EU-SILC 2006 in Estonia: Intermediate Quality Report

Tallinn 2007

CONTENTS

1. COMMON CROSS-SECTIONAL EUROPEAN UNION INDICATORS

Table 1.1. Common cross-sectional European Union indicators and their standard errors, 2006

Indicator	Value	Standard error
PORTFOLIO OF OVERARCHING INDICATORS		
At-risk-of-poverty threshold – one person household	34153	406
At-risk-of-poverty threshold – household with 2 adults and 2 children	71720	852
At-risk-of-poverty rate after social transfers: Total	18.3	0.6
At-risk-of-poverty rate after social transfers: Male	16.3	0.6
At-risk-of-poverty rate after social transfers: Female	20.0	0.7
At-risk-of-poverty rate after social transfers: 0-17 total	20.1	1.1
At-risk-of-poverty rate after social transfers: 18-64 total	16.0	0.6
At-risk-of-poverty rate after social transfers: 18-64 male	15.2	0.7
At-risk-of-poverty rate after social transfers: 18-64 female	16.7	0.7
At-risk-of-poverty rate after social transfers: 65+ total	25.1	1.7
At-risk-of-poverty rate after social transfers: 65+ male	13.7	1.8
At-risk-of-poverty rate after social transfers: 65+ female	30.8	2.0
Relative median at-risk-of-poverty gap: Total	22.1	1.0
Relative median at-risk-of-poverty gap: Male	26.5	1.5
Relative median at-risk-of-poverty gap: Female	20.1	1.0
Relative median at-risk-of-poverty gap: 0-17	27.8	2.0
Relative median at-risk-of-poverty gap: 18-64 total	27.8	1.4
Relative median at-risk-of-poverty gap: 18-64 male	29.1	1.7
Relative median at-risk-of-poverty gap: 18-64 female	25.6	1.5
Relative median at-risk-of-poverty gap: 65+ total	11.2	0.9
Relative median at-risk-of-poverty gap: 65+ male	10.7	1.7
Relative median at-risk-of-poverty gap: 65+ female	11.5	1.0
In-work at-risk-of-poverty rate: Total	7.5	0.4
In-work at-risk-of-poverty rate: Male	6.3	0.5
In-work at-risk-of-poverty rate: Female	8.8	0.6
S80/S20 income quintile share ratio	5.5	0.1
Relative median income ratio	0.73	-
Aggregate replacement ratio: Total	0.47	-
Aggregate replacement ratio: Male	0.40	-
Aggregate replacement ratio: Female	0.54	-
At-risk-of-poverty rate before social transfers (incl. pensions): Total	38.0	0.6
At-risk-of-poverty rate before social transfers (incl. pensions): Male	35.3	0.6
At-risk-of-poverty rate before social transfers (incl. pensions): Female	40.4	0.6
At-risk-of-poverty rate before social transfers (incl. pensions): 0-17	34.0	1.2
At-risk-of-poverty rate before social transfers (incl. pensions): 18-64 total	27.8	0.6
At-risk-of-poverty rate before social transfers (incl. pensions): 18-64 male	26.6	0.7
At-risk-of-poverty rate before social transfers (incl. pensions): 18-64 female	28.8	0.7
At-risk-of-poverty rate before social transfers (incl. pensions): 65+ total	82.1	1.0
At-risk-of-poverty rate before social transfers (incl. pensions): 65+ male	81.9	1.5
At-risk-of-poverty rate before social transfers (incl. pensions): 65+ female	82.2	1.1

2. ACCURACY

2.1. Sample design

As 2006 operation was the third round of EU-SILC in Estonia, the sample comprised of three parts:

1. The first part consists of households selected for the survey in 2004 and followed up in 2005 and 2006 (in total 3850 households including fresh split-off households). Initially this part consisted of 4 rotational groups, one of which was to be dropped after 2004 operation and another after 2005 operation. However, due to the smaller than expected response rates, it was decided to keep all rotational groups in the sample up to 2006.
2. The second part consists of households selected for the survey in 2005 and followed-up in 2006 (in total 648 household including fresh split-off households).
3. Sample of 2495 households introduced into the survey in 2006.

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

Sub-sections 2.1.1 – 2.1.5 describe the design of new sub-sample. Sampling design of 2004 and 2005 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 01.01.2006, 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 2006

Stratum h	Counties	R_g	n_g	n_g/R_g %
Large	Tallinn, Harju, Ida-Viru, Lääne-Viru, Pärnu, Tartu	859,577	1 498	0.17
Small	Jõgeva, Järva, Lääne, Põlva, Rapla, Saare, Valga, Viljandi, Võru	287,880	914	0.32
Hiiu	Hiiu	9159	83	0.91

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, 2006

Year a rotational group started	Rotational group	Initial sample size in 2006	Nr of split-off households	Final sample size
2004	1	925	42	967
	2	945	36	981
	3	906	35	941
	4	929	32	961
2005	5	626	22	648
2006	6	2495	.	2495
Total		6826	167	6993

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 March to June 2006. 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 June), while the new part was allocated into the first three months (March-May) 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 4% of households was interviewed after the official end of fieldwork period in July 2006.

2.1.7. Renewal of sample: Rotational groups

The sample consists of 6 rotational groups:

1. 4 rotational groups from 2004 (groups 1 - 4);
2. 1 rotational group from 2005 (group 5);
3. new sub-sample (group 6).

2.1.8. Weightings

2.1.8.1 Introduction

Weighting scheme was generally in line with documents V. Verma „EU-SILC weighting procedures: an outline" and J.-M. Museux „Weighting and estimation for the EU-SILC rotational design", with some peculiarities due to modified rotational scheme. This section will describe in detail the actual algorithm used.

The sample of year 2006 consists of three sub-samples to be weighted independently and combined thereafter:

s_3	households started in 2004 and their split-offs, participate for the third time (= 4 rotational groups)
s_2	households started in 2005 and their split-offs, participate for the second time (= one rotational group)
s_1	households started in 2006, participate for the first time (= one rotational group)

First, we need to calculate base weights of year 2006 for each sub-sample. Since weighting procedure was different in 2005, base weights of 2005 also need to be recalculated. In the following years, no recalculation will be needed, besides correction for attrition.

In what follows we describe the procedure of obtaining base weights of 2006 for each sub-sample independently, without making any special reference to the weights of 2005.

2.1.8.2 Calculation of base weights

2.1.8.2.1. Sub-sample s_1

This sub-sample is a usual random sample from population and it does not depend on other sub-samples. The 2006 base weights for this sub-sample are usual first-year cross-sectional weights. With these weights s_1 represents the cross-sectional population of 2006. Calculation of first-year cross-sectional weights is done in three steps:

1. Design weight of household h in stratum g is the inverse of inclusion probability:

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

where p_h is household size, i.e. number of persons aged 14+ in the household,

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.

