EU-SILC 2008 in Estonia: Intermediate Quality Report

Tallinn 2009

CONTENTS

1. COMMON CROSS-SECTIONAL EUROPEAN UNION INDICATORS

Table 1.1. Common cross-sectional European Union indicators and their standard errors, 2008

Indicator	Value	Standard error
PORTFOLIO OF OVERARCHING INDICATORS		
At-risk-of-poverty threshold – one person household	**52076**	**553**
At-risk-of-poverty threshold – household with 2 adults and 2 children	**109359**	**-**
At-risk-of-poverty rate after social transfers: Total	**19.5**	**0.6**
At-risk-of-poverty rate after social transfers: Male	16.5	0.7
At-risk-of-poverty rate after social transfers: Female	22.0	0.7
At-risk-of-poverty rate after social transfers: 0-15 total	17.2	1.2
At-risk-of-poverty rate after social transfers: 16-24 total	15.7	1.1
At-risk-of-poverty rate after social transfers: 25-49 total	12.5	0.7
At-risk-of-poverty rate after social transfers: 50-64 total	19.5	1.1
At-risk-of-poverty rate after social transfers: 65+ total	39.0	1.7
At-risk-of-poverty rate after social transfers: 16+ total	19.9	0.6
At-risk-of-poverty rate after social transfers: 16-64 total	15.1	0,6
At-risk-of-poverty rate after social transfers: 0-64 total	15.5	0,6
At-risk-of-poverty rate after social transfers: 0-15 male	17.8	1.5
At-risk-of-poverty rate after social transfers: 16-24 male	15.4	1.5
At-risk-of-poverty rate after social transfers: 25-49 male	12.5	0.9
At-risk-of-poverty rate after social transfers: 50-64 male	19.0	1.5
At-risk-of-poverty rate after social transfers: 65+ male	24.6	2.2
At-risk-of-poverty rate after social transfers: 16+ male	16.2	0.7
At-risk-of-poverty rate after social transfers: 16-64 male	14.7	0.7
At-risk-of-poverty rate after social transfers: 0-64 male	15.4	0.7
At-risk-of-poverty rate after social transfers: 0-15 female	16.5	1.5
At-risk-of-poverty rate after social transfers: 16-24 female	16.1	1.4
At-risk-of-poverty rate after social transfers: 25-49 female	12.5	0.8
At-risk-of-poverty rate after social transfers: 50-64 female	20.0	1.4
At-risk-of-poverty rate after social transfers: 65+ female	46.1	1.9
At-risk-of-poverty rate after social transfers: 16+ female	23.0	0.7
At-risk-of-poverty rate after social transfers: 16-64 female	15.4	0.7
At-risk-of-poverty rate after social transfers: 0-64 female	15.6	0.7
At-risk-of-poverty rate after social transfers by most frequent activity status and gender		
At-risk-of-poverty rate after social transfers: At work total	7.3	0.4
At-risk-of-poverty rate after social transfers: Not at work total	38.0	1.2
At-risk-of-poverty rate after social transfers: Unemployed total	60.6	4.1
At-risk-of-poverty rate after social transfers: Retired total	43.2	1.8
At-risk-of-poverty rate after social transfers: Other inactive total	29.3	1.3
At-risk-of-poverty rate after social transfers: At work male	6.0	0.5
At-risk-of-poverty rate after social transfers: Not at work male	35.7	1.6
At-risk-of-poverty rate after social transfers: Unemployed male	61.6	4.7
At-risk-of-poverty rate after social transfers: Retired male	30.6	2.5
At-risk-of-poverty rate after social transfers: Other inactive male	34.4	2.1
At-risk-of-poverty rate after social transfers: At work female	8.6	0.6
At-risk-of-poverty rate after social transfers: Not at work female	39.4	1.3
At-risk-of-poverty rate after social transfers: Unemployed female	57.0	8.1
At-risk-of-poverty rate after social transfers: Retired female	48.6	1.9
At-risk-of-poverty rate after social transfers: Other inactive female	25.8	1.5
At-risk-of-poverty rate after social transfers by accommodation tenure status		
At-risk-of-poverty rate after social transfers: Owner	19.3	0.6
At-risk-of-poverty rate after social transfers: Rent	23.8	3.3

At-risk-of-poverty rate after social transfers by household type		
At-risk-of-poverty rate after social transfers:single, < 65 years	30.9	2.4
At-risk-of-poverty rate after social transfers:single, 65+ years	79.5	2.4
At-risk-of-poverty rate after social transfers:single, male	42.1	3.3
At-risk-of-poverty rate after social transfers:single, female	57.7	2.2
At-risk-of-poverty rate after social transfers:single, total	52.1	1.9
At-risk-of-poverty rate after social transfers:2 adults, no children, both < 65	10.0	1.2
At-risk-of-poverty rate after social transfers:2 adults, no children, at least one 65+	14.7	1.8
At-risk-of-poverty rate after social transfers:other households without children	9.2	1.5
At-risk-of-poverty rate after social transfers:single parent, at least one child	39.9	3.6
At-risk-of-poverty rate after social transfers:2 adults, 1 child	10.6	1.3
At-risk-of-poverty rate after social transfers:2 adults, 2 children	10.7	1.5
At-risk-of-poverty rate after social transfers:2 adults, 3+ children	21.6	3.0
At-risk-of-poverty rate after social transfers:other households with children	9.2	1.4
At-risk-of-poverty rate after social transfers:households without children	24.6	0.9
At-risk-of-poverty rate after social transfers:households with children	15.1	0.8
Relative median at-risk-of-poverty gap: Total	**20.3**	**0.8**
Relative median at-risk-of-poverty gap: Male	23.8	1.4
Relative median at-risk-of-poverty gap:Female	19.3	0.8
Relative median at-risk-of-poverty gap:0-15 total	24.2	2.3
Relative median at-risk-of-poverty gap:16-64 total	26.5	1.4
Relative median at-risk-of-poverty gap:65+ total	14.8	0.8
Relative median at-risk-of-poverty gap:16+ total	19.6	0.8
Relative median at-risk-of-poverty gap:16-64 male	29.3	2.0
Relative median at-risk-of-poverty gap:65+ male	13.4	1.5
Relative median at-risk-of-poverty gap:16+ male	23.7	1.4
Relative median at-risk-of-poverty gap:16-64 female	24.6	1.5
Relative median at-risk-of-poverty gap:65+ female	15.5	0.9
Relative median at-risk-of-poverty gap:16+ female	18.6	0.8
Dispersion around the at-risk-of-poverty threshold:		
Dispersion around the at-risk-of-poverty threshold: 40%	5.5	0.4
Dispersion around the at-risk-of-poverty threshold: 50%	11.5	0.5
Dispersion around the at-risk-of-poverty threshold: 70%	27.8	0.6
At-risk-of-poverty rate before social transfers, excl. old-age and survivors' benefits: Total	**24.7**	**0.6**
At-risk-of-poverty rate before social transfers (excl. pensions): Male	21.9	0.7
At-risk-of-poverty rate before social transfers (excl. pensions): Female	27.2	0.7
At-risk-of-poverty rate before social transfers (excl. pensions): 0-15 total	27.0	1.4
At-risk-of-poverty rate before social transfers (excl. pensions): 16-64 total	20.0	0.6
At-risk-of-poverty rate before social transfers (excl. pensions): 65+ total	41.3	1.7
At-risk-of-poverty rate before social transfers (excl. pensions): 16-64 male	19.5	0.7
At-risk-of-poverty rate before social transfers (excl. pensions): 65+ male	26.7	2.2
At-risk-of-poverty rate before social transfers (excl. pensions): 16-64 female	20.5	0.7
At-risk-of-poverty rate before social transfers (excl. pensions): 65+ female	48.6	1.9
S80/S20 income quintile share ratio	**5.0**	**0.1**
Inequality of income distribution: Gini coefficient	**30.9**	**0.6**
At-risk-of-poverty rate before social transfers (incl. old-age and survivors' benefits): Total	**36.3**	**0.5**
At-risk-of-poverty rate before social transfers (incl. pensions): Male	32.6	0.7
At-risk-of-poverty rate before social transfers (incl. pensions): Female	39.5	0.6
At-risk-of-poverty rate before social transfers (incl. pensions): 0-15 total	30.0	1.3
At-risk-of-poverty rate before social transfers (incl. pensions): 16-64 total	25.8	0.6
At-risk-of-poverty rate before social transfers (incl. pensions): 65+ total	83.9	0.8
At-risk-of-poverty rate before social transfers (incl. pensions): 16-64 male	24.6	0.7
At-risk-of-poverty rate before social transfers (incl. pensions): 65+ male	81.7	1.0
At-risk-of-poverty rate before social transfers (incl. pensions): 16-64 female	26.9	1.5
At-risk-of-poverty rate before social transfers (incl. pensions): 65+ female	85.0	1.1

