EU-SILC 2007 in Estonia: Final Quality Report

Tallinn 2009

CONTENTS

INTRODUCTION

The EU-SILC survey in Estonia started in 2004. In the first year, a sample of 6000 households was selected for the survey. This households were randomly divided into four rotational groups. According to the original rotational scheme, one of the groups was to be dropped in 2005 and another in 2006, but due to lower than expected response rate, it was decided to keep all the rotational groups in the sample. New sub-samples were introduced into the survey in 2005, 2006. In 2007 two rotational groups from the 2004 initial sample were dropped and a new subsample was introduced. Thus the 2007 sample consists of five rotational groups (two started in 2004, one started in 2005, 2006 and 2007 respectively.). Present report concerns mostly longitudinal part of the survey including four rotational groups, two of which retained from 2004, one from 2005 and one from 2006. Unless specially mentioned, all tables in the report use data of these four sub-samples only.

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

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

Longitudinal indicators are not available.

2. ACCURACY

2.1. Sample design

2.1.1. Type of sampling design
Not to be provided after first wave.

2.1.2. Sampling units
Not to be provided after first wave.

2.1.3. Stratification and sub-stratification criteria
Not to be provided after first wave.

2.1.4. Sample size and allocation criteria
Not to be provided after first wave.

2.1.5. Sample selection schemes
Not to be provided after first wave.

2.1.6. Sample distribution over time
Not to be provided after first wave.

2.1.7. Renewal of sample: Rotational groups
Not to be provided after first wave.

2.1.8. Weightings

Weighting scheme was generally in line with documents V. Verma „EU-SILC weighting procedures: an outline" and J.-M. Museux „Weighting and estimation for the EU-SILC rotational design".
Longitudinal database of 2007 contains three subsamples of households:

S4	Households introduced into the survey in 2004 and their split-offs. Year 2007 is their fourth year in the survey. These housholds form two rotational groups (3 and 4) . [1]
S3	Households introduced into the survey in 2005 and their split-offs. They form one rotational group. Year 2007 is their third year in the survey.
S2	Households introduced into the survey in 2006 and their split-offs. They form also one rotational group. Year 2007 is their second year in the survey.

2.1.8.1. Design factor
Not to be provided after first wave.

2.1.8.2. Non-response adjustments

Personal base weights of 2006 are corrected for non-response for 2007 wave. This was done independently for each sub-sample s_2, s_3, s_4. Persons and households no longer in scope in 2007 were excluded prior to the correction as they are not considered as non-response. Correction for non-response was done with the help of logistic regression model with tenure status, household equivalised income, number of children in the household, urbanization status and county of place of residence, age, gender as auxiliary variables. The model was weighted on the base weights of 2006. According to the model response probability r_i of person i for year 2007, given he/she had responded in 2006, was estimated. Weight corrections were done according to same procedure as in 2006 and 2007.

2.1.8.3. Adjustment to external data

Calibration was performed using population data from Estonial Population Register according to sex, age-groups, county and urbanization.

2.1.8.4. Final longitudinal weight

The basis for calculating longitudinal weights are 2007 base weights for sub-samples corrected for non-response. There are three longitudinal sets of interest in year 2007:

- Longitudinal set of two year duration, involving data from year 2006 and 2007. Longitudinal weight to be used for this set is RB062 . All sub-samples S2 and S3 and S4 contribute to this set. Base weights of S2, S3 and S4 were multiplied by a factor according to the size of a sub-sample to combine the subsamples into one set.

- Longitudinal set of three year duration, involving data from year 2005, 2006 and 2007. Longitudinal weight to be used for this set is RB063. Only sub-samples S3 and S4 contribute to this set. Base weights of S3 and S4 were multiplied by a factor according to the size of a sub-sample to combine the subsamples into one set.

---

[1] For two other rotational groups (1 and 2) of the 2004 sample 2006 was their last year in the survey and they are excluded from the 2007 longitudinal database.