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).

2. Correction for non-response was done with logistic model with design weights computed at step 1. Auxiliary variables were: county group (4 groups), sex of address-person, age of address-person, 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 of responding households are thus $d_h^* = d_h / r_h$.

3. Non-response corrected weights were further calibrated to reproduce known population totals. 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).

Calibration totals originated from demographic statistics published by Statistics Estonia with institutionalized persons deducted. After calibration we get correction factor g_h for weight d_h^*. Final base weight of sub-sample s_1 is $w_{1h} = d_h^* g_h$. Every person i in household h, receives a weight of his/her household: $w_{1i} = w_{1h}, i \in h$.

2.1.8.2.2.. Sub-sample s_2

This sub-sample was taken in 2005 and just as s_1 it is a random sample independent from others. First, we go back to year 2005 and calculate base weights for year 2005. As this is the first year s_2 was in survey, base weights are usual first-year cross-sectional weights. With these weights s_2 represents target population of 2005. Calculation of 2005 cross-sectional weights was done similarly to sub-sample s_1 (with simpler model for non-response correction, county group only). Household cross-sectional weight is assigned to each of its members. Let us denote these weights by w_i (with i meaning a person).

To get base weights for 2006 we now need to correct for attrition that has happened in s_2 in year 2006. Correction for attrition is done on person-level, i.e. corrected weights of persons within one household no longer need to be constant. Prior to any corrections we need to exclude from consideration persons that became out-of-scope in 2006 as they are not considered as non-response. Out-of-scope are persons that were dead by 2006, became institutionalized or had left the country for longer period. After excluding out-of-scope persons, attrition was modelled using a logistic model. Auxiliary variables included household tenure status, household equivalised income, urbanization status of place of residence, county group, person's social status, age, ethnic nationality. With logistic

model we get an estimate of response probability r_i of person i for year 2006, given he/she had responded in 2005 (i.e. his/her household had responded in 2005). Base weights for year 2006 for a person i in sub-sample s_2 are thus $w_{2i} = w_i / r_i$.

Note that base weights defined above do not cover all individuals:

- 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.

2.1.8.2.3. Sub-sample s_3

Sub-sample s_3 was selected in 2004 and to get base weights for the year 2006 we need to make two transitions: from 2004 to 2005 and from 2005 to 2006. In year 2004 this was the only sample in the survey and thus cross-sectional weights of 2004 (at the personal level, RB050) are 2004 base weights for this sub-sample, then ones we start with.

Sub-sample s_3 consists of 4 rotational groups and base weights need to be calculated for these groups independently, as they will not be treated uniformly in the future. Therefore, 2004 base weight for person i in sub-sample s_3 is $w_i = 4 \cdot RB050_i$.

Next step is to correct for attrition that happened in 2005. Prior to that we exclude from consideration persons no longer in scope in 2005 (see 2.1.8.2.2). Correction for attrition is again done with the help of logistic model with tenure status, household equivalised income, urbanization status of place of residence, social status, age, sex, county, ethnic nationality and household's assessment to its ability to make ends meet as auxiliary variables. In this way we get an estimate for response probability r_i of person i in 2005 given a person responded in 2004 (note that henceforth we use same notation as in 2.1.8.2.2, but meaning should be clear from the context). The 2005 base weight of person i in sub-sample s_3 is thus $w_i^* = w_i / r_i$. After extending these base weights to other individuals (see 2.1.8.2.2) we can proceed to the next correction. Now we need to correct for attrition that happened in 2006. Here, similarly to the previous transition, persons no longer in scope in 2006 are excluded first and then a logistic model is used to estimate 2006 response probability r_i^* given a person responded in 2005. Preliminary base weights for 2006 are thus $w_{3i} = w_i^* / r_i^*$. After extension to other individuals, one more correction needs to be made.

In particular, there is a group of households, which were interviewed in 2006, but skipped the round of 2005, so called returnees. Since in 2005 the base weights of other households were adjusted to take their absence into account, returnees cannot be re-assigned a positive weight without adjusting the weights of other individuals. To calculate base weights for returnees, we first estimate the response probability r_i^{**} of a person in 2006 given he/she had responded in 2004 (note the difference from response probability r_i^*). Base weight 2006 for returnees is thus $w_{3i} = w_i / r_i^{**}$. As the final step we need to reduce the base weights of non-returnees. Let t_3 be the sum of base weights w_{3i} of non-returnees, and t_3^* the sum of weights w_{3i} of returnees. The final adjusted base weight of year 2006 for sub-sample s_3 is then

$$w_{3i}^* = \begin{cases} w_{3i}, & \text{if } i \text{ is returnee} \\ \dfrac{t_3 - t_3^*}{t_3} \cdot w_{3i}, & \text{if } i \text{ is not returnee} \end{cases}$$

With such correction the sum of w_{3i}^* is the same as for non-returnees, t_3.

After steps 1, 2 and 3 we have a base weight defined for every person in 2006 sample. This weight is specific to rotational group, i.e. every rotational group with its base weight is representative of the whole population (with exception of immigrant households for older rotational groups).

2.1.8.3. Calculation of cross-sectional weights

To calculate cross-sectional weights of year 2006, we need to combine together different sub-samples.

As households consisting of fresh immigrants are not present in older sub-samples, there is a potential danger to under-estimate this group of households and individuals and a special attention should be paid to them. But no immigrant household we present in either of 2005 and 2006 new samples, so this step is skipped.

The next step is to average individual base weights within a household. As a result we get a base weight for each household in the sample (denoted by w_{1h}, w_{2h} and w_{3h}^{*} for the first, second and third sub-sample respectively). To get preliminary household cross-sectional weight we combine household base weight according to the following scheme:

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

where n_i is number of responding households in sub-sample s_i, i=1,2,3 (in year 2006).

Thus, base weight of each sub-sample is reduced according to the share of corresponding sub-sample in the overall sample size.

Final stage is calibration of preliminary household cross-sectional weights $w_h^{together}$ to external totals. Calibration is 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).

Calibration totals originated from demographic statistics published by Statistics Estonia with institutionalized persons deducted. After calibration we get final cross-sectional household weight DB090.

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) are computed in two steps: first, correction is made for within-household non-response (weights RB050 are divided by the within-household empirical response rate of household members aged 16 and over), next, resulting weights are calibrated to the same totals as DB090.

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).