Table 1.2. Equivalised disposable income

Indicator	Value (EEK)
Median equalised yearly disposable income	86 793
Mean equalised yearly disposable income	99 090

2. ACCURACY

2.1. Sample design

As the 2008 operation was the fifth round of EU-SILC in Estonia, the sample comprised of four parts:

1. The first part consists of households selected for the survey in 2005 and followed up in 2006, 2007 and 2008. (in total 543 households including fresh split-off households).

2. The second part consists of households selected for the survey in 2006 and followed-up in 2007 and 2008 (in total 1597 household including fresh split-off households).

3. The third part consists of households selected for the survey in 2007 and followed-up in 2008 (in total 1586 household including fresh split-off households).

4. Sample of 2421 households introduced into the survey in 2008.

In what follows we call parts 1, 2 and 3 together replications or old part of the sample and part 4 a new part of the sample.

Sub-sections 2.1.1 – 2.1.5 describe the design of new sub-sample. Sampling design of 2005, 2006 and 2007 sub-samples can be found in quality reports of respective years. They were also originally selected by unequal probability design, similar to one used for selection of new sub-sample.

Unequal probability design is likely to have negative effect on sample efficiency, and research on the possibilities of improving the design has been carried out and will continue in the future. So far, however, no suitable frame for selecting addresses has been found.

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 are shown in Table 2.1. R_g stands for the number of persons aged 14 and over living in stratum g as at 01.01.2008, n_g is the sample size of the stratum g and n_g/R_g (%) is the sampling fraction in the corresponding stratum.

Table 2.1. Stratification of the new part of the sample by counties, Estonian EU-SILC 2008

Stratum h	Counties	R_g	n_g	n_g/R_g %
Large	Tallinn, Harju, Ida-Viru, Lääne-Viru, Pärnu, Tartu	860,907	1,461	0.17
Small	Jõgeva, Järva, Lääne, Põlva, Rapla, Saare, Valga, Viljandi, Võru	286,676	878	0.31
Hiiu	Hiiu	9208	82	0.89

Next table shows sample size by rotational group: the initial sample size, number of split-off households and final sample size.

Table 2.2. Sample size by rotational group, 2008

Year a rotational group started	Rotational group	Initial sample size in 2007	Nr of split-off households	Final sample size
2005	5	532	11	543
2006	6	1542	55	1597
2007	7	1527	59	1586
2008	8	2421	-	2421
Total		6022	125	6147

2.1.5. Sample selection schemes

Systematic sampling of address-persons with foregoing sample sizes in each stratum. 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. The fieldwork period was from February to May 2008. For the convenience of fieldwork administration, the old part of the sample was equally allocated into the whole fieldwork period (with slightly smaller sample size in May), while the new part was allocated into the first three months (February-April) only. When allocating households into the months of fieldwork period, uniform workload of interviewers was targeted. Due to lack of interviewers in some areas, ca 5% of households was interviewed after the official end of fieldwork period in June 2008.

2.1.7. Renewal of sample: Rotational groups

The sample consists of 4 rotational groups:
1. 1 rotational group from 2005 (group 5);
2. 1 rotational group from 2006 (group 6);
3. 1 rotational group from 2007 (group 7)
4. new sub-sample (group 8).

2.1.8. Weighting

The sample of year 2008 consists of four sub-samples to be weighted independently and combined thereafter for final cross-sectional weight.

s_4	households started in 2005 and their split-offs, participate for the fourth time (= two rotational groups)
s_3	households started in 2006 and their split-offs, participate for the third time (= one rotational groups)
s_2	households started in 2007 and their split-offs, participate for the second time (= one rotational group)

s_1	households started in 2008, participate for the first time (= one rotational group)

For a new sub-sample s_1 the base weights are calculated from the beginning. For other sub-samples base weights for year 2008 are calculated from base weights of 2007 corrected for attrition between 2006 and 2007. Calculation of weights is done according to same procedure as in 2006 and 2007. For details, see 2006 Intermediate Quality Report, section 2.1.8

2.1.8.1.Design factor

Design weights DB080 were calculated to all households of the new part of the sample (subsample s_1). 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 were calculated for all households of the new sample (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

The weights are corrected for non-response as follows: For s_1 (new subsample) personal base weights are calculated as household design weights corrected for non-response. Correction for non-response was done with logistic regression model predicting the response probability of the household. Auxiliary variables were: sex and age of address-person, county group and urbanization status (rural/urban). County and urbanization status were defined with respect to the real place of residence where possible and according to PR otherwise.With this model, the response probability r_h was estimated for each household (responding and non-responding). Non-response corrected weights are thus $d_h^{*} = d_h / r_h$.

