EU-SILC 2010 in Estonia: Final Quality Report

Tallinn 2012

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, 2009 and 2010 sample structure is as originally planned: it consists of four rotational groups (in 2010 - one from each year 2007-2010). In longitudinal dimension, sample contains three panels:

- 4-year panel (rotation group 3, i.e. DB075=3) started in 2007;
- 3-year panel (rotation group 4, i.e. DB075=4) started in 2008;
- 2-year panel (rotation group 1, i.e. DB075=1) started in 2009;

New part introduced in 2010 constitutes rotational group 2 (DB075=2).

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, 2010

Age groups	Total	Males	Females
Total	9,9	7,8	11,7
Less than 18 years	9,5	9,7	9,4
From 18 to 24 years	8,1	7,4	8,7
From 18 to 64 years	8,3	7,1	9,3
18 years or over	10,0	7,4	12,0
From 25 to 49 years	5,5	3,9	6,9
From 50 to 64 years	13,2	13,0	13,4
65 years or over	15,7	8,5	19,4

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, %			
		2007	**2008**	**2009**	**2010**
Large	Tallinn, Harju, Ida-Viru, Lääne-Viru, Pärnu, Tartu	0.17	0.17	0.16	0.17
Small	Jõgeva, Järva, Lääne, Põlva, Rapla, Saare, Valga, Viljandi, Võru	0.31	0.31	0.30	0.31
Hiiu	Hiiu	0.85	0.89	0.84	0.87

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 2007-2010.

Table 2.2. Allocation of sample into months, 2007 – 2010

Year	New part					Longitudinal part				
	Feb	Mar	April	May	June	Feb	Mar	April	May	June
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%	
2010	35%	35%	30%	0%		35%	35%	14%	16%	

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 2007-2010). In longitudinal dimension, sample contains three panels:
- 4-year panel (rotation group 3, i.e. DB075=3) started in 2007;

- 3-year panel (rotation group 4, i.e. DB075=4) started in 2008;
- 2-year panel (rotation group 1, i.e. DB075=1) started in 2009;

New part introduced in 2010 constitutes rotational group 2 (DB075=2).

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 2010 contains three subsamples of households:

S4	Households introduced into the survey in 2007 and their split-offs. Year 2010 is their fourth year in the survey.
S3	Households introduced into the survey in 2008 and their split-offs. Year 2010 is their third year in the survey.
S2	Households introduced into the survey in 2009 and their split-offs. Year 2010 is their second year in the survey.

Sub-sample of households introduced into the survey in 2010 doesn't contribute to longitudinal database of 2010. 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 2007 were corrected 3 times etc as shown below.

S4: 2007 -> 2008 ->2009 ->2010 (3 corrections)
S3: 2008 ->2009 ->2010 (2 corrections)

S2: 2009 ->2010 (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 2010 base weights for sub-samples corrected for non-response. There are three longitudinal sets of interest in year 2010:

- Longitudinal set of two year duration, involving data from year 2009 and 2010. 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 2008, 2009 and 2010. 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 2007, 2008, 2009 and 2010. Longitudinal weight to be used for this set is RB064. Only the sub-sample S4 contributes to this set.

Children born between interviews of 2009 and 2010 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 2007-2010 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, 2010

Income components	Mean	Number of observations		Standard error
		Before imputation	After imputation	
Total household gross income (HY010)	192935	92	4972	3024
Total disposable household income (HY020)	163982	737	4972	2424
Total disposable household income before social transfer other than old-age and survivors' benefits (HY022)	148341	1598	4972	2322
Total disposable household income before social transfers including old-age and survivors' benefits (HY023)	120978	1909	4972	2514
Net income components at household level				
Imputed rent (HY030N)	20941	0	4709	305
Income from rental of a property or land (HY040N)	13983	127	150	2876
Family/ children related allowances (HY050N)	22161	6	1882	1252
Social exclusion not elsewhere classified (HY060N)	9461	0	86	1633
Housing allowances (HY070N)	8731	0	120	924
Regular inter-household cash transfers received (HY080N)	23165	0	233	1948
Alimonies received (compulsory + voluntary) (HY081N)	22314	0	100	1947
Interest, dividends, profit from capital investments in incorporated business (HY090N)	2497	8	2207	577
Interest repayments on mortgage (HY100N)	21759	0	512	994
Income received by people aged under 16 (HY110N)	2598	45	80	803
Regular taxes on wealth (HY120N)	668	0	3572	25
Regular inter-household cash transfers paid (HY130N)	19902	0	250	1656
Alimonies paid (compulsory + voluntary) (HY131N)	27789	0	69	4743
Tax on income and social contributions, net (HY140N)	19902	0	.	0
Repayments/ receipts for tax	-5639	2031	2401	160