Distributional characteristics of final household cross-sectional weights (DB090) by stratum and household size are presented in Table 2.3.

Table 2.3. Distributional characteristics of final household cross-sectional weights by stratum and household size, 2006

Stratum	Household size	Mean	Std	CV
1	**1**	245.1	115.3	47.1
	2	140.2	71.4	50.9
	3	125.0	68.3	54.6
	4	104.4	51.2	49.0
	5	84.8	45.5	53.6
	6	66.4	35.1	52.8
	7	54.0	22.2	41.2
	8	51.3	16.8	32.6

	9	41.5	19.6	47.3
	10	25.1	.	.
	12	40.9	.	.
2	**1**	108.6	49.9	46.0
	2	56.3	24.7	43.9
	3	51.3	26.4	51.5
	4	45.1	21.6	48.0
	5	33.5	16.2	48.3
	6	28.7	12.6	43.9
	7	21.0	9.7	46.4
	8	22.1	10.0	45.2
	9	24.9	13.2	53.0
	10	20.0	3.6	17.8
	11	18.7	.	.
	13	8.0	.	.
3	**1**	36.8	17.0	46.1
	2	19.8	10.2	51.7
	3	14.4	6.0	41.6
	4	12.8	6.0	46.3
	5	11.4	5.9	51.4
	6	10.1	3.4	33.8
	7	7.1	2.1	29.7
	9	2.8	1.0	35.7
All		99.3	83.0	83.6

2.1.8.2. 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.

Recent simulation study showed that design effect of at-risk-of-poverty rate was 1.2 and for mean equivalized income 1.25. In this simulation study a self-weighting sub-sample was sub-selected from the new part of the sample of 2006. To compare variances, a simple random sample of the same size was also selected from the new part (thus reproducing the sampling design used in the survey). Variances of at-risk-of-poverty rate and mean equavalised income were calculated from both of the samples. Design effect was calculated as the ratio of these two variances.

Of total 6993 households in the sample, data of 5631 household were accepted for the final database. In these households 13,007 persons aged 16+ were interviewed. Effective sample size is thus 4693 households and 10,839 persons (according to Commission regulation we use design effect of at-risk-of-poverty rate). 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 2006 was 82 households, which makes the proportion of the over-coverage in the new part of the sample 3.3% and of the whole sample 1.2% (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 2006 2.9% of all records referring to persons aged 14+ were dropped before selecting the sample).

In the new part of the sample of 2006 there were 144 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 2.4. Reasons for over-coverage in the new part of the sample, 2006

Frame error	Number of households	Proportion in the frame error (%)
Total, of which	82	100.0
Address person was dead	6	7.3
Address person has left Estonia	56	68.3
Address person was staying in an institution	19	23.2
Address person was surveyed through one of his/her household members	1	1.2

2.3.2. Measurement and processing errors

2.3.2.1. Measurement errors

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

The questionnaires were drawn up following the international practises in collecting income data. Also, where possible questions from the existing surveys carried out by the Statistics Estonia and known to be valid and reliable, were used. Pilot surveys were carried out in 2002 and 2003 with the main aim of testing the questionnaires. The results were thoroughly analysed and feedback sessions with interviewers were carried out. The questionnaires were modified accordingly for the use in the main operation in 2004. The experience from the first two waves of the survey was further used to improve the questionnaire for the 2006 operation. The main modifications in 2006 concerned employee income and self-employment income where income brackets were added to those unable or unwilling to

provide a precise answer, question on income from bank accounts was also more fleshed out and income brackets were added. Also improved were questions on child-care, family benefits and unemployment benefits.

All returning interviewers attended a day long training session in small groups. During the training, the EU-SILC team briefed the interviewers on all aspects relating to the fieldwork organisation, the questionnaire and general interviewing techniques. Special emphasis was placed on survey questions about income – types of income, their more common amounts and recipients. A separate session was held on tracing and specifics of assigning household and person numbers in the longitudinal survey. Interviewers new to EU-SILC attended a 2 day training session, which included a thorough overview of questionnaires and practical exercises.

Overall, 58 interviewers were responsible for conducting the interviews. The household (gross sample) – interviewer ratio was 120 households per interviewer.

2.3.2.2. Processing errors

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

The data for 2006 operation was collected using CAPI. The data-entry program was written in Blaise and contained most of the checks. This way, most of the errors could already be corrected during an interview. The data-entry controls were of 4 major types:

1) Checks of consistency between different answers. These included, but were not limited to following instances:
    a. whether a household or a person who according to other data should have received a certain type of income reported it or not (e.g. whether households with children received family benefits, employed persons received wages and so on);
    b. whether answers provided to different non-monetary deprivation items agreed with each other;
    c. whether the relationships in the household matrix were consistent with each other as well as with the age and sex of the household members;
    d. whether the difference between the starting and finishing time of the interview was too short or too long and so on.
2) Lower and upper bounds of income variables. These checks were developed based on data collected in the previous wave as well as administrative information.
3) Tracing checks. These controls were implemented to ensure that all split-off households and new household members were assigned correct split numbers and person numbers respectively.
4) Checks with information from the previous year. These controls concerned demographic data, information on educational level and labour status as well as the calendar of activities.

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

1) All the errors suppressed by interviewers were activated and checked.
2) All remarks made by interviewers in the data entry-program were read through and where necessary relevant corrections were made.
3) All split-off households as well as all households from which at least one member had left were scrutinised one by one.
4) All answers where other-option was used, were read through and assigned to a correct category when necessary.

Table 2.5 presents the number of different types of errors that were detected during fieldwork, including concurrent in-office checks. 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 situation is unusual but correct to add a remark to the data entry-program explaining this error. In assessing the quality of interviewers work, not adding a remark to otherwise correct situation is also counted as an error. Altogether 5654 errors in 20% questionnaires were detected, one third of which required a call-back. Most common types of errors were those discovered during concurrent in-office checks. Also frequent were missing remarks to errors in correct situations. All other types errors were not as wide-spread, although not correcting an error and not understanding a question also

occurred in over 500 instances. More than half of the errors required a call-back in the case of incorrectly recorded times and insufficient information for coding. Data entry mistakes, mistakes uncovered during in-office checking and incorrect use of category Other were the types of mistake where call-backs were least likely.