For replications (sub-samples s_2, s_3, s_4) personal base weights of 2007 are corrected for non-response. Correction for non-response was done with the help of logistic regression model with tenure status, household equivalised income, urbanization status and county of place of residence, social status, age, gender, ethnic nationality, household's assessment to its ability to make ends meet and number of children in household as auxiliary variables. Persons and households no longer in scope in 2008 were excluded prior to the correction as they are not considered as non-response. With the logistic model the response probability r_i of person i for year 2008, given he/she had responded in 2007, was estimated. Correction for non-response for replications is done on person-level, i.e. corrected weights of persons within one household no longer need to be constant. For sub-samples s_3 and s_4 also a correction for returnees was made. Corrections were done according to same procedure as in 2006 and 2007. For details, see 2006 Intermediate Quality Report.

The obtained base weights were extended on the new members of the household as follows:

- children born to sample women get the base weight of the mother;
- persons moving into sample household from outside the survey population receive the average of base weights of existing household members;
- persons moving into sample households from other non-sample households in the population receive zero base weight.

2.1.8.3. Adjustments to external data

Non-response corrected weights were calibrated to reproduce known population totals. Calibration totals originated from demographic statistics published by Statistics Estonia with institutionalized persons deducted. Calibration was made on the following auxiliaries:
- County (with capital Tallinn forming separate stratum);
- Sex (male/female) x age group (0-12, 13-14, 15, 16-19, 20-24, 25-29, 30-34, …, 65-69, 70-74, 75+);
- Urbanization status (rural/urban).

After calibration we get correction factor g_h for weight. $w_h = d_h^* g_h$. Calibration was done using the Bascula module of Blaise.

2.1.8.4. Final cross-sectional weight

To calculate household cross sectional weight DB090 different sub-samples were combined together. At first individual base weights within a household were averaged. As a result we get a base weight for each household in the sample (denoted by w_{1h}, w_{2h}, w_{3h}, and w_{4h} for the first, second, third and fourth sub-sample respectively). To get preliminary household cross-sectional weight we combine household base weight according to the following scheme:

$$w_h^{together} = \begin{cases} w_{1h} \cdot n_1/(n_1 + n_2 + n_3 + n_4), h \in s_1 \\ w_{2h} \cdot n_2/(n_1 + n_2 + n_3 + n_4), h \in s_2 \\ w_{3h} \cdot n_3/(n_1 + n_2 + n_3 + n_4), h \in s_3 \\ w_{4h} \cdot n_4/(n_1 + n_2 + n_3 + n_4), h \in s_4 \end{cases}$$

where n_i is number of responding households in sub-sample s_i, i=1,2,3,4. Thus, base weight of each sub-sample is reduced according to the share of corresponding sub-sample in the overall sample size. The preliminary household cross-sectional weights $w_h^{together}$ were calibrated to population totals (as personal weights of household members).

Personal cross-sectional weight of a person (RB050) is equal to the cross-sectional weight DB090 of its household.

Personal cross-sectional weights for all household members aged 16 and over (PB040) coincide in 2007 with weights RB050, as within-household non-response is imputed.

To get cross-sectional weights for child care (RL070), weights RB050 are adjusted in one-year age groups to reproduce number of children in the population.

Weights were not scaled at any step; the amount of correction at each step of weight computation procedure was carefully checked (no extreme correction factors appeared); at each calibration step, trimming was applied (with bounds at most 0.5 to 1.8).

2.1.8. Substitution
No substitution was used.

2.2. Sampling errors

2.2.1. Standard error and effective sample size

Variance estimation of the common cross-sectional EU indicators was done using the Deville linearization approach, with the help of linearization macros provided by Eurostat. After linearization the variance estimates were computed using the Bascula module of Blaise. All sub-samples were treated as if they were freshly selected. Variances were computed at the final stage of weighting procedure (2.1.8.3) together with final calibration.

Standard errors of the common cross-sectional indicators broken down by background variables are shown in Table 1.1.Of total 6,147 households in the sample, data of 4,744 household were accepted for the final database. In these households 10,761 persons aged 16+ were interviewed. Effective sample size is thus 3,953 households and 8,968 persons. (According to Commission regulation we use here design effect of at-risk-of-poverty rate, which was 1.2 according to the recent simulation study. Details can be found in 2007 Intermediate Quality Report.) Minimum requirements are thus satisfied (3500 households and 7750 persons).

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 2008 was 108 households, which makes the proportion of the over-coverage in the new part of the sample 4.5% and of the whole sample 1.8% (Table 2.4).

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 2007 2.9% of all records referring to persons aged 14+ were dropped before selecting the sample).

In the new part of the sample of 2007 there were 161 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 2.3. Reasons for over-coverage in the new part of the sample, 2008

Frame error	Number of households	Proportion in the frame error (%)
Total, of which	108	100
Address person was dead	14	13.0
Address person has left Estonia	76	70.4
Address person was staying in an institution	18	16.7
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 ESS questionnaire has been drafted following international experience in collecting income data. Where possible questions and wordings from Statistics Estonia's previous surveys, the reliability and validity of which had been checked in practise, were used. In 2007 the questionnaire was supplemented using the experience from the past three waves. The main corrections in the household questionnaire were adding in questions about production of foodstuffs for own consumption and questions allowing the calculation of savings from imputed rent. In the personal questionnaire the main developments in 2007 were adding questions about education obtained since the previous interview for the longitudinal panel, allowing the choice to report wage income as yearly or monthly and net or gross, adding questions about non-monetary income from wage labour and a separate block of income questions for entrepreneurs. The social benefit questions were also updated and additional checkpoints created to ask respondents the questions that concern their situation specifically. The questions on child-care, family benefits and unemployment benefits were also improved.

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

To reduce interviewer-induced measurement errors, the training programme was conducted in 4 smaller groups of about 15 people, with emphasis on practical work and discussion of mistakes from previous years. All returning interviewers attended a day long training session. During the training, the EU-SILC survey manager briefed the interviewers on all updates in the questionnaires, discussed previous years' errors, tracing rules 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 in a simulated situation using their laptop. At the end of the training session, each interviewer received personal feedback about their mistakes in the previous wave. 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.

Overall, 53 interviewers were responsible for conducting the interviews. The household– interviewer ratio was 90 households per interviewer.

2.3.2.2. Processing errors

Checking the data was done in three stages: data-entry checks during the interview, additional in-office checks during fieldwork and lastly data cleaning.

The data for 2008 operation was collected using CAPI. The data-entry program was written in Blaise and contained most of the consistency checks. In 2008, the Blaise consistency checks underwent further development, with many new logical checks creating error messages in described situations put in place. In Statistics Estonia, interviewers are required to react in some form to all error messages that occur during interviewing. The solution is either to correct an erroneous situation or if the situation is unusual but correct, add a remark to the data entry-program explaining this error. When assessing the quality of an interviewer's work, not adding a remark to an actually correct situation tat prompts an error message is also counted as an error. These logical checks allow to correct most of the errors already during an interview.

