EU-SILC 2009 in Estonia: Final Quality Report

Tallinn 2011

CONTENTS

INTRODUCTION

The EU-SILC survey in Estonia started in 2004. In the first year, a sample of 6000 households was selected for the survey. These households were randomly divided into four rotational groups. According to the original rotational scheme, one of the groups was to be dropped in 2005 and another in 2006, but due to lower than expected response rate, it was decided to keep all the rotational groups in the sample. New sub-samples were introduced into the survey in 2005 and 2006. In 2007 two rotational groups from the 2004 initial sample were dropped and a new subsample was introduced. Two remaining rotational groups from 2004 were dropped in 2008. Thus the 2008 and 2009 sample structure is as originally planned: it consists of four rotational groups (in 2009 - one from each year 2006-2009). In longitudinal dimension, sample contains three panels:

- 4-year panel (rotation group 2, i.e. DB075=2) started in 2006;

- 3-year panel (rotation group 3, i.e. DB075=3) started in 2007;

- 2-year panel (rotation group 4, i.e. DB075=4) started in 2008;

New part introduced in 2009 constitutes rotational group 1 (DB075=1).

Report follows as much as possible recommendations of two documents: Regulation No 28/2004 as regards the detailed content of intermediate and final quality reports and Technical document on intermediate and final quality reposts (EU-SILC 132/04).

Report is compiled by Erika Taidre and Julia Aru, Statistics Estonia.

1. COMMON LONGITUDINAL EUROPEAN UNION INDICATORS BASED ON THE LONGITUDINAL COMPONENT OF EU-SILC

At persistent at-risk-of-poverty rate by gender and age (%)

This indicator is computed as the percentage of the population living in households where the equivalised disposable income was below the 60% threshold for the current year and at least two out of the preceding three years.

Table 1.1. At persistent at-risk-of-poverty rate by gender and age, 2009

Age groups	Total	Males	Females
Total	12,8	11,5	13,8
Less than 18 years	13,5	11,6	15,3
From 18 to 24 years	8,2	8,3	8,1
From 18 to 64 years	9,2	10,1	8,4
18 years or over	12,7	10,6	14,3
From 25 to 49 years	7,9	8,2	7,6
From 50 to 64 years	12,2	15,0	9,9
65 years or over	13,2	15,6	10,8

2.1. Sample design

In sections 2.1.1 to 2.1.7 we describe the sample design of first wave of each rotational group. Sample design has stayed the same in all years, so we describe it only once.

2.1.1. Type of sampling design

The design used is one-stage stratified unequal probability sampling of households, with a household selected with probability proportional to the number of persons aged 14+ in it. It is because a sample of persons aged 14+ (so called address-persons) is selected first with equal probabilities within strata, and then the household of the selected person is identified, and all eligible persons in the household are interviewed. Stratification is done by geographical region (see 2.1.3).

2.1.2. Sampling units

One stage sampling design was used. Households are regarded as sampling units although selection was made using the sample of address-persons.

2.1.3. Stratification and sub-stratification criteria

Geographical stratification was used. The counties (and capital Tallinn) were grouped into three strata by the population size:
1. big counties: Tallinn, Harju (excluding Tallinn), Ida-Viru, Lääne-Viru, Pärnu, Tartu;
2. small counties: Jõgeva, Järva, Lääne, Põlva, Rapla, Saare, Valga, Viljandi, Võru;
3. Hiiu County formed a separate stratum as the smallest county with the population size times smaller of the next smallest.

2.1.4. Sample size and allocation criteria

Inclusion probabilities of address-persons in different strata and years are shown in Table 2.1.

Table 2.1. Stratification of the new part of the sample by counties, Estonian EU-SILC

Stratum h	Counties	Sampling fraction, %			
		2006	2007	2008	2009
Large	Tallinn, Harju, Ida-Viru, Lääne-Viru, Pärnu, Tartu	0.17	0.17	0.17	0.16
Small	Jõgeva, Järva, Lääne, Põlva, Rapla, Saare, Valga, Viljandi, Võru	0.32	0.31	0.31	0.30
Hiiu	Hiiu	0.91	0.85	0.89	0.84

2.1.5. Sample selection schemes

Systematic sampling of address-persons in each stratum separately. For households this procedure results in unequal probability sampling with inclusion probabilities proportional to household size (number of persons aged 14+ in it).

2.1.6. Sample distribution over time

Fixed income reference period was used and therefore the sample was not principally divided into months or weeks. Although for the convenience of fieldwork administration, the sample was distributed along the fieldwork period. Table 2.2 shows allocation of new part and longitudinal part of the sample into months of fieldwork period in different years. This allocation is prior to the start of fieldwork and actual month of interview may be different. The table also shows changes in official fieldwork period in 2006-2009.

Table 2.2. Allocation of sample into months, 2006 – 2009

Year	New part					Longitudinal part				
	Feb	Mar	April	May	June	Feb	Mar	April	May	June
2006		25%	25%	25%	25%		25%	26%	26%	23%
2007		35%	29%	36%	0%		26%	26%	26%	21%
2008	34%	33%	33%	0%		27%	27%	27%	19%	
2009	34%	34%	32%	0%		24%	27%	23%	26%	

For the households in longitudinal part, the survey month was chosen as close as possible to the month the household was interviewed in preceding year.
In every year, there are ca 4-6% of households that are interviewed after the official end of fieldwork period, due to lack of interviewers in some areas.

2.1.7. Renewal of sample: Rotation groups

The sample consists of four rotational groups (one from each year 2006-2009). In longitudinal dimension, sample contains three panels:

- 4-year panel (rotation group 2, i.e. DB075=2) started in 2006;
- 3-year panel (rotation group 3, i.e. DB075=3) started in 2007;
- 2-year panel (rotation group 4, i.e. DB075=4) started in 2008;

New part introduced in 2009 constitutes rotational group 1 (DB075=1).

2.1.8. Weightings

Weighting scheme was generally in line with documents V. Verma „EU-SILC weighting procedures: an outline" and J.-M. Museux „Weighting and estimation for the EU-SILC rotational design".

Longitudinal database of 2009 contains three subsamples of households:

S4	Households introduced into the survey in 2006 and their split-offs. Year 2009 is their fourth year in the survey.
S3	Households introduced into the survey in 2007 and their split-offs. Year 2009 is their third year in the survey.
S2	Households introduced into the survey in 2008 and their split-offs. Year 2009 is their second year in the survey.

Sub-sample of households introduced into the survey in 2009 doesn't contribute to longitudinal database of 2009. For them, we describe the computation of design factor only (section 2.1.8.1).

2.1.8.1. Design factor

Design weights DB080 are defined only for the first year of each sub-sample.

Computation of design weights draws on following considerations resulting from the sampling design: Inclusion probabilities for a household depend on how many possibilities are there to reach this household. As address persons are used and inclusion probabilities are equal for all members aged 14+ of the household, the inclusion probability of a household is proportional to selection probability of an address-person in his stratum and to the number of members aged 14+ in the household. Accordingly, design weights for households are calculated as

$$d_h = \frac{N_g}{n_g \, p_h}$$

where n_g is the number of address-persons selected in stratum g, N_g is the number of persons aged 14+ in stratum g as stated in the Population Register (PR) at the moment of sample selection and p_h is household size, i.e. number of persons aged 14+ in the household.

Design weights can be calculated for both responded and non-responded. For non-responders, p_h is the number of persons aged 14+ registered to the address of address-person according to the Population Register. If it was greater that 8, household size was stochastically imputed (with county as an auxiliary variable). A household living in one address can be handled as a cluster. Inclusion probabilities are thus equal for all members aged 14+ of the household.

2.1.8.2. Non-response adjustments

Longitudinal weights RB062, RB063 and RB064 are calculated on the basis of base weights of each sub-sample. In the year a sub-sample is introduced into the survey, a base weight is the design factor (section 2.1.8.1) , corrected for non-response (logistic model with region as well as age and sex of

address-person) and calibrated to population totals (section 2.1.8.3). After that from year to year base weights for each sub-sample are only corrected for attrition. That means, that the weights of sub-sample introduced into the survey in 2006 were corrected 3 times etc as shown below.

S4: 2006 -> 2007 ->2008 ->2009 (3 corrections)
S3: 2007 -> 2008 -> 2009 (2 corrections)
S2: 2008 -> 2009 (1 correction).

Every year, correction for attrition is done using the same principles:
Correction for attrition is brought out independently for each sub-sample s_2, s_3, s_4. Persons and households no longer in scope are excluded prior to the correction as they are not considered as non-response. Correction for attrition was done with the help of logistic regression model with tenure status, household equivalised income, number of children in the household, urbanization status and county of place of residence, age, gender as auxiliary variables. The model was weighted on the base weights of previous year. According to the model response probability r_i of person i, given he/she had responded in previous year, was estimated.

2.1.8.3. Adjustment to external data

Calibration was performed using population data from Estonia Population Register according to sex, age-groups, county and urbanization. Persons in institutions were deducted from population figures prior to calibration.

2.1.8.4. Final longitudinal weight

The basis for calculating longitudinal weights is 2009 base weights for sub-samples corrected for non-response. There are three longitudinal sets of interest in year 2009:

- Longitudinal set of two year duration, involving data from year 2008 and 2009. Longitudinal weight to be used for this set is RB062. All sub-samples S2 and S3 and S4 contribute to this set. Base weights of S2, S3 and S4 were multiplied by a factor according to the size of a sub-sample to combine the subsamples into one set.

- Longitudinal set of three year duration, involving data from year 2007, 2008 and 2009. Longitudinal weight to be used for this set is RB063. Only sub-samples S3 and S4 contribute to this set. Base weights of S3 and S4 were multiplied by a factor according to the size of a sub-sample to combine the subsamples into one set.

