eurostat

SILC_ESQRS_A_BE_2014_0000

National Reference Metadata in ESS Standard for Quality Reports Structure (ESQRSSI)
Compiling agency: Statistics Belgium
Time Dimension: 2014-A0
Data Provider: BE1
Data Flow: SILC_ESQRS_A:1.1

Eurostat metadata

Reference metadata

For any question on data and metadata, please contact: EUROPEAN STATISTICAL DATA SUPPORT

1. Contact

Top

1.1. Contact organisation	Statistics Belgium
1.2. Contact organisation unit	DTS
1.5. Contact mail address	

2. Introduction

Top

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.

Finally it is the combination of the previous intermediate and final quality reports therefore it is worth mentioning that it refers to both the cross sectional and the longitudinal data.

3. Quality management - assessment

Top

Not requested by Reg. 28/2004

4. Relevance

Top

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. Accuracy depends on sample size, sampling design effects that, sampling errors and non sampling errors need to be taken into account. Sampling error refers to the variability that occurs at random because of the use of a sample rather than a c phases of the data collection and production process.

5.1. Accuracy - overall

In terms of precision requirements, the EU-SILC framework regulation as well the Commission Regulation on sampling and tracing rules refers respectively, to the effective sample si effective sample size combines sample size and sampling design effect which depends on sampling design, population structure and non-response rate.

5.2. Sampling error

EU-SILC is a complex survey involving different sampling design in different countries. In order to harmonize and make sampling errors comparable among countries, Eurostat (with chosen to apply the "linearization" technique coupled with the “ultimate cluster” approach for variance estimation. Linearization is a technique based on the use of linear approximatio 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 cc 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 flexil be generalized to calculate variance of the differences of one year to another .

The main hypothesis on which the calculations are based is that the "at risk of poverty" threshold is fixed. According to the characteristics and availability of data for different countrie information. In particular, countries have been split into four groups:

1)BE, BG, CZ, IE, EL, ES, FR, IT, LV, HU, NL, PL, PT, RO, SI, UK and HR whose sampling design could be assimilated to a two stage stratified type we used DB050 (primary strat Unit) for cluster specification;
2) DE, EE, CY, LT, LU, AT, SK, FI, CH whose sampling design could be assimilated to a one stage stratified type we used DB050 for strata specification and DB030 (household ID)
3) DK, MT, SE, IS, NO, whose sampling design could be assimilated to a simple random sampling, we used DB030 for cluster specification and no strata;

In case Eurostat methodology is not accepted by your country, please describe the methodology used at national level for computing the estimates.

5.2.1. Sampling error - indicators

	1. AROPE			2. AROP			3. SMD			4. LWI		
	valeur	st_err	half_length	valeur	st_err	half_length	valeur	st_err	half_length	valeur	st_err	half_length
1. ALL	21.20	0.75	1.50	15.50	0.65	1.30	5.90	0.56	1.10	14.60	0.80	1.60
2. Sex : Male	20.90	0.88	1.80	15.00	0.78	1.60	6.20	0.70	1.40	14.30	0.91	1.80
3. Sex : Female	21.60	0.76	1.50	15.90	0.67	1.30	5.60	0.50	1.00	14.90	0.82	1.60
4. Age : 0-17	23.20	1.47	2.90	18.80	1.31	2.60	6.80	0.95	1.90	13.10	1.13	2.30
5. Age : 18-64	21.60	0.87	1.70	14.20	0.72	1.40	6.50	0.65	1.30	15.10	0.81	1.60
6. Age : 65+	17.30	1.40	2.80	16.10	1.36	2.70	2.40	0.45	0.90	.	.	.

Annexes:
Annex sampling errors 2014

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 frame.
- Measurement errors: errors that occur at the time of data collection. There are a number of sources for these errors such as the survey instrument, the information system, the inte
- Processing errors: errors in post-data-collection processes such as data entry, keying, editing and weighting
- Non-response errors: errors due to an unsuccessful attempt to obtain the desired information from an eligible unit. Two main types of non-response errors are considered:

1. – Unit non-response: refers to absence of information of the whole units (households and/or persons) selected into the sample
1. – Item non-response: refers to the situation where a sample unit has been successfully enumerated, but not all required information has 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 do not exist in practice
- Under-coverage: refers to units not included in the sampling frame
- Misclassification: refers to incorrect classification of units that belong to the target population

In Belgium, the sampling frame is the Central Population Register.

As there was a period of one month between the drawing of households and the survey itself, over-coverage, under-coverage and misclassification could be happen.

Over-coverage: Persons who died before the survey. Households who moved outside Belgium before the survey. Address is not the principal residence.

Under-coverage: Immigrants who came in Belgium before the survey. Persons who moved from a household to create a new household. Diplomats exempt from an inscription in the r

Misclassification: Household who moved from a region in Belgium to another region of Belgium.

The size of coverage errors is not available but it was obviously small.

5.3.1.1. Over-coverage - rate

[information not available]