The primary data-entry consistency controls were of 6 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/should not have received a certain type of income reported it or not (e.g. whether households with children received family benefits, retired people (or people below retirement age) received pensions, employed persons received wages and so on);

   b. 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;

   c. Whether an educational level attained was possible below a certain age, or educational levels were possible in said combinations for given years;

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

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

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

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

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

   i. checks for correct survey area, interviewer code and personal numbers matching household numbers.

2) Lower and upper bounds of income variables (incl. benefits). 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 not allowing for occupations to be written on too general a scale for coding. (e.g. salesperson, cleaner)

5) Checks for goods produced for own consumption, for instance quantities;

6) 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 the 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 scrutinized one by one.

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

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

6) Errors in coding were gone through.

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

8) Consistency between time reported working under socio-economic status and months that salary was received.

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

10) Demographic information in the interviewers' reports was compared to the data recorded in the electronic questionnaires.

All mistakes found during 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 of the EU-SILC team and the interviewers' supervisors.

The number of primary consistency errors increased slightly after a dramatic drop in 2007. In 2006 there had been a total of 5654 errors, in 2007 the number had fallen to 1677 and in 2008 the total number of errors was 1779. Out of all the errors in 2008, 52% (934 cases) required callback and clarification with the interviewer or interviewee. In 2007, 60% of cases had required callback.

As can be seen from table 2.5, the most common types of errors 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.

Table 2.5. Interviewer errors and their processing, 2008

Type of error	Number of errors detected	Share of errors requiring a call-back
No remark explaining unusual situation	307	86%
Interviewer made an error, but did not correct it	403	58%
Interviewer's remark does not explain unusual situation	69	84%

Data not sufficient for coding	73	83%
Starting and finishing times recorded incorrectly	57	77%
Use of category Other, while a suitable category exists	337	12%
In-office checks	201	69%
Interviewer has misunderstood a question	255	44%
Data entry mistake	0	-
Not interviewers error	24	67%
Not codified	18	83%
Total	1779	

The third and final stage of data checks involved later in-office data cleaning. The controls implemented at this stage involved further checks of data consistency, consistency across time, and of extreme income values and as a final step the Eurostat data-checks. Extreme values for all income components as well as total income were checked and handled on a case-by-case basis.

2.3.3. Non-response errors

2.3.3.1. Achieved sample size

Data for 4744 households were accepted for database and used in analysis. This makes the overall share of complete household interviews accepted for the database 77.2%. On personal level, the share of complete personal interviews within the households accepted for the database was 99.2% – 10,761 interviews of possible 10,851. Income data for remaining 90 persons who didn't completed personal interview was imputed by closest neighbour full record imputation (RB250=14).

For rotational group breakdown see 2.3.3.3.

2.3.3.2. Unit non-response

Indicator	Total sample	New part
Address contact rate (Ra)	0.924	0.863
Proportion of complete household interviews accepted for the database (Rh)	0.855	0.725
Proportion of complete personal interviews within the households accepted for the database (Rp)	0.992	0.992
Household non-response rates (NRh)	21.0	37.4
Individual non-response rates (NRp)	0.8	0.8
Overall individual non-response rates ($*NRp$)	21.6	37.9

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 2.1.1):

DB120=21	DB120=22
▪ 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 ▪ 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 accesses due to poor weather conditions etc

2.3.3.3. Distribution of households by 'record of contact at address' (DB120), by 'household questionnaire result' (DB130) and by 'household interview acceptance' (DB135) for each rotational group and for the total

Table 2.6. Distribution of households by 'record of contact at address' (DB120) for each rotational group and in total, 2008

Record of contact at address	Rotational group 5		Rotational group 6		Rotational group 7		Rotational group 8		Total	
	N	%	N	%	N	%	N	%	N	%
Total (DB120=11 to 23)	543	100.0	1597	100.0	1586	100.0	2421	100.0	6147	100.0
Address contacted (DB120=11)	520	95.8	1533	96.0	1502	94.7	1995	82.4	5550	90.2
Address non-contacted (DB120=21 to 23)	23	4.2	64	4.0	84	5.3	426	17.6	597	9.7
Total address non-contacted (DB120=21 to 23)	23	100.0	64	100.0	84	100.0	426	100.0	597	100.0
Address cannot be located (DB120=21)	15	65.2	49	76.6	70	83.3	303	71.1	437	73.2
Address unable to access (DB120=22)	0	0.0	1	1.6	4	4.8	15	3.5	20	3.4
Address does not exist or is non-residential address or is unoccupied or not principal residence (DB120=23)	8	34.8	14	21.9	10	11.9	108	25.4	140	23.4

Table 2.7. Distribution of addresses contacted by 'household questionnaire result' (DB130) and by household interview acceptance (DB135), 2008

Household questionnaire result	Rotational group 5		Rotational group 6		Rotational group 7		Rotational group 8		Total	
	N	%	N	%	N	%	N	%	N	%
Total (DB130=11 to 24)	520	100.0	1533	100.0	1502	100.0	1995	100.0	5550	100.0
Household questionnaire completed (DB130=11)	498	95.8	1421	92.7	1391	92.6	1447	72.5	4757	85.7
Interview not completed (DB130= 21 to 24)	22	4.2	112	7.3	111	7.4	548	27.5	793	14.3
Total interview not completed (DB130=21 to 24)	22	100.0	112	100.0	111	100.0	548	100.0	793	100.0
Refusal to co-operate (DB130=21)	14	63.6	71	63.4	80	72.1	407	74.3	572	72.1
Entire household temporarily away for duration of fieldwork (DB130=22)	2	9.1	12	10.7	12	10.8	46	8.4	72	9.1
Household unable to respond (illness, incapacity, etc) (DB130=23)	2	9.1	9	8.0	12	10.8	44	8.0	67	8.4
Other (DB130=24)	4	18.2	20	17.9	7	6.3	51	9.3	82	10.3
Household questionnaire completed (DB135=1 to 2)	498	100.0	1421	100.0	1391	100.0	1447	100.0	4757	100.0
Interview accepted to database (DB135=1)	497	99.8	1414	99.5	1386	99.6	1447	100.0	4744	99.7

| Interview rejected (DB135=2) | 1 | 0.2 | 7 | 0.5 | 5 | 0.4 | 0 | 0 | 13.0 | 0.3 |

Table 2.8. Distribution of household members by Respondent Status (RB245) and rotational group, 2008

Responent Status	Rotational group 5		Rotational group 6		Rotational group 7		Rotational group 8		Total	
	N	%	N	%	N	%	N	%	N	%
Total (RB245=1 to 4)	1326	100.0	3965	100.0	3792	100.0	3949	100.0	13032	100.0
Current household members aged 16 and over (RB245 = 1)	1135	85.6	3332	84.0	3139	82.8	3245	82.2	10851	83.3
Selected respondent (RB245=2)	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Not selected respondent (RB245=3)	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Not eligible person (RB245=4)	191.0	14.4	633.0	16.0	653.0	17.2	704.0	17.8	2181	16.7