Income components	Mean	Number of observations		Standard error
		Before imputation	After imputation	
adjustment (HY145N)				
Value of goods produced for own consumption(HY170N)	3266	0	2332	77
Gross income components at household level				
Imputed rent (HY030G)	23136	0	4780	292
Income from rental of a property or land (HY040G)	17700	13	150	3641
Family/ children related allowances (HY050G)	25679	1515	1882	1558
Social exclusion not elsewhere classified (HY060G)	9461	76	86	1633
Housing allowances (HY070G)	8731	102	120	924
Regular inter-household cash transfers received (HY080G)	23165	215	233	1948
Alimonies received (compulsory + voluntary) (HY081G)	22314	88	100	1947
Interest, dividends, profit from capital investments in incorporated business (HY090G)	3080	33	2207	727
Interest repayments on mortgage (HY100G)	21759	0	512	994
Income received by people aged under 16 (HY110G)	2684	0	80	875
Regular taxes on wealth (HY120G)	668	3424	3572	25
Regular inter-household cash transfers paid (HY130G)	19902	245	250	1656
Alimonies paid (compulsory + voluntary) (HY131G)	27789	66	69	4743
Tax on income and social contributions, gross (HY140G)	35516	0	3885	649
Value of goods produced for own concumption(HY170G)	3266	0	2332	77
Net income components at personal level				
Employee cash or near cash income (PY010N)	101463	5092	5948	1328
Non-cash employee income (PY020N)	12173	726	1364	630
Company car (PY021N)	39558	0	226	1094
Contributions to individual private pension plans (PY035N)	6057	628	738	267
Cash benefits or losses from self employment (PY050N)	16030	143	814	1613
Pension from individual private plans (PY080N)	8235	27	28	2451
Unemployment benefits (PY090N)	23802	342	702	1428
Old-age benefits (PY100N)	55316	1997	2883	262
Survivors' benefits (PY110N)	19668	115	118	1002
Sickness benefits (PY120N)	4469	640	1097	243

Income components	Mean	Number of observations		Standard error
		Before imputation	After imputation	
Disability benefits (PY130N)	29568	0	898	559
Education-related benefits (PY140N)	9646	0	467	832
Gross income components at personal level				
Employee cash or near cash income (PY010G)	124501	233	5948	1701
Non-cash employee income (PY020G)	15409	0	1364	798
Company car (PY021G)	50073	0	226	1385
Employer's social insurance contributions (PY030G)	42323	0	5851	573
Contributions to individual private pension plans (PY035G)	20577	542	877	2431
Cash benefits or losses from self employment (PY050G)	8235	0	28	2451
Value of goods produced for own consumption (PY070G)	28381	188	702	1734
Pension from individual private plans (PY080G)	56862	0	2883	309
Unemployment benefits (PY090G)	19924	0	118	1104
Old-age benefits (PY100G)	5656	0	1097	308
Survivors' benefits (PY110G)	29568	887	898	559
Sickness benefits (PY120G)	9646	428	467	832
Disability benefits (PY130G)	20577	542	877	2431
Education-related benefits (PY140G)	8235	0	28	2451

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, 2010

	Mean	Number of observations		Standard error
		Before imputation	After imputation	
Subclasses by household size				
1 household member	82137	406	1074	2430
2 household member	117786	57	1020	3367
3 household members	109090	11	1347	2104
4 and more	107872	263	1531	2227
Population by age group				
<25	104262	85	4538	1545
25-34	133424	63	1371	3932
35-44	113916	85	1768	3066
45-54	103744	138	1979	2290
55-64	106020	209	1529	2781
65+	81509	568	2289	1117
Population by sex				
Male	109290	491	6350	1612
Female	103409	657	7124	1230

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 2010 was 104 households, which makes the proportion of the over-coverage in the new part of the sample 4,3% and of the whole sample 1,6% (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 2010 ca 2% of all records referring to persons aged 14+ were dropped before selecting the sample).

In the new part of the sample of 2010 there were 151 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 non-response reason 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, 2010

Frame error	Number of households	Proportion in the frame error (%)
Total, of which	104	100.0
Address person was dead	11	10.6
Address person has left Estonia	81	77.8
Address person was staying in an institution	9	8.7
Address person was surveyed through one of his/her household members	3	2.9

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.

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 guidelines

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

In 2010, the questions used to determine using child care services reformulated to better meet Eurostat's guidelines.

Other notable modifications in 2010 concerned the following variables:

1) The question about total housing cost (including electricity, water, gas and heating) was reformulated to better meet Eurostat's guidelines.

2) The question about capacity to face unexpected financial expenses was reformulated to better meet Eurostat's guidelines.

3) The question about volunteer work for free last week was added.

4) The question about type of work contract was added.

5) The questions about pension from individual private plans were reformulated to better meet Eurostat's guidelines

6) The questions about intra-household sharing of resources were added

7) The questions about material deprivation were excluded

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 2010 the training program was conducted in 4 smaller groups of about 15 people, similarly to 2007 - 2009, 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 2010 overall, 63 interviewers were responsible for conducting the interviews. The household–interviewer ratio was 79 households per interviewer.
In 2009 and 2008 overall, 53 interviewers were responsible for conducting the interviews. The household– interviewer ratio was 90 households per interviewer.
In 2007 overall, 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.
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.
In 2009 the total number of errors was 1939. In 2010 the number had fallen to 1856.