5.3.2. Measurement error

Cross sectional data

Source of measurement errors	Building process of questionnaire	Interview training	
-survey instrument -information system -interviewer -mode of collection (CAPI interview)	The questionnaire of the SILC2013 survey is the result of several steps: -For building up the questionnaire we took the blue print questionnaire of Eurostat as the basis (documents SILC055, SILC065 and EU-SILC65/02 Addendum II). The order of the questions and the groups (themes of) questions is taken from this blue print. The majority of the questions are almost literally copied (and translated), other questions are changed, however, because experiences in Belgium gave better results posing the questions in another way (The questionnaires were developed in collaboration with the universities that have the experience of the ECHP/PSBH project in Belgium). -After each survey an evaluation of the questionnaire was made (detection of the problematic or difficult to answer questions based on the comments of the interviewers and on a study of the item non-response). When building up the SILC2013 questionnaire we took account of this evaluation.	**Interview training (Number of training days and information on the intensity and efficiency of interview training)** Overall we had the impression that the working-experience of the interviewers with EU-SILC starts to pay of. In our opinion the basis data has improved since 2010. All new interviewers have to follow a two day formation. All trained interviewers followed a formation for an hour and half. They both had to complete a test-interview before they could download their data. So we can be sure they can completely manage the use of the PC and that they know the questionnaire before they go on the field. A training group for new interviewers consisted of minimum 5 to maximum 20 interviewers, and according to the size of the training group there were 1 or 2 trainers. Even though the accent was given to the practical side of the training (getting to know the questions and mastering the CAPI-program by imitating interview situations), three manuals were distributed and explained during the training: -A general manual (‘Manuel general aux enquêteurs’) containing information about the objectives of the survey, the organisation of the survey, legal and administrative aspects around the survey, fieldwork aspect (how to contact the household, how to introduce oneself, who answers which questions, time delays, …) and the content of the questionnaires. -A second manual (‘Manuel contenu’) with all kinds of additional explanations and examples for certain questions/answers. -A third manual (‘Manuel CAPI’) about the use of the portable PC for the SILC Computer Assisted Personal Interviews and about the data entry program itself. The first day of the training there was half a day for learning about and discussing the first two manuals. In the afternoon the trainees received	**• Skills testing before star** Interviewers were selected centralised for all the surve basic curriculum vitae is pr surveys they have experien specific unit at Statistics Be selection of the interviewer of the interviewers. They t areas to cover for SILC. Th areas several interviewers a no candidates. Note that in second or casual occupatio **• Skills control during the** During the fieldwork we co their completed questionna some trained interviewers t negative) resulting from the interviewer so they could in **• Number of households b** Groups of secondary unit strata. Most of the interv several interviewers also

Cross sectional data

Source of measurement errors	Building process of questionnaire	Interview training
		their laptop and got to know the survey and the tool to carry out the interview in practice. One test-interview was simulated collectively. The second day of the training a small part of the time was dedicated to testing to send the data electronically after carrying out the interview. All the rest of the day interviewers practiced several interviews and interview situations with each other on the basis of household profiles that were given. There was also a lot of time for questions and discussions in between these test-interviews. At the end of the training sessions the instructors had a good image on the degree in which each interviewer ameliorated during the training and on the degree in which they mastered the work. For certain interviewers two days of training was more than enough to master the work, for others it was necessary that they practiced some more at home on specific aspects of carrying out this survey (for example using of the CAPI-program itself, working on the content of the survey, …). They were recommended to do so before carrying out their first real interview. They were often also recommended to start interviewing one-person households. A training group for trained interviewers consisted maximum 30 interviewers with two trainers. The accent was also given on the content: questions that changed, the module 2013 and questions, which are misunderstood by the interviewers. We made an extra manual for trained interviewers. The trained interviewers obtained four manuals, the same three as the new interviewers, and a fourth ('Modifications du questionnaire : module 2013') about the module, changed questions and questions misunderstood by the interviewers.

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 main types of non-response errors are considered:

1) Unit non-response which refers to the absence of information of the whole units (households and/or persons) selected into the sample. According the Commission Regulation 28/2

- **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 contacted addresses

- **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 were completed and accepted for the database

- **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 individual non-response rates will be calculated for 'the selected 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 required information has been obtained.

5.3.3.1. Unit non-response - rate

Address contact rate (Ra)*			Complete household interviews (Rh)*			Complete personal interviews (Rp)*			Household Non-response rate (NRh)*			Individual non-re (NRp)*	
A*	B*	C*	A*	B*	C*	A*	B*	C*	A*	B*	C*	A*	B*
0.978	0.9639	0.9899	0.6517	0.4338	0.8302	0.9896	0.98	0.9938	36.26	58.19	17.81	1.04	2

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

A* = Total sample; B* = New sub-sample; C* = Panel 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 Commission Regulation No 1982/2003. Item non-respons household and personal level.

Annexes:

Annex Item non response 2014

5.3.3.2.1. Item non-response rate by indicator

In the table below an overview including interpretation for the non-response is presented.

Indicator	**Achieved** sample size (number of individuals)	**Non-response**
Mean Equivalised disposable income	14346	0
Risk of poverty threshold: one person household	1821	0
Risk of poverty threshold: household with 2 adults and 2 dependent children	2612	0
Risk of poverty rate by age	14346	0
Risk of poverty rate by gender	14346	0
Risk of poverty rate by most frequent activity	11291	170
Risk of poverty rate by household type	14346	0
Risk of poverty rate by household type: Single households	1821	0
Risk of poverty rate by tenure status	11459	2

Risk of poverty rate by work intensity of the household	12079	2267
Dispersion around at risk poverty threshold	14346	0
Relative median risk-of-poverty gap by age and gender	14346	0
Risk-of-poverty rate by age and gender before all transfers (including pensions)	14346	0
S80/S20 quintile share ratio	14346	0
Gini coefficient	14346	0

Item non-response and number of observations at unit level of the common cross-sectional European Union indicators and for equivalised disposable income.