Table 2.9. Distribution of household members by Data Status (RB250) and rotational group, 2008

Data Status	Rotational group 5		Rotational group 6		Rotational group 7		Rotational group 8		Total	
	N	%	N	%	N	%	N	%	N	%
Total (RB250=11 to 33)	1135	100.0	3332	100.0	3139	100.0	3245	100.0	10851	100.0
Information completed only from interview (RB250 = 11)	1130	99.6	3308	99.3	3105	98.9	3218	99.2	10761	99.2
Information completed only from registers (RB250 = 12)	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Information completed from both (RB250 = 13)	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Information completed from full-record imputation (RB250=14)	5	0.4	24	0.7	34	1.1	27	0.8	90	0.8
Individual unable to respond (RB250=21)	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0

Individual failed to return self-completed questionnaire (RB250=22)	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Refusal to cooperate (RB250=23)	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Person temporarily away and no proxy available (RB250=31)	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
No contact for other reasons (RB250=32)	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Information not completed, reason unknown (RB250=33)	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0

Table 2.10. Distribution of household members by Type of Interview (RB260) and rotational group, 2008

Responent Status	Rotational group 5		Rotational group 6		Rotational group 7		Rotational group 8		Total	
	N	%	N	%	N	%	N	%	N	%
Total (RB260=1 to 5)	1130	100.0	3308	100.0	3105	100.0	3218	100.0	10761	100.0
Face to face interview - PAPI (RB260 = 1)	28	2.5	72	2.2	96	3.1	105	3.3	301	2.8
Face to face interview - CAPI (RB260 = 2)	946	83.7	2751	83.2	2600	83.7	2689	83.6	8986	83.5
CATI, telephone interview (RB260=3)	2	0.2	4	0.1	10	0.3	11	0.3	27	0.3
Self-administered by respondent (RB260=4)	0	0	1	0.0	0	0	0	0	1	0.0
Proxy interview (RB260=5)	154	13.6	480	14.5	399	12.9	413	12.8	1446	13.4

2.3.3.4. Distribution of substituted units

Substitution was not used.

2.3.3.5. Item non-response

The following table shows 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.

Income values imputed by full-record imputation are included.

A value obtained by gross/net conversion was not considered as non-response.

Table 2.11. Distribution of item non-response, household-level variables, 2008

Income variable	% of hhs having received an amount		% of hhs with missing values (before imputation)		% of hhs with partial information (before imputation)	
	Count	%	Count	%	Count	%
Total household gross income (HY010)	4739	99.9	50	1.1	2462	51.2
Total disposable household income (HY020)	4738	99.8	47	1.0	2564	54.1
Total disposable household income before social transfer other than old-age and survivors' benefits (HY022)	4706	99.2	122	2.6	2334	49.6
Total disposable household income before social transfers including old-age and survivors' benefits (HY023)	4452	93.8	286	6.4	2160	48.5
Net income components at household level						
Imputed rent (HY030N)	4604	97.0	4604	100.0	0	0.0
Income from rental of a property or land (HY040N)	91	1.9	7	7.7	0	0.0

Family/ children related allowances (HY050N)	1899	40.0	0	0.0	3	0.2
Social inclusion not elsewhere classified (HY060N)	79	1.7	0	0.0	0	0.0
Housing allowances (HY070N)	90	1.9	5	5.6	0	0.0
Regular inter-household cash transfers received (HY080N)	177	3.7	**7**	4.0	0	0.0
Alimonies received, compulsory and voluntary (HY081N)	113	2.4	4	3.5	0	0.0
Interest, dividends, profit from capital investments in incorporated business (HY090N)	2206	46.5	2077[1]	94.2	8	0.8
Interest repayments on mortgage (HY100N)	378	8.0	133	35.2	0	0.0
Income received by people aged under 16 (HY110N)	91	1.9	27	29.7	0	0.0
Regular taxes on wealth (HY120N)	3193	67.3	134	4.2	0	0.0
Regular inter-household cash transfers paid (HY130N)	173	3.6	7	4.0	0	0.0
Alimonies paid, compulsory and voluntary (HY131N)	59	1.2	4	6.8	0	0.0
Tax on income and social contributions, net (HY140N)[2]	0	0.0	0	-	0	-
Repayments/ receipts for tax adjustment (HY145N)	1993	42.0	257	12.9	43	2.2

[1] Of which 2055 are such that the only capital income of household is dividends from Estonian banks, and these are imputed based on the interval provided by respondent.

[2] Tax on income is not collected. This variable is fully computed at Statistics Estonia based on person's and household's income and taxes paid. Computed values is assumed to be gross, net values are set to zeroes in the database.

Gross income components at household level						
Imputed rent (HY030G)	4604	97.0	4604	100.0	0.0	100.0
Income from rental of a property or land (HY040G)	91	1.9	7	7.7	0	0.0
Family/ children related allowances (HY050G)	1899	40.0	0	0.0	3	0.2
Social inclusion not elsewhere classified (HY060G)	79	1.6	0	0.0	0	0.0
Housing allowances (HY070G)	90	1.9	5	5.6	0	0.0
Regular inter-household cash transfers received (HY080G)	177	3.7	7	4.0	0	0.0
Alimonies received, compulsory and voluntary (HY081G)	113	2.4	4	3.5	0	0.0
Interest, dividends, profit from capital investments in incorporated business (HY090G)	2206	46.5	2077	94.2	17	0.8
Interest repayments on mortgage (HY100G)	378	8.0	133	35.2	0	0.0
Income received by people aged under 16 (HY110G)	91	1.9	27	29.7	0	0.0
Regular taxes on wealth (HY120G)	3193	67.3	134	4.2	0	0.0
Regular inter-household cash transfers paid (HY130G)	173	3.6	7	4.0	0	0.0
Alimonies paid, compulsory and voluntary (HY131G)	59	1.2	4	6.8	0	0.0
Tax on income and social contributions, gross (HY140G)	3560	75.0	3560	100.0	0	0.0

Table 2.12. Distribution of item non-response, person-level variables, 2008

Income variable	% of persons 16+ having received an amount		% of persons with missing values (before imputation)		% of persons with partial information (before imputation)	
	Count	%	Count	%	Count	%
Net income components at personal level						
Employee cash or near cash income (PY010N)	6279	57.9	515	8.2	37	0.6
Non-cash employee income (PY020N)	1456	13.4	421	28.9	221	15.1
Company car (PY021N)[3]	242	2.2	242	100.0	0	0.0
Contributions to individual private pension plans (PY035N)	783	7.2	83	10.6	0	0.0
Cash benefits or losses from self employment (PY050N)	706	6.5	89	12.6	3	0.4
Value of goods produced by own-consumption (PY070N)	6935	6.4	6935	100.0	0	0.0
Pension from individual private plans (PY080N)	2	0.0	0	0.0	0	0.0
Unemployment benefits (PY090N)	89	0.8	9	10.1	0	0.0
Old-age benefits (PY100N)	2671	24.6	53	2.0	3	0.1
Survivor's benefits (PY110N)	106	1.0	3	2.8	0	0.0
Sickness benefits (PY120N)	988	9.1	379	38.4	0	0.0
Disability benefits (PY130N)	747	6.9	15	2.0	0	0.0
Education-related benefits (PY140N)	267	2.5	13	4.9	0	0.0
Gross income components at personal level						
Employee cash or near cash income (PY010G)	6279	57.9	515	8.2	37	0.7

---

[3] 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.