Table 2.5. Interviewer errors and their processing, 2006

Type of error	Number of errors detected	Share of errors requiring a call-back
No remark explaining unusual situation	1178	28
Interviewer made an error, but did not correct it	556	46
Interviewer's remark does not explain unusual situation	451	26
Data not sufficient for coding	167	63
Starting and finishing times recorded incorrectly	38	74
Use of category Other, while a suitable category exists	250	12
In-office checks	1408	15
Interviewer has misunderstood a question	520	30
Data entry mistake	442	17
Not interviewers error	644	97
Total	5654	34

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

2.3.3. Non-response errors

2.3.3.1. Achieved sample size

Data for 5631 households were accepted for database and used in analysis. This makes the overall share of complete household interviews accepted for the database 80.5%. The contact rate was 90.5% and the household response rate (given contact is achieved) was 89.0%. On personal level, the share of complete personal interviews within the households accepted for the database was 98.6% – 13,007 interviews of possible 13,187.

For rotational group breakdown see 2.3.3.3.

2.3.3.2. Unit non-response

The final response rates for the total sample were as follows:

Household non-response rate *NRh* = 17.9

Individual non-response rate *NRp* = 1.4

Overall non-response rate **NRp* = 18.2

Response rates for the new part of the sample were:

Household non-response rate *NRh* = 31.8

Individual non-response rate *NRp* = 2.0

Overall non-response rate **NRp* = 33.1

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	▪ 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

▪ Given address does not exist ▪ Address can be located, but no contact can be made since nobody is at home	

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, 2006

Record of contact at address	Rotational group 1		Rotational group 2		Rotational group 3		Rotational group 4		Rotational group 5		Rotational group 6		Total	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Total (DB120=11 to 23)	967	100	981	100	941	100	961	100	648	100	2495	100	6993	100
Address contacted (DB120=11)	919	95	941	96	899	96	916	95	609	94	2044	82	6328	90
Address non-contacted (DB120=21 to 23)	48	5	40	4	42	4	45	5	39	6	451	18	665	10
Total address non-contacted (DB120=21 to 23)	48	100	40	100	42	100	45	100	39	100	451	100	665	100
Address cannot be located (DB120=21)	38	79	27	68	32	76	33	73	30	77	359	80	519	78
Address unable to access (DB120=22)	0	0	1	3	1	2	1	2	0	0	10	2	13	2
Address does not exist or is non-residential address or is unoccupied or not principal residence (DB120=23)	10	21	12	30	9	21	11	24	9	23	82	18	133	20

Table 2.7. Distribution of addresses contacted by 'household questionnaire result' and by household interview acceptance, 2006

Household questionnaire result	Rotational group 1		Rotational group 2		Rotational group 3		Rotational group 4		Rotational group 5		Rotational group 6		Total	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Total (DB130=11 to 24)	919	100	941	100	899	100	916	100	609	100	2044	100	6328	100
Household questionnaire completed (DB130=11)	872	95	882	94	847	94	864	94	564	93	1651	81	5680	90
Interview not completed (DB130= 21 to 24)	47	5	59	6	52	6	52	6	45	7	393	19	648	10
Total interview not completed (DB130=21 to 24)	47	100	59	100	52	100	52	100	45	100	393	100	648	100
Refusal to co-operate (DB130=21)	31	66	38	64	39	75	32	62	38	84	291	74	469	72
Entire household temporarily away for duration of fieldwork (DB130=22)	8	17	4	7	4	8	6	12	4	9	42	11	68	10
Household unable to respond (illness, incapacity, etc) (DB130=23)	6	13	13	22	7	13	10	19	2	4	41	10	79	12
Other (DB130=24)	2	4	4	7	2	4	4	8	1	2	19	5	32	5
Household questionnaire completed (DB135=1 to 2)	872	100	882	100	847	100	864	100	564	100	1651	100	5680	100
Interview accepted to database (DB135=1)	861	99	871	99	838	99	854	99	561	99	1646	100	5631	99
Interview rejected (DB135=2)	11	1	11	1	9	1	10	1	3	1	5	0	49	1

Table 2.8. Distribution of household members by Respondent Status and rotational group (RB245), 2006

Responent Status	Rotational group 1		Rotational group 2		Rotational group 3		Rotational group 4		Rotational group 5		Rotational group 6		Total	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Total (RB245=1 to 4)	2351	100	2507	100	2418	100	2400	100	1519	100	4645	100	15840	100
Current household members aged 16 and over (RB245 = 1)	1962	83	2111	84	2037	84	2018	84	1270	84	3789	82	13187	83
Selected respondent (RB245=2)	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
Not eligible person (RB245=4)	389	17	396	16	381	16	382	16	249	16	856	18	2653	17

Table 2.9. Distribution of household members by Data Status and rotational group (RB250), 2006

Data Status	Rotational group 1		Rotational group 2		Rotational group 3		Rotational group 4		Rotational group 5		Rotational group 6		Total	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Total (RB250=11 to 33)	1962	100.0	2111	100.0	2037	100.0	2018	100.0	1270	100.0	3789	100.0	13187	100.0
Information completed only from interview (RB250 = 11)	1941	98.9	2081	98.6	2011	98.7	2002	99.2	1257	99.0	3715	98.1	13007	98.6
Information completed only from registers (RB250 = 12)	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Information completed only from both (RB250 = 13)	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Individual unable to respond (RB250=21)	0	0.0	3	0.1	1	0.1	2	0.1	1	0.1	2	0.1	9	0.1
Individual failed to return self-completed questionnaire (RB250=22)	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Refusal to cooperate (RB250=23)	13	0.7	18	0.9	20	1.0	8	0.4	10	0.8	60	1.6	129	1.0
Person temporarily away and no proxy available (RB250=31)	5	0.3	5	0.2	5	0.3	5	0.3	1	0.1	4	0.1	25	0.2
No contact for other reasons (RB250=32)	3	0.2	4	0.2	0	0.0	1	0.1	1	0.1	8	0.2	17	0.1
Information not completed, reason unknown (RB250=33)	0	0.0	0	0.0	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 and rotational group (RB260), 2006

Responent Status	Rotational group 1		Rotational group 2		Rotational group 3		Rotational group 4		Rotational group 5		Rotational group 6		Total	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Total (RB245=1 to 5)	1941	100.0	2078	100.0	2006	100.0	1996	100.0	1256	100.0	3710	100.0	12987	100.0
Face to face interview - PAPI (RB260 = 1)	30	1.6	30	1.4	37	1.8	27	1.4	25	2.0	60	1.6	209	1.6
Face to face interview - CAPI (RB260 = 2)	1783	91.9	1922	92.5	1854	92.4	1826	91.5	1152	91.7	3465	93.4	12002	92.4
CATI, telephone interview (RB260=3)	4	0.2	6	0.3	4	0.2	3	0.2	1	0.1	3	0.1	21	0.2
Self-administered by respondent (RB260=4)	1	0.1	0	0.0	3	0.2	1	0.1	0	0.0	0	0.0	5	0.0
Proxy interview (RB260=5)	123	6.3	120	5.8	108	5.4	139	7.0	78	6.2	182	4.9	750	5.8

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.