As can be seen from table 2.6, in 2010 the most common types of errors were those discovered during concurrent in-office checks and the use of category 'other', while a suitable category existed. Thanks to the continuing development of primary logical checks in the data entry program, in 2010 was only 3 data entry mistakes and 3 cases when interviewer's remark does not explain unusual situation. Starting and finishing times recorded incorrectly mistakes have decreased to zero.

Out of all the errors in 2010 41% (758 cases) required callback and clarification with the interviewer or interviewee. In 2009, 49% of cases had required callback.

Table 2.6. Interviewer errors and their processing, 2010

Type of error	Number of errors detected	Share of errors requiring a call-back
No remark explaining unusual situation	251	71.31%
Interviewer made an error, but did not correct it	369	50.41%
Interviewer's remark does not explain unusual situation	3	100.00%
Data not sufficient for coding	48	43.75%
Starting and finishing times recorded incorrectly	0	0,00%
Use of category Other, while a suitable category exists	507	20.32%
In-office checks	515	45.63%
Interviewer has misunderstood a question	90	30.00%
Data entry mistake	3	66.67%
Not interviewers error	70	2.86%
Total	1856	40.84%

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 2010 (i.e. contains only 2-, 3- and 4-year panels of 2010, not new sample).

Table 2.8. Achieved sample size by panel

	2-year panel (rotation gr 1)	3-year panel (rotation gr 4)	4-year panel (rotation gr 3)
Number of households for which an interview is accepted for the database (DB135=1)	1139	1203	1183
Number of persons 16 years or older who are members of the households for which the interview is accepted for the database, of which	2519	2731	2690
sample persons (RB100=1)	2461	2588	2497
co-residents (RB100=2)	58	143	193

2.3.3.2. Unit non-response

Table 2.9. Household wave response rate by panel

Number	Total (rotation groups 1, 4, 3)	2-year panel (rotation gr 1)	3-year panel (rotation gr 4)	4-year panel (rotation gr 3)
A. Number of hhs passed on to 2010 from 2009 (DB110=1,2,3,4,5,6,7,11,10)	3762	1205	1294	1263
B. Number of hhs added to the sample due to spli-off (DB110=8)	141	42	61	38
C. Number of hhs out-of-scope or non-existent (DB110=3,4,5,6 or DB120=23)	43	9	19	15
D. Number of hhs with interview accepted (DB135=1)	3525	1139	1203	1183
E. Number of hhs with interview accepted in 2009	3675	1205	1221	1249
Wave response rate D/(A+B-C)	0,91	0,92	0,,90	0,92
Achieved sample size ratio (D/E)	0,96	0,95	0,99	0,95

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

Number	Total (rotation groups 1, 4, 3)	2-year panel (rotation gr 1)	3-year panel (rotation gr 4)	4-year panel (rotation gr 3)
A. Number of hhs passed on to 2010 from 2009 (DB110=1,2,3,4,5,6,7,11,10)	2499	1205	1294	NA
B Number of hhs out-of-scope or non-existent among those in row A (DB110=3,4,5,6)	28	9	19	NA
C. Number of initial hhs to be	2332	1132	1200	NA

passed on to 2011				
D. Number of split-off hhs to be passed on to 2011	71	31	40	NA
Longitudinal follow-up rate C/(A-B)	0,94	0,95	0,94	NA
Follow-up ratio (C+D)/(A-B)	0,97	0,97	0,97	NA

Table 2.11. Wave response rate for persons by panel

	Total (rotation groups 1, 4, 3)	2-year panel (rotation gr 1)	3-year panel (rotation gr 4)	4-year panel (rotation gr 3)
A. Number of sample persons aged 16 and over passed on to 2010 from 2009	8320	2712	2875	2733
B. Number of persons (among those in A) belonging to households no longer in scope (DB110 = 3,4,5,6)	40	8	19	13
C. Number of persons (among those in A) no longer in scope in existing households (RB110=6 or RB120=2,3).	92	24	33	35
D. Number of sample persons completed the personal interview (RB100=1 and RB250=11 to 13).	7499	2445	2574	2480
E. Number of completed personal interviews (RB250=11 to 13)	7884	2502	2711	2671
F. Number of completed personal interviews in 2009	8204	2639	2809	2756
Wave response rate, D/(A-B-C)	0,92	0,91	0,91	0,92
Achieved sample size ratio, E/F	0,96	0,95	0,97	0,97

Since longitudinal component does not contain any other sample persons than those passed on to 2010 from 2009, **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
- 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

DB120=22
- 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

is at home

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	3903	3553	143	6	14	22	0	2	141	0	2	20
Rotation gr 1	1247	1152	43	1	3	5	0	0	42	0	0	1
Rotation gr 4	1355	1210	55	4	6	9	0	0	61	0	1	9
Rotation gr 3	1301	1191	45	1	5	8	0	2	38	0	1	10

RECORD OF CONTACT AT ADDRESS

	Total	DB120				
		11	21	22	23	Missing
Total (DB110 = 2,8,9)	284	197	23	63	1	0
Rotation gr 1	85	62	6	17	0	0
Rotation gr 4	116	78	12	26	0	0
Rotation gr 3	83	57	5	20	1	0

