

EU-SILC 2009 in Estonia: Intermediate Quality Report

Tallinn 2010

CONTENTS

1. COMMON CROSS-SECTIONAL EUROPEAN UNION INDICATORS	3
2. ACCURACY	10
2.1. Sample design	10
2.1.1. Type of sampling design	10
2.1.2. Sampling units.....	10
2.1.3. Stratification and sub-stratification criteria	10
2.1.4. Sample size and allocation criteria	10
2.1.5. Sample selection schemes	11
2.1.6. Sample distribution over time.....	11
2.1.7. Renewal of sample: Rotational groups	11
2.1.8. Weighting	11
2.1.8. Substitution	13
2.2. Sampling errors	13
2.2.1. Standard error and effective sample size	13
2.3. Non-sampling errors	14
2.3.1. Sampling frame and coverage errors.....	14
2.3.2. Measurement and processing errors	15
2.3.3. Non-response errors	18
2.4. Mode of data collection.....	25
2.5. Interview duration	25
3. COMPARABILITY	25
3.1. Basic concepts and definitions	25
3.1.1. The reference population	25
3.1.2. The private household definition	25
3.1.3. The household membership	25
3.1.4. The income reference period used	25
3.1.5. The period for taxes on income and social insurance contributions	25
3.1.6. The reference period for taxes on wealth	26
3.1.7. The lag between the income reference period and current variables.....	26
3.1.8. The total duration of the data collection of the sample	26
3.1.9. Basic information on activity status during the income reference period.....	26
3.2. Components of income.....	26
3.2.1. Differences between the national definitions and standard EU-SILC definitions	26
3.2.2. The source or procedure used for the collection of income variables	27
3.2.3. The form in which income variables at component level have been obtained	27
3.2.4. The method used for obtaining income target variables in the required form	28
4. COHERENCE.....	28
4.1. Comparison of income target variables and number of persons who receive income from each 'income component', with external sources.....	28
4.2. Comparison of other target variables with external sources	31

1. COMMON CROSS-SECTIONAL EUROPEAN UNION INDICATORS

Table 1.1. Common cross-sectional European Union indicators and their standard errors, 2009

Subpopulation	Estimate	Standard error	Design effect	Nr of values
At-risk-of-poverty rate, total	19,7	0,6	1,32	13542
<i>At-risk-of-poverty rate by gender and age</i>				
At-risk-of-poverty rate , after social transfers: Age 0-17	20,5	1,4	1,28	2751
At-risk-of-poverty rate , after social transfers: Age 18-24	15,6	1,2	1,48	1833
At-risk-of-poverty rate , after social transfers: Age 25-49	14,1	0,8	1,30	4275
At-risk-of-poverty rate , after social transfers: Age 50-64	19	1,1	1,25	2365
At-risk-of-poverty rate , after social transfers: Age 65+	33,9	1,7	1,27	2286
At-risk-of-poverty rate , after social transfers: Age 18+	19,5	0,6	1,32	10759
At-risk-of-poverty rate , after social transfers: Age 18-64	15,8	0,6	1,32	8473
At-risk-of-poverty rate , after social transfers: Age 0-64	16,8	0,7	1,31	11224
At-risk-of-poverty rate , after social transfers: Male	17,5	0,7	1,34	6361
At-risk-of-poverty rate , after social transfers: Female	21,6	0,8	1,30	7181
At-risk-of-poverty rate , after social transfers: Male Age 0-17	21,6	1,8	1,27	1411
At-risk-of-poverty rate , after social transfers: Male Age 18-24	14,9	1,5	1,57	946
At-risk-of-poverty rate , after social transfers: Male Age 25-49	13,7	1	1,30	2056
At-risk-of-poverty rate , after social transfers: Male Age 50-64	21,2	1,6	1,32	1110
At-risk-of-poverty rate , after social transfers: Male Age 65+	18,9	2,3	1,40	816
At-risk-of-poverty rate , after social transfers: Female Age 0-17	19,2	1,7	1,31	1340
At-risk-of-poverty rate , after social transfers: Female Age 18-24	16,4	1,5	1,39	887
At-risk-of-poverty rate , after social transfers: Female Age 25-49	14,5	0,9	1,31	2219
At-risk-of-poverty rate , after social transfers: Female Age 50-64	17,3	1,2	1,19	1255
At-risk-of-poverty rate , after social transfers: Female Age 65+	41,3	1,9	1,23	1470