- Longitudinal set of four year duration, involving data from year 2006, 2007, 2008 and 2009. Longitudinal weight to be used for this set is RB064. Only the sub-sample S4 contributes to this set.

Children born between interviews of 2008 and 2009 and persons who moved into sample household from outside received zero weight.

2.1.8.8. Final household cross-sectional weight
Final cross-sectional households weights DB090 for years 2006-2009 were recalculated in a way to correspond only to sub-samples contributing to corresponding year in longitudinal component.

2.1.9. Substitution
No substitution has been used.

2.2. Sampling errors

The following table reports the mean, the number of observations (before and after imputations) and the standard error for different income components. Zeroes are excluded from calculations for income components, but included for total household income. Net/gross conversion is also considered as imputation. All four rotational groups are used for the table.

Standard errors were calculated using Jacknife Repeated Replication method with programs provided by Eurostat.

Number of observations before imputation for total variables is very small since these variables include sub-components that are totally imputed and also there is allways some amount of net/gross conversion.

Table 2.3. Number of observations and standard error of different income components, 2009

| Income components | Mean | Number of observations | | Standard error |
		Before imputation	After imputation	
Total household gross income (HY010)	207004	92	4965	3326
Total disposable household income (HY020)	174334	824	4965	2577
Total disposable household income before social transfer other than old-age and survivors' benefits (HY022)	161672	1674	4965	2485
Total disposable household income before social transfers including old-age and survivors' benefits (HY023)	135857	1995	4965	2592
Net income components at household level				
Imputed rent (HY030N)	30051	0	4820	360
Income from rental of a property or land (HY040N)	20868	107	136	3981
Family/ children related allowances (HY050N)	20655	2	1932	1158
Social exclusion not elsewhere classified (HY060N)	8251	0	108	676
Housing allowances (HY070N)	5332	0	88	684
Regular inter-household cash transfers received (HY080N)	22664	0	203	1631
Alimonies received (compulsory + voluntary) (HY081N)	23134	0	117	1749
Interest, dividends, profit from capital investments in incorporated business (HY090N)	3483	10	1940	802
Interest repayments on mortgage (HY100N)	31197	0	469	1363
Income received by people aged under 16 (HY110N)	4686	46	80	1223
Regular taxes on wealth (HY120N)	598	0	3563	21
Regular inter-household cash transfers paid (HY130N)	21197	0	209	1881
Alimonies paid (compulsory	28715	0	67	3574

| Income components | Mean | Number of observations | | Standard error |
		Before imputation	After imputation	
+ voluntary) (HY131N)				
Tax on income and social contributions, net (HY140N)	.	.	.	.
Repayments/ receipts for tax adjustment (HY145N)	-4342	1954	2342	318
Gross income components at household level				
Imputed rent (HY030G)	32609	0	4827	372
Income from rental of a property or land (HY040G)	26415	22	136	5040
Family/ children related allowances (HY050G)	23908	1606	1932	1455
Social exclusion not elsewhere classified (HY060G)	8251	82	108	676
Housing allowances (HY070G)	5332	77	88	684
Regular inter-household cash transfers received (HY080G)	22664	189	203	1631
Alimonies received (compulsory + voluntary) (HY081G)	23134	106	117	1749
Interest, dividends, profit from capital investments in incorporated business (HY090G)	4296	61	1940	1013
Interest repayments on mortgage (HY100G)	31197	0	469	1363
Income received by people aged under 16 (HY110G)	4732	0	80	1252
Regular taxes on wealth (HY120G)	598	3438	3563	21
Regular inter-household cash transfers paid (HY130G)	21197	203	209	1881
Alimonies paid (compulsory + voluntary) (HY131G)	28715	66	67	3574
Tax on income and social contributions, gross (HY140G)	40714	0	3857	881
Net income components at personal level				
Employee cash or near cash income (PY010N)	108254	5387	6450	1436
Non-cash employee income (PY020N)	10251	680	1406	509
Company car (PY021N)	19536	0	236	574
Contributions to individual private pension plans (PY035N)	6004	685	790	226
Cash benefits or losses from self employment (PY050N)	18451	153	737	1650
Value of goods produced for own consumption (PY070N)	1676	0	5688	33
Pension from individual private plans (PY080N)	14554	6	7	3854
Unemployment benefits (PY090N)	19649	125	226	2044
Old-age benefits (PY100N)	52599	1925	2848	243

Income components	Mean	Number of observations		Standard error
		Before imputation	**After imputation**	
Survivors' benefits (PY110N)	18601	93	96	895
Sickness benefits (PY120N)	4177	873	1397	182
Disability benefits (PY130N)	28242	0	883	547
Education-related benefits (PY140N)	10352	0	355	1458
Gross income components at personal level				
Employee cash or near cash income (PY010G)	133185	313	6450	1870
Non-cash employee income (PY020G)	12976	0	1406	645
Company car (PY021G)	24730	0	236	726
Employer's social insurance contributions (PY030G)	43352	0	6189	600
Contributions to individual private pension plans (PY035G)	6004	685	790	226
Cash benefits or losses from self employment (PY050G)	22393	493	812	1928
Value of goods produced for own consumption (PY070G)	1676	0	5688	33
Pension from individual private plans (PY080G)	14554	0	7	3854
Unemployment benefits (PY090G)	23414	54	226	2478
Old-age benefits (PY100G)	53790	0	2848	279
Survivors' benefits (PY110G)	18623	0	96	903
Sickness benefits (PY120G)	5287	0	1397	230
Disability benefits (PY130G)	28242	863	883	547
Education-related benefits (PY140G)	10352	343	355	1458

The following table provides the same information for the equivalised disposable income broken down by sex, age groups and household size.

Table 2.4. Number of observations and standard error of mean equivalised disposable income, 2009

	Mean	Number of observations		Standard error
		Before imputation	**After imputation**	
Subclasses by household size				
1 household member	85823	419	1072	3595
2 household member	126459	69	982	3589
3 household members	114237	23	1391	2393
4 and more	116601	313	1520	3148
Population by age group				
<25	112365	104	4616	1902
25-34	145959	84	1306	5128
35-44	120032	71	1822	3210
45-54	111672	164	2036	1956
55-64	112554	266	1476	2947
65+	80438	656	2286	1372
Population by sex				
Male	116395	583	6361	1867
Female	109657	762	7181	1442

2.3. Non-sampling errors

2.3.1. Sampling frame and coverage errors

Sampling frame for selection of the new part of the sample was the Population Register of Estonia. This is the document-based register of Estonian citizens and those having a living permission. Records of the register are updated both in real-time and regularly from administrative sources. The register data originates from local governments, civilian registry offices, county councils, courts, Citizenship and Migration Board and other governmental organisations.

Frame error is considered to be an over-coverage error if address-person did not actually belong to target population, i.e.

- was dead;
- had moved to another county;
- stayed in an institution permanently (had been there over half a year);
- was surveyed through one of his/her household members;

All households classified under DB120=23 are considered to constitute over coverage error. The amount of this error in the new part of the sample in 2009 was 99 households, which makes the proportion of the over-coverage in the new part of the sample 4.3% and of the whole sample 1.5% (Table 2.5).

Since there is no registration law in Estonia, people do not need to show their actual addresses in the Population Register. For that reason the register contains some amount of records without any address and for some part of records the address shown is not correct. Records without an address or incomplete address were dropped out of the register before selecting the sample (for example, in 2009 3.3% of all records referring to persons aged 14+ were dropped before selecting the sample).

In the new part of the sample of 2009 there were 261 address-persons those address in the population register was definitely wrong and no information on new address could be obtained from neighbours. According to national classification, this includes the following reasons for non-contact:

- Address-person does not live at given address, no information on new address available;
- Address-person has moved to another address, no information on new address available;
- Given address does not exist.

It does not seem reasonable to assume that these persons do not belong to target population nor constitute frame over-coverage. Above mentioned reasons for non-contact are currently classified under DB120=21.

Due to absence of registration law in Estonia, there is also some under-coverage of persons and households present in the population register. Investigations made by the Sampling Working Group of HBS in 1999 showed that on average under-coverage of addresses in the population register may reach 5-6%. Degree of under-coverage of households is much more difficult to asses, since even if a person is missing from Population Register or his/her address is incorrect or not precise enough, a household could be reached through another household member. Assuming that all persons living permanently in Estonia are registered in the Population Register and considering the amount of imprecise addresses in PR, the under-coverage of households may be at most 1-1.5%.

Table 2.5. Reasons for over-coverage in the new part of the sample, 2009

Frame error	Number of households	Proportion in the frame error (%)
Total, of which	99	100
Address person was dead	18	18.0
Address person has left Estonia	69	70.0
Address person was staying in an institution	12	12.0
Address person was surveyed through one of his/her household members	0	0.0

2.3.2. Measurement and processing errors

2.3.2.1. Measurement errors

The measurement errors can stem from the questionnaire (its wording, design etc), the interviewees, the interviewers and the data collection method. While it is impossible to avoid this type of errors completely, steps were taken to reduce them as much as possible.

The questionnaires were drawn up in 2004 following the international practices in collecting income data. Also, where possible questions from the existing surveys carried out by the Statistics Estonia and known to be valid and reliable, were used. Pilot surveys were carried out in 2002 and 2003 with the main aim of testing the questionnaires. The results were thoroughly analysed and feedback sessions with interviewers were carried out. The questionnaires were modified accordingly for the use in the main operation.

The questionnaire has been modified every year using the experience from the previous waves of the survey. The main modifications in 2005 concerned self-employment income, child-care, change of job and different types of social insurance payments.