HOUSEHOLD QUESTIONNAIRE RESULT

	Total	DB130					
		11	21	22	23	24	Missing
Total (DB110=1 or DB120=11)	3750	3527	181	9	27	6	0
Rotation gr 1	1214	1139	64	4	7	0	0
Rotation gr 4	1288	1204	60	4	16	4	0
Rotation gr 3	1248	1184	57	1	4	2	0

HOUSEHOLD INTERVIEW ACCEPTANCE

	Total	DB135		
		1	2	Missing
Total (DB130=11)	3527	3525	2	0
Rotation gr 1	1139	1139	0	0
Rotation gr 4	1204	1203	1	0
Rotation gr 3	1184	1183	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	9839	9072	130	206	95	227	67	42	0

Rotation gr 1	3114	2864	48	64	38	74	15	11	0
Rotation gr 4	3420	3134	46	82	34	84	28	12	0
Rotation gr 3	3305	3074	36	60	23	69	24	19	0

MOVED TO

	Total	RB120			
		1	2	3	4
Total	227	199	6	22	0
Rotation gr 1	74	65	1	8	0
Rotation gr 4	84	77	2	5	0
Rotation gr 3	69	57	3	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 according to the biggest proportion in the resulting value.

Whole sample (4 rotation groups) is used for the tables (i.e. this is a copy from 2010 intermediate quality report).

Table 2.12. Distribution of item non-response, household-level variables, 2010

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)	4965	99.9	61	1.2	836	16.8	3983	80.2
Total disposable household income (HY020)	4968	99.9	47	0.9	2641	53.2	1547	31.1
Total disposable household income before social transfer other than old-age and survivors' benefits (HY022)	4926	99.1	93	1.9	2408	48.9	873	17.7
Total disposable household income before social transfers including old-age and survivors' benefits (HY023)	4634	93.2	227	4.9	2259	48.7	577	12.5
Net income components on household level								
Imputed rent (HY030N)	4709	94.7	0	0	0	0	4709	100.0

Income from rental of a property or land (HY040N)	150	3	10	6.7	0	0	13	8.7
Family/ children related allowances (HY050N)	1882	37.9	1	0.1	30	1.6	1845	98
Social inclusion not elsewhere classified (HY060N)	86	1.7	0	0	0	0	86	100
Housing allowances (HY070N)	120	2.4	0	0	0	0	120	100
Regular inter-household cash transfers received (HY080N)	233	4.7	0	0	0	0	233	100
Alimonies received, compulsory and voluntary (HY081N)	100	2	0	0	0	0	100	100
Interest, dividends, profit from capital investments in incorporated business (HY090N)	2207	44.4	1	0	2	0.1	2196	99.5
Interest repayments on mortgage (HY100N)	512	10.3	261	51	0	0	0	0
Income received by people aged under 16 (HY110N)	80	1.6	23	28.8	1	1.3	11	13.8
Regular taxes on wealth (HY120N)	3572	71.8	0	0	0	0	3572	100
Regular inter-household cash transfers paid (HY130N)	250	5	0	0	0	0	250	100
Alimonies paid, compulsory and voluntary (HY131N)	69	1.4	0	0	0	0	69	100
Repayments/ receipts for tax adjustment (HY145N)	2401	48.3	249	10.4	121	5	0	0
Value of goods produced for own consumption (HY170N)	2332	46.9	2332	100.0	0	0	0	0
Gross income components on household level								
Imputed rent (HY030G)	4780	96.1	4780	100.0	0	0	0	0
Income from rental of a property or land (HY040G)	150	3	0	0	0	0	137	91.3
Family/ children related allowances (HY050G)	1882	37.9	5	0.3	21	1.1	341	18.1
Social inclusion not elsewhere classified (HY060G)	86	1.7	10	11.6	0	0	0	0
Housing allowances (HY070G)	120	2.4	18	15	0	0	0	0
Regular inter-household cash transfers received (HY080G)	233	4.7	18	7.7	0	0	0	0
Alimonies received, compulsory and voluntary (HY081G)	100	2	12	12	0	0	0	0
Interest, dividends, profit from capital investments in incorporated business (HY090G)	2207	44.4	2059	93.3	25	1.1	90	4.1
Interest repayments on	512	10.3	0	0	0	0	512	100

	Count	%	Count	%	Count	%	Count	%
mortgage (HY100G)								
Income received by people aged under 16 (HY110G)	80	1.6	0	0	0	0	80	100
Regular taxes on wealth (HY120G)	3572	71.8	148	4.1	0	0	0	0
Regular inter-household cash transfers paid (HY130G)	250	5	5	2	0	0	0	0
Alimonies paid, compulsory and voluntary (HY131G)	69	1.4	3	4.3	0	0	0	0
Tax on income and social contributions, gross (HY140G)	3885	78.1	0	0	0	0	0	0
Value of goods produced for own consumption (HY170G)	2332	46.9	0	0	0	0	2332	100.0