- Longitudinal set of four year duration, involving data from year 2004, 2005, 2006 and 2007. Longitudinal weight to be used for this set is RB064. Only the sub-sample S4 contributes to this set.

Children born between interviews of 2005 and 2006 and persons who moved into sample household from outside received zero weight.

2.1.8.8. Final household cross-sectional weight

Final cross-sectional households weights DB090 for year 2007 wave were recalculated in a way to correspond only to sub-samples S2, S3 and S4. That is, without taking new households of 2007 into account. Thus, weight DB090 is different from similar weight in cross-sectional database of 2007. For 2004, 2005 and 2006 waves the cross-sectional weights were recalculated taking into account that two rotational groups (1 and 2) of the 2004 are excluded from the 2007 longitudinal database as 2006 was their last year in the survey.

2.1.9. Substitution

No substitution has been used.

2.2. Sampling errors

The following table reports the mean, the number of observations (before and after imputations) and the standard error for different income components.

Table 2.1. Number of observations and standard error of different income components, 2007

| Income components | Mean[2] | Number of observations | | Standard error |
		Before imputation[3]	After imputation	
Total household gross income (HY010)	156811	1701	3620	5954
Total disposable household income (HY020)	131695	1848	3620	4556
Total disposable household income before social transfer other than old-age and survivors' benefits (HY022)	123763	2429	3620	4550
Total disposable household income before social transfers including old-age and survivors' benefits (HY023)	104543	2438	3620	4618
Net income components at household level				
Imputed rent (HY030N)	27306	120	3620	445
Income from rental of a property or land (HY040N)	100	3620	3620	37
Family/ children related allowances (HY050N)	4156	3613	3620	235
Social exclusion not elsewhere classified (HY060N)	22	3620	3620	11
Housing allowances	86	3611	3620	15

---

[2] Zeros are included in calculations.
[3] Imputation includes both fully and partially missing values of national components of the income variable. Net/gross conversion is not considered as imputation.

Income components	Mean[2]	Number of observations		Standard error
		Before imputation[3]	After imputation	
(HY070N)				
Regular inter-household cash transfers received (HY080N)	632	3617	3620	82
Interest, dividends, profit from capital investments in incorporated business (HY090N)	599	2297	3620	153
Interest repayments on mortgage (HY100N)	1855	3505	3620	178
Income received by people aged under 16 (HY110N)	48	3611	3620	12
Regular taxes on wealth (HY120N)	264	3543	3620	12
Regular inter-household cash transfers paid (HY130N)	663	3618	3620	86
Tax on income and social contributions, net (HY140N)	0	3620	3620	0
Repayments/ receipts for tax adjustment (HY145N)	-944	3467	3620	100

Gross income components at household level				
Imputed rent (HY030G)	29288	120	3620	458
Income from rental of a property or land (HY040G)	130	3620	3620	48
Family/ children related allowances (HY050G)	4627	3613	3620	292
Social exclusion not elsewhere classified (HY060G)	22	3620	3620	11
Housing allowances (HY070G)	86	3611	3620	15
Regular inter-household cash transfers received (HY080G)	632	3617	3620	82
Interest, dividends, profit from capital investments in incorporated business (HY090G)	748	2297	3620	198
Interest repayments on mortgage (HY100G)	1855	3505	3620	178
Income received by people aged under 16 (HY110G)	48	3611	3620	12
Regular taxes on wealth (HY120G)	264	3543	3620	12
Regular inter-household cash transfers paid (HY130G)	663	3618	3620	86
Tax on income and social contributions, gross (HY140G)	24189	957	3620	1397
Net income components at personal level				
Employee cash or near cash income (PY010N)	49346	8219	8532	1103
Non-cash employee income (PY020N)	1200	8375	8532	86
Contributions to individual private pension plans (PY035N)	372	8476	8532	30
Cash benefits or losses from self employment (PY050N)	2445	8461	8532	1785
Pension from individual private plans (PY080N)	6	8532	8532	5
Unemployment benefits (PY090N)	86	8523	8532	16
Old-age benefits (PY100N)	9592	8500	8532	211
Survivors' benefits (PY110N)	90	8531	8532	13
Sickness benefits (PY120N)	291	8345	8532	31
Disability benefits (PY130N)	1230	8519	8532	72
Education-related benefits (PY140N)	241	8524	8532	60
Gross income components at personal level				
Employee cash or near cash income (PY010G)	60747	8219	8532	1415
Non-cash employee income (PY020G)	1558	8375	8532	112
Employer's social insurance contributions (PY030G)	19810		8532	472
Contributions to individual private pension plans (PY035G)	372	8476	8532	30