The main modifications in 2006 concerned employee income and self-employment income where income brackets were added to those unable or unwilling to provide a precise answer, the question on income from bank accounts was more fleshed out and income brackets were added. The questions on child-care, family benefits and unemployment benefits were also improved.

Other notable modifications in the 2006 questionnaire were as follows:

a) Clarifying the type of work contract

b) Making it easier for the respondent to declare their incomes by giving both month/year and gross/net options

c) Providing intervals to report income and profit

d) Making benefits lists more precise by listing possible benefits (unemployment and family) and thus running more accurate primary tests and improving reporting everything

e) Breaking down non-monetary income components into separate questions and checking if their value was counted into the original reported income

f) Removing redundant questions in use for filtering and adding new filter questions that proved necessary in the experience of previous years

g) Clarifying self-employment income, change of job and different types of social insurance payments' questions.

The main modifications in 2007 concerned the inclusion of questions about own consumption, and those necessary for calculating imputed rent in the household questionnaire. In the personal questionnaire the most important improvements concerned the inclusion of the 'education obtained since previous interview' for the longitudinal respondents, simplifying the salary questions by giving the respondent more options for naming it in time and gross/net categories, adding questions about the use of a company car and other non-cash employee income and developing income questions for self-employed persons and entrepreneurs.

Other notable modifications in the 2007 questionnaire were as follows:

1) Personal ID number of household member responsible for dwelling split between owners and tenants

2) Adding in cost of utilities

3) Developing mortgage payment and interest questions

4) Simplifying for the respondent questions about income from rent of property or land

5) Updating lists of social benefits and including question about alimony paid and received

6) Adding in questions to filter out information on incomes, employer social contributions etc for temporary workers and entrepreneurs.

In 2008, the questions used to determine a respondent's level of education were improved. Previously a person had to choose their level of education from a long list of official names, resulting in considerable errors, especially with vocational education. In 2008 these questions were redesigned for more accuracy and less respondent-induced errors.

Other notable modifications in 2008 concerned the following variables:

1) Questions about managerial duties for current and last job were added

2) Socio-economic statuses were prefilled for respondents who had answered the personal questionnaire the previous year for the months they had already provided answers for

3) An additional question was added regarding pensions paid by the local government and the conscript allowance paid to young men serving time in the armed forces

4) The restriction in activities due to health problems was reformulated to better meet Eurostat's guidlines

5) Households in the panel were asked if their dwelling, its number of rooms or amount of square feet had changed since the previous year and if not the information about the dwelling was prefilled from the previous year's data to reduce respondent burden

6) A question concerning the condition of the dwelling was added

7) The ownership of the dwelling question was split so that two variables would no longer be asked in the same question, along with a new question about the type of rental contact (oral or written) for the dwelling

8) Questions about a household's inability to make rent, utilities, mortgage, loan or hire purchase (re)payments on time were added

9) Mortgage information was filtered to be asked only for the main dwelling's construction, purchase or repairs

10) Questions about income earned from abroad were added into the questionnaire for two years

In 2009, the questions used to determine a respondent's level of education were improved. Previously a person had to choose their level of education from a long list of official names, resulting in considerable errors. In 2009 these questions were redesigned for more accuracy and less respondent-induced errors.
Other notable modifications in 2009 concerned the following variables:

1) The question about ability to make ends meet was reformulated to better meet Eurostat's guidelines

2) The question about lowest monthly income to make ends meet reformulated to better meet Eurostat's guidelines

3) The question about financial burden of the total housing cost reformulated to better meet Eurostat's guidelines

4) The question about burden of the repayment of debts from hire purchases or loans reformulated to better meet Eurostat's guidelines

5) The question about noise from neighbors or from the street reformulated to better meet Eurostat's guidelines

6) The question about pollution, grime or other environmental problems reformulated to better meet Eurostat's guidelines

7) The questions about material deprivation were added

To reduce the measurement error stemming from the data collection method, CAPI was introduced as a data collection method from 2005 operation onwards. The main source of errors in the questionnaires in the 2004 operation resulted from routing mistakes and inconsistencies between questions. CAPI eliminates the former type of error and considerably reduces the latter, as the data-entry program includes several checks. As a result, the need to make callbacks declined and the

quality of the information obtained this way was increased due to a remarkably faster pace of the whole cycle.

Secondary data editing procedures were improved by getting skilled personnel to work through the logical inconsistencies with the interviewers. This was aided by secondary logical checks in SAS after the questionnaires had arrived into Statistics Estonia's databases. All the secondary testing was done during the fieldwork period which officially ended when all inconsistencies had been resolved. After the fieldwork period, tertiary data checks were run to check for longitudinal inconsistencies, such as different jobs one year and the next while the respondent claims not to have changed jobs.

In 2005, all interviewers attended a two-day training session in small groups. In 2006, the training session lasted four days and interviewers were divided into four smaller groups to allow for a more efficient learning environment. During the training sessions mistakes from the previous years were discussed, followed by a separate block about seeking out the previous waves' respondents and assigning household/personal numbers to new and split off households and their members. One whole day was dedicated to going through the questionnaires and their tough spots with the interviewers. New interviewers also underwent training of general IT skills and data-entry program specific instruction in order for them to be able to work with CAPI. Interviewers were also tested as in previous years on their factual knowledge as well as simulated interview situations.

In 2009 the training program was conducted in 4 smaller groups of about 15 people, similarly to 2006 - 2008, with more emphasis on practical work and discussion of mistakes from previous years as in preceding interviewer trainings. All returning interviewers attended a day long training session. During the training, the EU-SILC team briefed the interviewers on all renewals in the questionnaires, discussed previous years' errors and tracing and specifics of assigning household and person numbers in the longitudinal survey. Practical work sessions were conducted in groups of five and each interviewer had to conduct a model interview regarding for a simulated situation using their laptop. At the end of the training session, each interviewer received personal feedback about their mistakes the previous year. Interviewers new to EU-SILC attended a 2 day training session, which included a thorough overview of questionnaires and practical exercises as well as all the topics covered with returning interviewers.

In 2009 and 2008 overall, 53 interviewers were responsible for conducting the interviews. The household– interviewer ratio was 90 households per interviewer.

In 2007 ooverall, 58 interviewers were responsible for conducting the interviews. The household (gross sample) – interviewer ratio was 82 households per interviewer in 2007. The ratio was 103 households per interviewer in 2004, 90 households per interviewer in 2005, 96 households in 2006 and 82 households in 2007.

2.3.2.2. Processing errors

In 2004, the interviews were carried out using PAPI and the data was entered centrally. The data-entry program was written in Blaise and contained most of the logical checks. The checks included, but were not limited to routing checks, consistency between different answers and upper and lower bounds for income variables. The most common mistake made by the interviewers was failing to mark an answer to one or more question or sub-questions (74% of all mistakes). Other most frequent types of error were marking the answer so that the correct answer remained unclear (for example by using wrong codes) and inconsistencies between answers provided to different questions (accounting to 7% and 6% of all mistakes respectively). The questions that were most prone to mistakes were:

1) Enforced lack of durables (missing answers due to unclear questionnaire layout);
2) Family benefits (inconsistent with household composition);
3) Relationship between household members (implausible relationships recorded in the household matrix);
4) Number of years in employment (inconsistencies with the time of taking up the first job);
5) Limitations in daily activities for health reasons (sometimes skipped despite there being no routing);
6) Calendar of activities (inconsistencies with other data).

20% of all questionnaires contained one error or more. Social Statistics Department personnel checked all errors discovered in the course of data entry. Errors that could be corrected using other data in the questionnaire or external data were corrected in office. The errors that could not be solved this way were forwarded to the interviewers' network, who consulted with the interviewer and when necessary made call-backs to the household. 20% of all errors (4% of all questionnaires) were forwarded for call-backs.

Not all checks could feasibly be implemented during the data-entry; so further data cleaning was carried out at a later stage using SAS. These checks were mainly targeted to detecting extreme income values and data-entry mistakes. Finally, the Eurostat data-checks were also implemented.

In 2005, the checking of the data consisted of 3 stages: the data-entry checks during interview, additional in-office checks during fieldwork and later data cleaning.

As mentioned above, the data for 2005 operation was collected using CAPI. The data-entry program was written in Blaise and contained most of the checks. This way, most of the errors could already be corrected during the interview. The data-entry controls were of 4 major types:

1) Checks of consistency between different answers. These included, but were not limited to following instances:

   a. whether a household or a person who according to other data should have received a certain type of income reported it or not (e.g. whether households with children received family benefits, employed persons received wages and so on);

   b. whether answers provided to different non-monetary deprivation items agreed with each other;

   c. whether the relationships in the household matrix were consistent with each other as well as with the age and sex of the household members;

   d. whether the difference between the starting and finishing time of the interview was too short or too long and so on.

2) Lower and upper bounds of income variables. These checks were developed with regard to data collected in the previous wave as well as administrative information.

3) Tracing checks. These controls were implemented to ensure that all split-off households and new household members were assigned correct split numbers and person numbers respectively.

4) Checks with information from the previous year. These controls concerned demographic data, information on educational level and labour status as well as the calendar of activities.

The in-office staff promptly checked the questionnaires that were electronically transmitted to the central office. This stage included following controls:

1) All the errors suppressed by interviewers were activated and checked;

2) All remarks made by interviewers in the data entry-program were read through and where necessary relevant corrections were made.

3) All split-off households as well as all households from which at least one member had left were scrutinised one by one.

4) Demographic information in the interviewers' reports, which were still filled out on paper, was compared to the data recorded in the electronic questionnaires.

5) Additionally, a few questions (child care, place of residence) had to be screened due to mistakes in the data-entry program.

The third and final stage involved later in-office data cleaning. The controls implemented at this stage involved further checks of data consistency and of extreme income values and as a final step the Eurostat data-checks. The checks of data consistency were mainly concerned with non-income variables, such as education. Also extreme values for all income components as well as total income were checked.