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

Table 2.11. Distribution of item non-response, household-level variables, 2006

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)	5620	99.8	44	0.8	1936	34.4
Total disposable household income (HY020)	5624	99.9	16	0.3	2053	36.5
Total disposable household income before social transfer other than old-age and survivors' benefits (HY022)	5569	98.9	40	0.7	841	15.1
Total disposable household income before social transfers including old-age and survivors' benefits (HY023)	5197	92.3	145	2.8	730	14.0
Net income components at household level						
Income from rental of a property or land (HY040N)	92	1.6	2	2.2	0	0.0
Family/ children related allowances (HY050N)	2358	41.9	0	0.0	0	0.0
Social inclusion not elsewhere classified (HY060N)	41	0.7	0	0.0	0	0.0
Housing allowances (HY070N)	129	2.3	9	7.0	0	0.0
Regular inter-household cash transfers received (HY080N)	195	3.5	5	2.6	0	0.0
Interest, dividends, profit from capital investments in incorporated business (HY090N)	1602	28.4	1333[1]	83.2	100	6.2
Income received by people aged under 16 (HY110N)	104	1.8	13	12.5	6	5.8
Regular taxes on wealth (HY120N)	3602	64.0	87	2.4	0	0.0
Regular inter-household cash transfers paid (HY130N)	274	4.9	3	1.1	0	0.0
Tax on income and social	0	0.0	0[2]	0.0	0	0.0

[1] Of which 1321 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.

contributions, net (HY140N)						
Repayments/ receipts for tax adjustment (HY145N)	2027	36.0	124	6.1	30	1.5
Gross income components at household level						
Income from rental of a property or land (HY040G)	92	1.6	2	2.2	0	0.0
Family/ children related allowances (HY050G)	2358	41.9	0	0.0	0	0.0
Social inclusion not elsewhere classified (HY060G)	41	0.7	0	0.0	0	0.0
Housing allowances (HY070G)	129	2.3	9	7.0	0	0.0
Regular inter-household cash transfers received (HY080G)	195	3.5	5	2.6	0	0.0
Interest, dividends, profit from capital investments in incorporated business (HY090G)	1602	28.4	1333[1)]	83.2	100	0.0
Income received by people aged under 16 (HY110G)	104	1.8	13	12.5	6	5.8
Regular taxes on wealth (HY120G)	3602	64.0	87	2.4	0	0.0
Regular inter-household cash transfers paid (HY130G)	274	2.9	3	1.1	0	0.0
Tax on income and social contributions, gross (HY140G)	4179	74.2	4179[2)]	100.0	0	0.0

Table 2.12. Distribution of item non-response, person-level variables, 2006

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)	7099	54.6	439	6.2	39	0.5
Non-cash employee income (PY020N)	246	1.9	246[3]	100.0	0	0.0
Contributions to individual private pension plans (PY035N)	690	5.3	84	12.2	10	1.4
Cash benefits or losses from self employment (PY050N)	1007	7.7	156	15.5	15	1.5
Pension from individual private plans (PY080N)	6	0.0	1	16.7	1	16.7
Unemployment benefits (PY090N)	187	1.4	2	1.1	7	3.7
Old-age benefits (PY100N)	3220	24.8	10	0.3	1	0.0
Survivor's benefits (PY110N)	119	0.9	2	1.7	0	0.0
Sickness benefits (PY120N)	851	6.5	167[4)]	19.6	0	0.0

[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.

[3] Company car only. PY020N was calculated as the number of months a company car was used multiplied by 2000.

Disability benefits (PY130N)	839	6.5	7	0.8	0	0.0
Education-related benefits (PY140N)	284	2.2	7	2.5	1	0.4
Gross income components at personal level						
Employee cash or near cash income (PY010G)	7099	54.6	439	6.2	39	0.5
Non-cash employee income (PY020G)	246	1.9	246[3]	100.0	0	0.0
Employer's social insurance contributions (PY030G)	6845	52.6	6845[5]	100.0	0	0.0
Contributions to individual private pension plans (PY035G)	690	5.3	84	12.2	10	1.4
Cash benefits or losses from self employment (PY050G)	1012	7.8	162	16.0	15	1.5
Pension from individual private plans (PY080G)	6	0.0	1	16.7	1	16.7
Unemployment benefits (PY090G)	187	1.4	2	1.1	7	3.7
Old-age benefits (PY100G)	3220	24.8	10	0.3	1	0.0
Survivor's benefits (PY110G)	119	0.9	2	1.7	0	0.0
Sickness benefits (PY120G)	851	6.5	167[4]	19.6	0	0.0
Disability benefits (PY130G)	839	6.5	7	0.8	0	0.0
Education-related benefits (PY140G)	284	2.2	7	2.5	1	0.4

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, 2006

	Number of sample observations (achieved sample size)	**Number of sample observations not taken into account due to item non-response**	**Non-response at individual level (if applicable)**	**Non-response at household level (number of households)**
At-risk-of-poverty rate after social transfers				
Total[6]	15,755	29	NA	1229
By age and gender				
men total[6]	7358	29	NA	1229
women total[6]	8397	29	NA	1229
0-15 years[6]	2595	29	NA	1229
16-24 years	2904	29	NA	1229
25-49 years	4972	29	NA	1229
50-64 years	2835	29	NA	1229
65+ years	2449	29	NA	1229
16+ years	13,160	29	NA	1229
16-64 years	10,711	29	NA	1229
0-64 years[6]	13,306	29	NA	1229

[4] Of which for 156 persons the number of days he/she received sickness benefits is known. In this case average wage per day is calculated, corrected according to sickness benefits paying order and multiplied by the number of days.
[5] Not collected. Calculated on the basis of collected or imputed wages.
[6] Children born in 2006 are excluded (56 persons in total).