Subpopulation	Estimate	Standard error	Design effect	Nr of values
At-risk-of-poverty rate , after social transfers: Male Age 18+	16,4	0,7	1,36	4950
At-risk-of-poverty rate , after social transfers: Female Age 18+	22,1	0,8	1,29	5841
At-risk-of-poverty rate , after social transfers: Male Age 18-64	15,9	0,7	1,35	4112
At-risk-of-poverty rate , after social transfers: Female Age 18-64	15,7	0,7	1,29	4361
At-risk-of-poverty rate , after social transfers: Male Age 0-64	17,3	0,8	1,33	5545
At-risk-of-poverty rate , after social transfers: Female Age 0-64	16,4	0,7	1,29	5711
<i>At-risk-of-poverty rate by household type</i>				
At-risk-of-poverty rate , after social transfers: One person hh under 65 years	30,9	2,6	1,23	496
At-risk-of-poverty rate , after social transfers: One person hh 65years and over	71,3	2,8	1,16	576
At-risk-of-poverty rate , after social transfers: One person hh, male	41,8	3,4	1,24	328
At-risk-of-poverty rate , after social transfers: One person hh, female	52,7	2,4	1,29	744
At-risk-of-poverty rate , after social transfers: One person hh, total	48,9	1,9	1,27	1072
At-risk-of-poverty rate , after social transfers: 2 adults, nodependant children, both adults under 65 years	9,5	1,1	1,04	1348
At-risk-of-poverty rate , after social transfers: 2 adults, nodependant children, at least one adult 65 years or more	11,6	1,4	1,08	1366
At-risk-of-poverty rate , after social transfers: Other hh without dependant children	7,6	1,2	1,10	1545
At-risk-of-poverty rate , after social transfers: Single parent hh, one or more dependant children	38,7	3,9	1,35	702
At-risk-of-poverty rate , after social transfers: 2 adults, one dependant child	12,3	1,5	1,14	1659
At-risk-of-poverty rate , after social transfers: 2 adults, two dependant children	14,0	1,7	1,23	2040
At-risk-of-poverty rate , after social transfers: 2 adults, three or more dependant children	28,6	3,3	1,23	1250
At-risk-of-poverty rate , after social transfers: Other hh with dependant children	11,8	1,5	1,09	2509
At-risk-of-poverty rate , after social transfers: Hh without dependant children	21,9	0,9	1,29	5331
At-risk-of-poverty rate , after social transfers: Hh with dependant children	17,7	0,9	1,29	8160
<i>At-risk-of-poverty rate by accomodation tenure status</i>				

Subpopulation	Estimate	Standard error	Design effect	Nr of values
At-risk-of-poverty rate , after social transfers: Owner or rent free	19,5	0,6	1,31	13103
At-risk-of-poverty rate , after social transfers: Tenant	25,2	3,8	1,30	437
<i>At-risk-of-poverty rate by main activity status</i>				
At-risk-of-poverty rate , after social transfers: Employed	8,1	0,4	1,26	5848
At-risk-of-poverty rate , after social transfers: Unemployed	55,3	3,7	1,34	375
At-risk-of-poverty rate , after social transfers: Retired	37,9	1,8	1,26	2356
At-risk-of-poverty rate , after social transfers: Other inactive	23,1	1,6	1,48	1429
At-risk-of-poverty rate , after social transfers: Employed, Male	6,5	0,5	1,17	2928
At-risk-of-poverty rate , after social transfers: Unemployed, Male	56,3	4,5	1,35	264
At-risk-of-poverty rate , after social transfers: Retired, Male	24,6	2,7	1,38	785
At-risk-of-poverty rate , after social transfers: Other inactive, Male	22,2	2,1	1,50	697
At-risk-of-poverty rate , after social transfers: Employed, Female	9,7	0,6	1,31	2920
At-risk-of-poverty rate , after social transfers: Unemployed, Female	52,5	5,8	1,27	111
At-risk-of-poverty rate , after social transfers: Retired, Female	43,9	2	1,22	1571
At-risk-of-poverty rate , after social transfers: Other inactive, Female	23,8	2	1,46	732
At-risk-of-poverty rate , after social transfers: Not employed	35,7	1,3	1,34	4204
At-risk-of-poverty rate , after social transfers: Not employed, Male	30,2	1,8	1,44	1771
At-risk-of-poverty rate , after social transfers: Not employed, Female	39,3	1,5	1,28	2433
<i>At risk-of-poverty rate by work intensity</i>				
At-risk-of-poverty rate , after social transfers: hh without dependent children, w=0	48,0	1,9	1,23	2115
At-risk-of-poverty rate , after social transfers: hh without dependent children, 0<w<1	13,7	1,6	1,31	1400
At-risk-of-poverty rate , after social transfers: hh without dependent children, w=1	3,6	0,5	1,30	1816
At-risk-of-poverty rate , after social transfers: hh with dependent children, w=0	72,7	7,9	1,83	295
At-risk-of-poverty rate , after social transfers: hh with dependent children, 0<w<0,5	62,6	5,1	1,34	546