Table 2.13. Distribution of item non-response, person-level variables, 2010

	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)	5948	53	573	9.6	29	0.5	254	4.3
Non-cash employee income (PY020N)	1364	12.2	466	34.2	172	12.6	0	0
Company car (PY021N)[1]	226	2	226	100	0	0	0	0
Contributions to individual private pension plans (PY035N)	738	6.6	106	14.4	4	0.5	0	0
Cash benefits or losses from self employment (PY050N)	814	7.3	43	5.3	2	0.2	626	76.9
Pension from individual private plans (PY080N)	28	0.2	1	3.6	0	0	0	0
Unemployment benefits (PY090N)	702	6.3	48	6.8	18	2.6	294	41.9
Old-age benefits (PY100N)	2883	25.7	70	2.4	12	0.4	804	27.9
Survivor's benefits (PY110N)	118	1.1	3	2.5	0	0	0	0
Sickness benefits (PY120N)	1097	9.8	457	41.7	0	0	0	0
Disability benefits (PY130N)	898	8	0	0	0	0	898	100
Education-related benefits (PY140N)	467	4.2	5	1.1	0	0	462	98.9
Employee cash or near cash income (PY010G)	5948	53	78	1.3	0	0	5637	94.8
Non-cash employee income (PY020G)	1364	12.2	26	1.9	0	0	1338	98.1
Company car (PY021G)	226	2	3	1.3	0	0	223	98.7
Employer's social insurance contributions (PY030G)	5851	52.2	5851	100.0	0	0	0	0
Contributions to individual private pension plans	738	6.6	0	0	0	0	738	100.0

---

[1] 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 4000.

(PY035G)								
Cash benefits or losses from self employment (PY050G)	877	7.8	71	8.1	3	0.3	261	29.8
Pension from individual private plans (PY080G)	28	0.2	0	0	0	0	28	100
Unemployment benefits (PY090G)	702	6.3	31	4.4	0	0	483	68.8
Old-age benefits (PY100G)	2883	25.7	16	0.6	0	0	2867	99.4
Survivor's benefits (PY110G)	118	1.1	1	0.8	0	0	117	99.2
Sickness benefits (PY120G)	1097	9.8	11	1	0	0	1086	99
Disability benefits (PY130G)	898	8	11	1.2	0	0	0	0
Education-related benefits (PY140G)	467	4.2	35	7.5	4	0.9	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), 2010

HOUSEHOLD MEMBERS 16+ (RB245=1 to 3)

	Total	RB250=11	=12	=13	=14	=21	=22	=23	=31	=32	=33
Total	7940	7884	0	0	56	0	0	0	0	0	0
Rotation gr 2	2519	2502	0	0	17	0	0	0	0	0	0
Rotation gr 3	2731	2711	0	0	20	0	0	0	0	0	0
Rotation gr 4	2690	2671	0	0	19	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	7546	7499	0	0	47	0	0	0	0	0	0
Rotation gr 2	2461	2445	0	0	16	0	0	0	0	0	0
Rotation gr 3	2588	2574	0	0	14	0	0	0	0	0	0
Rotation gr 4	2497	2480	0	0	17	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	394	385	0	0	9	0	0	0	0	0	0
Rotation gr 2	58	57	0	0	1	0	0	0	0	0	0
Rotation gr 3	143	137	0	0	6	0	0	0	0	0	0
Rotation gr 4	193	191	0	0	2	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	7884	85	5833	29	0	1937	0

Rotation gr 2	2502	20	1901	14	0	567	0
Rotation gr 3	2711	27	1998	6	0	680	0
Rotation gr 4	2671	38	1934	9	0	690	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	7499	83	5621	25	0	1770	0
Rotation gr 2	2445	20	1867	11	0	547	0
Rotation gr 3	2574	26	1918	5	0	625	0
Rotation gr 4	2480	37	1836	9	0	598	0

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

	Total	RB260=1	=2	=3	=4	=5	Missing
Total	385	2	212	4	0	167	0
Rotation gr 2	57	0	34	3	0	20	0
Rotation gr 3	137	1	80	1	0	55	0
Rotation gr 4	191	1	98	0	0	92	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: legal marital status, social-economic status, household size, tenure status, usage of car in hh, whether hh can get 4300 kr in case of sudden need, household cope, lowest monthly income for household to make ends meet, number of children in the hh, geographical location (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 2010 EU-SILC the Real Estate data of year 2009 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 10 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 (PY020N), 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 4 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 for 2009 and 2010 (in 2008 prices were derived from the Household Budget Survey, which was not available in 2009 and 2010).
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).

2.5.5. Imputation of income variables

Where possible, data from previous years was used for imputation. Data of 2009 was used only if household or person received particular kind of income in 2009 and analysis showed that these two incomes are sufficiently closely related. If analysis indicated no correlation between the incomes of 2009 and 2010, values were not used in imputation.

Details on the number of values forwarded from 2009 to 2010 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 2010, i.e. including new sub-sample.