men 16-24 years	1458	29	NA	1229
men 25-49 years	2391	29	NA	1229
men 50-64 years	899	29	NA	1229
men 65+ years	899	29	NA	1229
men 16+ years	6017	29	NA	1229
men 16-64 years	5118	29	NA	1229
men 0-64 years[6]	6459	29	NA	1229
women 16-24 years	1446	29	NA	1229
women 25-49 years	2581	29	NA	1229
women 50-64 years	1566	29	NA	1229
women 65+ years	1550	29	NA	1229
women 16+ years	7143	29	NA	1229
women 16-64 years	5593	29	NA	1229
women 0-64 years[6]	6847	29	NA	1229
By most frequent activity status and gender				
employed	6674	240	164	1229
unemployed	478	240	164	1229
retired	2672	240	164	1229
other inactive	2961	240	164	1229
men, employed	3395	240	164	1229
men, unemployed	274	240	164	1229
men, retired	902	240	164	1229
men, other inactive	1250	240	164	1229
women, employed	3279	240	164	1229
women, unemployed	204	240	164	1229
women, retired	1770	240	164	1229
women, other inactive	1711	240	164	1229
By household type[7]				
single, < 65 years	564	29	NA	1229
single, 65+ years	575	29	NA	1229
single, male	342	29	NA	1229
single, female	797	29	NA	1229
single, total	1139	29	NA	1229
2 adults, no children, both < 65	1402	29	NA	1229
2 adults, no children, at least one 65+	1460	29	NA	1229
other households without children	1840	29	NA	1229
single parent, at least one child	781	29	NA	1229
2 adults, 1 child	2016	29	NA	1229
2 adults, 2 children	2256	29	NA	1229
2 adults, 3+ children	1447	29	NA	1229
other households with children	3388	29	NA	1229
households without children	5841	29	NA	1229
households with children	9888	29	NA	1229

[7] Persons in households where it was impossible to determine household type are excluded (82 persons).

By accommodation tenure status				
owner or rent-free	15,171	29	NA	1229
tenant	640	29	NA	1229
At-risk-of-poverty threshold				
Median of the equivalised disposable household income	15,811	29	NA	1229
At-risk-of-poverty threshold - total	15,811	29	NA	1229
Inequality of income distribution S80/S20 income quintile share ratio	15,811	29	NA	1229
Relative median at-risk-of-poverty gap				
Total	3048	29	NA	1229
By age and gender				
men total	1326	29	NA	1229
women total	1722	29	NA	1229
0-15 years	633	29	NA	1229
16-64 years	1925	29	NA	1229
65+ years	490	29	NA	1229
16+ years	2415	29	NA	1229
men, 16-64 years	878	29	NA	1229
men, 65+ years	105	29	NA	1229
men, 16+ years	983	29	NA	1229
women, 16-64 years	1047	29	NA	1229
women, 65+ years	385	29	NA	1229
women, 16+ years	1432	29	NA	1229
Dispersion around the risk-of-poverty threshold				
40%	15,811	29	NA	1229
50%	15,811	29	NA	1229
70%	15,811	29	NA	1229
At-risk-of-poverty rate before social transfers except old-age and survivors' benefits –total[6]	15,755	29	NA	1229
At-risk-of-poverty rate before social transfers including old-age and survivors' benefits – total[6]	15,755	29	NA	1229
Gini coefficient	15,811	29	NA	1229
Mean equivalised disposable income	15,811	29	NA	1229

Notes:

Item non-response: number of eligible persons in households with missing HY025 (29) + number of questionnaires with no information on most frequent activity status, when applicable to indicator (211).

Non-response on individual level: individual questionnaire missing (in households with non-missing HY025), when applicable to indicator (164).

Non-response on household level: interview not competed, DB130=21 to 24 (648) + interview rejected, DB135=2 (49) + address cannot be located, DB120=21 (519) + address unable to access, DB120=22 (13).

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: 51 minutes and 12 seconds (interview duration for 5 households and 13 persons was not known). 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 last calendar year (2005).

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 (last 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 (last 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 March till June, but had to be extended by a further one month due shortage of interviewers. Thus, data was collected during a 5 month period, although only 3.7% of households were interviewed during in the final fifth month.

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
Income received by people under 16 other than wages and salaries is not recorded. Nevertheless, it is extremely unlikely that this omission affects the estimation of total household income in any meaningful way (See 3.2.1.8.).

3.2.1.2. Total disposable household income
See 3.2.1.1.

3.2.1.3. Total disposable household income, before social transfers other than old-age and survivors' benefits
See 3.2.1.1.

3.2.1.4. Total disposable household income, before social transfers including old-age and survivors' benefits
See 3.2.1.1.

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. Also, cost of minor repairs could not be subtracted and only gross value is therefore available.

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
Only wages and salaries received by people under 16 were recorded, other types of income received by these persons were not collected. Yet as the vast majority of income received by people under 16 is likely to be in the form of wages and salaries, the effect of this omission on the total income is likely to be small. Also, 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
Variable was not recorded.

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 calculated gender pay gap.

3.1.1.15. All other variables not listed above
There were no divergences from common definitions.

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 summarises 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, HY090, HY120, HY130, PY035, PY130, PY140), i.e. where gross equals net. The only exception is interest, dividends and profit from capital investments in unincorporated businesses, which were collected in gross. 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 2006 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			
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 2006 data to various external sources, including EU-SILC 2005, the National Accounts (NA), the Household Budget Survey (HBS), the Labour Force Survey (LFS), wage statistics and social protection statistics.

HBS is a continuous survey of households, which has been carried since 1996. Annual sample size is approximately 4500 households. HBS is designed to collect information on income and expenditure of households. Data on income is gathered using a diary, where households record all income received during one month. Questionnaires are administered using CAPI. HBS was the source of Laeken indicators up until EU-SILC.

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. Up until 2005, LFS is carried out using CAPI. LFS is the main source for labour market information.

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

Table 4.1 compares the aggregate amounts of income components in EU-SILC and National Accounts. The total wages and salaries are by 5% lower in EU-SILC compared to NA. That is to be expected, given that NA figure also includes non-cash employee income. Income from rental of property or land is substantially higher in EU-SILC, whereas the property income is considerably underestimated. The total transfers are 16% higher in EU-SILC.

Table 4.1. Total income in EU-SILC and NA in millions of kroons, income year 2005

Type of income	Income component in EU-SILC	Total in EU-SILC	Total in NA	Distributive transaction in NA
Wages and salaries	PY010G	57,624.6	60,762.5	D11
Income from rental	HY040G	100.4	30.4	D45
Property income	HY090G	328.8	4,192.2	D41+D42
Transfers	PY090G + PY100G + PY110G + PY120G + PY130G + PY140G + HY050G	13,603.1	11,772.5	D621

Next, EU-SILC income data is compared component by component to income data from HBS and administrative sources for income year 2005. Table 4.2 presents the comparisons by average amounts and Table 4.3 by number of recipients. Only the income components where definitions are similar enough to warrant comparisons are presented here.