Cash benefits or losses from self employment (PY050G)	3183	8461	8532	2318
Pension from individual private plans (PY080G)	7	8532	8532	6
Unemployment benefits (PY090G)	102	8523	8532	19
Old-age benefits (PY100G)	9666	8500	8532	214
Survivors' benefits (PY110G)	90	8531	8532	13
Sickness benefits (PY120G)	342	8345	8532	38
Disability benefits (PY130G)	1230	8519	8532	72
Education-related benefits (PY140G)	241	8524	8532	60

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

Table 2.2. Number of observations and standard error of mean equivalised disposable income, 2007

	Mean	Number of observations		Standard error
		Before imputation[4]	After imputation	
Subclasses by household size				
1 household member	61992	485	735	3097
2 household member	80697	1035	2078	1901
3 household members	87989	754	1982	1625
4 and more	92937	1127	3737	5151
Population by age group				
<25	89902	646	1918	6981
25-34	107026	302	988	3501
35-44	85153	463	1383	2019
45-54	88423	548	1505	7031
55-64	76071	510	1125	1623
65+	57271	932	1613	890
Population by sex				
Male	87644	1511	3954	2863
Female	79851	1890	4578	2412

2.3. Non-sampling errors

2.3.1. Sampling frame and coverage errors

Not to be provided after first wave.

2.3.2. Measurement and processing errors

2.3.2.1. Measurement errors

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

The questionnaires were drawn up in 2004 following the international practices in collecting income data. Also, where possible questions from the existing surveys carried out by the Statistics Estonia

---

[4] Imputation includes both
 fully and partially missing values of national components of the income variable. Net/gross conversion is not considered as imputation.

and known to be valid and reliable, were used. Pilot surveys were carried out in 2002 and 2003 with the main aim of testing the questionnaires. The results were thoroughly analysed and feedback sessions with interviewers were carried out. The questionnaires were modified accordingly for the use in the main operation.

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

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

Other notable modifications in the 2006 questionnaire were as follows:

a) Clarifying the type of work contract

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

c) Providing intervals to report income and profit

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

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

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

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

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

Other notable modifications in the 2007 questionnaire were as follows:

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

2) Adding in cost of utilities

3) Developing mortgage payment and interest questions

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

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

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

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

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

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

In 2007 the training programme was conducted in 4 smaller groups of about 15 people, similarly to 2006, with more emphasis on practical work and discussion of mistakes from previous years as in preceding interviewer trainings. All returning interviewers attended a day long training session. During the training, the EU-SILC team briefed the interviewers on all renewals in the questionnaires, discussed previous years' errors and tracing and specifics of assigning household and person numbers in the longitudinal survey. Practical work sessions were conducted in groups of five and each interviewer had to conduct a model interview regarding for a simulated situation using their laptop. At the end of the training session, each interviewer received personal feedback about their mistakes the previous year. Interviewers new to EU-SILC attended a 2 day training session, which included a thorough overview of questionnaires and practical exercises as well as all the topics covered with returning interviewers.