Subpopulation	Estimate	Standard error	Design effect	Nr of values
At-risk-of-poverty rate , after social transfers: hh with dependent children, $0,5 \leq w < 1$	17,2	1,7	1,31	3410
At-risk-of-poverty rate , after social transfers: hh with dependent children, $w=1$	10,4	1,0	1,28	3909
At-risk-of-poverty rate , before social transfers including pensions, total	25,9	0,6	1,30	13542
At-risk-of-poverty rate , before social transfers including pensions:Age 0-17	29,7	1,6	1,29	2751
At-risk-of-poverty rate , before social transfers including pensions:Age 18-24	21,5	1,4	1,41	1833
At-risk-of-poverty rate , before social transfers including pensions:Age 25-49	20,2	0,9	1,30	4275
At-risk-of-poverty rate , before social transfers including pensions:Age 50-64	25,6	1,3	1,22	2365
At-risk-of-poverty rate , before social transfers including pensions:Age 65+	36,4	1,6	1,26	2286
At-risk-of-poverty rate , before social transfers including pensions:Age 18-64	22	0,6	1,30	8473
At-risk-of-poverty rate , before social transfers including pensions:Male	23,6	0,7	1,31	6361
At-risk-of-poverty rate , before social transfers including pensions:Female	27,7	0,8	1,29	7181
At-risk-of-poverty rate , before social transfers including pensions:Male Age 0-17	30,8	2,1	1,27	1411
At-risk-of-poverty rate , before social transfers including pensions:Male Age 18-24	19,9	1,7	1,46	946
At-risk-of-poverty rate , before social transfers including pensions:Male Age 25-49	19,4	1,1	1,28	2056
At-risk-of-poverty rate , before social transfers including pensions:Male Age 50-64	28,6	1,7	1,28	1110
At-risk-of-poverty rate , before social transfers including pensions:Male Age 65+	20,8	2,4	1,37	816
At-risk-of-poverty rate , before social transfers including pensions:Female Age 0-17	28,5	1,8	1,31	1340
At-risk-of-poverty rate , before social transfers including pensions:Female Age 18-24	23,1	1,8	1,37	887
At-risk-of-poverty rate , before social transfers including pensions:Female Age 25-49	21	1	1,32	2219
At-risk-of-poverty rate , before social transfers including pensions:Female Age 50-64	23,2	1,4	1,17	1255
At-risk-of-poverty rate , before social transfers including pensions:Female Age 65+	44	1,8	1,22	1470
At-risk-of-poverty rate , before social transfers including pensions:Male Age 18+	21,8	0,7	1,32	4950
At-risk-of-poverty rate , before social transfers including pensions:Female Age 18+	27,6	0,7	1,29	5841

Subpopulation	Estimate	Standard error	Design effect	Nr of values
At-risk-of-poverty rate , before social transfers including pensions:Male Age 18-64	21,9	0,8	1,32	4112
At-risk-of-poverty rate , before social transfers including pensions:Female Age 18-64	22	0,8	1,29	4361
At-risk-of-poverty rate , before social transfers including pensions:Male Age 0-64	24	0,9	1,30	5545
At-risk-of-poverty rate , before social transfers including pensions: Female Age 0-64	23,4	0,8	1,29	5711
At-risk-of-poverty rate , before social transfers excluding pensions, total	37,5	29,2	1,29	13542
At-risk-of-poverty rate , before social transfers excluding pensions:Age 0-17	32,1	1,6	1,28	2751
At-risk-of-poverty rate , before social transfers excluding pensions:Age 18-24	24	1,4	1,39	1833
At-risk-of-poverty rate , before social transfers excluding pensions:Age 25-49	22,9	0,9	1,28	4275
At-risk-of-poverty rate , before social transfers excluding pensions:Age 50-64	36,5	1,4	1,22	2365
At-risk-of-poverty rate , before social transfers excluding pensions:Age 65+	83,7	1	1,18	2286
At-risk-of-poverty rate , before social transfers excluding pensions:Age 18-64	27	0,7	1,29	8473
At-risk-of-poverty rate , before social transfers excluding pensions:Male	34,1	0,8	1,29	6361
At-risk-of-poverty rate , before social transfers excluding pensions:Female	40,4	0,8	1,30	7181
At-risk-of-poverty rate , before social transfers excluding pensions:Male Age 0-17	33,1	2,2	1,27	1411
At-risk-of-poverty rate , before social transfers excluding pensions:Male Age 18-24	22,4	1,7	1,43	946
At-risk-of-poverty rate , before social transfers excluding pensions:Male Age 25-49	22,2	1,2	1,26	2056
At-risk-of-poverty rate , before social transfers excluding pensions:Male Age 50-64	36,7	1,8	1,27	1110
At-risk-of-poverty rate , before social transfers excluding pensions:Male Age 65+	80,4	1,5	1,20	816
At-risk-of-poverty rate , before social transfers excluding pensions:Female Age 0-17	31	1,8	1,30	1340
At-risk-of-poverty rate , before social transfers excluding pensions:Female Age 18-24	25,7	1,9	1,35	887
At-risk-of-poverty rate , before social transfers excluding pensions: Female Age 25-49	23,5	1	1,31	2219
At-risk-of-poverty rate , before social transfers excluding pensions: Female Age 50-64	36,4	1,6	1,19	1255
At-risk-of-poverty rate , before social transfers excluding pensions: Female Age 65+	85,3	1,1	1,17	1470