Table 4.2. Average amounts of income components by source of information, income year 2005

Income component	EU-SILC	HBS	Other sources*
Person-level components			
Net cash or near-cash employee income (PY010N)	70,768	66,065	77,160
Net non-cash employee income (PY020N)	16,843	17,492	
Net old-age benefits (PY100N)	32,398	30,551	28,968
Net sickness benefits (PY120N)[8]	4,207	2,036	1,924
Net disability benefits (PY130N)	17,362	13,806	
Household-level components			
Contributions to individual private pension plans (PY035N)	6,122	5,339	
Net cash profits or losses from self-employment (PY050N)	13,924	45,894	
Net survivor's benefits (PY110N)	13,240	13,975	15,846
Net income from rental of property or land (HY040N)	12,067	44,453	
Family/children related allowances (HY050N)	10,724	11,641	
Housing allowances (HY070N)	6,066	13,112	
Regular inter-household cash transfer received (HY080N)	14,156	18,894	
Net property income (HY090N)	1,420	3,096	
Regular inter-household transfers paid (HY130N)	12,390	15,162	
Total disposable household income (HY020)	107,329	94,448	
Total disposable household income, before social transfers other than old-age and survivor's benefits (HY022)	102,080	91,152	
Total disposable household income, before social transfers including old-age and survivor's benefits (HY023)	92,554	93,426	
Total disposable per capita income	48,629	41,739	

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

Turning to the cash employee income first, the average amount is by 7% higher in EU-SILC compared to HBS. The corresponding figure from wage statistics is considerably higher, however. When comparing the number of people receiving wages and salaries, it appears that there are some 100,000

[8] Monthly in EU-SILC, per leave in HBS and administrative sources.

persons more in EU-SILC who report this type of income than in HBS. The corresponding figure in wage statistics is lower still. The difference with wage statistics is to be expected, given that the latter refer to the full-time equivalent and the unofficial work relationships are not included. The EU-SILC – HBS difference in the number of recipients can probably be traced to the survey design. As HBS yearly figures are derived from monthly data, shorter employment spells that are concentrated to a few months in a year (mainly summer) are under reported in HBS, while recall errors may be causing over-estimation of amounts received in EU-SILC.

Table 4.3. Number of recipients of income components by source of information, income year 2005

Income component	**EU-SILC**	**HBS**	**Other sources***
Person-level components			
Net cash or near-cash employee income (PY010N)	653,971	554,778	496,277
Net non-cash employee income (PY020N)	26,099	11,797	
Net unemployment benefits (PY090N)	16,004	30,556	
Net old-age benefits (PY100N)[9]	286,167	286,996	296,082
Net sickness benefits (PY120N)	80,244	13,292	
Net disability benefits (PY130N)	65,049	99,362	
Household-level components			
Contributions to individual private pension plans (PY035N)	65,717	31,497	
Net cash profits or losses from self-employment (PY050N)	55,585	46,118	
Net survivor's benefits (PY110N)	7,020	11,509	9,312
Net income from rental of property or land (HY040N)	6,321	5,435	
Family/children related allowances (HY050N)	190,720	173,031	187,397
Housing allowances (HY070N)	12,604	10,304	
Regular inter-household cash transfer received (HY080N)	21,187	12,572	
Net property income (HY090N)	183,962	15,739	
Regular inter-household transfers paid (HY130N)	28,999	6,282	

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

The differences in non-cash employee income (i.e. company cars), however, are much more substantial. The number of recipients is considerably higher in EU-SILC as opposed to HBS, while average amount is about the same. This is again related to survey design. While in EU-SILC the question about the use made of company cars is posed directly to interviewees, in the case of HBS the respective question is in the diary and may go unnoticed by the household. Also, the methods used to estimate the benefit from the car are different (taxation approach in EU-SILC vs. self-estimation in HBS).

In the case of the unemployment benefits, the definitions that can be used differ between the sources and thus only the number of recipients can be only tentatively compared. The number of receivers is substantially lower in EU-SILC. Comparison on subcomponent level to administrative sources reveals that both surveys underestimate the number of recipients.

The average amounts of old-age benefits received are somewhat higher in EU-SILC as opposed to HBS, which in turn provides higher estimates than administrative statistics. The number of receivers is almost identical in two surveys, which is about 10,000 less than could be expected from administrative statistics. It must be taken into consideration, however, that the average amount in EU-SILC also includes other old-age benefits and disability benefits paid to retired persons that are not taken into account in the other sources. Underestimation of number of recipients is probably related to some below retirement age persons failing to report superannuated pensions in the case of EU-SILC and misclassification of old-age benefits as disability benefits in HBS.

The number of people receiving sickness benefits is, given the seasonal nature of the component, underestimated in HBS. The amounts are also higher in EU-SILC, despite the HBS figure including some lump-sum family benefits. Neither the number of recipients nor the average amounts paid is available from the 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 under reported in EU-SILC. It is likely that respondents do not separate smaller amounts from wages and salaries.

[9] Benefits received from abroad are excluded from survey estimates.

The average disability benefits received by people are higher in EU-SILC, while the number of recipients is lower. This is related to the fact that disability benefits paid to people in retirement age have been added to the old-age benefits.

Although contributions to individual private pension plans as well as self-employment income are person-level variables in EU-SILC, the comparison with HBS is only possible on household level. Tables 4.2 and 4.3 show that while the average amounts of payments to private pension plans are similar in both surveys, the number households making payments is twice as high in EU-SILC. In the case of self-employment income the number of households in recipience is greater in EU-SILC, whereas the average amounts are substantially higher in HBS. This is again related to irregular nature of this income component as well as different ways of collecting self-employment income.

Survivors' benefits as best compared at household level, as they are usually paid to a household as a whole. The average amount received matches rather well between surveys. The number of recipients differs, however, being substantially lower in EU-SILC. The administrative figure lies between these two values, indicating underestimation in EU-SILC and overestimation in HBS.

The number of households receiving rental income is very similar in two surveys, but the amounts are substantially higher in HBS. In the case of property income, however, the number of recipients is more than ten times as high in EU-SILC. This is due to small interest payments to bank account being much better represented in EU-SILC than in HBS. Cash transfers paid to and received from other households inhibit a different pattern: the average amounts are higher in HBS, whereas number of households in recipience is considerably higher in EU-SILC.

The figures for family benefits should be compared with some caution, because HBS data is missing maternity benefits and administrative sources are also in part short of the same benefit as far as the number of recipients is concerned. Given these qualifications, it appears that data from all three sources is reasonably similar. The same cannot be said about the housing benefits, however. The differences that emerge are probably related to the irregular nature of this component.

These component level differences between EU-SILC and HBS also translate into differences in total income. As can be seen from Table 4.2, the average disposable income is by 14% higher in EU-SILC as compared to the respective figure from HBS.