5.3.4. Processing error

Question number	Control
Contact form	
Column 21, 22, 23 and 24	You can't combine father, mother or being spouse with 'being younger than 12 years".
Column 8,21 and 22	It's not possible to combine being 'female' and being 'father'. It's not possible to combine being 'male' and being 'mother'.
Column 21 and 22	Mother and father have to be older than their children (and at least being older than 12 years).
Column 21, 22, 23, 24	Parents of the spouses or of the partners must be different.
Column 23, 24	You can't mix 'spouse 'and 'partner'. Must choose one of both for the couple.
Household questionnaire	
H5 and H7:	It is not possible to combine H5,code 6 with H7 code 2, 3, 4, 5, 6, 7, 8, 9, 10
H13	Enter a numeric value between 1900 and 2011
H19	The first of the reimbursement must be between 1954 and 2008 (included). The year of the first purchase must be at the same time or later than the date of buying.
H27 category g, H45 category g:	Code 1 is only possible if at question H5, code 3,4,5,6 or 7
H44	Not possible to answer more than 12 months
H95	Persons have to be between the age of 11 and 23 (included) to obtain a scholarship for secondary school
H97	Persons have to be between the age of 16 and 99 (included) to obtain a scholarship for higher education
Individual questionnaire	
Question I6, I7 and I8	You can't combine code 2 of questions I6 and I7 with code 1, 2, 3, 4 and 10 of the question I8.
Question I6 , I7 and I8	You can't combine code 1 of question I6 or question I7 with code 5, 6, 7, 8, 9 and 11 of the question I8.
Question I13 and I14:	You can't combine code 1,2,3,4 and 10 question in I13 with code 2 and 3 in question I14
Question I13 et I16	You can't combine code 1, 2, 3, 4 and 10 of the question I13 with code 1, 2 of the question I16.
Question I14 and I16	You can't combine code 2 or 3 of the question I14 and code 3 or 4 of the question I16.
Question I21 and I22	You can't combine code 1,2,3,4 or 10 in question I21 with code 2 or 3 in question I22.
Question I21 and I29.	You can't combine code 1, 2, 3, 5, 6 of the question I29 with the code 1, 2, 3, 4 or 10 of the question I21.
Question I29 and I22	You can't combine code 7 of the question I 29 with code 2 or 3 of the question I22.
Question I37	Age has to be less than current age and not less than 8 year.
Question I38	Number of years can't be higher than current age minus the age mentioned in question I37.
Question I 52, I 92.	Can't be higher than 12 months.
Question I 116	Can't enter a year which is before date of birth.
Question I25 (I26) (gross income) and question I27 (I28) (net income)	Amounts given in question I25 can't be higher than the amounts given in the question I27. ***Remark*** *: Ditto for the questions I47 (I48) and i50 (I51), I53 and I54, I55 and I56, I90 and I91, and I93 and I94, I98_A, B, C, D, E, F, G, H and I99 and I102_A, B, C, D, E and I115_ A, B, C, D, E and I116_ A, B*
Question I25 and I 26	If the person didn't give an exact amount at the question I25, please go to the question I26. ***Remark*** *: Ditto for the question I27 and I28; I47 and I48; I50 and I51*

5.3.4.1. Imputation - rate

Not requested by Reg. 28/2004

5.3.4.2. Common units - proportion

5.3.5. Model assumption error

5.3.6. Data revision

5.3.6.1. Data revision - policy

5.3.6.2. Data revision - practice

5.3.6.3. Data revision - average size

5.3.7. Seasonal adjustment

6. Timeliness and punctuality

Top

Not requested by Reg. 28/2004

6.1. Timeliness

Not requested by Reg. 28/2004

6.1.1. Time lag - first result

6.1.2. Time lag - final result

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

Top

Not requested by Reg. 28/2004

7.1. Dissemination format - News release

7.2. Dissemination format - Publications

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

7.5. Documentation on methodology

Not requested by Reg. 28/2004

7.5.1. Metadata completeness - rate

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

8. Comparability

Top

According to the Regulation (EC) No 1177/2003 of the European Parliament and of the Council concerning EU-SILC: "Comparability of data between Member States shall be a fundamental objective and shall be pursued through the development of methodological studies from the outset of EU-SILC data collection, carried out in close collaboration between the Member States and Eurostat".

Although the best way for keeping the comparability of data is to apply the same methods and definitions of variables, small departures of the definitions given by Eurostat are allowed in EU-SILC. In this way, the mentioned Regulation in its article 16th says: "Small departures from common definitions, such as those relating to private household definition and income reference period, shall be allowed, provided they affect comparability only marginally. The impact of comparability shall be reported in the quality reports."

8.1. Comparability - geographical

8.1.1. Asymmetry for mirror flow statistics - coefficient

8.1.2. Reference population

Reference population	Private household definition	Household membership
The reference population is all citizens living officially living at Belgian territory (population de jure). This means that the source of our sample is the central population register. This Register includes all private households and their current members residing in the territory. Persons living in collective households and in institutions are excluded from the target population.	The definition of household that Eurostat recommends is used. Household is defined as a person living alone or a group of people who live together in the same dwelling and share expenditures including the joint provision of the essentials of living.	The definition of household membership is the same as mentioned in the Eurostat document EU-SILC065/03 about the description of target variables (Chapter 'Units'). All household members of 16 year and older at the end of the income reference period, are selected for a personal interview.

8.1.3. Reference Period

Period for taxes on income and social insurance contributions	Income reference periods used	Reference period for taxes on wealth	Lag between the income ref period and current variables
A fixed twelve-month period, namely the previous calendar year. For SILC 2014, the period is the year 2013.	A fixed twelve-month period, namely the previous calendar year. For SILC 2014, the income reference period is the year 2013.	n.a.	The income reference period is the previous calendar year (year 2013) and the current variables refer to the fieldwork period (April-December 2014). Therefore the lag is at minimum 4 months and at maximum 12 months.

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)	Total disposable hh income before all social transfers (HY023)
F	F (remark: We didn't take count of HY120G, because regular taxes on wealth do not exist in Belgium.)	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)	Interest paid on mortgage (HY100)	Income received by people aged under 16 (HY110)	Regular taxes on wealth (HY120)	Regular inter-hh transfers paid (HY130)
F	F	F	F	F	F	F	F	F	NC	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)	Survivors benefits (PY110)	Sickness benefits (PY120)	Disability benefits (PY130)	Education-related allowances (PY140)	Gross monthly earnings for employees (PY200)
F	F	F	F	F	F	F	F	F	F/L/P/NC	F/L/P/NC	F/L/P/NC	F/L/P/NC

The source or procedure used for the collection of income variables

The collection of the income variables is by interview.Belgiumhas no income variables collected from registers for the survey of 2011.