Subpopulation	Estimate	Standard error	Design effect	Nr of values
At-risk-of-poverty rate , before social transfers excluding pensions: Male Age 18+	34,4	0,8	1,29	4950
At-risk-of-poverty rate , before social transfers excluding pensions: Female Age 18+	42,3	0,8	1,30	5841
At-risk-of-poverty rate , before social transfers excluding pensions: Male Age 18-64	26	0,8	1,30	4112
At-risk-of-poverty rate , before social transfers excluding pensions: Female Age 18-64	27,8	0,8	1,28	4361
At-risk-of-poverty rate , before social transfers excluding pensions: Male Age 0-64	27,7	0,8	1,29	5545
At-risk-of-poverty rate , before social transfers excluding pensions: Female Age 0-64	28,5	0,9	1,29	5711
Median equivalised disposable income	97150	711,92	1,33	13542
At-risk-of-poverty threshold	58290	427,15	1,33	13542
At-risk-of-poverty threshold, one person hh	58290	427,15	1,33	13542
At-risk-of-poverty threshold, hh 2 adults 2 dependent children	58290	427,15	1,33	13542
S80/S20	5,009	0,1259	1,52	13542
Relative median at-risk-of-poverty gap	17	3,3	1,33	2819
Relative median at-risk-of-poverty gap: Age 0-17	19,8	1,1	1,33	684
Relative median at-risk-of-poverty gap: Age 18-24	25,5	5,4	1,44	334
Relative median at-risk-of-poverty gap: Age 25-49	21,1	4,1	1,38	721
Relative median at-risk-of-poverty gap: Age 50-64	24,9	1,9	1,25	469
Relative median at-risk-of-poverty gap: Age 65+	11,4	0,9	1,21	601
Relative median at-risk-of-poverty gap: Age 18-64	23,1	2,3	1,36	1524
Relative median at-risk-of-poverty gap: Age 18+	16,7	0,9	1,33	2125
Relative median at-risk-of-poverty gap: Male	20,7	1,6	1,34	1206
Relative median at-risk-of-poverty gap:Female	15,5	1,2	1,33	1613
Relative median at-risk-of-poverty gap: Male Age 0-17	20,6	3,3	1,33	353
Relative median at-risk-of-poverty gap: Male Age 18-24	24,1	3,8	1,43	168
Relative median at-risk-of-poverty gap: Male Age 25-49	25,3	4	1,36	330
Relative median at-risk-of-poverty gap: Male Age 50-64	28,6	1,9	1,27	232
Relative median at-risk-of-poverty gap: Male Age more then 64	8,7	2,8	1,21	115

Subpopulation	Estimate	Standard error	Design effect	Nr of values
Relative median at-risk-of-poverty gap: Male, Age 18-64	26,7	2,2	1,36	730
Relative median at-risk-of-poverty gap: Male, Age 18+	20,8	3	1,34	845
Relative median at-risk-of-poverty gap: Female Age 0-17	19,6	3	1,32	331
Relative median at-risk-of-poverty gap: Female Age 18-24	28	3,5	1,44	166
Relative median at-risk-of-poverty gap: Female Age 25-49	19	3,9	1,41	391
Relative median at-risk-of-poverty gap: Female Age 50-64	20,2	4,7	1,23	237
Relative median at-risk-of-poverty gap: Female Age more then 64	11,9	1,2	1,21	486
Relative median at-risk-of-poverty gap: Female, Age 18-64	20,7	2,3	1,36	794
Relative median at-risk-of-poverty gap: Female, Age 18+	14,7	0,8	1,33	1280
Median income below the at-risk-of-poverty threshold	48359	400,71	1,33	2819
Dispersion around the risk-of-poverty threshold- 40%	5,3	0,4	1,31	13542
Dispersion around the risk-of-poverty threshold- 50%	10,2	0,5	1,25	13542
Dispersion around the risk-of-poverty threshold- 70%	27,5	0,7	1,30	13542
Gini coefficient	31,4	0,5	1,78	13542
Mean equivalised disposable income	112755	1566,4	1,67	13542

Notes:

In domains by age, children born in 2009 are excluded (32 persons).

In domains by tenure status, persons in household with unknown tenure status are excluded (2 persons).

In domains by main activity status, persons with personal questionnaire imputed (88 persons) and those for whom it was impossible to define main activity status (1212 persons) were excluded.

In domains by household type, persons in households where it was impossible to determine the type (51 persons) were excluded.

2. ACCURACY

2.1. Sample design

The 2009 sample comprised of four parts:

1. The first part consists of households selected for the survey in 2006 and followed up in 2007, 2008 and 2009. (in total 1506 households including fresh split-off households).
2. The second part consists of households selected for the survey in 2007 and followed-up in 2008 and 2009 (in total 1481 household including fresh split-off households).
3. The third part consists of households selected for the survey in 2008 and followed-up in 2009 (in total 1503 household including fresh split-off households).
4. Sample of 2325 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 2006, 2007 and 2008 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 01.01.2009, 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 2009

Stratum h	Counties	R_g	n_g	n_g/R_g %
Large	Tallinn, Harju, Ida-Viru, Lääne-Viru, Pärnu, Tartu	861376	1404	0.16
Small	Jõgeva, Järva, Lääne, Põlva, Rapla, Saare, Valga, Viljandi, Võru	285261	843	0.30
Hiiu	Hiiu	9233	78	0.84

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, 2009

Year a rotation group started	Rotational group	Initial sample size in 2009	Nr of split-off households	Final sample size
2006	2	1459	47	1506
2007	3	1425	56	1481
2008	4	1447	56	1503
2009	1	2325	0	2325
Total		6656	159	6815

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 2009. 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 9% of households was interviewed after the official end of fieldwork period in June 2009.

2.1.7. Renewal of sample: Rotational groups

The sample consists of 4 rotational groups:

1. 1 rotational group from 2006 (DB075= 2);
2. 1 rotational group from 2007 (DB075= 3);
3. 1 rotational group from 2008 (DB075= 4);
4. new sub-sample (DB075=1).