In conclusion, the coherence between EU-SILC and HBS is not easy to assess due to different definitions and extrapolation of monthly income to yearly income. When an omission is made to these instances, it appears, that the number of people receiving a benefit is in most cases higher in EU-SILC, whereas the amounts received are either the same or moderately lower in HBS. Sickness and unemployment benefits, as small and non-salient types of income seem to suffer from a recall bias in EU-SILC, while the other income components are reasonably coherent with HBS and administrative data.

Table 4.4 compares the mean and number of recipients of most income components in EU-SILC 2006 to the estimates from 2005 operation. Changes that emerge are, in general, in line with what could be expected. While the average salary has increased by 18%, the increase in wage receivers has been more modest – 5%. Both of these changes reflect the general growth of economy at large. Value of company cars has not changed much, since the tax approach is used and corresponding legislation has remained the same. The number of those benefiting from the use of a company car has increased, however, reflecting too the favourable economic situation. Amounts paid to private pension schemes and number of people making these payments have both increased, which is to be expected given that funds are actively campaigning for more people to join up. Number of persons who engage in self-production is greater, but the average income they report is smaller than in the last year. Although the drop is not very large and could be a result of sample fluctuations.

Turning to individual level social transfers now, the average amounts and the number of persons receiving unemployment benefits has increased despite the drop in overall unemployment rate. This is due to redesign of respective questions in the questionnaire, which has resulted in much better capture of different unemployment benefits. Old-age benefits have risen, while the number of receivers has not changed much. The survivors' and sickness benefits have gone through much greater and unexpected changes, with both amounts increasing substantially and the number of people receiving sickness benefits also increasing by 29%. Average disability and education-related benefits have both increased while the number of recipients has not changed, which is in line with what could be expected.

Of household level variables the average amounts have not changed much for any of the components with the exception of property income. The questionnaire for 2006 operation was changed to better

capture small interest payments from bank accounts. Greater validity achieved by this change is evident from the drop in average amounts received and an enormous increase in the number of households reporting this type of income. The number of households receiving rental income has increase by one third, but the figures concerned here are so small that it could well be due to sample fluctuations. Average amount of housing benefits has increased that can be explained by wage growth, which has resulted in fewer households being eligible for these benefits. Number of households making and receiving transfers from other households has also decreased, but no plausible explanation is available for this. Average gross income has increased by 17% while average disposable income increased by 19%, which is mainly driven by increase in employee cash income.

Table 4.4. Mean and number of recipients of income components in EU-SILC 2005 and 2006

	Mean		Number of recipients	
	2005	2006	2005	2006
Individual level components				
PY010N	60,212	70,768	624,902	653,971
PY020N	17,165	16,843	22,319	26,099
PY035N	4,538	5,207	63,390	78,039
PY050N	11,475	9,587	64,937	73,750
PY090N	4,813	7,693	12,826	16,004
PY100N	27,771	32,398	303,057	296,346
PY110N	8,856	12,533	11,960	10,964
PY120N	2,763	3,580	62,161	80,244
PY130N	14,925	17,362	65,631	65,049
PY140N	9,073	10,659	18,487	18,782
Household level components				
HY040N	11,758	12,067	4,653	6,321
HY050N	10,424	10,724	190,368	190,720
HY070N	5,783	6,066	15,870	12,604
HY080N	13,827	14,156	27,402	21,187
HY090N	4,989	1,420	21,650	183,962
HY110N	2,679	2,428	6,725	5,770
HY120N	555	485	300,951	308,450
HY130N	13,278	12,390	34,867	28,999
HY145N	2,790	2,672	164,034	195,274
HY010	109,938	128,581		
HY020	90,483	107,329		
HY022	85,985	102,080		
HY023	79,335	92,554		

4.2. Comparison of other target variables with external sources

The differences in the share of household possessing consumer durables are negligible in the case of telephones and cars; the difference in the case of TVs is also within the standard error of estimate (Table 4.5). For Washing machines and personal computers, however, EU-SILC provides somewhat higher estimates than HBS.

Table 4.5. Share of households in possession of various consumer durables based on EU-SILC and the HBS, 2006

Consumer durable	**EU-SILC**	**HBS**
Telephone, including mobile phone	94.5	95.1
TV	97.7	95.4
Washing machine	86.3	82.6
Car	48.2	48.1
Personal computer	48.1	45.1

Table 4.6 presents the distribution of households by dwelling type, revealing that the differences between two surveys are non-existent regarding this variable.

In Table 4.7 the distribution of population aged 16-74 derived from EU-SILC and LFS is compared. Most of the differences are minor, with the only exceptions being ISCED levels 4 and 5. There are somewhat less people with post-secondary non-tertiary education according to EU-SILC and more people with first stage of tertiary education. Given that the questions used in the two surveys are identical, this must be due to sample fluctuations.

Table 4.6. Households by the type of dwelling based on EU-SILC and the HBS, 2006

Type of dwelling	EU-SILC	HBS
Detached house	26.2	25.7
Semi-detached or terraced house	3.5	3.7
Apartment or flat	69.3	68.7
Some other kind of accommodation	(1.1)*	1.8
Total	100.0	100.0

* Unreliable estimate, based on 20-39 sample observations.

Table 4.7. Distribution of population aged 16-74 by ISCED level, based on the EU-SILC and the LFS, 2006

ISCED level	EU-SILC	LFS
0 Pre-primary education	0.3	0.4
1 Primary education	2.9	2.7
2 Lower secondary education	17.8	18.4
3 (Upper) secondary education	44.5	45.3
4 Post-secondary non tertiary education	8.0	5.6
5 First stage of tertiary education	25.9	27.2
6 Second stage of tertiary education	(0.3)*	0.4
Total	100.0	100.0

* Unreliable estimate, based on 20-39 sample observations.

Finally, Table 4.8 presents the comparison of population aged 16 or over by current activity status. The differences that can be observed between the two data sources are relatively minor and may be due to misclassification to ‘other inactive’ category in HBS.

Table 4.8. Distribution of population aged 16 and over by self-defined activity status based on EU-SILC and the HBS, 2006

Activity status	EU-SILC	HBS
Working full-time	52.9	53.6
Working part-time	3.6	3.2
Unemployed	4.3	4.1
Pupil, student	8.8	8.9
In retirement	21.8	21.6
Permanently disabled	3.7	4.3
Fulfilling domestic tasks and care responsibilities	4.7	3.2
Other inactive	(0.1)*	1.1
Total	100.0	100.0

* Unreliable estimate, based on 20-39 sample observations.