Table 2.16. Percentage of imputed cases by income component in national questionnaire, 2010

Code	Description	National code	Description	Total number of persons/ households having received an amount	No of missing values	No of values taken from previous year	Method of imputation, comments
PY010N	Employee cash or near cash income	H01N	Net wages in a year	5879	493	66	Random regression within given interval, amount per month
		H07B	Total amount of additional payments that had not been taken into consideration in net wages	470	29	0	Random regression
PY020N	Non-cash employee income	H21A2	Approximate value of motor fuel compensated by employer	175	24	7	Random regression
		H21B2	Approximate value of use of public transportation compensated by employer	63	29	4	Random regression
		H21C2	Approximate value of food at work received free or at a reduced price	336	135	29	Random regression
		H21D2	Approximate value of using company housing free of charge or on favorable terms	48	35	14	Random regression
		H21N2	Approximate value of housing costs compensated by employer	9	6	0	Median
		H21E2	Approximate value of (mobile) phone or postal services compensated by employer	372	58	23	Random regression
		H21F2	Approximate value of health services compensated by employer	196	63	0	Random regression
		H21G2	Approximate value of training not related to work but compensated by employer	11	1	0	Median
		H21H2	Approximate value of sporting possibilities free of charge or at reduced prices	296	101	34	Random regression

Code	Description	National code	Description	Total number of persons/ households having received an amount	No of missing values	No of values taken from previous year	Method of imputation, comments
		H21I2	Approximate value of holiday trip paid by employer	21	3	0	Median
		H21J2	Approximate value of foodstuffs	93	42	0	Random regression
		H21K2	Approximate value of leasing or loan at reduced interest rate	12	7	0	Median
		H21L2	Approximate value of use of equipment and/or other tools (e.g. power saw, lawnmower, etc.)	96	36	0	Random regression
		H21M2	Approximate value of other non-cash income from labour	15	3	0	Median
		H18, H20	Number of months a company car was used	See PY021N			
PY021N	Company car	H18	Possibility to use company car	5879	7	0	Logical deduction
		H20	Number of months a company car was used	223	10	0	Median
PY035N	Contributions to individual private pension plans	HK2	Payments made into the 3rd pillar of pension insurance	632	89	25	Random regression
		HK5	Payments made to the collecting insurance	132	13	0	Median
PY050N	Cash benefits or losses from self employment	H27N	Amount of loss from self-employment	13161	6	0	Random regression
		HC1	Interval of H28N	26	0	0	Random regression
		H28N	Net profit from self-employment	122	26	0	Random regression within given interval
		H35B	Net amount of royalties, remuneration or payment under contract for creative or scientific work	62	4	0	Random regression, amount per month
		H46A2	Income from private provision of fee-charging services to other persons or households	419	56	0	Median

Code	Description	National code	Description	Total number of persons/ households having received an amount	No of missing values	No of values taken from previous year	Method of imputation, comments
		H46B2	Income from the sale of own-produced consumer goods (e.g. handicrafts, souvenirs, etc.)	46	3	0	Median
		H46C2	Income from the sale of own-produced foodstuffs (e.g. pies, waffles, shashlik, etc.)	8	1	0	Median
		H46D2	Income from intermediate commercial transactions	15	3	0	Median
		H46E2	Income from agricultural or forestry activities	204	17	0	Random regression
		H46F2	Income from other unregistered self-employment	9	1	0	Median
PY080N	Pension from individual private plans	HK3A	Payments from the 3rd pillar of pension insurance	12	0	0	
		HK6A	Payments from collecting insurance	16	1	0	Median
PY090N	Unemployment benefits	H55A	Amount of unemployment benefit or any other benefits relating to unemployment	294	41	0	Random regression, amount per month
		HF6A	Amount of unemployment benefits paid by Estonian Unemployment Insurance Fund	359	44	0	Random regression, amount per month
		HF	Amount of redundancy payment in 2006 and/or 2007	3807	469	0	Random regression
PY100N	Old-age	H51A	Amount of old-age benefits	2867	65	47	Random regression, amount per month

Code	Description	National code	Description	Total number of persons/households having received an amount	No of missing values	No of values taken from previous year	Method of imputation, comments
	benefits	H52A	Amount of pension for incapacity for work or any other benefits relating to disability	1323	17	10	Random regression, amount per month
		HF_P	Lump-sum benefits upon retirement from employer	3807	469	0	Random regression
		HF10	Did You receive an one-off benefit paid by the city of Tallinn to pensioners	546	2	0	Logical deduction
PY110N	Survivors' benefits	H53A	Amount of survivors' pension or any other benefits relating to the loss of a provider	121	2	0	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	1086	446	153	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	12	1	0	Median
		H57B2	Amount of scholarship or grant awarded by a fund or organisation located in Estonia	23	2	0	Median
		H57D2	Amount of student loan written-off to a certain extent	85	16	8	Random regression
		H57G2	Education allowance	364	14	0	Random regression