As in 2005, the data for the 2006 operation was collected using CAPI. In case of each consistency check in the data entry programme the interviewer had to check if the situation was correct, if not, correct it, if yes, make an explanatory remark. All remarks and suppressed consistency errors were manually checked during the secondary in-office data editing procedure.

The primary data-entry consistency controls were the same as in 2005 with some additions:

1) Under checks of consistency between different answers. Some new checks were added for the following instances:

   a. Whether benefits reported to have been received were logical in the age and gender dimensions. For instance student benefits for over 50 year-olds, income taxes for under 15 year-olds, maternity leave and childbirth allowances for men etc.

   b. Whether an educational level attained was possible below a certain age.

   c. whether reported taxes or medical benefits received were consistent with income

   d. Membership in pension plans checked by year of birth to see if legally bound to have joined pension pillar.

   e. Checks for correct survey area, interviewer code and personal numbers matching household numbers.

2) Checks for correct survey area, interviewer code and household and personal numbers matching.

3) Checks not allowing for occupations to be written on too general a scale for coding. (e.g. salesperson, cleaner)

As in the previous year, the in-office staff checked the questionnaires that were electronically transmitted to the central office. In addition to the previous year's controls, six new ones were added:

1) All category 'other' answers were gone through to see if they could be classified under one of the given options.

2) Additionally paid income tax was checked in-household to check for double-reporting.

3) Errors in coding.

4) Study benefits were checked by possibility of obtaining them in the school the respondent attended and legally set amounts.

5) Consistency between time reported working under socio-economic status and months that salary was received. Also time spent in prison.

6) Reported amounts of family benefits were checked compared with eligibility based on the structure of the family and benefit levels set out in legislation.

In 2006, 5685 household and 13418 personal questionnaires arrived in the Statistics Estonia base. Of them 1031 household and 2734 personal questionnaires had mistakes in them. This means that interviewers made mistakes in about a fifth of all the questionnaires: 18% of household and 20% of personal questionnaires were imperfect.

In all the materials combined a total of 5587 mistakes were registered, 4943 of those were counted as interviewers' errors. Mistakes were sent for clarification and specification for two reasons:

1) The situation was so indistinct that the data could not be made sense of based on existing information (such as info from previous waves, other information in the form)

2) The errors were repetitive and through clarification interviewers received additional training. In this case fieldwork managers were consulted separately.

All mistakes found through the secondary in-office data editing were put up in a shared excel table, and had to be clarified with the interviewer or interviewee by the end of the fieldwork period. This was done in co-operation by the EU-SILC team and the Data Gathering department's Fieldwork Managers.

The third stage of data checks was carried out similarly to the 2005 operation.

In 2007, the Blaise consistency checks underwent further extensive development, with many new logical checks creating error messages in described situations put in place.

The primary data-entry consistency controls belonged to the same 4 major types as in previous years. Some new controls included:

1) Under checks of consistency between different answers, whether households not in an electrical grid could have electrical appliances, or households not connected to a sewerage could have a shower etc ;

2) Checks for goods produced for own consumption, for instance their quantities;

The number of primary consistency errors dropped dramatically in 2007 after a special training given to interviewers about the necessity to give an explanatory remark for every consistency check that pops up in Blaise. In 2006 there had been a total of 5654 errors, in 2007 the number had fallen to 1677. There was no such training at the end of 2007, and in 2008 there was a small increase in the number of errors, totaling at 1779. Out of all the errors in 2007, 60% (998 cases) required callback and clarification with the interviewer or interviewee.

In 2008 had to do with interviewers not correcting their mistakes after an error code had prompted them to do so, not making remarks when they were needed, and the use of category 'other', while a suitable category existed. In 2007 not correcting mistakes and using 'other' when unnecessary were also the most common error types. Not making remarks, however, has returned as a major problem in 2008, as it had been in 2006 before a special training devoted to the subject. This training, undergone in 2007 for all interviewers, decreased the frequency of this problem considerably, but the effect lasted for only a year.
The call-back rates usually range from 65-85%, with the use of category 'other' as a major exception. Data entry mistakes have decreased to zero thanks to the continuing development of primary logical checks in the data entry program.

As can be seen from table 2.6, in 2009 similar the previous year, the most common types of errors had to do with interviewers not correcting their mistakes after an error code had prompted them to do so, not making remarks when they were needed, and the use of category 'other', while a suitable category existed.
The call-back rates usually range from 65-85%, with the use of category 'other' as a major exception. Data entry mistakes have decreased to zero thanks to the continuing development of primary logical checks in the data entry program.

Table 2.6. Interviewer errors and their processing, 2009

Type of error	Number of errors detected	Share of errors requiring a call-back
No remark explaining unusual situation	376	86%
Interviewer made an error, but did not correct it	497	57%
Interviewer's remark does not explain unusual situation	3	100%
Data not sufficient for coding	92	52%
Starting and finishing times recorded incorrectly	1	100%
Use of category Other, while a suitable category exists	571	26%
In-office checks	158	78%
Interviewer has misunderstood a question	95	19%
Data entry mistake	0	-
Not interviewers error	146	3%
Total	1939	

The secondary in-office data checks and tertiary checks were the same as in the previous waves.

2.3.3. Non-response errors

2.3.3.1. Achieved sample size

The following numbers are taken from the longitudinal database for the year 2009 (i.e. contains only 2-, 3- and 4-year panels of 2009, not new sample).

Table 2.8. Achieved sample size by panel

	2-year panel (rotation gr 4)	3-year panel (rotation gr 3)	4-year panel (rotation gr 2)
Number of households for which an interview is accepted for the database (DB135=1)	1249	1221	1290
Number of persons 16 years or older who are members of the households for which the interview is accepted for the database, of which	2825	2780	3035
sample persons (RB100=1)	2740	2660	2832
co-residents (RB100=2)	85	120	203

2.3.3.2. Unit non-response

Table 2.9. Household wave response rate by panel

Number	Total (rotation groups 2, 3, 4)	2-year panel (rotation gr 4)	3-year panel (rotation gr 3)	4-year panel (rotation gr 2)
A. Number of hhs passed on to 2009 from 2008 (DB110=1,2,3,4,5,6,7,11,10)	4331	1447	1425	1459
B. Number of hhs added to the sample due to spli-off (DB110=8)	159	56	56	47
C. Number of hhs out-of-scope or non-existent (DB110=3,4,5,6 or DB120=23)	58	21	16	21
D. Number of hhs with interview accepted (DB135=1)	3760	1249	1221	1290
E. Number of hhs with interview accepted in 2008	4247	1447	1386	1414
Wave response rate D/(A+B-C)	0,85	0,84	0,83	0,87
Achieved sample size ratio (D/E)	0,89	0,86	0,88	0,91

Table 2.10. Household longitudinal follow-up rate and follow-up ratio by panel

Number	Total (rotation groups 2, 3, 4)	2-year panel (rotation gr 4)	3-year panel (rotation gr 3)	4-year panel (rotation gr 2)
A. Number of hhs passed on to 2009 from 2008 (DB110=1,2,3,4,5,6,7,11,10)	2872	1447	1425	NA[1]
B Number of hhs out-of-scope or non-existent among those in row A (DB110=3,4,5,6)	37	21	16	NA
C. Number of hhs to be passed on to 2010	2474	1250	1224	NA

---

[1] Rotational group 2 will be dropped from the sample in 2010, longitudinal follow-up rate not applicable.

D. Number of split-off hhs to be passed on to 2010	72	38	34	NA
Longitudinal follow-up rate C/(A-B)	0,87	0,88	0,87	NA
Follow-up ratio (C+D)/(A-B)	0,90	0,90	0,89	NA

Table 2.11. Wave response rate for persons by panel

	Total (rotation groups 2, 3, 4)	2-year panel (rotation gr 4)	3-year panel (rotation gr 3)	4-year panel (rotation gr 2)
A. Number of sample persons aged 16 and over passed on to 2009 from 2008	9838	3303	3242	3293
B. Number of persons (among those in A) belonging to households no longer in scope (DB110 = 3,4,5,6)	83	32	21	30
C. Number of persons (among those in A) no longer in scope in existing households (RB110=6 or RB120=2,3).	123	36	43	44
D. Number of sample persons completed the personal interview (RB100=1 and RB250=11 to 13).	8183	2727	2639	2817
E. Number of completed personal interviews (RB250=11 to 13)	8581	2809	2756	3016
F. Number of completed personal interviews in 2008	10761	3218	3105	3308
Wave response rate, D/(A-B-C)	0,85	0,84	0,83	0,88
Achieved sample size ratio, E/F	0,80	0,87	0,89	0,91

Since longitudinal component does not contain any other sample persons than those passed on to 2009 from 2008, **longitudinal follow-up rate for persons** coincides with wave response rate for persons. Non-responding persons in responding households have been imputed by full-record imputation and marked RB250=14, so causes of non-response are not indicated in the database.

In reporting these non-response rates we assume that all non-contacted households other than those coded as DB120=23 are in fact existing. This seems to be a reasonable assumption since codes DB120=21 and DB120=22 include the following non-contact reasons according to national classification (see the meaning of the term "address-person" in Intermediate Quality Report):

DB120=21	**DB120=22**
▪ Address-person does not live at given address and no information is available on new address ▪ Address-person has moved to another address, no information on new address available ▪ Given address does not exist ▪ Address can be located, but no contact can be made since nobody is at home	▪ The house given is located but given address can not be accessed (due to locked doors or gates, etc) ▪ Address of address-person can not be accessed due to poor weather conditions etc

2.3.3.3. Distribution of households by household status (DB110), by record of contact at address (DB120), by household questionnaire result (DB130) and by household interview acceptance (DB135).