Non-cash employee income (PY020G)	1456	13.4	421	28.9	221	15.1
Company car (PY021G)	242	2.2	242	100.0	0	0.0
Employer's social insurance contributions (PY030G)	5979	55.1	5979	100.0	0	0.0
Contributions to individual private pension plans (PY035G)	783	7.2	83	10.6	1	0.1
Cash benefits or losses from self employment (PY050G)	706	6.5	89	12.6	3	0.4
Value of goods produced by own-consumption (PY070G)	6935	6.4	6935	100.0	0	0.0
Pension from individual private plans (PY080G)	2	0.0	0	0.0	0	0.0
Unemployment benefits (PY090G)	89	0.8	9	10.1	0	0.0
Old-age benefits (PY100G)	2671	24.6	53	2.0	3	0.1
Survivor's benefits (PY110G)	106	1.0	3	2.8	0	0.0
Sickness benefits (PY120G)	988	9.1	379	38.4	0	0.0
Disability benefits (PY130G)	747	6.9	15	2.0	0	0.0
Education-related benefits (PY140G)	267	2.5	13	4.9	0	0.0

2.3.3.6. Item non-response and number of observations at unit level of the common cross-sectional EU indicators

Table 2.13. Total item non-response and number of observations in the sample at unit level of the common cross-sectional EU indicators based on the cross-sectional component of EU–SILC, for equivalised disposable income, 2008

	Number of sample observations (achieved sample size)	Number of sample observations not taken into account due to item non-	Non-response at individual level (if applicable)	Non-response at household level (number of households)

		response		
At-risk-of-poverty rate after social transfers				
Total[4]	12999	0	NA	1263
By age and gender				1263
men total[4]	6091	0	NA	1263
women total[4]	6908	0	NA	1263
0-15 years[4]	2148	0	NA	1263
16-24 years	2321	0	NA	1263
25-49 years	4187	0	NA	1263
50-64 years	2224	0	NA	1263
65+ years	2119	0	NA	1263
16+ years	10851	0	NA	1263
16-64 years	8732	0	NA	1263
0-64 years[4]	10880	0	NA	1263
men 16-24 years	1202	0	NA	1263
men 25-49 years	2001	0	NA	1263
men 50-64 years	1038	0	NA	1263
men 65+ years	764	0	NA	1263
men 16+ years	5005	0	NA	1263
men 16-64 years	4241	0	NA	1263
men 0-64 years[4]	5327	0	NA	1263
women 16-24 years	1119	0	NA	1263
women 25-49 years	2186	0	NA	1263

---

[4] Children born in 2007 are excluded (33 persons in total).

women 50-64 years	1186	0	NA	1263
women 65+ years	1355	0	NA	1263
women 16+ years	5846	0	NA	1263
women 16-64 years	4491	0	NA	1263
women 0-64 years[4]	5553	0	NA	1263
By most frequent activity status and gender				1263
employed	5728	290	NA	1263
unemployed	219	290	NA	1263
retired	2205	290	NA	1263
other inactive	2409	290	NA	1263
men, employed	2923	290	NA	1263
men, unemployed	166	290	NA	1263
men, retired	728	290	NA	1263
men, other inactive	1041	290	NA	1263
women, employed	2805	290	NA	1263
women, unemployed	53	290	NA	1263
women, retired	1477	290	NA	1263
women, other inactive	1368	290	NA	1263
By household type[5]				1263
single, < 65 years	479	0	NA	1263
single, 65+ years	519	0	NA	1263
single, male	305	0	NA	1263
single, female	693	0	NA	1263
single, total	998	0	NA	1263
2 adults, no children, both < 65	1240	0	NA	1263

[5] Persons in households where it was impossible to determine household type are excluded (13 persons).

2 adults, no children, at least one 65+	1290	0	NA	1263
other households without children	1365	0	NA	1263
single parent, at least one child	708	0	NA	1263
2 adults, 1 child	1644	0	NA	1263
2 adults, 2 children	2092	0	NA	1263
2 adults, 3+ children	1310	0	NA	1263
other households with children	2339	0	NA	1263
households without children	4893	0	NA	1263
households with children	8093	0	NA	1263
By accommodation tenure status				1263
owner or rent-free	12571	0	NA	1263
tenant	461	0	NA	1263
At-risk-of-poverty threshold				1263
Median of the equivalised disposable household income	13032	0	NA	1263
At-risk-of-poverty threshold - total	13032	0	NA	1263
Inequality of income distribution S80/S20 income quintile share ratio	13032	0	NA	1263
Relative median at-risk-of-poverty gap				1263
Total	2607	0	NA	1263
By age and gender				1263
men total	1089	0	NA	1263
women total	1518	0	NA	1263
0-15 years	459	0	NA	1263
16-64 years	1489	0	NA	1263
65+ years	659	0	NA	1263

16+ years	2148	0	NA	1263
men, 16-64 years	703	0	NA	1263
men, 65+ years	153	0	NA	1263
men, 16+ years	856	0	NA	1263
women, 16-64 years	786	0	NA	1263
women, 65+ years	506	0	NA	1263
women, 16+ years	1292	0	NA	1263
Dispersion around the risk-of-poverty threshold				1263
40%	13032	0	NA	1263
50%	13032	0	NA	1263
70%	13032	0	NA	1263
At-risk-of-poverty rate before social transfers except old-age and survivors' benefits –total	12999	0	NA	1263
At-risk-of-poverty rate before social transfers including old-age and survivors' benefits – total [6]	12999	0	NA	1263
Gini coefficient	13032	0	NA	1263
Mean equivalised disposable income	13032	0	NA	1263

Notes:

Item non-response: number of questionnaires with no information on most frequent activity status, when applicable to indicator (317).

Non-response on individual level: not present in 2008 since income information in missing questionnaires is imputed

Non-response on household level: interview not competed, DB130=21 to 24 (793) + interview rejected, DB135=2 (13) + address cannot be located, DB120=21 (437) + address unable to access, DB120=22 (20).

2.4. Mode of data collection

Distribution of household members aged 16 and over by Data Status and by Type of Interview can be found in Tables 2.9 and 2.10 in Section 2.3.3.3.

2.5. Interview duration

Mean interview duration per household: 50 minutes and 12 seconds . Thus, mean interview duration per household is lower than the one-hour limit set in Regulation 1177/2003.