2.1.8. Weighting

The sample of year 2009 consists of four sub-samples to be weighted independently and combined thereafter for final cross-sectional weight.

s_4	households started in 2006 and their split-offs, participate for the fourth time
s_3	households started in 2007 and their split-offs, participate for the third time
s_2	households started in 2008 and their split-offs, participate for the second time
s_1	households started in 2009, participate for the first time

For a new sub-sample s_1 the base weights are calculated from the beginning. For other sub-samples base weights for year 2009 are calculated from base weights of 2008 corrected for attrition between 2008 and 2009. Calculation of weights is done according to same procedure as in 2005-2008.

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 than 8, household size was stochastically imputed (with county as an auxiliary variable).

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 2008 are corrected for attrition. Correction for attrition 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 2009 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 2009, given he/she had responded in 2008, 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 2005-2008.

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 2009 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 Jackknife Repeated Replication method with programs provided by Eurostat (project director Giulio GHELLINI, Project 2008/s 105-140310, accessed in CIRCA).

Standard errors of the common cross-sectional indicators broken down by background variables are shown in Table 1.1.

Of total 6815 households in the sample, data of 4965 household were accepted for the final database. In these households, 11220 persons aged 16+ were interviewed. Effective sample size is thus 3761 households and 8500 persons. (According to Commission regulation we use here design effect of at-

risk-of-poverty rate, which was 1.32 according to the Table 1.1). 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 2009 was 99 households, which makes the proportion of the over-coverage in the new part of the sample 4,3% and of the whole sample 1,5% (Table 2.3).

Since there is no registration law in Estonia, people do not need to show their actual addresses in the Population Register. For that reason the register contains some amount of records without any address and for some part of records the address shown is not correct. Records without an address or incomplete address were dropped out of the register before selecting the sample (for example, in 2009 ca 2% of all records referring to persons aged 14+ were dropped before selecting the sample).

In the new part of the sample of 2009 there were 179 address-persons whose 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 assess, 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, 2009

Frame error	Number of households	Proportion in the frame error (%)
Total, of which	99	100.0
Address person was dead	18	18.2
Address person has left Estonia	12	12.1
Address person was staying in an institution	69	69.7
Address person was surveyed through one of his/her household members	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 questions about managerial duties for current and last job were added and 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. 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.

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 neighbours or from the street reformulated to better meet Eurostat's guidelines
- 6) The question about pollution, grime or other environmental problems reformulated to better meet Eurostat's guidelines
- 7) The questions about material deprivation were added

To reduce 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 2009 operation was collected using CAPI. The data-entry program was written in Blaise and contained most of the consistency checks. In 2009, 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 that 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. In 2008 the total number of errors was 1779 and in 2009 the number of errors was 1939.

Out of all the errors in 2009, 49% (953 cases) required callback and clarification with the interviewer or interviewee. In 2008, 52% of cases had required callback.

As can be seen from table 2.5, the most common types of errors in 2009 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 2008 not correcting mistakes, using 'other' when unnecessary and not making remarks were also the most common error types.

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, 2009

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

The 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 4965 households were accepted for database and used in analysis. This makes the overall share of complete household interviews accepted for the database 72,9%. On personal level, the share of complete personal interviews within the households accepted for the database was 99,2% – 11220 interviews of possible 11308. Income data for remaining 88 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 (<i>Ra</i>)	0,89	0,79
Proportion of complete household interviews accepted for the database (<i>Rh</i>)	0,84	0,68
Proportion of complete personal interviews within the households accepted for the database (<i>Rp</i>)	0,99	0,99
Household non-response rates (<i>NRh</i>)	25,4	45,9
Individual non-response rates (<i>NRp</i>)	0,8	1,1
Overall individual non-response rates (<i>*NRp</i>)	26,0	46,5

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
<ul style="list-style-type: none"> ▪ 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 	<ul style="list-style-type: none"> ▪ 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 2.4. Distribution of households by 'record of contact at address' (DB120) for each rotational group and in total, 2009

Record of contact at address	Rotation group 2		Rotation group 3		Rotation group 4		Rotation group 1		Total	
	N	%	N	%	N	%	N	%	N	%
Total (DB120=11 to 23)	1506	100	1481	100	1503	100	2325	100	6815	100
Address contacted (DB120=11)	1396	92.7	1369	92.4	1405	93.5	1766	76.0	5936	87.1
Address non-contacted (DB120=21 to 23)	110	7.3	112	7.6	98	6.5	559	24.0	879	12.9
Total address non-contacted (DB120=21 to 23)	110	100.0	112	100.0	98	100.0	559	100.0	879	100.0
Address cannot be located (DB120=21)	86	78.2	89	79.5	72	73.5	424	75.8	671	76.3
Address unable to access (DB120=22)	3	2.7	7	6.3	5	5.1	36	6.4	51	5.8
Address does not exist or is non-residential address or is unoccupied or not principal residence (DB120=23)	21	19.1	16	14.3	21	21.4	99	17.7	157	17.9

Table 2.5. Distribution of addresses contacted by 'household questionnaire result' (DB130) and by household interview acceptance (DB135), 2009