	Total	DB110										
		1	2	3	4	5	6	7	8	9	10	11
Total	4490	3992	72	8	20	30	0	167	159	0	0	42
Rotation gr 2	1506	1344	26	4	7	10	0	54	47	0	0	14
Rotation gr 3	1481	1310	20	1	6	9	0	64	56	0	0	15
Rotation gr 4	1503	1338	26	3	7	11	0	49	56	0	0	13

RECORD OF CONTACT AT ADDRESS

	Total	DB120				
		11	21	22	23	Missing
Total (DB110 = 2,8,9)	231	178	52	1	0	0
Rotation gr 2	73	52	21	0	0	0
Rotation gr 3	76	59	16	1	0	0
Rotation gr 4	82	67	15	0	0	0

HOUSEHOLD QUESTIONNAIRE RESULT

	Total	DB130					
		11	21	22	23	24	Missing
Total (DB110=1 or DB120=11)	4170	3765	296	31	43	35	0
Rotation gr 2	1396	1292	72	8	17	7	0
Rotation gr 3	1369	1223	108	4	15	19	0
Rotation gr 4	1405	1250	116	19	11	9	0

HOUSEHOLD INTERVIEW ACCEPTANCE

	Total	DB135		
		1	2	Missing
Total (DB130=11)	3765	3760	5	0
Rotation gr 2	1292	1290	2	0
Rotation gr 3	1223	1221	2	0
Rotation gr 4	1250	1249	1	0

2.3.3.4. Distribution of persons by membership status (RB110)

MEMBERSHIP STATUS

	Total	Current household members				No current household members			Missing
		RB110=1	=2	=3	=4	RB110=5	=6	=7	
Total	10723	9887	101	259	83	270	85	38	0
Rotation gr 2	3716	3446	33	71	26	96	29	15	0
Rotation gr 3	3449	3177	32	85	28	90	31	6	0
Rotation gr 4	3558	3264	36	103	29	84	25	17	0

MOVED TO

	Total	RB120			
		1	2	3	4
Total	270	218	8	44	0
Rotation gr 2	96	74	0	22	0
Rotation gr 3	90	73	4	13	0
Rotation gr 4	84	71	4	9	0

2.3.3.5. Item non-response

The following tables show the amount of item non-response for income variables (among households whose interview was accepted for the database):

- percentage of persons/households having received an amount (other than 0),
- percentage of households for which no information for appropriate income variable was obtained from the questionnaire (missing values) and
- Percentage of households for which partial information (not all the questions required) for appropriate income variable was obtained from the questionnaire.

A value obtained by gross/net conversion was not considered as imputed.
Income values imputed by full-record imputation are included.

Cases with both partial imputed and net/gross conversion were classified acoording to biggest proportion in the resulting value.

Whole sample (4 rotation groups) is used for the tables (i.e. this is a copy from 2009 intermediate quality report).

Table 2.12. Distribution of item non-response, household-level variables, 2009

Income variable	Hhs having received an amount		Hhs with missing values		Hhs with partial missing		Hhs with gross/net conversion	
	Count	%	Count	%	Count	%	Count	%
Total household gross income (HY010)	4954	99.8	61	1.2	692	14	4120	83.2
Total disposable household income (HY020)	4957	99.8	49	1	2213	44.6	1879	37.9
Total disposable household income before social transfer other than old-age and survivors' benefits (HY022)	4929	99.3	105	2.1	1995	40.5	1191	24.2
Total disposable household income before social transfers including old-age and survivors' benefits (HY023)	4611	92.9	193	4.2	2011	43.6	766	16.6
Net income components on household level								
Imputed rent (HY030N)	4820	97.1	4820	100.0	0	0	0	0
Income from rental of a property or land (HY040N)	136	2.7	7	5.1	0	0	22	16.2
Family/ children related	1932	38.9	0	0	14	0.7	1916	99.2

allowances (HY050N)								
Social inclusion not elsewhere classified (HY060N)	108	2.2	0	0	0	0	108	100
Housing allowances (HY070N)	88	1.8	0	0	0	0	88	100
Regular inter-household cash transfers received (HY080N)	203	4.1	0	0	0	0	203	100
Alimonies received, compulsory and voluntary (HY081N)	117	2.4	0	0	0	0	117	100
Interest, dividends, profit from capital investments in incorporated business (HY090N)	1940	39.1	1	0.1	0	0	1929	99.4
Interest repayments on mortgage (HY100N)	469	9.4	269	57.4	0	0	0	0
Income received by people aged under 16 (HY110N)	80	1.6	12	15	5	6.3	17	21.3
Regular taxes on wealth (HY120N)	3563	71.8	0	0	0	0	3563	100
Regular inter-household cash transfers paid (HY130N)	209	4.2	0	0	0	0	209	100
Alimonies paid, compulsory and voluntary (HY131N)	67	1.3	0	0	0	0	67	100
Repayments/ receipts for tax adjustment (HY145N)	2342	47.2	274	11.7	114	4.9	0	0
Gross income components on household level								
Imputed rent (HY030G)	4827	97.2	4827	100.0	0	0	0	0
Income from rental of a property or land (HY040G)	136	2.7	0	0	0	0	114	83.8
Family/ children related allowances (HY050G)	1932	38.9	1	0.1	14	0.7	311	16.1
Social inclusion not elsewhere classified (HY060G)	108	2.2	25	23.1	1	0.9	0	0
Housing allowances (HY070G)	88	1.8	11	12.5	0	0	0	0
Regular inter-household cash transfers received (HY080G)	203	4.1	14	6.9	0	0	0	0
Alimonies received, compulsory and voluntary (HY081G)	117	2.4	11	9.4	0	0	0	0
Interest, dividends, profit from capital investments in incorporated business (HY090G)	1940	39.1	1776	91.5	22	1.1	81	4.2
Interest repayments on mortgage (HY100G)	469	9.4	0	0	0	0	469	100
Income received by people aged under 16 (HY110G)	80	1.6	0	0	0	0	80	100
Regular taxes on wealth (HY120G)	3563	71.8	125	3.5	0	0	0	0

	Count	%	Count	%	Count	%	Count	%
Regular inter-household cash transfers paid (HY130G)	209	4.2	6	2.9	0	0	0	0
Alimonies paid, compulsory and voluntary (HY131G)	67	1.3	1	1.5	0	0	0	0
Tax on income and social contributions, gross (HY140G)	3857	77.7	3857	100.0	0	0	0	0

Table 2.13. Distribution of item non-response, person-level variables, 2009

	Persons having received an amount		Persons with missing values		Persons with partial missing		Persons with gross/net conversion	
	Count	%	Count	%	Count	%	Count	%
Employee cash or near cash income (PY010N)	6450	57.0	643	10.0	39	0.6	381	5.9
Non-cash employee income (PY020N)	1406	12.4	532	37.8	194	13.8	0	0
Company car (PY021N)[2]	236	2.1	236	100.0	0	0	0	0
Contributions to individual private pension plans (PY035N)	790	7.0	0	0	1	0.1	0	0
Cash benefits or losses from self employment (PY050N)	737	6.5	43	5.8	2	0.3	539	73.1
Value of goods produced by own-consumption (PY070N)	5688	50.3	45	0.8	0	0	5643	99.2
Pension from individual private plans (PY080N)	7	0.1	1	14.3	0	0	0	0
Unemployment benefits (PY090N)	226	2.0	6	2.7	9	4.0	86	38.1
Old-age benefits (PY100N)	2848	25.2	62	2.2	9	0.3	852	29.9
Survivor's benefits (PY110N)	96	0.8	3	3.1	0	0	0	0
Sickness benefits (PY120N)	1397	12.4	524	37.5	0	0	0	0
Disability benefits (PY130N)	883	7.8	4	0.5	0	0	879	99.5
Education-related benefits (PY140N)	355	3.1	3	0.8	0	0	352	99.2
Employee cash or near cash income (PY010G)	6450	57.0	65	1.0	0	0	6072	94.1
Non-cash employee income (PY020G)	1406	12.4	18	1.3	0	0	1388	98.7
Company car (PY021G)	236	2.1	1	0.4	0	0	235	99.6
Employer's social insurance contributions (PY030G)	6189	54.7	0	0	0	0	0	0
Contributions to individual private pension plans (PY035G)	790	7.0	0	0	0	0	0	0
Cash benefits or losses from self employment (PY050G)	812	7.2	37	4.6	0	0	282	34.7
Value of goods produced by own-consumption (PY070G)	5688	50.3	5688	100.0	0	0	0	0
Pension from individual private plans (PY080G)	7	0.1	0	0	0	0	7	100.0

---

[2] Non-cash income from company car is not collected from respondent, but imputed as the number of months a company car was used multiplied by 2000.