Code	Description	National code	Description	Total number of persons/ households having received an amount	No of missing values	No of values taken from previous year	Method of imputation, comments
HY040N	Income from rental of a property or land	D09N	Net income from rental of property	150	9	0	Random regression
HY050N	Family/ children related allowances	D11B	Parental benefit received in the previous calendar year	320	19	0	Random regression
		D11F	Child allowance received in the previous calendar year	1870	7	0	Logical deduction
		D11D	Maternity leave benefits received in the previous calendar year	111	16	0	Random regression
		D11E	Net adoption allowance received in the previous calendar year	0	0	0	
		D11H	Childbirth allowance received in the previous calendar year	149	2	0	Logical deduction
		D11I	Child care allowance received in the previous calendar year	416	16	0	Random regression
		D11J	Single parent allowance received in the previous calendar year	158	1	0	Logical deduction
		D11K	Conscript's child allowance received in the previous calendar year	1	0	0	
		D11L	Allowance for a parent of seven or more children received in the previous calendar year	9	0	0	
		D11M	Local government child support received in the previous calendar year	138	3	0	Median
		D11N	Adoption allowance received in the previous calendar year	2	0	0	
		D11O	Allowance for child under guardianship received in the previous calendar year	19	0	0	

Code	Description	National code	Description	Total number of persons/ households having received an amount	No of missing values	No of values taken from previous year	Method of imputation, comments
HY060N	Social inclusion not elsewhere classified	H58A2	Amount of other support/benefit/pension not mentioned above	52	1	0	Median
		H58B2	Amount of other support/benefit/pension not mentioned above	4	0	0	
		HF8A	The net amount of conscript's benefit in the previous calendar year	42	8	0	Median
HY070N	Housing allowances	D03A	Amount of subsistence benefit	120	18	0	Random regression
HY080N	Regular inter-household cash transfers received	D16A	Amount of regular payments from other households	137	7	0	Random regression
		DA2A	Total alimony received previous calendar year	See HY081N			
HY081N	Alimonies received (compulsory + voluntary)	DA2A	Total alimony received previous calendar year	100	12	5	Random regression
HY090N	Interest, dividends, profit from capital investments in incorporated business	H49B2	Interest income from securities (shares, bonds)	9	2	0	Random regression
		H49C2	Dividend income from securities (i.e. shares, bonds)	91	18	2	Random regression
		HE2	Interest income from deposits in a bank – interval	3807	469	0	Random regression
		HE1A	Interest income from deposits in a bank – amount	3807	3734	0	Random number from given interval for lower intervals, hot-deck in last two intervals

Code	Description	National code	Description	Total number of persons/ households having received an amount	No of missing values	No of values taken from previous year	Method of imputation, comments
HY100N	Interest repayments on mortgage	D08B_E	Mortgage interest paid last year	512	78	0	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	37	3	0	Median
HY120N	Regular taxes on wealth	D10A	Amount of tax on land or any other property tax paid	3574	148	72	Random regression
HY130N	Regular inter-household cash transfers paid	D14A	Amount of regular payments to other households	181	2	0	Median
		DA1A	Total alimony paid previous calendar year	See HY131N			
HY131N	Alimonies paid (compulsory + voluntary)	DA1A	Total alimony paid previous calendar year	69	3	0	Random regression
HY145N	Repayments/receipts from tax adjustment	H64A	Income tax return for the income received in the previous year	2939	340	99	Random regression
		H63	Additional amount of tax paid on income in the previous calendar year	116	26	0	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 (CASI).

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 2009 and 2010 and to income data from administrative sources for income years 2008 and 2009. 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 decrease of about 15% in incomes from 2008 to 2009. (In Estonia, the income reference year is one year behind the survey year so the 2010 survey gives us annual incomes for 2009, and the 2009 survey gave us the incomes of 2008). The wage statistics figures show a similar decrease although the incomes reported through wage statistics by companies indicate systematically lower total wage incomes. The under-representation is around 12%, aka wage statistics incomes make up 88% 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 2008 and 2009

Income component	2008		2009	
	EU-SILC	Administrative sources*	EU-SILC	Administrative sources*
Cash or near-cash employee income (PY010N)	75,663,148	68,368,093	66,099,253	58,374,890
Old-age benefits (PY100G)	15,296,438	15,591,630	16,221,551	16,948,500
Gross survivor's benefits (PY110G)	873,759	2,052,940	199,74	238,97
Disability benefits (PY130G)	2,077,324	2,975,720	2,202,905	3,345,940

* 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 underreporting 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 2008 and 2009

Income component	2008		2009	
	EU-SILC	**Other sources***	**EU-SILC**	**Other sources***
Cash or near-cash employee income (PY010N)	698,938	543,432	651,463	488,166
Old-age benefits (PY100G)	284,371	293,4	285,281	297,273
Survivor's benefits (PY110G)	7,06	12,98	10,025	12,51
Disability benefits (PY130G)	73,554	73,11	74,504	79,48

* 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 2010 to the estimates from the 2009 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 2010 data actually refers to the incomes of 2009. EU-SILC in Estonia collects the respondent's annual income from the previous calendar year. Within a year the average salary decreased by 7%, while the number of wage receivers decreased 7%. Administrative data confirms the survey results.
At the same time, the number of people receiving unemployment benefits increased by 225%.The reason is that in Estonia the economic crisis started from 2008. Administrative data confirms the survey results. The increase in the mean of PY090N was 21%.
Most income components show a decrease from 2009 to 2010 in line with the decreasing salaries. Benefits from self-employment deceased considerably but the number of entrepreneurs increased. This has to do with sample fluctuations- there are very few self-employed people in the sample and fluctuations in their business dealings have a big effect on variable PY050. The number of entrepreneurs seems to fluctuate between survey years, which also hint to a relatively big pool of short-lived businesses.