Household questionnaire result	Rotation group 2		Rotation group 3		Rotation group 4		Rotation group 1		Total	
	N	%	N	%	N	%	N	%	N	%
Total (DB130=11 to 24)	1396	100.0	1369	100.0	1405	100.0	1766	100.0	5936	100.0
Household questionnaire completed (DB130=11)	1292	92.6	1223	89.3	1250	89.0	1207	68.3	4972	83.8
Interview not completed (DB130= 21 to 24)	104	7.4	146	10.7	155	11.0	559	31.7	964	16.2
Total interview not completed (DB130=21 to 24)	104	100.0	146	100.0	155	100.0	559	100.0	964	100.0
Refusal to co-operate (DB130=21)	72	69.2	108	74.0	116	74.8	449	80.3	745	77.3
Entire household temporarily away for duration of fieldwork (DB130=22)	8	7.7	4	2.7	19	12.3	26	4.7	57	5.9
Household unable to respond (illness, incapacity, etc) (DB130=23)	17	16.3	15	10.3	11	7.1	49	8.8	92	9.5
Other (DB130=24)	7	6.7	19	13.0	9	5.8	35	6.3	70	7.3
Household questionnaire completed (DB135=1 to 2)	1292	100.0	1223	100.0	1250	100.0	1207	100.0	4972	100.0
Interview accepted to database (DB135=1)	1290	99.8	1221	99.8	1249	99.9	1205	99.8	4965	99.9
Interview rejected (DB135=2)	2	0.2	2	0.2	1	0.1	2	0.2	7	0.1

Table 2.6. Distribution of household members by Respondent Status (RB245) and rotational group, 2009

Respondent Status	Rotation group 2		Rotation group 3		Rotation group 4		Rotation group 1		Total	
	N	%	N	%	N	%	N	%	N	%
Total (RB245=1 to 4)	3576	100,0	3322	100,0	3432	100,0	3212	100,0	13542	100,0
Current household members aged 16 and over (RB245 = 1)	3035	84,9	2780	83,7	2825	82,3	2668	83,1	11308	83,5
Selected respondent (RB245=2)	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
Not eligible person (RB245=4)	541	15,1	542	16,3	607	17,7	544	16,9	2234	16,5

Table 2.7. Distribution of household members by Data Status (RB250) and rotational group, 2009

Data Status	Rotation group 2		Rotation group 3		Rotation group 4		Rotation group 1		Total	
	N	%	N	%	N	%	N	%	N	%
Total (RB250=11 to 33)	3035	100.0	2780	100.0	2825	100.0	2668	100.0	11308	100.0
Information completed only from interview (RB250 = 11)	3016	99.4	2756	99.1	2809	99.4	2639	98.9	11220	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)	19.0	0.6	24.0	0.9	16.0	0.6	29.0	1.1	88.0	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 2.8. Distribution of household members by Type of Interview (RB260) and rotational group, 2009

Responent Status	Rotation group 2		Rotation group 3		Rotation group 4		Rotation group 1		Total	
	N	%	N	%	N	%	N	%	N	%
Total (RB260=1 to 5)	3016	100.0	2756	100.0	2809	100.0	2639	100.0	11220	100.0
Face to face interview - PAPI (RB260 = 1)	27	0.9	44	1.6	48	1.7	52	2.0	171	1.5
Face to face interview - CAPI (RB260 = 2)	2257	74.8	2058	74.7	2132	75.9	2085	79.0	8532	76.0
CATI, telephone interview (RB260=3)	6	0.2	5	0.2	3	0.1	4	0.2	18	0.2
Self-administered by respondent (RB260=4)	0	0	1	0.0	1	0.0	1	0.0	3	0.0
Proxy interview (RB260=5)	726	24.1	648	23.5	625	22.2	497	18.8	2496	22.2

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.

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

Table 2.9. Distribution of item non-response, household-level variables, 2009

Income variable	Hhs having received an amount		Hhs with missing values		Hhs with partial missing		Hhs with gross/net conversion	
	Count	%	Count	%	Count	%	Count	%
Total household gross income (HY010)	4954	99.8	61	1.2	692	14	4120	83.2
Total disposable household income (HY020)	4957	99.8	49	1	2213	44.6	1879	37.9
Total disposable household income before social transfer other than old-age and survivors' benefits (HY022)	4929	99.3	105	2.1	1995	40.5	1191	24.2
Total disposable household income before social transfers including old-age and survivors' benefits (HY023)	4611	92.9	193	4.2	2011	43.6	766	16.6
Net income components on household level								
Imputed rent (HY030N)	4820	97.1	4820	100.0	0	0	0	0
Income from rental of a property or land (HY040N)	136	2.7	7	5.1	0	0	22	16.2
Family/ children related allowances (HY050N)	1932	38.9	0	0	14	0.7	1916	99.2
Social inclusion not elsewhere classified (HY060N)	108	2.2	0	0	0	0	108	100
Housing allowances (HY070N)	88	1.8	0	0	0	0	88	100
Regular inter-household cash transfers received (HY080N)	203	4.1	0	0	0	0	203	100
Alimonies received, compulsory and voluntary (HY081N)	117	2.4	0	0	0	0	117	100
Interest, dividends, profit from	1940	39.1	1	0.1	0	0	1929	99.4