The form in which income variables at component level have been obtained

Areas	Qr. Block	Target Variable	Unit of measurement	Tax or tax-exempt	If taxable, how the amount is recorded
Employee Income	PY010	Gross Employee Cash or near cash Income in reference period	Individual level	Taxable	Net + gross
	PY020	Gross Non-Cash Employee income (company car, mail tickets)	Individual level	Not taxable (mail tickets are not taxable for the employee and can not be deducted from taxes by the employer) (the company car itself is not taxable but the kilometres that are done for job/work distances and for private distances are taxed: there is always a minimum of 5000 km taxed)	
Self-employment Income	PY050	Gross Cash Income benefits/Losses from self-employment (including profit/loss from unincorporated enterprise, royalties)	Individual level	Taxable For losses, this means a deduction from taxes of this amount can be done on other income posts of that year or on income of the next year)	Net OR gross
Imputed rent[1]	HY030	imputed rent	Household level	-	
Property income	HY090	Interest, dividends, profit from capital investments in unincorporated business	Individual level	Taxable	Net
	HY040	Income from rental of property or land	Household level	Taxable	Gross
	PY080	Regular pension from Private (non-ESSPROS) schemes	Individual level	Taxable	Gross (for the major part of the pensions)
Current transfer received Social benefits: ESSPROS	HY050	Family-related allowances: parental leave benefits	Individual level	Taxable	Net + gross
		Family-related allowances:	Household level	Not taxable	
	HY060	Social assistance	Individual level	Not taxable	
	HY070	Housing allowances	Household level	Not taxable	
	PY090	Unemployment Benefits	Individual level	Taxable	Net + gross
	PY100	Old-age benefits	Individual level	Taxable	Net + gross
	PY110	Survivor's Benefits	Individual level	Taxable	Net + gross
	PY120	Sickness Benefits	Individual level	Taxable	Net + gross

The method used for obtaining target variables in the required form

See information on control, correction, imputation and creation of the gross target variables.

The source or procedure used for the collection of income variables	The form in which income variables at component level have been obtained					The method used for obtaining target variables in the required form
	PY130	Invalidity Benefits	Individual level	Taxable	Net + gross	
	PY140	Education-related Allowances	Household level	Not taxable		
Regular inter household transfer received	HY080	Regular inter-household cash transfers received	Household level	Not taxable, but taxed if alimentation	Gross	
Other income received	HY110	Income received by people aged under 16	Household level	Not taxable		
Interest payments	HY100	Interest repayments on mortgage	Household level	Taxable, this means a deduction from taxes can be done	Gross	
Current transfers paid	HY130	Regular inter-household cash transfers paid	Household level	Not taxable or deductible, but taxed if alimentation	Gross	

[1] Information on that component is asked because it is important to know if :
- an owner is taxed regarding his tenure status (specific tax on property income)
- a 'rent-free' tenant could be taxed on behalf of the accommodation's owner

8.2. Comparability - over time

The results of the Belgian EU-SILC 2014 operation are very similar to those of the 2013 operation.

8.2.1. Length of comparable time series

8.3. Comparability - domain

9. Coherence

Top

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

9.1.1. Coherence - sub annual and annual statistics

9.1.2. Coherence - National Accounts

9.2. Coherence - internal

10. Cost and Burden

Top

We checked the number of minutes to complete the household questionnaire (by the means of flags in the CAPI program). The household questionnaire took about 20 minutes, a single individual questionnaire 10 minutes. Mean total interview duration is 39 minutes.

11. Confidentiality

Top

11.1. Confidentiality - policy

Not requested by Reg. 28/2004

11.2. Confidentiality - data treatment

Not requested by Reg. 28/2004

12. Statistical processing

Top

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

In Belgium, the sampling frame is the Central Population Register. This Register includes all private households and their current members residing in the territory. Persons living in collective households and in institutions are excluded from the target population.
The Central Population Register of 1 February was used.

Updating actions: Central Population Register is updated two times during a month. The changes were communicated to the interviewers.

12.1.1. Sampling design and procedure

Type of sampling design

The Belgian EU-SILC 2014 survey is based on a stratified 2-stage sampling scheme in 2004, followed by rotation since 2005. Rotation allows to replace roughly one fourth of the sample each year. Hence, households (ignoring split-offs) participating in 2014 have been drawn for participation since 2011 2012, 2013 and 2014.

Stratification and sub stratification criteria

The main stratification criterion is the NUTS2 level. The 11 sampling strata are the 10 Belgian provinces (5 in Flanders – coded BE21-BE25 – and 5 in Wallonia – coded BE31 to BE35) and the Brussels Capital Region (BE10).Further implicit stratification is obtained by sorting PSUs (sub-municipalities) on mean income and sorting SSUs (households) in selected PSUs on age of reference person.

Sample selection schemes

Sampling units and 2-stage sampling in 2004

In 2004, when organizing EU-SILC for the first time (ignoring the pilot survey in 2003), 2-stage sampling has been applied in each sampling stratum.

Stage 1 – Primary Sampling Units

The primary sampling units (PSUs) in stage 1 are the municipalities, or parts thereof in the larger ones. In each stratum, the PSUs in the frame are first descendingly sorted by average income; next, a fixed number of times a PSU is drawn according to a systematic PPS (probability proportional to size) selection scheme, where size is measured as the number of private households. This systematic sampling method generally causes some PSUs being selected repeatedly (e.g. Schaerbeek, a rather large municipality in stratum BE10, turns out to be drawn 6 times). In total, i.e. in all 11 sampling strata together, 275 PSU draws were made in 2004, once and for all (i.e. for the whole duration of EU-SILC).

Stage 2 – Secondary Sampling Units

The secondary sampling units (SSUs) in stage 2 are private households. According to each single PSU draw, a group (generally of fixed size) of households is selected in this stage; notice that a group of households corresponds to each PSU draw.

In 2004, 40 households have been selected for each PSU draw (i.e. in each group); e.g. in Schaerbeek, 6 times 40 households were drawn. Systematic selection of households has been applied, after sorting the households in selected PSUs by age of reference person. Within each group, the selected households were numbered 1 to 40; households 1-10 constitute the first rotational group or replication, households 11-20 constitute the second rotational group or replication, and so on. The first replication was meant to participate in 2004 only, the second until 2005, and so on.

The initial household sample in 2004 was self-weighting, by the combination of (systematic) PPS sampling of sub-municipalities (PSUs) – size of PSUs being the number of private households – and (systematic) sampling of private households (SSUs), as explained.

Renewal of the sample by rotation, since 2005

Since 2005, a rotation scheme has been applied. Details for each year, from 2005 to 2011, can be found in the corresponding Quality Reports .

The rotation pattern is such that the overlap between samples in any two successive years is roughly 75%, and that the sample is completely renewed after 4 years. Hence four replications or rotational groups in each year, one of which is replaced the year after. Since 2005, each new replication remains in the survey during the next 4 years, and since 2007, each of the four replications is in the survey during four consecutive years.