Table 4.3. Mean (EEK-s) and number of recipients of income components in EU-SILC 2009 and 2010

	Mean		Number of recipients	
	2009	2010	2009	2010
Individual components	level			
PY010N	108254	101463	698938	651463
PY020N	10251	12173	182429	176378

PY035N	6004	6057	95614	84317
PY050N	20493	17594	54448	65553
PY090N	19649	23802	23436	76162
PY100N	52599	55316	284371	285281
PY110N	18601	19668	7060	10025
PY120N	4177	4469	165250	127791
PY130N	28242	29568	73554	74504
PY140N	10352	9646	30703	45441
Household	*level*			
components				
HY040N	20868	13983	11937	13203
HY050N	20655	22161	180377	176746
HY070N	5332	8731	10422	12240
HY080N	22664	23165	25563	31350
HY090N	3483	2497	255340	283952
HY110N	4686	2598	5251	5795
HY120N	598	668	397103	400789
HY130N	21197	19902	27661	32572
HY145N	-4342	-5638	266545	270067
HY010	208150	193679		
HY020	175302	164679		
HY022	165382	152483		
HY023	172048	153506		

Household level variables reflect changes in line with personal level variables.
HY040N decreased, despite the fact that similar numbers of people are receiving income from renting your property or land. Rents were low in 2009 (which includes benefits), so it is logical that the amount of income earned from renting a property is reduced.
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 increased and the amount of recipient households has increased. This must have been due to the economic crisis.
The number of households receiving and paying transfers from other households has increased and the sums paid have increased. This might have something to do with people having greater financial possibilities for helping their relatives with larger sums than before.
More people had to pay taxes on wealth, but the amount went up only a bit and has not changed substantially.

The drop in incomes earned by people aged under 16, despite the increase in overall recipients, is difficult to explain. Perhaps children worked for a smaller period of time, for instance one month during the summer vacation instead of two, because they could rely on more money from their households thanks to increased household incomes. Since the questionnaire does not specify how long the child worked and doing what, it is not possible to venture more than a guess.

Total household income decreased by approximately 8% in 2009. The decreases stem from the lower wages and other income components, most of which have gone down.

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, 2007-2010

ISCED level	2007		2008		2009		2010	
	EU-SILC	LFS	EU-SILC	LFS	EU-SILC	LFS	EU-SILC	LFS
0 Pre-primary education	0,6	0,5	0.5	0.4	.2*	0.5	,2*	0,4
1 Primary education	4,2	3,5	6.5	3.3	2,20	2,10	2,1	1,6
2 Lower secondary education	18,6	17,9	22.4	18.3	16,70	17,40	15,7	16,7
3 (Upper) secondary education	44,1	45,8	47.2	44.8	49.0	45.4	49	46
4 Post-secondary non tertiary education	7,1	5,2	2.6	5.2	3.0	4,90	3,2	5,6
5 First stage of tertiary education	24,9	26,8	20.6	27.8	28,60	29,60	29,4	29,5
6 Second stage of tertiary education	(0,4)**	(0,3)**	(0.3)***	(0,2)***	0.4	(0.2)*	0,4	,2*
Total	99,9	100.0	99.9	100.0	99.9	99.9	100	100

* 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 and 2010 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 and 2010 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.

Activity status	2007		2008		2009**		2010**	
	EU-SILC	HBS	EU-SILC	Safety survey	EU-SILC	LFS	EU-SILC	LFS
Working full-time	54.2	50.8	53.5	53.8	53.6	52.1	49,6	49,5
Working part-time	3.9	3.2	3.8	4.4	5.4	5.4	5,6	5,3
Unemployed	3.2	2.8	3.1	7.9	8.3	8.9	11,5	10,8
Pupil, student	8.8	12.3	8.7	9.5	9.4	9.6	9,5	10,2
In retirement	21.7	21.5	21.6	15.4	14.2	13.8	14,7	14,5
Permanently disabled	3.7	4.1	3.7	4	4.1	4.6	4,4	4,5
Fulfilling domestic tasks and care responsibilities	4.4	1.5	4.6	4.3	4.9	5.3	4,5	4,9
Other inactive	(0.0)*	3.4	(0.2)*	(0.7)*	.2*	.3*	.2*	.3*
Total	99.9	99.6	99.2	100	99.9	99.7	100	100

Table 4.5. Distribution of population aged 16 and over (2009 and 2010 aged 16-74) by self-defined activity status based on longitudinal EU-SILC, HBS, Safety survey and LFS, 2007-2010

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