3. COMPARABILITY

3.1. Basic concepts and definitions

3.1.1. The reference population

Persons living in collective households are included in the reference population. The share of persons who are living in collective households and who are not at the same time members of some other private household is likely to be very low. Additionally, there is no feasible way to estimate their share in the total population. Thus, the exclusion of these persons is unlikely to affect the comparability and reliability of the estimates.

3.1.2. The private household definition

There were no divergences from the common definition.

3.1.3. The household membership

There were no divergences from the common definition.

3.1.4. The income reference period used

There were no divergences from the common definition. The income reference period was the previous calendar year (2007).

3.1.5. The period for taxes on income and social insurance contributions

There were no divergences from the common definition. Tax on income and social insurance contributions, as well as tax repayments and receipts refer to the income received during the income reference period (previous calendar year).

3.1.6. The reference period for taxes on wealth

There were no divergences from the common definition. Taxes on wealth paid during the income reference period (previous calendar year) were recorded.

3.1.7. The lag between the income reference period and current variables

The lag between the income reference period and current variables ranges from 3 to 7 months, thus not exceeding 8 months stipulated in the regulation.

3.1.8. The total duration of the data collection of the sample

Data collection was planned to last from February till June, but had to be extended by a month due to shortage of interviewers and low response rates. Thus, data was collected during a 6 month period, although the extension of the fieldwork period did not provide an improvement in the overall response rates by more than a few percentage points.

3.1.9. Basic information on activity status during the income reference period

There were no divergences from the common definition.

3.2. Components of income

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

3.2.1.1. Total household gross income
There were no divergences from the common definition.

3.2.1.2. Total disposable household income
There were no divergences from the common definition.

3.2.1.3. Total disposable household income, before social transfers other than old-age and survivors' benefits
There were no divergences from the common definition.

3.2.1.4. Total disposable household income, before social transfers including old-age and survivors' benefits
There were no divergences from the common definition.

3.2.1.5. Imputed rent
There were no divergences from the common definition. User cost method was employed, as the share of market rents is very small. External data used for modelling refers to survey year and not income year. As sale prices have been rising quickly, imputed rent value may consequently be overestimated compared to other income variables.

3.2.1.6. Interest, dividends, profit from capital investments in unincorporated business
There were no divergences from the common definition.

3.2.1.7. Interest paid on mortgages
There were no divergences from the common definition.

3.2.1.8. Income received by people aged under 16
Survivors' benefits received by people aged 15 or less are recorded under variable PY110 (see below).

3.2.1.9. Cash or near-cash employee income
There were no divergences from the common definition.

3.2.1.10. Cash profits or losses from self-employment (including royalties)
There were no divergences from the common definitions. Profits or losses reported in annual accounts for tax purposes were recorded. In the case of unregistered self-employment, the respondents were asked to estimate the income received this way.

3.2.1.11. Value of goods produced for own consumption
There were no divergences from the common definition. Most quantities were imputed from answers provided by respondents and unit costs were taken from the Household Budget Survey. Production costs were deducted from the total price thus obtained for own-consumption goods, and the profits were transferred to the personal level. The transfer was done by dividing the household aggregate characteristic by all members of the household aged 16 or over who answered the personal questionnaire. This value was added to their record as variable PY070N.
3.2.1.12. Unemployment benefits
There were no divergences from the common definition.

3.2.1.13. Survivors' benefits
If more than one household member is eligible for survivors' benefits, the individual benefits are, by default, combined and paid as a single sum to one household member. Due to infeasibility of dividing the survivors' benefit received between household members, the whole benefit is recorded only for the household member to whose account it was transferred. This can marginally affect variable HY110 (income received by those under 16), but has no effect on total household income.

3.2.1.14. Gross monthly earnings for employees
Variable was not recorded, as EU-SILC is not used to calculate the gender pay gap.

3.1.1.15. All other variables not listed above
There were no additional divergences.

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

Income variables were collected via face-to-face interviews at component or where applicable at sub-component level.

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

Table 3.1 summarizes mode in which different income variables were collected. It should be noted that where collection of only gross values is indicated designate in fact income components, which are not taxable (HY060, HY070, HY080, HY100, HY120, HY130, PY035, PY130, PY140), i.e. where gross equals net. Variables HY040,

HY110, PY010 and PY050 were collected as either net or gross, depending on which was easier for the respondent to report. The remaining variables were collected only in net.

Table 3.1. Mode of collection for gross income variables in Estonian EU-SILC 2008 operation

Income component	Collected gross	Collected net of tax and social contributions	Mixed mode net/gross
HY040			X
HY050		X	
HY060	X		
HY070	X		
HY080	X		
HY090		X	
HY100	X		
HY110			X
HY120	X		
HY130	X		
HY140		X	
HY145		X	
PY010			X
PY020		X	
PY035	X		
PY050			X
PY080		X	
PY090		X	
PY100		X	
PY110		X	
PY120		X	
PY130	X		
PY140	X		

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

Where only net values were collected or only net or gross value was recorded, the corresponding net and gross values were calculated on the basis of recorded values. Conversion algorithms were created on the basis of the local tax system. Information as to which taxes were paid on income components were also collected and taken into account in conversions.

4. COHERENCE

This section will compare the EU-SILC 2008 data to various external sources, including EU-SILC 2007, National Accounts (NA), the Labour Force Survey (LFS), the Safety Survey, wage statistics and social protection statistics. 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 is carried out according to the common EU methodology since 1995. The yearly sample size is about 12,000 working aged persons. From 2006, LFS is carried out using CAPI. LFS is the main source for labour market information.

The Safety survey is a nonrecurrent survey for Statistics Estonia carried out on commission from the Ministry of Social affairs and the Ministry of Justice. The sample was about 6000 persons and the survey was conducted using CAPI. The fieldwork took place between November 2008 and May 2009. The Safety survey is a source of crime, security and abuse statistics.

The Immigrant Population survey 2008, ii kvartal märts-juuni. Ca 5000.

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 Table 4.1, EU-SILC income data is compared component by component to income data from administrative sources for income year 2007. Table 4.1 presents the comparisons by total amounts and Table 4.2 by number of recipients. Only the income components where definitions are similar enough to warrant comparisons are presented here.

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

Income component	EU-SILC	Other sources*
Net cash or near-cash employee income (PY010N)	68,405,162	59,452,036
Gross old-age benefits (PY100G)	12,900,000	12,704,570
Gross sickness benefits (PY120G)[6]	428,000	1,694,150
Gross disability benefits (PY130G)	1,539,193	2,411,970
Gross survivor's benefits (PY110G)	141,000	195,680

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

Turning to the cash employee income first, the figure from wage statistics is almost 9 billion kroons lower than its EU-SILC equivalent. The difference was exactly the same with 2006 incomes. When comparing the number of people receiving wages and salaries, it appears that there are almost 150,000 persons more in EU-SILC who report

---

[6] Monthly in EU-SILC, per leave in administrative sources.

this type of income than in wage statistics. Again, the difference is very close to what it was in 2006. This difference with wage statistics is to be expected, given that wage statistics use full-time equivalents and not actual people as their units and that unlike EU-SILC unofficial work relationships are not included. That is to say, EU-SILC also catches part-time employment and unofficial earnings, making the amounts received higher and the number of recipients larger. In wage statistics, PY020G is included in PY010G and could not be separated from it for individual analysis. The data concerning wage statistics comes from in-house sources, not administrative registers.