capital investments in incorporated business (HY090N)								
Interest repayments on mortgage (HY100N)	469	9.4	269	57.4	0	0	0	0
Income received by people aged under 16 (HY110N)	80	1.6	12	15	5	6.3	17	21.3
Regular taxes on wealth (HY120N)	3563	71.8	0	0	0	0	3563	100
Regular inter-household cash transfers paid (HY130N)	209	4.2	0	0	0	0	209	100
Alimonies paid, compulsory and voluntary (HY131N)	67	1.3	0	0	0	0	67	100
Repayments/ receipts for tax adjustment (HY145N)	2342	47.2	274	11.7	114	4.9	0	0
Gross income components on household level								
Imputed rent (HY030G)	4827	97.2	4827	100.0	0	0	0	0
Income from rental of a property or land (HY040G)	136	2.7	0	0	0	0	114	83.8
Family/ children related allowances (HY050G)	1932	38.9	1	0.1	14	0.7	311	16.1
Social inclusion not elsewhere classified (HY060G)	108	2.2	25	23.1	1	0.9	0	0
Housing allowances (HY070G)	88	1.8	11	12.5	0	0	0	0
Regular inter-household cash transfers received (HY080G)	203	4.1	14	6.9	0	0	0	0
Alimonies received, compulsory and voluntary (HY081G)	117	2.4	11	9.4	0	0	0	0
Interest, dividends, profit from capital investments in incorporated business (HY090G)	1940	39.1	1776	91.5	22	1.1	81	4.2
Interest repayments on mortgage (HY100G)	469	9.4	0	0	0	0	469	100
Income received by people aged under 16 (HY110G)	80	1.6	0	0	0	0	80	100
Regular taxes on wealth (HY120G)	3563	71.8	125	3.5	0	0	0	0
Regular inter-household cash transfers paid (HY130G)	209	4.2	6	2.9	0	0	0	0
Alimonies paid, compulsory and voluntary (HY131G)	67	1.3	1	1.5	0	0	0	0
Tax on income and social contributions, gross (HY140G)	3857	77.7	3857	100.0	0	0	0	0

Table 2.10. Distribution of item non-response, person-level variables, 2009

	Persons having received an amount		Persons with missing values		Persons with partial missing		Persons with gross/net conversion	
	Count	%	Count	%	Count	%	Count	%
Employee cash or near cash income (PY010N)	6450	57.0	643	10.0	39	0.6	381	5.9
Non-cash employee income (PY020N)	1406	12.4	532	37.8	194	13.8	0	0
Company car (PY021N) ¹	236	2.1	236	100.0	0	0	0	0
Contributions to individual private pension plans (PY035N)	790	7.0	0	0	1	0.1	0	0
Cash benefits or losses from self employment (PY050N)	737	6.5	43	5.8	2	0.3	539	73.1
Value of goods produced by own-consumption (PY070N)	5688	50.3	45	0.8	0	0	5643	99.2
Pension from individual private plans (PY080N)	7	0.1	1	14.3	0	0	0	0
Unemployment benefits (PY090N)	226	2.0	6	2.7	9	4.0	86	38.1
Old-age benefits (PY100N)	2848	25.2	62	2.2	9	0.3	852	29.9
Survivor's benefits (PY110N)	96	0.8	3	3.1	0	0	0	0
Sickness benefits (PY120N)	1397	12.4	524	37.5	0	0	0	0
Disability benefits (PY130N)	883	7.8	4	0.5	0	0	879	99.5
Education-related benefits (PY140N)	355	3.1	3	0.8	0	0	352	99.2
Employee cash or near cash income (PY010G)	6450	57.0	65	1.0	0	0	6072	94.1
Non-cash employee income (PY020G)	1406	12.4	18	1.3	0	0	1388	98.7
Company car (PY021G)	236	2.1	1	0.4	0	0	235	99.6
Employer's social insurance contributions (PY030G)	6189	54.7	0	0	0	0	0	0
Contributions to individual private pension plans (PY035G)	790	7.0	0	0	0	0	0	0
Cash benefits or losses from self employment (PY050G)	812	7.2	37	4.6	0	0	282	34.7
Value of goods produced by own-consumption (PY070G)	5688	50.3	5688	100.0	0	0	0	0
Pension from individual private plans (PY080G)	7	0.1	0	0	0	0	7	100.0
Unemployment benefits (PY090G)	226	2.0	15	6.6	0	0	157	69.5
Old-age benefits (PY100G)	2848	25.2	9	0.3	0	0	2839	99.7
Survivor's benefits (PY110G)	96	0.8	0	0	0	0	96	100.0
Sickness benefits (PY120G)	1397	12.4	14	1.0	0	0	1383	99.0
Disability benefits (PY130G)	883	7.8	20	2.3	0	0	0	0
Education-related benefits (PY140G)	355	3.1	12	3.4	0	0	0	0

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 2000.

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

Number of observations for each subpopulation is indicated in Table 1.1.

Item non-response – see notes to Table 1.1

Non-response on individual level: not present in 2009 since income information in missing questionnaires is imputed

Non-response on household level is 1693 households = interview not completed, DB130=21 to 24 (964) + interview rejected, DB135=2 (7) + address cannot be located, DB120=21 (671) + address unable to access, DB120=22 (51).