At the start of 2014, the replication that is in the survey since 2010, is entirely (i.e. irrespective of whether the households are responding or not) dropped. The three replications which entered into the survey in 2011, 2012 and 2013, respectively, are retained (including their split-offs); the households belonging to these three replications will be designated ‘old’ hereafter.

The supplementary sample, i.e. the new replication that replaces the just dropped replication, is obtained by selecting, for each PSU draw, a fixed number of new households from the corresponding PSU. This selection is done again by systematic sampling, after sorting the households in each PSU on age of reference person. The number of new households for each PSU draw, is determined by considering some (expected) attrition of old households, some (expected) nonresponse for new households, and the required/desired minimum and maximum numbers of responding households, given some precision and budget constraints.

Hence, the (cross-sectional) sample of SILC 2014 consists of

- “old” households: drawn between 2011 and 2013; and
- “new” households: drawn in 2014, staying until 2017

12.1.2. Sampling unit

2-stage sampling .

Stage 1 – Primary Sampling Units

The primary sampling units (PSUs) in stage 1 are the municipalities, or parts thereof in the larger ones. In each stratum, the PSUs in the frame are first descendingly sorted by average income; next, a fixed number of times a PSU is drawn according to a systematic PPS (probability proportional to size) selection scheme, where size is measured as the number of private households. This systematic sampling method generally causes some PSUs being selected repeatedly (e.g. Schaerbeek, a rather large municipality in stratum BE10, turns out to be drawn 6 times). In total, i.e. in all 11 sampling strata together, 275 PSU draws were made in 2004, once and for all (i.e. for the whole duration of EU-SILC).

Stage 2 – Secondary Sampling Units

The secondary sampling units (SSUs) in stage 2 are private households. According to each single PSU draw, a group (generally of fixed size) of households is selected in this stage; notice that a group of households corresponds to each PSU draw.

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 per nuts2 for 2014 :

NUTS2	Name	Old hh	New hh	Total hh	Accepted hh (DB135=1)
		27	104	131	0
BE10	Brussels	845	747	1592	997
BE21	Antwerpen	731	659	1390	854
BE22	Limburg	388	252	640	415
BE23	Oost-Vlaanderen	593	498	1091	678
BE24	Vlaams-Brabant	489	405	894	551
BE25	West-Vlaanderen	517	329	846	631
BE31	Brabant Wallon	138	126	264	167
BE32	Hainaut	689	590	1279	809
BE33	Liège	479	393	872	558
BE34	Luxembourg	130	92	222	159
BE35	Namur	170	153	323	202
Total	**Belgium**	5196	4348	9544	6021

12.2. Frequency of data collection

The survey is lead each year.

12.3. Data collection

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.

see Annex "Data collection"

Annexes:
Annex Data collection

12.4. Data validation

12.5. Data compilation

12.5.1. Weighting procedure

Recall that, for the first year of the panel (=SILC 2004 in Belgium), the computation of weights involved three stages (described in 134-04)

(a) initial weights
(b) weights corrected for nonresponse
(c) final (calibrated) weights

For 2014, a distinction has to be made between

"old" households i.e. households that contain at least one sample person who took part in 2013, and had to be surveyed again in 2014 according to the rotation and tracing rules (excluding the outgoing fourth) (household composition may have changed, whence quotations marks)

"new" households i.e. households that were drawn for the first time in 2014, among those households not containing any sample person already drawn before

This distinction pertains to initial weights and nonresponse correction

Since the "old" households are selected indirectly from the 2011, 2012 or 2013 samples, and household composition may have changed, some kind of "weight sharing" must be applied to determine the (2014) initial weights, or rather base weights. On the other hand, "new" households have their own inclusion probability, whose inverse gives the initial weights;

For the "old" households, (2014) nonresponse=attrition can be linked with (2013) SILC information. For the "new" households, all we can rely upon to explain initial nonresponse is auxiliary information from the Population Register (household size, urban/rural character) and the Financial Statistics (median fiscal income by municipality:)

On the other hand,

Calibration can be done together for "old" and "new" households. With respect to our 2004 model, we decided in 2005 to relax the constraints (basically, calibrating at NUTS1-level instead of NUTS2), in order to decrease the standard deviation of weights.

1. Initial weights for the new households

Belgium chose to draw the Primary Sampling Units (= municipalities or parts thereof) "forever", and to rotate the Secondary Sampling Units (=households) within the selected PSU's.

The 2004 PPS two-stage sampling design was self-weighting within each stratum h: x denoting any households in municipality X), we had (in 2004)

P (x drawn) = P(x drawn|X drawn) . P(X drawn) = n_h/N_X . N_X/N_h . g_h = n_h/N_H . g_h, where

n_h	denotes	the number of households to be drawn in the (selected) PSU (viz. 40)
N_X		the number of households in the PSU (in 2004)
N_h		the number of households in the stratum (in 2004)
g_h		the number of PSU's drawn in the stratum.

(This is an oversimplification, since PSU are drawn with repetition; the selection probability for a PSU should be replaced by the expectation of selection multiplicity, and the term 40 by a multiple depending on the selection multiplicity…but the idea is the same).

In 2013, the picture has become

P (x drawn) = P(x drawn|X drawn) . P(X drawn) = m_h/M_X . N_X/N_h . g_h, where

m_h	is	the number of households to be drawn in the (selected) PSU (depending on h)
M_X	is	the number of households in the PSU (in 2014)

The factor N_X/M_X indicates the increase-decrease in inclusion probabilities in PSU X (still assuming X has been drawn) between 2014 and 2004.

Now it would seem logical to replace N_X by a smaller number, to account for the households[1] already drawn in 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012 or 2013 whence immunized from being drawn again in 2014.