Unemployment benefits (PY090G)	226	2.0	15	6.6	0	0	157	69.5
Old-age benefits (PY100G)	2848	25.2	9	0.3	0	0	2839	99.7
Survivor's benefits (PY110G)	96	0.8	0	0	0	0	96	100.0
Sickness benefits (PY120G)	1397	12.4	14	1.0	0	0	1383	99.0
Disability benefits (PY130G)	883	7.8	20	2.3	0	0	0	0
Education-related benefits (PY140G)	355	3.1	12	3.4	0	0	0	0

2.4. Mode of data collection

Table 2.14. Distribution of household members aged 16 and over in responded households by data status (RB250), 2009

HOUSEHOLD MEMBERS 16+ (RB245=1 to 3)

	Total	RB250=11	=12	=13	=14	=21	=22	=23	=31	=32	=33
Total	8640	8581	0	0	59	0	0	0	0	0	0
Rotation gr 2	3035	3016	0	0	19	0	0	0	0	0	0
Rotation gr 3	2780	2756	0	0	24	0	0	0	0	0	0
Rotation gr 4	2825	2809	0	0	16	0	0	0	0	0	0

SAMPLE PERSONS 16+ (RB245= 1 to 3 and RB100=1)

	Total	RB250=11	=12	=13	=14	=21	=22	=23	=31	=32	=33
Total	8232	8183	0	0	49	0	0	0	0	0	0
Rotation gr 2	2832	2817	0	0	15	0	0	0	0	0	0
Rotation gr 3	2660	2639	0	0	21	0	0	0	0	0	0
Rotation gr 4	2740	2727	0	0	13	0	0	0	0	0	0

CO-RESIDENTS 16+ (RB245= 1 to 3 and RB100=2)

	Total	RB250=11	=12	=13	=14	=21	=22	=23	=31	=32	=33
Total	408	398	0	0	10	0	0	0	0	0	0
Rotation gr 2	203	199	0	0	4	0	0	0	0	0	0
Rotation gr 3	120	117	0	0	3	0	0	0	0	0	0
Rotation gr 4	85	82	0	0	3	0	0	0	0	0	0

Table 2.15. Distribution of household members aged 16 and over in responded households by type of Interview (RB260), 2009

HOUSEHOLD MEMBERS 16+ (RB245= 1 to 3) and RB250= 11 or 13

	Total	RB260=1	=2	=3	=4	=5	Missing
Total	8581	119	6447	14	2	1999	0
Rotation gr 2	3016	27	2257	6	0	726	0
Rotation gr 3	2756	44	2058	5	1	648	0
Rotation gr 4	2809	48	2132	3	1	625	0

SAMPLE PERSONS MEMBERS 16+ (RB245= 1 to 3, RB100=1) and RB250= 11 or 13

	Total	RB260=1	=2	=3	=4	=5	Missing

Total	8183	116	6212	14	1	1840	0
Rotation gr 2	2817	26	2129	6	0	656	0
Rotation gr 3	2639	42	1995	5	1	596	0
Rotation gr 4	2727	48	2088	3	0	588	0

CO-RESIDENTS 16+ (RB245= 1 to 3, RB100=2) and RB250= 11 or 13

	Total	RB260=1	=2	=3	=4	=5	Missing
Total	398	3	235	0	1	159	0
Rotation gr 2	199	1	128	0	0	70	0
Rotation gr 3	117	2	63	0	0	52	0
Rotation gr 4	82	0	44	0	1	37	0

2.5. Imputation procedure

2.5.1. Imputation for within-household non-response

Data of non-responding persons in responding households was imputed by full record imputation. A donor for imputation was the nearest neighbour as defined by distance function. Distance function used was the sum of absolute values of differences between the values for non-responding persons and responding persons. Variables used in distance function were: household size, usage of car in hh, usage of personal computer in hh, whether hh receives family related benefits, whether hh can get 5000 kr in case of sudden need, geographical location, household cope (relevance of auxiliary variables was controlled by logistic regression for response status). Variables were accounted for with different weight. In case there were several persons with equal distance, donor was selected randomly among them.

Variables copied from the donor are income variables in P-file (PY-variables).

2.5.2. Imputed rent

For calculating imputed rent the following method was used, worked out by E.-M.-Tiit in 2006.

Based on registered prices of dwellings in the Real Estate data-base (REDB) regression models predicting the price of m^2 of dwelling by its location, type, size and quality were built. For 2009 EU-SILC the Real Estate data of year 2008 was used. For houses and flats two different regression models were built.

Imputed rent was calculated by multiplying the estimated price of the dwelling (based on the m2 price predicted above and the size of the dwelling) by the rate of 12-month EURIBOR (average of the 9 last years).

2.5.3. Company cars

In the personal questionnaire, each employee was requested to report whether he or she had an option to use a company car for private ends during the previous calendar year or not. Those reporting the use were further asked to indicate the number of months the car was used, as well as the make, model and year of issue of the car. Since there is no reliable information on used care prices in Estonia, the construction of depreciation model was not possible and the conversion using tax rules was used instead. For each person reporting a benefit from the company car, the special benefit tax paid by the employer on the use of the car is recorded.

2.5.4. Value of goods produced for own consumption

Variable PY070N was calculated in 5 steps:
1. Questionnaire included questions on the amount of different goods produced for own consumption, but answers included many missing values. In most cases respondents provided an interval if they didn't know the exact value. So, as the first step, exact amounts of goods were imputed with hot-deck method within given intervals.
2. Prices of goods derived for 2008 were corrected with corresponding consumer price index (in 2008 prices were derived from the Household Budget Survey, which was not available in 2009).
3. Total value of goods produced for own consumption was calculated using amounts and prices above.
4. Production costs were deducted from total value above (using index derived from Household Budget Survey).
5. Resulting value was equally divided between all household members.

2.5.5. Imputation of income variables

Where possible, data from previous years was used for imputation. Data of 2008 was used only if household or person received particular kind of income in 2008 and analysis showed that these two incomes are sufficiently closely related. If analysis indicated no correlation between the incomes of 2008 and 2009, values were not used in imputation.

Details on the number of values forwarded from 2008 to 2009 are given in Table 2.16.

If missing value could not be imputed with data from previous year, the following methods were used (in this order):

- Logical deduction of value, based on other data in questionnaire;

- Imputation with median or average, when only single values were missing;

- When exact value was missing but respondent provided an interval, the values were imputed with hot-deck method or random regression with IveWare within this interval;

- Random regression with IveWare;

In general, empirical bounds of values present in the dataset were used in IVEware to bound imputed values. For some income components, amount per month was imputed and then converted into amount per year.

If an income component was collected only net, then missing net values were imputed and then converted to gross using net/gross conversion algorithm, where necessary. Respectively, if an income component was collected only gross, then a gross value was imputed and then converted to net.

For income components, where respondent could choose whether to provide a value net or gross (PY010, PY050 etc), gross values were converted to net prior to imputation. Missing values are thus imputed as net.

Net/gross and gross/net conversion algorithms were based on local tax system.

Following table provides numbers of values imputed for each income component by method of imputation. Numbers are given for the full sample of 2009, i.e. including new sub-sample.

Table 2.16. Percentage of imputed cases by income component in national questionnaire, 2009

Code	Description	National code	Description	Total number of persons/ households having received an amount	No of values from previous year	No of imputed values	Method of imputation, comments
INCOME COMPONENTS ON PERSONAL LEVEL							
PY010N	Employee cash or near cash income	HA5	Interval of H01N	6400	0	83	Random regression
		H01N	Net wages in a year	6400	111	579	Random regression within given interval, amount per month
		H07B	Total amount of additional payments that had not been taken into consideration in net wages	938	0	38	Random regression
PY020N	Non-cash employee income	H21A2	Approximate value of motor fuel compensated by employer	201	7	44	Random regression
		H21B2	Approximate value of use of public transportation compensated by employer	104	7	34	Random regression
		H21C2	Approximate value of food at work received free or at a reduced price	336	33	165	Random regression
		H21D2	Approximate value of using company housing free of charge or on favorable terms	65	10	53	Random regression
		H21N2	Approximate value of housing costs compensated by employer	17	0	15	Random regression
		H21E2	Approximate value of (mobile) phone or postal services compensated by employer	375	25	63	Random regression
		H21F2	Approximate value of health services compensated by employer	135	0	74	Random regression
		H21G2	Approximate value of training not related to work but compensated by employer	16	0	6	Random regression
		H21H2	Approximate value of sporting possibilities free of charge or at reduced prices	321	32	119	Random regression

Code	Description	National code	Description	Total number of persons/ households having received an amount	No of values from previous year	No of imputed values	Method of imputation, comments
		H21I2	Approximate value of holiday trip paid by employer	46	0	12	Random regression
		H21J2	Approximate value of foodstuffs	106	0	51	Random regression
		H21K2	Approximate value of leasing or loan at reduced interest rate	9	0	3	Random regression
		H21L2	Approximate value of use of equipment and/or other tools (e.g. power saw, lawnmower, etc.)	88	0	36	Random regression
		H21M2	Approximate value of other non-cash income from labour	16	0	4	Median
		H18, H20	Number of months a company car was used	See PY021N			
PY021N	Company car	H18	Possibility to use company car	6400	0	1	Logical deduction
		H20	Number of months a company car was used	235	0	4	Logical deduction
PY035N	Contributions to individual private pension plans	HK2	Payments made into the 3rd pillar of pension insurance	677	30	91	Random regression
		HK5	Payments made to the collecting insurance	130	1	8	Random regression
PY050N	Cash benefits or losses from self employment	H27N	Amount of loss from self-employment	44	0	7	Random regression
		HC1	Interval of H28N	169	0	8	Random regression
		H28N	Net profit from self-employment	169	0	31	Random regression within given interval
		H35B	Net amount of royalties, remuneration or payment under contract for creative or scientific work	50	0	3	Random regression, amount per month
		H46A2	Income from private provision of fee-charging services to other persons or households	341	0	23	Random regression

Code	Description	National code	Description	Total number of persons/ households having received an amount	No of values from previous year	No of imputed values	Method of imputation, comments
		H46B2	Income from the sale of own-produced consumer goods (e.g. handicrafts, souvenirs, etc.)	21	0	1	median
		H46C2	Income from the sale of own-produced foodstuffs (e.g. pies, waffles, shashlik, etc.)	4	0	0	
		H46D2	Income from intermediate commercial transactions	10	0	4	median
		H46E2	Income from agricultural or forestry activities	185	0	6	Random regression
		H46F2	Income from other unregistered self-employment	4	0	0	
PY080N	Pension from individual private plans	HK3A	Payments from the 3rd pillar of pension insurance	3	0	1	Logical deduction
		HK6A	Payments from collecting insurance	4	0	0	
PY090N	Unemployment benefits	H55A	Amount of unemployment benefit or any other benefits relating to unemployment	86	0	19	Random regression, amount per month
		HF6A	Amount of unemployment benefits paid by Estonian Unemployment Insurance Fund	85	0	17	Random regression, amount per month
		HF	Amount of redundancy payment in 2006 and/or 2007	130	0	18	Random regression
PY100N	Old-age benefits	H51A	Amount of old-age benefits	2840	45	63	Random regression, amount per month