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 2.7 and 2.8 in Section 2.3.3.3.

2.5. Interview duration

Mean interview duration per household: 48 minutes and 52 seconds (household and personal interviews together) . 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 (2008).

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 2009 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 2009 data to various external sources, including EU-SILC 2008, 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.

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 2008. 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 2008

Income component	EU-SILC	Other sources*
Net cash or near-cash employee income (PY010N)	75,663,148	68,368,093
Gross old-age benefits (PY100G)	15,296,438	15,591,630
Gross sickness benefits (PY120G) ²	873,759	2,052,940
Gross disability benefits (PY130G)	2,077,324	2,975,720
Gross survivor's benefits (PY110G)	131,471	230,400

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

² Monthly in EU-SILC, per leave in administrative sources.

Turning to the cash employee income first, the figure from wage statistics is about 7 billion kroons lower than its EU-SILC equivalent. The difference was little less than in income year 2007 when the difference was almost 9 billion kroons. When comparing the number of people receiving wages and salaries, it appears that there are almost 156,000 persons more in EU-SILC who report this type of income than in wage statistics. The difference is close to what it was in 2007. 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 two 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.

The numbers of people having reported to receive disability benefits in EU-SILC and number of recipients according to national data sources are very similar. The administrative records number should be somewhat higher though as the number given 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 as best compared at household level, as they are usually paid to a household as a whole. The administrative figure is higher than the EU-SILC figure, indicating underestimation in EU-SILC.

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 2008

Income component	EU-SILC	Other sources*
<i>Person-level components</i>		
Net cash or near-cash employee income (PY010N)	698,938	543,432
Old-age benefits (PY100)	284,371	293,400
Disability benefits (PY130)	73,554	73,110
Survivor's benefits (PY110G)	7,060	12,980

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

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

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

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

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

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

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

	Mean		Number of recipients	
	2008	2009	2008	2009
<i>Individual level components</i>				
PY010N	99056	108254	690573	698938
PY020N	9219	10251	172318	182429
PY035N	5308	6004	92736	95614
PY050N	20331	20493	51762	54448
PY090N	13218	19649	9189	23436
PY100N	43838	52599	289604	284371
PY110N	15964	18601	8846	7060
PY120N	3326	4177	100303	165250
PY130N	23312	28242	66024	73554
PY140N	8921	103351	19675	30703
<i>Household level components</i>				
HY040N	12788	20868	7720	11937
HY050N	15274	20655	182718	180377
HY070N	5859	5332	10000	10422
HY080N	20989	22664	23703	25563
HY090N	2112	3483	289877	255340
HY110N	2991	4686	6763	5251
HY120N	477	598	348412	397103
HY130N	18689	21197	20927	27661
HY145N	-3074	-10915	228958	11700
HY010	183157	208150		
HY020	153603	175302		
HY022	147552	165382		
HY023	152713	172048		

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

For HY040N the increase in the total amount is in line with more people obtaining an income from renting out their property or land. Rent prices were on the rise in 2008 (to which the income belongs) so it is logical that the amount of income earned from renting out property would increase and since the market was growing, more people would rent out their property.

Family allowances have increased, in compliance with increases of national benefit levels, most notably the parental benefit which is tied to incomes. Larger parental benefits and more people becoming eligible for larger sums along with rising incomes amounted for a noticeable increase in the overall amount despite the fall in numbers of recipients (which may well be due to sample fluctuations).

The average amount of housing allowances has decreased which can be explained by an overall increase in wages, leaving fewer households eligible, and for smaller sums. The number of households receiving and paying transfers from other households has increased. This might have something to do with more help from former household members living and working abroad and sending money home or just a more favourable economic setting where people have greater financial possibilities for helping their relatives.

An increase in income collected through HY090 reflects a positive situation on the financial markets, with more people investing and larger returns that have to do with overall economic growth.

More people had to pay taxes on wealth and the amount went up but not changed substantially.

The general economic picture should also account for the increase in income received by people aged under 16, the average salary increased but there was not as much short-term and summer jobs as last year.

Total household income increased by approximately 13% in 2008. The increases stem from the higher wages and other income components, most of which have gone up 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. 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 EU-SILC and the LFS, 2009

ISCED level	EU-SILC	LFS
0 Pre-primary education	.2*	0.5
1 Primary education	2.2	2.1
2 Lower secondary education	16.7	17.4
3 (Upper) secondary education	49.0	45.4
4 Post-secondary non tertiary education	3.0	4.9
5 First stage of tertiary education	28.6	29.6
6 Second stage of tertiary education	0.4	.2*
Total	99.8	99.8

*Unreliable estimate, based on 35-44 sample observations

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

Table 4.5. Distribution of population aged 16-74 by self-defined activity status based on EU-SILC and the LFS, 2009

Activity status	EU-SILC	LFS
Working full-time	53.6	52.1
Working part-time	5.4	5.4
Unemployed	8.3	8.9
Pupil, student	9.4	9.6
In retirement	14.2	13.8
Permanently disabled	4.1	4.6
Fulfilling domestic tasks and care responsibilities	4.9	5.3
Conscript	.2*	.2*
Other inactive	(0.0)*	.1*
Total	99.9	99.6

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