However, the following argument shows that (assuming momentarily that X has been drawn and that the population figures N_X and M_X remain stable) matters are not so easy:

P(x drawn in 2014) =
(P(x drawn in 2014|x drawn before) . P(x drawn before)) +
(P(drawn in 2014|x not drawn before) . P(x not drawn before),

the first term vanishes and the second equals $n_h/(M_X\text{-}b)$. $(N_X\text{-}b)/N_h$, where b denotes the number of hh already drawn; since both fraction terms are much larger than b (at least 900 in all selected PSU's), the ratio $(N_X\text{-}b)/(M_X\text{-}b)$ is (close to 1, and) very close to N_X/M_X. Since the term b is an approximation anyway, we chose to stick to m_h/M_X . N_X/N_h. g_h as inclusion probabilities, and its inverse for initial weights **INIwei=DB080**. Note that, with this concept of DB080, the "new" hh correspond to the total Belgian population (some 4,5 millions private hh); before calibrating, theses weights will be scaled down "to make room" for the old hh; recovering the strange hh means that the sum of the pre-calibration weights will be slightly larger than 4,5 millions (average of g-weights slightly less than 1)

2. Nonresponse correction for the new households

Following Eurostat's suggestion (see Document 065, WEIGHTING II. WEIGHTING FOR THE FIRST YEAR OF EACH SUB-SAMPLE), we replaced the homogeneous response groups (based on household size crossed with urbanity) ratio by a multiple regression model (based on the same dummy variables). By "responding", we mean only those households whose results were accepted (DB135=1). Since 2009 we used logistic regression.

The file was split by NUTS1 and the following variables were used

- Everywhere: Household size, recoded into the four values "one", "two", "three" and "four or more" (so three dummies)
- Out of Brussels: DB100 = urbanity
- In Brussels = BE10: median fiscal income of municipality

The regression produced a new variable "expresp", allowing us to define

NRwei = INIwei/expresp

3. Attrition for the old households

Before "sharing" the 2013 weights, a correction for attrition should be introduced. This year, we elected to perform this correction at the level of individuals, since a 2013 sample person either stays in the panel or leaves it (rotated out, left population, noncontact, refusal or inability to respond, while the structure of a household can change. Note that all household characteristics (e.g. HH021) can be distributed to the members.
We separated the "Children" (for which only basic personal information from the R-file and the distributed H-file is available) from the "Adults" (present in the 2013 P-file as well), i.e. those persons born in 1998 or before.

In the children's model, the following predictors (all, except the last, from the 2013 file – although this does not matter much for group A) were used, grouped by type :

1. individual demographic information: age from RB080, sex = RB090,
2. housing information: dwelling type = HH010 and tenure = HH020
3. household type: a limited number of dummies, as there is at least one dependent child;
4. monetary indicators: we refrained from taking the equivalised income (outliers), but took a transform of it, as well as the dummy "poor or not" and the subjective ability to make ends meet = HS120
5. sampling and rotation: number of years in panel (from DB075) and urbanisation (=DB100)
6. one variable (paradata) related to fieldwork in 2013 (computed from HB040 and HB050)

For the adults, the same predictors were used, and moreover :

1. variables from the P-file (related to education level and health);
2. country of birth (dummy Belgium Yes/No)

were integrated.

We used logistic regression.

4. Weight sharing

We followed Eurostat's recommendation "EU-SILC weighting procedures: an outline" and shared the calibrated 2013 weights, after correcting for attrition (instead of the initial weights, see Lavallée).
This can be illustrated by an imaginary example, dealing simultaneously with fusions (persons *A&B* in same 2013 hh, *C* in another 2013 hh, so "fusion" in the sense of DB110 occurs), new members (a baby like *E* or already in population like *D*); we focus on the 2013 hh, what happened to those who co-resided with *A* and *B* or with *C* in 2013 (left or split) is irrelevant!
Note that

- RB050 = weight 2013: same for *A* & *B*, vacuous for *D* and *E*
- Newi: in general a bit larger than RB050; *A*'s differs from *B*'s (attrition correction at individual level)
- Somwe = 950+1000+850 involves only *A*, *B* and *C*
- *Weiind:* = ¼ * **somwe** (*A B C D* : four contribute to the denominator)[2]

Person in 2014 hh	*A*	*B*	*C*	*D*	*E*
RB110 (2013)	1	1	2	3	4
RB050 (weight 2013)	800	800	600	---	---
Newi = Weight 2014 (after attrition correction)	950	1000	850	---	---
Somwe (sum Newi over 2014 hh)	2800	2800	2800	2800	2800
Weiind	700	700	700	700	700

Weiind will be injected as "initial" weight in the final calibration job.

5. Calibration

We first put the pieces together: weiind is defined as

- (new = started in 2014) :

initial weight, corrected for initial nonresponse, scaled, see 2.1.8.1)

- (old = took part in 2013) :

2013 weight, corrected for attrition and weight sharing if necessary, see 2.1.8.4)

- (back = did not take part in 2013 but before) :

initial weight, no correction.

Type	# ind	Mean of weiind
NEW	4340	626.42
OLD	9436	847.55
BACK	604	554.82
Total	14380	768.52

In terms of persons, the weiind statistics were
Recall that 11 *sampling* strata were used (provinces= NUTS2); we use 3 *extrapolation* strata (the 3 NUTS1 regions BRUssels=BE1, VLAanderen=BE2 and WALlonia=BE3)
Calibration model was adapted in 2012. From this year we take 2 additional individuals variables into account for our model : BIT status and Social integration benefits status.
In 2014, our calibration model is the following :

VLA, WAL:
SIZE4+(AGE8XSEX2)+PROV5 +statbit3 +RIS2 --> 23 individual[3] + 4 household constraints
BRU:
SIZE4+(AGE8XSEX2) + statbit3 +RIS2 --> 19 individual + 4 household constraints
Prov = province where interviewed
Statbit3 = BIT status (unemployed – worker – inactive)
RIS2 = receiving social integration benefits (yes – no).