Code	Description	National code	Description	Total number of persons/ households having received an amount	No of values from previous year	No of imputed values	Method of imputation, comments
		H52A	Amount of pension for incapacity for work or any other benefits relating to disability	1360	8	19	Random regression, amount per month
		HF_P	Lump-sum benefits upon retirement from employer	5	0	1	Random regression
		HF10	Did You receive an one-off benefit paid by the city of Tallinn to pensioners	503	0	0	
PY110N	Survivors' benefits	H53A	Amount of survivors' pension or any other benefits relating to the loss of a provider	100	2	3	Median, amount per month
		HF10	Did You receive an one-off benefit paid by the city of Tallinn to pensioners	See PY100N			
PY120N	Sickness benefits	H54A	Amount of sickness benefits or any other benefits relating to health	1383	129	510	Random regression, amount per day
PY130N	Disability benefits	H52A	Amount of pension for incapacity for work or any other benefits relating to disability	See PY100N			
		HF10	Did You receive an one-off benefit paid by the city of Tallinn to pensioners	See PY100N			
PY140N	Education-related benefits	H57A2	Amount of state stipendium	17	0	3	Median
		H57B2	Amount of scholarship or grant awarded by a fund or organisation located in Estonia	15	0	1	Median
		H57D2	Amount of student loan written-off to a certain extent	65	1	10	Random regression
		H57G2	Education allowance	325	0	6	Random regression

Code	Description	National code	Description	Total number of persons/ households having received an amount	No of values from previous year	No of imputed values	Method of imputation, comments
INCOME COMPONENTS ON HOUSEHOLD LEVEL							
HY040N	Income from rental of a property or land	D09N	Net income from rental of property	136	0	6	Random regression
HY050N	Family/ children related allowances	D11B	Parental benefit received in the previous calendar year	294	0	5	Logical deduction
		D11C	Total amount of other benefits received in the previous calendar year	1932	0	0	
		D11D	Maternity leave benefits	91	0	5	Logical deduction
		D11E	Net adoption allowance received last year	3	0	0	
HY060N	Social inclusion not elsewhere classified	H58A2	Amount of other support/benefit/pension not mentioned above	35	0	1	Median
		H58B2	Amount of other support/benefit/pension not mentioned above	0	0	0	0
		HF8A	The net amount of conscript's benefit in the previous calendar year	76	0	26	Random regression)
HY070N	Housing allowances	D03A	Amount of subsistence benefit	88	0	10	Random regression
HY080N	Regular inter-household cash transfers received	D16A	Amount of regular payments from other households	89	1	4	Logical deduction
		DA2A	Total alimony received previous calendar year	See HY081N			
HY081N	Alimonies received (compulsory + voluntary)	DA2A	Total alimony received previous calendar year	117	0	11	Logical deduction

Code	Description	National code	Description	Total number of persons/ households having received an amount	No of values from previous year	No of imputed values	Method of imputation, comments
HY090N	Interest, dividends, profit from capital investments in incorporated business	H49B2	Interest income from securities (shares, bonds)	4	0	0	
		H49C2	Dividend income from securities (i.e. shares, bonds)	87	0	13	Random regression)
		HE2	Interest income from deposits in a bank – interval	3266	0	290	Random regression)
		HE1A	Interest income from deposits in a bank – amount	3375	0	3269	hot-deck within a given interval
HY100N	Interest repayments on mortgage	D08E_A	Mortgage interest paid last year	469	0	210	When interest is not reported, details about mortgage are requested and interest is deducted analytically
HY110N	Income received by people aged under 16	D19A	Income received by children aged 16 or less	36	0	1	median
HY120N	Regular taxes on wealth	D10A	Amount of tax on land or any other property tax paid	3565	56	122	Random regression
HY130N	Regular inter-household cash transfers paid	D14A	Amount of regular payments to other households	146	0	5	Random regression
		DA1A	Total alimony paid previous calendar year	See HY131N			
HY131N	Alimonies paid (compulsory + voluntary)	DA1A	Total alimony paid previous calendar year	67	0	1	Random regression

Code	Description	National code	Description	Total number of persons/ households having received an amount	No of values from previous year	No of imputed values	Method of imputation, comments
HY145N	Repayments/re ceipts fro tax adjustment	H64A	Income tax return for the income received in the previous year	2921	98	380	Random regression
		H63	Additional amount of tax paid on income in the previous calendar year	112	0	22	Random regression

3. COMPARABILITY

3.1. Basic concepts and definitions

There were no changes in basic concepts and definitions from the first wave.

3.2. Components of income

3.2.1. Differences between the national definitions and standard EU-SILC definitions

From 2006, Estonia began to calculate imputed rent and HY100 (interest repayments on mortgage). For imputed rent we use the user cost method which is a nationally developed calculation method. For HY100 we use standard Eurostat definitions but there is a great deal of imputation involved in the actual data assembly.

There were no changes in the source or procedure used for the collection of other income variables from 2006.

3.2.2. The source or procedure used for the collection of income variables

There were no changes in the source or procedure used for the collection of income variables from the first wave.

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

There were no changes in the form in which income variables at component level have been collected from the first wave.

3.2.4. The method used for obtaining income target variables in the required form

There were no changes in the source or procedure used for the collection of income variables from the first wave.

3.3. Tracing rules

There were no differences between the national tracing rules and the standard EU-SILC tracing rules.

4. COHERENCE

This section will compare the longitudinal EU-SILC data to various external sources, including the *National Accounts* (NA), the *Household Budget Survey* (HBS), the *Labour Force Survey* (LFS), the *Safety survey*, wage statistics and social protection statistics.

The HBS is a continuous survey of households, which has been carried out since 1996. The yearly sample size was approximately 4500 households. The HBS is designed to collect information on income and expenditure of households. Data on income is gathered using a diary, where a household records all income received during one month. The HBS was the source of Laeken indicators up until EU-SILC. A comparison with HBS is not possible between 2008-2009 as the HBS underwent a redesign in this period and no fieldwork was carried out.

The LFS is a continuous survey, which has been carried out according to the common EU methodology since 1995. The yearly sample size is about 12,000 working aged persons. From 2006 LFS was switched fully into CAPI with face-to-face interviews. The LFS is the main source for labour market information.

The Safety survey sample size was about 7300 persons. Data was gathered from November 2008 to May 2009 using to conduct face to face interviews with a laptop (CAPI) and self-filled questionnaires (CAS).

Wage statistics have in their current form been continuously calculated since 1992. All enterprises employing 50 persons or more are obliged to provide data. A sample is drawn from smaller enterprises. Wage data is used to calculate hourly and monthly wages, both gross and net, as well as labour costs. All figures have been converted into full-time units.

4.1. Comparison of income target variables and number of persons who receive income from each 'income component', with external sources

In the following paragraphs, EU-SILC income data is compared component by component between the waves 2008 and 2009 and to income data from administrative sources for income years 2007 and 2008. Table 4.1 presents the comparisons by total amounts received/spent and Table 4.2 by number of recipients.

As illustrated in table 4.1, the total cash incomes from employment indicate a rise of about 11% in incomes from 2007 to 2008. (In Estonia, the income reference year is one year behind the survey year so the 2009 survey gives us annual incomes for 2008, and the 2008 survey gave us the incomes of 2007) The wage statistics figures show a similar increase although the incomes reported through wage statistics by companies indicate systematically lower total wage incomes. The under-representation is around 11%, aka wage statistics incomes make up 89% of the total wage incomes reported by wage-earners themselves. This can partly be due to unreported wages paid to temporary employees, or failing to report other cash benefits by companies and also from PY010 containing a wider set of income components than the definition for wage statistics. Wage statistics also refer to the full-time equivalent of income and exclude part-time work contracts; unofficial work relationships are also excluded. In EU-SILC data both are included. The data concerning wage statistics comes from in-house sources, not administrative registers. The numbers of recipients (table 4.2) for PY010G also have a comparability problem because in wage statistics the number of recipients is calculated in full time units and not actual persons.

Table 4.1. Total amounts of income components by source of information (in kroons), income years 2007 and 2008

Income component	2007		2008	
	EU-SILC	**Administrative sources***	**EU-SILC**	**Administrative sources***
Cash or near-cash employee income (PY010N)	68,405,162	59,452,036	75,663,148	68,368,093
Old-age benefits (PY100G)	12,900,000	12,704,570	15,296,438	15,591,630
Gross survivor's benefits (PY110G)	141,000	195,680	873,759	2,052,940
Disability benefits (PY130G)	1,539,193	2,411,970	2,077,324	2,975,720

* Wage statistics in the case of PY010 and administrative sources for other variables.

Variable PY100G demonstrates a pretty good fit between survey data and administrative data, with slight fluctuations between under-or over-reporting in EU-SILC. EU-SILC also includes pension benefits received from abroad, which tend to be higher than national benefits. There are, however, very few people in the sample that receive pensions from abroad. The Estonian state at the same time pays old age benefits to its citizens residing abroad while the EU-SILC survey does not have people currently living abroad in its sample.

