1

Table 5: AROPE indicator (2013) standard error and 95% confidence intervals Source: Eurostat

The same approach has been used to calculate variance of net change over two different years. In order to monitor the process towards agreed policy goals, particularly in the context of the Europe 2020 strategy, users are particularly interested in the evolution of social indicators. However, interpreting differences between point estimates at different wave may be misleading. It is therefore necessary to estimate the standard error for these differences in order to judge whether or not the observed differences are statistically significant.

Estimated standard errors and confidence intervals (based on normality assumption) for net changes in the AROPE between 2008 and 2013 are shown in Table 6. If a confidence interval does not include 0, it can be assumed the difference in the AROPE between 2008 and 2013 is statistically significant (at a given level of confidence).

	AROPE - 2008	AROPE - 2013	(2013) - (2008)	Estimated standard error (% points)	Confidence interval – Lower Bound	Confidence interval – Upper Bound	Is the difference significant?
EU-27	23.8	24.5	0.7	0.2	0.2	1.2	Y
Belgium	20.8	20.8	0.0	1.2	-2.3	2.3	N
Bulgaria	44.8	48.0	3.2	1.6	0.0	6.4	Y
Czech Republic	15.3	14.6	-0.7	0.8	-2.3	0.9	N
Denmark	16.3	18.9	2.6	1.2	0.2	5.0	Y
Germany	20.1	20.3	0.2	0.5	-0.7	1.1	N
Estonia	21.8	23.5	1.7	1.0	-0.2	3.6	N
Ireland	23.7	29.5	5.9	1.5	3.0	8.7	Y
Greece	28.1	35.7	7.6	1.2	5.3	9.9	Y
Spain	24.5	27.3	*				
France	18.5	18.1	-0.4	0.8	-1.9	1.1	N
Croatia	:	29.9	**				
Italy	25.3	28.4	3.1	0.7	1.7	4.5	Y
Cyprus	23.3	27.8	4.5	1.2	2.1	6.9	Y
Latvia	34.2	35.1	0.9	1.5	-2.1	3.9	N
Lithuania	27.6	30.8	3.2	1.6	0.0	6.4	N
Luxembourg	15.5	19.0	3.5	1.4	0.8	6.2	Y
Hungary	28.2	33.5	5.3	1.0	3.3	7.3	Y
Malta	20.1	24.0	3.9	1.1	1.7	6.1	Y
Netherlands	14.9	15.9	1.0	1.2	-1.4	3.4	N
Austria	20.6	18.8	-1.8	1.0	-3.7	0.1	N
Poland	30.5	25.8	-4.7	0.8	-6.2	-3.2	Y
Portugal	26.0	27.4	1.4	1.4	-1.4	4.2	N
Romania	44.2	40.4	-3.8	1.7	-7.1	-0.5	Y
Slovenia	18.5	20.4	1.9	0.7	0.5	3.3	Y
Slovakia	20.6	19.8	-0.8	1.0	-2.7	1.1	N
Finland	17.4	16.0	-1.4	0.6	-2.6	-0.2	Y
Sweden	14.9	16.4	1.5	0.7	0.1	2.9	Y
United Kingdom	23.2	24.8	*				

*) break in time series

**) no data in 2008

Table 6: Estimated standard errors for estimators of net change in the AROPE between 2008 and 2013 Source: Eurostat

Tracing rules

EU-SILC is composed of two components – the cross-sectional and the longitudinal one. The main objective of the longitudinal component is to study changes over time at individual level, such as transitions from school to work and from work to retirement, flows into and out of economic activity and work and, above all, changes in the level of income and poverty of individuals and households. It should be noted that one of the most important EU-SILC indicators – “At persistent-risk-of poverty” is based on longitudinal component. In each country the longitudinal

component of EU-SILC consists of one or more panels or subsamples (four subsamples in the recommended four-year rotational design). For each panel/ sub-sample, sample households and sample persons (see explanations in Table 7) representing the target pop-
ulation at the time of its selection are followed for a minimum period of four years on the basis of specific tracing rules.

	Sample of households (dwellings)	Sample of persons
Sample household	Households (all households in the dwelling) that were selected for the survey	Households that contain at least one sample person
Sample person	All persons from sample household	Persons selected for the survey
Co-resident	Appears from the second wave onwards: A person who joined a household containing at least one sample person	Persons from sample households that were not selected for the survey

Table 7: Definition of sample household, sample person and co-resident

The objective of the tracing rules is to reflect any changes in the target population drawn in the initial sample and to follow up individuals over time. In order to study changes over time at the individual level, all sample persons (members of the panel/subsample at the time of its selection) should be followed up over time, despite the fact that they may move to a new location during the life of the panel/subsample. However, in the EU-SILC implementation some restrictions are applied owing to cost and other practical reasons. Only those persons staying in one private household or moving from one to another in the national territory are followed up. Sample persons moving to a collective household or to an institution, moving to national territories not covered in the survey, or moving abroad (to a private household, collective household or institution, within or outside the EU), would normally not be traced. The only exception would be the continued tracing of those moving temporarily (for an actual or intended duration of less than six months) to a collective household or institution within the national territory covered, as they are still considered as household members.

The longitudinal sample must also remain representative of all age groups in the population. This means than in principle, persons of all ages should be followed up. However, in view of cost and other practical considerations, separate follow-up may be restricted to persons above a certain age. The minimum EU-SILC requirements are for a follow-up of individuals in the longitudinal sample for a period of four years. For panels of such short duration, it is acceptable (in view of cost and other practical reasons) to separately follow-up only persons aged 14 or over at the time of selection of the initial sample for a panel. The table 8 presents details on follow-up of sample persons, sample households and co-residents.

Units	Actions
Sample persons	
Persons move to a private household within the national territory covered in the survey	Followed to the new location of the household
Other persons temporarily away but who are still considered as members of the household	Covered in the household they belong to
Persons who are no longer members of the private household or who have moved outside the national territory covered in the survey	Dropped from the survey
Sample households	
Non-enumerated for a single year due to impossibility of locating the address, the address being non-residential or unoccupied, lost (no information on what happened to the household), or the household refusing to co-operate	Dropped from the survey
Non-contacted in the first year of the panel or non-contacted two consecutive years due to the impossibility to access the address because the whole household is temporarily away or is unable to respond due to incapacity or illness or for other reasons	Dropped from the survey
Co-residents	
Living in a household containing at least one sample person	Followed to the new location
Living in a household not containing any sample person	Dropped from the survey

Table 8: Rules for the follow-up of sample persons, sample households and co-residents

See also

- EU statistics on income and living conditions (EU-SILC) methodology (overview of all articles)

Main tables

- Income and living conditions (t_ilc)

Database

- Living conditions and welfare (livcon), see:

Income and living conditions (ilc)

Dedicated section

Income and living conditions (ilc)

Publications

- Comparative EU quality reports

Methodology

- Income and living conditions (ilc) (ESMS metadata file — ilc_esms)
- Methodological guidelines and description of EU-SILC target variables

View this article online at http://ec.europa.eu/eurostat/statistics-explained/index.php/EU_statistics_on_income_and_living_conditions_(EU-SILC)_methodology_-_sampling