Individual constraints 32=16+11+3+2 (age*sex + prov+statbit+RIS ;
note that each province belongs to one single region
(extrapolation stratum), for the other two regions, the
total is set to 0 and the condition is vacuous)

Household constraints (size: "1", "2", "3 or "4 & more",)

Calibration type (after some trials and errors…): truncated

6. Final cross-sectional weights

Statistics

	N	Minimum	Maximum	Mean	Std. Dev.
Final weights	6021	44.81	9014.22	799.45	458.24

[1] Perhaps a bit less (households that vanished already subtracted) or a bit more (split households, both components of which stayed in PSU, should be subtracted twice)
[2] Do we abide by the Eurostat rules (starting from base weights, it is unclear whether "their" attrition correction precedes or follows weight sharing) ?
There remain some additional categories of persons to be considered:
-Children born to sample women. They receive the weight of the mother (this assumes that the baby belongs to his/her mother's hh)
-Persons moving into sample households from outside the survey population. They receive the average of base weights of existing household members (vacuous here, as RB110 enables us to identify the newborns, but not the immigrants or the –few- persons moving from a collective to a private hh)
-Persons moving into sample households from other non-sample households in the population – these are "co-residents" and are given zero base weight.
[3] Five provinces and 16 age*sex categories, but sum over provinces = sum over age*sex

12.5.2. Estimation and imputation

Preceding important remark

In contrast to 2004 and as 2005 – from 2006 onwards (so also in 2013) the calendar question (i40 in the questionnaire) was presented to every respondent rather the only those who indicated that had been a change in their social-economic position. It enabled us to assess and check much thoroughly the link between the social-economic position and the income variables. Notably for the self-employed this resulted in a substantive number of cases (being identified as being self-employed) who would be otherwise (and who were to some extent in 2004) not identified as being self-employed. These cases mainly concern people in jobs 'somewhere on the bridge' between being self-employed and employee but who nevertheless indicated in the calendar that they were self-employed.

1.1.1. Overall strategy: Emphasis on internal information and integration of outlier detection- , imputation- and control-phases.

Between 2012 and 2013 there was no major changes in our overall strategy.

1.1.1.1. Emphasis on internal information.

We can't emphasise enough that to correct and impute our data (for any variable) we relied:

a) as much as possible on internal information present in the data itself

b) on formal and legal sources of information and

c) only as final resort turned to statistical procedures (random imputations for ex.)

1.1.1.2. An integrated strategy.

As it was the case for previous SILC-surveys we used from SILC-2013 again an 'integrated approach' to organise the detection of outliers and the imputations. Crucial to the understanding of our way of working are the concepts of what we call 'vertical' and 'horizontal integration'.

By 'vertical integration' we mean that the phases of outlier detection and imputation were done together for each variable separately (1) rather than that both phases were done separately for all variables together (2). The differences between (1) – the way we did things for SILC 2004 - and (2) the way it was done for SILC 2003 – are subtle but nevertheless more than semantics, especially when combined with horizontal integration.

By horizontal integration we mean that information for each respondent on one variable was checked against information on another variable or another source. Information on the monthly gross income for example was – if both possible and applicable- checked with information on the net income, the yearly income, the current income (if no changes had occurred), the household income, other 'proxi'- variables to income (status etc…) and very important external sources of information like legislation.

The interplay between what we call vertical and horizontal integration leads to a dynamic strategy: variables are checked for outliers and inconsistencies, variables are compared to each other and corrected, (corrected) variables are immediately imputed consistently to the information in other (also corrected) variables – and this several times repeated.

We believe that the emphasis of this strategy on consistency of internal information for respondents throughout the survey and the use of external sources of information (legislation) is a far more successful way of detecting outliers and imputing missing values compared to methods of screening for outliers entirely based on (univariate) distributional features of variables (box-plot methods for example) and imputation methods mainly based on statistical probability models (IVE for example).

Outlier detection:

The shift in strategy also implies – of course - a shift in the techniques that are used. As far as the outlier detection concerns there is far less emphasis on univariate - purely distributional related methods like box-plots but more emphasis on inconsistency checks. For the income variables these checks were done in 2 ways: a) comparison of ratio's between variables and b) comparison of the relative position of a respondent's answer on one variable to its position on another variable.

a) Comparison of ratio's between variables:

Comparison of the ratio between two inputs on comparable income variables is a straightforward way to detect outliers. Atypical large or small ratios between gross and net variants of income variables are obviously an indication of 'something being wrong'.

b) Comparison of relative positions on income variables:

The central issue in this procedure is the comparison of two income variables by comparison of the normal scores calculated for each case on both variables, after log-transformation. The log-transformation is necessary to normalize the otherwise poisson-distributed income variables.

The inputs of both comparable incomes are considered to be consistent if both normal scores are within predefined boundaries (for example -1,96 and 1,96) and/or the difference between the normal scores is limited (less than 1,96).

There is an indication of bias if the input of one of the incomes for a case is situated within 'normal boundaries' (-1.96 – 1.96) but the other input is not and/or if the difference between the two normal scores differ substantially (>1.96). In fact, the entire procedure consist out of 4 steps:

- Identification of the variables to be compared.
- Log-transformations, normality checks, calculation of means and standard deviations.
- Calculation of normal scores.
- Consistency control and identification of inconsistencies.

c) Other techniques :

There was explicitly more emphasis on the above techniques but this does not imply that the 'conventional' box-plot method was not used at all. In this method input outside the interval below were considered to be outliers:

[First Quartile – 1,5 * (Third Quartile – First Quartile) ; Third Quartile +

1,5 * (Third Quartile – First Quartile)]

Furthermore and as already mentioned, where applicable and usable legal maximums and minimums were also used to some extent.

Finally, we also checked for outliers via controls on a 'case to case' base in which we maximally used information of proxi-variables like professional status and other variables. In this process manifest errors in proxi- and/or other variables associated with the income variables were also removed/corrected (for example 'the number of months').

Imputation

We did no longer make use of IVE. Instead we a) corrected (not imputed – in fact) a greater number of cases and if correction was not desirable or possible, but information on a directly comparable variable was present anyway (see section on internal information above), we b) resorted to direct imputation, via a regression model.

a) Corrections.

Corrections were also mainly done on basis of information in other comparable variables. Gross-net ratio of 12 - yearly income entered as monthly or vice versa - lead to simple corrections of the gross or the net, for example.

b) Regressions.

If correction was not desirable or possible but information on a directly comparable variable was present anyway, we resorted to direct imputation, via a regression model, of the variable for which input was missing. Below we describe how this was done for net –gross imputation, which were the most prevalent instances of that sort. The method was extended, however, to other imputations (imputations of the reference year income based on the current income, for example).