Another condition to be considered is that the administrative data includes the institutionalised population whereas the EU-SILC survey does not include institutionalised people in its sample—e.g.

hospitalized, in retirement homes and imprisoned people. This difference influences old-age and disability benefits especially.

Survivors' benefits are usually paid to a household as a whole. The administrative figure indicates significant and systematic underestimation in EU-SILC both in total amounts and number of recipients. This is most likely due to the very small amounts of survivors' benefits, and some benefits included, such as the funeral allowance, not being separately asked in the questionnaire. Respondents hardly ever think to report the funeral allowance when asked to report any 'other' benefits and survivors' benefits are probably reported incorrectly and forgotten to report when they are not a significant source of income for the household.

Disability benefits too are underreported in EU-SILC. Although the number of recipients is not very different from what administrative accounts indicate. EU-SILC reports a larger number of recipients which is probably due to the fact that administrative records count the number of disability benefit recipients eligible for the benefit at a fixed moment in time (January first of the following year) but in EU-SILC everyone that has received the benefit at any time in the previous year is counted as a recipient.

The difference in total amounts paid is to a small extent related to the fact that disability benefits paid to people in retirement age have been added to the old-age benefits. But for the most part disability benefits, often small amounts, are not sufficiently captured by the survey. This is further indicated by the fact that the administrative records number should be somewhat lower than the survey's result since administrative information includes only disability and early retirement benefits. The numbers of recipients for care allowances and economic integration of the handicapped are not included for administrative records in Table 4.2, whereas the amounts received by them are included in Table 4.1.

Table 4.2. Number of recipients of income components by source of information, income years 2007 and 2008

Income component	2007		2008	
	EU-SILC	Other sources*	EU-SILC	Other sources*
Cash or near-cash employee income (PY010N)	690,633	543,060	698,938	543,432
Old-age benefits (PY100G)	289,604	293,640	284,371	293,400
Survivor's benefits (PY110G)	8,846	14,032	7,060	12,980
Disability benefits (PY130G)	66,024	70,498	73,554	73,110

* Wage statistics in the case of PY010 and administrative sources for other variables.

Table 4.3 compares the mean and number of recipients of most income components in EU-SILC 2009 to the estimates from the 2008 operation. Changes that emerge are, in general, in line with what could be expected. It should be noted that the fieldwork period ended in June and the 2009 data actually refers to the incomes of 2008. EU-SILC in Estonia collects the respondent's annual income from the previous calendar year. Within a year the average salary increased by 9%, while the number of wage receivers was more modest – 1%.

At the same time, the number of people receiving unemployment benefits increased by 155%.The reason is that in Estonia the economic crisis started in 2008.Administrative data confirms the survey results. The increase in the mean of PY090N was 49%.

Most income components show an increase from 2008 to 2009 in line with the rising salaries. Benefits also went up due to small increases in benefit rates, even though the number of recipients fell, for instance old-age and survivor's benefits.

Benefits from self employment increased. The number of entrepreneurs seems to fluctuate between survey years which hints to a relatively big pool of short-lived businesses.

Old-age benefits have increased due to increase the national pension, while the number of receivers decreased

Table 4.3. Mean (EEK-s) and number of recipients of income components in EU-SILC 2008 and 2009

	Mean		Number of recipients	
	2008	2009	2008	2009
Individual level components				
PY010N	99056	108254	690573	698938
PY020N	9219	10251	172318	182429
PY035N	5308	6004	92736	95614
PY050N	20331	20493	51762	54448
PY090N	13218	19649	9189	23436
PY100N	43838	52599	289604	284371
PY110N	15964	18601	8846	7060
PY120N	3326	4177	100303	165250
PY130N	23312	28242	66024	73554
PY140N	8921	10352	19675	30703
Household level components				
HY040N	12788	20868	7720	11937
HY050N	15274	20655	182718	180377
HY070N	5859	5332	10000	10422
HY080N	20989	22664	23703	25563
HY090N	2112	3483	289877	255340
HY110N	2991	4686	6763	5251
HY120N	477	598	348412	397103
HY130N	18689	21197	20927	27661
HY145N	-3074	-4342	228958	266545
HY010	183157	208150		
HY020	153603	175302		
HY022	147552	165382		
HY023	152713	172048		

Household level variables reflect changes in line with personal level variables.
For HY040N the increase in the total amount is in line with more people obtaining an income from renting out their property or land. Rent prices were on the rise in 2008 (to which the income belongs) so it is logical that the amount of income earned from renting out property would increase and since the market was growing, more people would rent out their property.
Family allowances have increased, in compliance with increases of national benefit levels, most notably the parental benefit which is tied to incomes. Larger parental benefits and more people becoming eligible for larger sums along with rising incomes amounted for a noticeable increase in the overall amount despite the fall in numbers of recipients (which may well be due to sample fluctuations).
The average amount of housing allowances has decreased which can be explained by an overall increase in wages, leaving fewer households eligible, and for smaller sums. The number of households receiving and paying transfers from other households has increased. This might have something to do with more help from former household members living and working abroad and sending money home or just a more favorable economic setting where people have greater financial possibilities for helping their relatives.
An increase in income collected through HY090 reflects a positive situation on the financial markets, with more people investing and larger returns that have to do with overall economic growth.
More people had to pay taxes on wealth and the amount went up but not changed substantially.
The general economic picture should also account for the increase in income received by people aged under 16, the average salary increased but there was not as much short-term and summer jobs as last year.
Total household income increased by approximately 13% in 2008. The increases stem from the higher wages and other income components, most of which have gone up.

4.2. Comparison of other target variables with external sources

In Table 4.4 the distribution of population aged 16-74 derived from EU-SILC and LFS is compared. Most of the differences are minor. The LFS does indicate a slightly better educated populace than EU-SILC. There are more people with post secondary education and less people with secondary or lower education. Given that the questions used in the two surveys are identical, this must be due to sample fluctuations.

Table 4.4. Distribution of population aged 16-74 by ISCED level, based on the cross-sectional EU-SILC and the LFS, 2006-2009

ISCED level	2006		2007		2008		2009	
	EU-SILC	LFS	EU-SILC	LFS	EU-SILC	LFS	EU-SILC	LFS
0 Pre-primary education	0.6	0.5	0.6	0.5	0.5	0.4	.2*	0.5
1 Primary education	4.8	4.1	4.2	3.5	6.5	3.3	2.2	2.1
2 Lower secondary education	18.7	18.2	18.6	17.9	22.4	18.3	16.7	17.4
3 (Upper) secondary education	43.0	44.5	44.1	45.8	47.2	44.8	49.0	45.4
4 Post-secondary non tertiary education	7.6	5.5	7.1	5.2	2.6	5.2	3.0	4.9
5 First stage of tertiary education	24.9	26.8	24.9	26.8	20.6	27.8	28.6	29.6
6 Second stage of tertiary education	0,3	0.4	(0.4)**	(0.3)**	(0.3)***	(0.2)***	0.4	(0.2)*
Total	100.0	100.0	99.9	100.0	99.9	100.0	99.9	99.9

* Very unreliable estimate, based on less than 20 sample observations
** Unreliable estimate, based on 35-44 sample observations.

Table 4.5 presents the longitudinal comparison of population aged 16 or over (2009 aged 16-74) by most frequent current activity status in EU-SILC, HBS, LFS and the Safety survey.

The differences that can be observed between the in EU-SILC and HBS data sources are relatively minor with the exceptions of students, domestic workers and people in the "other inactive" category. This indicates that domestic workers are frequently reported as "other inactive" in HBS.

Comparing the EU-SILC and Safety survey results from 2008 two facts should be noted. Firstly, the Safety survey was carried out from November 2008 - May 2009, meaning that the results are not strictly for 2008, as are those of EU-SILC, whose fieldwork period ranged from February to July 2008. Secondly the Safety survey falls well into the time of the world economic crisis, whereas EU-SILC is still fully reflective of the times of the economic boom. In this context, the figures show a remarkably great similarity of the socio-economic composition of the populace. The main difference is firstly the share of the unemployed, of whom there are almost 5% more in the Safety survey- a trend also reflected in administrative statistics of unemployment. Secondly, EU-SILC demonstrates a greater share of the population as in retirement. This difference of 6% points is probably caused by sampling.

In 2009 the Table 4.5 presents the comparison of population aged 16-74 or over by current activity status in EU-SILC and the LFS. Differences between data from the two surveys were small, mostly less than 1%. The differences that can be observed between the two data sources may be due to misclassification to 'other inactive' category in LFS.

Table 4.5. Distribution of population aged 16 and over (2009 aged 16-74) by self-defined activity status based on longitudinal EU-SILC, the HBS and Safety survey, 2006-2009

Activity status	2006		2007		2008		2009**	
	EU-SILC	HBS	EU-SILC	HBS	EU-SILC	Safety survey	EU-SILC	LFS
Working full-time	52.9	50.6	54.2	50.8	53.5	53.8	53.6	52.1
Working part-time	3.6	3,3	3.9	3.2	3.8	4.4	5.4	5.4
Unemployed	4.3	3.9	3.2	2.8	3.1	7.9	8.3	8.9
Pupil, student	8.8	10.7	8.8	12.3	8.7	9.5	9.4	9.6
In retirement	21.8	22.0	21.7	21.5	21.6	15.4	14.2	13.8
Permanently disabled	3.7	4.2	3.7	4.1	3.7	4	4.1	4.6
Fulfilling domestic tasks and care responsibilities	4.7	1.5	4.4	1.5	4.6	4.3	4.9	5.3
Other inactive	…*	3.6	(0.0)*	3.4	(0.2)*	(0.7)*	(0.2)*	(0.3)*
Total	100.0	100.0	99.9	99.6	99.2	100	99.9	99.7

* Extremely unreliable estimate, based on less than 20 sample observations
* * Population aged 16 -74