Variable PY100G demonstrates a good overlap of survey data and administrative data. EU-SILC also includes pension benefits received from abroad (although there are very few such cases in the sample), which tend to be higher than national benefits. The survey also includes other old-age benefits that are not taken into account in the national administrative sources (such as local benefits provided by the local government to pensioners residing in their municipality). These can offer some explanation for the somewhat higher total figure received from the survey. On the other hand, the Estonian state pays old age benefits to its citizens residing abroad while the EU-SILC survey does not have people currently living abroad in its sample. This should in turn make the figure from national accounts comparatively higher.
Neither the number of recipients nor the average amounts paid as sickness benefits are available from administrative sources. The only information that can be used is the number of leaves taken and the total amount of benefits paid, which are both times higher than the respective figures from EU-SILC. This suggests that sickness benefits are underreported by about four times in EU-SILC and PY120 does not reflect the actual situation in the population. It is likely that respondents do not separate smaller amounts from wages and salaries and the variable has a very big measurement error.

Disability benefits too are underreported in EU-SILC, and the number of recipients is smaller than administrative accounts indicate. To a small extent this is 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.

Survivors' benefits are usually paid to a household as a whole. The administrative figure indicates underestimation in EU-SILC both in total amounts and number of recipients. The extent of the underestimation is similar to the case of disability benefits.

The old age benefits paid to the institutionalized population are not included in the administrative data sources' total amounts presented in Table 4.1 but they are included in the number of recipients in Table 4.2. The latter explains the higher number of recipients according to national records. Underestimation of number of recipients is probably also related to some below retirement age persons failing to report superannuated pensions in the case of EU-SILC.

Table 4.2. Number of recipients of income components by source of information, income year 2007

Income component	EU-SILC	Other sources*
Person-level components		
Net cash or near-cash employee income (PY010N)	690,633	543,060
Old-age benefits (PY100)	289,604	293,640
Disability benefits (PY130)	66,024	70,498
Net survivor's benefits (PY110G)	8,846	14,032

* 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 2008 to the estimates from the 2007 operation. Changes that emerge are, in general, in line with what could be expected. Much like the previous year, 2008 was a year of considerable economic growth and increase in real salaries in Estonia. It should be noted that the fieldwork period ended in July, just before the beginning of the economic crisis and that the 2008 data actually refers to the incomes of 2007. EU-SILC in Estonia collects the respondent's annual income from the previous calendar year. Within a year the average salary increased by 18%, while the increase in wage receivers was more modest – 4%. At the same time, the number of people receiving unemployment benefits went down by 15% as many discouraged workers found employment and unemployment levels reached record lows. Administrative data confirms the survey results. The increase in the mean of PY090N has to do with a new unemployment benefit attached to one's earlier salary which is considerably lower than the fixed unemployment allowance—more people become eligible for this new unemployment benefit every year as it is dependent on the time a person has worked before losing their job. The effect of this benefit is pulling up the mean of PY090 each year.

Most income components show an increase from 2007 to 2008 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 with sickness and disability benefits and family and education-related allowances. Benefits from self-employment deceased considerably and the number of entrepreneurs also fell. 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 hints 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 2007 and 2008

	Mean		Number of recipients	
	2007	2008	2007	2008
Individual level components				
PY010N	84117	99056	666,494	690573
PY020N	8676	9219	162,593	172318
PY035N	5,111	5308	84,401	92736
PY050N	36,000	20331	60,224	51762
PY090N	10,566	13218	10,836	9189
PY100N	37,409	43838	287,568	289604
PY110N	12,739	15964	8,482	8846
PY120N	3,261	3326	102,774	100303
PY130N	19,903	23312	66,379	66024
PY140N	10,092	8921	22,284	19675

Household level components

HY040N	11,290	12788	5,726	7720
HY050N	12,498	15274	185,502	182718
HY070N	4,295	5859	11,790	10000
HY080N	15,138	20989	28,870	23703
HY090N	1,519	2112	238,307	289877
HY110N	4,473	2991	6,092	6763
HY120N	430	477	356,151	348412
HY130N	14,728	18689	33,360	20927
HY145N	-2,455	-3074	213,965	228958
HY010	155,991	183157		
HY020	130,759	153603		
HY022	125,883	147552		
HY023	130,453	152713		

Household level variables reflect changes in line with personal level variables.

For HY040N the small 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 2007 (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 increased while the amount of recipient households has decreased. This must have been due to a small increase in the local level benefit in some local municipalities. Less people receiving the benefit probably has to do with rising income levels, making some people on the borderline ineligible.

The number of households receiving and paying transfers from other households has decreased while 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.

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. Less people had to pay taxes on wealth, but the amount went up only a bit and has not changed substantially. More people are having to pay tax but on smaller sums.

The drop in incomes earned by people aged under 16, despite the small 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 increased by approximately 17% in 2007. The increases stem from the higher wages and other income components, most of which have gone up considerably.

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. The differences are not great but 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, and the opposite trend was the case last year, this must be due to sample fluctuations.

Table 4.4. Distribution of population aged 16-74 by ISCED level, based on the EU-SILC and the LFS, 2008

ISCED level	EU-SILC	LFS
0 Pre-primary education	,5	,4
1 Primary education	6,5	3,3
2 Lower secondary education	22,4	18,3
3 (Upper) secondary education	47,2	44,8
4 Post-secondary non tertiary education	2,6	5,2
5 First stage of tertiary education	20,6	27,8
6 Second stage of tertiary education	,3*	,2*
Total	99,9	100.0

* Unreliable estimate, based on 35-44 sample observations.

Table 4.5 presents the comparison of population aged 15 or over by current activity status in EU-SILC and the Safety survey. When comparing the results 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.

Table 4.5. Distribution of population aged 15 and over by self-defined activity status based on EU-SILC and the Safety survey, 2008

Activity status	EU-SILC	Safety survey
Working full-time	53,5	53,8
Working part-time	3,8	4,4
Unemployed	3,1	7,9
Pupil, student	8,7	9,5
In retirement	21,6	15,4
Permanently disabled	3,7	4,0
Fulfilling domestic tasks and care responsibilities	4,6	4,3
Conscript	,2*	,1*
Other inactive	(0.0)*	0.6
Total	99.2	99.4

* Unreliable estimate, based on less than 20 sample observations.