Missing values on gross income variables (PY010G, PY020G, … and components) were, if collected, imputed on the basis of the corresponding net variables (PY010N, PY020N, … and components). The implementation of this imputation procedure was quasi-similar for almost all (income) variables on which it was applied. The procedure implied 6-steps:

1) Identification of the 'reference cases' (both gross and net collected) and identification of the cases to be imputed (net collected – gross missing).
2) Calculation of the gross/net ratio for the reference cases. Cases with an extreme value on this ratio were excluded from further use in the procedure.
3) Curve estimation of the relation (regression model) between gross and net income. The best fitting model (linear, logarithmic, quadratic, exponential) was being implemented.
4) Implementation of the regression model for the reference cases to identify outliers.
5) Re-implementation of the regression model for the reference cases after removal of the outliers.
6) Actual imputation step: missing (gross) values are imputed on the basis of

a) net values and

b) the estimates for the relation between gross and net income assessed in the steps above.

In step 1 the cases of which both gross and net income were collected are identified. We refer to these cases as 'reference cases' (step 1). The relationship between their net and gross income serves as reference for the imputation of the gross incomes for the cases where only the net was collected (cases to be imputed).

To avoid bias in this imputation model atypical reference cases (both outliers and errors) were identified and removed at several steps in the procedure (step 2 and 4).

In step 2 (reference)cases for whom the ratio between gross and net income exceeded what can be considered typical for the taxation regime applicable to the income concerned, were excluded.

In the case of almost all variables the boundary value of this ratio was set at 2,5. This boundary was arbitrary chosen.

Scrutiny of the excluded cases, however, validates this value's potential to discriminate between incomes which were subjected to real(istic) taxation and outliers or errors.

The latter category seldom counted more than a few percent of the total population in the survey and their gross/net ratio often exceeded the 2,5 considerably.

Further exploration also revealed that the exclusion of these cases from the procedure results in a dramatic increase of the fit of the regression model on which the imputation is based.

In step 4 outliers in the regression model were identified and removed using default regression diagnostics.

The underlying probability model of the net-gross relation was assessed with SAS regression model or SAS logistic procedures (step 3). For most variables the linear model fitted the data well. For a few variables the fit of the quadratic model was slightly better, however. Overall, and we underline this, the fit was very good and R-squares very high (always > 0.85).

The estimates of this regression model (step 5) served as direct input for the implementation of the actual imputation (step 6).

c) Other techniques.

Although we preferred the techniques above we were in some instances forced to resort to other techniques (due to lack of information – for example).

For some cases we imputed median values calculated after categorising using relevant variables. Most of the median values imputed, were for example, calculated after categorisation for status.

1.1.2. Particular cases

Gross/Net imputations.

For a limited number of monetary variables a limited number of respondents had given only a value for the gross variant of the variable (the opposite – only net is given - occurred much more). For these cases a net value was imputed on basis of the gross using the Belgian rules of taxation. A small number of net- pensions and unemployment benefits were imputed in this way.

Imputation of 'total housing cost'

For the calculation of the total housing cost, we examined the current costs for small, average and large usage and used these amounts for both outlier detection and imputation, while taking into account other variables such as the number of household members and the household income. The cost for the water usage for example can be subdivided in subscriber money (fixed) and costs for the actual usage (variable). The cost for the usage of electricity depends largely whether the heating is electric or not: Singles in an apartment without electric heating consume approximately 600 kWh per year (~ 7 euro), while large consumers with accumulation warmth have an annual usage of approximately 20.000 kWh (~ 240 euro).

Imputation of partial unit non-response

The method chosen for Belgium was imputation of an income for each member of the household who did not answer the questionnaire. Imputation is based on the variable RB210 (basic activity status) of the individual given in the R-file. When the answer is missing or 4 (other inactive person), it is chosen not to impute any income. When available, we preferably used the longitudinal information's from 2012 for imputation. For the other cases the chosen method for imputation was imputation of a sub-category median based on age and sex. Net incomes were computed with a gross to net model, based on the imputed gross incomes.

Collection variable company Car

Since 2005, we decided to work with **the national rules of the tax authorities**. The benefit for individuals of using a company car for private goals was not directly assessed at the interview but afterwards calculated by applying the applicable taxation rules.

The fiscal benefit of all nature that a person has - due to disposition of a company car for private goals - is calculated by multiplying a fixed amount of kilometres driven for private use by a coefficient. To calculate the latest we need the fiscal cylinder capacity of the car. This fixed amount of kilometres driven for private use is for the tax authorities 5000 km if the distance home-work is less than 25 km, and 7500 if it's more than 25 km.

Since 2005, we asked directly the fiscal cylinder capacity and the distance between work and home. In case of non response of the cylinder capacity, we asked the mark, type and registration year of the car. Than we had to use an imputation method.

Imputation: To calculate the cylinder capacity, we did the following. We assumed that a company car is mostly diesel driven. We looked up for each mark, type and diesel engine what the corresponding cylinder capacity is. If we had several cylinder capacities for the type of the mark, we calculated the weighted mean of the cylinder capacity. If there is not diesel version for a type of car, we did the same logic but than for petrol.

Once we had that we could easily find the corresponding fiscal coefficient. Than we only had to multiply it by the fixed amount of kilometres driven for private use to obtain the fiscal benefit of all nature

Example:

Type of car	Fiscal cylinder capacity	Forfait	Distance home work	Fixed amount	Fiscal benefit of all nature
Smart fortwo	5	0,1898	< 25 km	5000	949 €
Smart fortwo	5	0,1898	> 25 km	7500	1423.5 €

Table 1 : example calculation company car
After we calculated the fiscal benefit of all nature for a whole year, we weighted it for respondents who didn't dispose for a whole year of the company car. **The fiscal benefit of all nature is a gross non-cash employee income**.

Imputed rent

From 2007 onwards a measure for 'imputed rent' needs to add to the data.
IN the QR-rapport for the 2007 we extensively reported on the method to calculated imputed rent. In the 2013 operation exactly the same method has been used. Results were very similar.

12.6. Adjustment

13. Comment

Top

National questionnaire is available in Circa BC at: https://circabc.europa.eu/
Please select EU SILC section and then select the folder called "06 National Questionnaire" in the library list.

Annexes:
HH questionnaire FR
HH questionnaire NL
P questionnaire FR
P questionnaire NL

Related metadata

Top

Annexes

Top

HH questionnaire FR
HH questionnaire NL
individual questionnaire FR
individual questionnaire NL