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

2.3.2.2. Processing errors

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

1) Enforced lack of durables (missing answers due to unclear questionnaire layout);

2) Family benefits (inconsistent with household composition);

3) Relationship between household members (implausible relationships recorded in the household matrix);

4) Number of years in employment (inconsistencies with the time of taking up the first job);

5) Limitations in daily activities for health reasons (sometimes skipped despite there being no routing);

6) Calendar of activities (inconsistencies with other data).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

3) Errors in coding.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

As can be seen from table 2.5, the most common types of errors in 2007 had to do with interviewers not correcting their mistakes after an error code had prompted them to do so, and the use of category 'other', while a suitable category existed. These were the categories with the least errors in 2006, and as the error numbers have not increased, it is clear that in all other categories, error counts have decreased considerably.

Table 2.5. Interviewer errors and their processing, 2007

Type of error	Number of errors detected	Share of errors requiring a call-back
No remark explaining unusual situation	28	10
Interviewer made an error, but did not correct it	485	249
Interviewer's remark does not explain unusual situation	57	16
Data not sufficient for coding	89	30
Starting and finishing times recorded incorrectly	9	0
Use of category Other, while a suitable category exists	429	391
In-office checks	105	83
Interviewer has misunderstood a question	226	163
Data entry mistake	16	16
Not interviewers error	195	22
Total	1639	980

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

2.3.3. Non-response errors

2.3.3.1. Achieved sample size

Number of households for which an interview is accepted for the database (DB135=1):   3620

Number of persons 16 years or older who are members of the households for which the interview is accepted for the database: 8532

      of which sample persons (RB100=1):   8184

      co-residents (RB100=2):   348

2.3.3.2. Unit non-response

Response rate for households

In total, 3934 households were passed on to 2007 from 2006. Another 138 households were added to the sample as a result of split-off of original households (DB110=8). Among them, 40 households were out-of-scope or non-existent in 2007 (DB110=3,4,5,6 or DB120=23). Interview of 3620 households was accepted to the database (DB135=1). **Wave response rate** is thus: 3620/(3934+138-40) = 90.0%.

Rotational groups 3 and 4 will be dropped from the sample in 2008 due to rotation. In rotational groups 5 and 6, 2214 households were passed on to 2007 from 2006. Among them, 23 households were out-of-scope or non-existent in 2007 (DB110=3,4,5,6 or DB120=23). Among those 2214 households, 2029 will be passed on to 2008. **Longitudinal follow-up rate** is thus: 2029/(2214 -23)=92.6%.

In addition to 2029 households passing on to 2008 among those passed on to 2007 from 2006 in rotational groups 5 and 6, there are 46 households among split-off households that will be passed on to 2008. **Follow-up ratio** is thus: (2029+46)/(2214-23)=94.7%.

In 2006 longitudinal component, interview of 3985 households was accepted for the database. **Achieved sample size ratio** is thus: 3620/3985 =0.91

Response rate for persons

In total, 9166 sample persons aged 16 and over were passed on to 2007 from 2006.
Among them, 53 persons belonged to households no longer in scope (DB110 = 3,4,5,6) and 106 persons were no longer in scope in existing households (RB110=6 or RB120=2,3). Among remaining 9007 sample persons 8120 were completed the personal interview (RB250=11 to 13). **Wave response rate for persons** is thus: 8120/(9166-53-106) =90.2%.

Wave response rate for co-residents selected in first wave cannot be calculated since all co-residents selected in first wave have not yet reached the age of 16 years.

Since longitudinal component does not contain any other sample persons than those passed on to 2007 from 2006, **longitudinal follow-up rate for persons** coincedes with wave response rate for persons: 8120/(9166-53-106) =90.2%.

In 2006, 8965 sample persons and 319 co-residents completed personal interview in longitudinal components. In 2007 8125 sample persons and 346 co-residents completed personal interview. **Achieved sample size ratio for sample persons** is thus: 8125/8965 = 0.91

Achieved sample size ratio for sample persons and co-residents is thus: (8125+346)/(8965+319) = 0,91

Achieved sample size ratio for co-residents selected in first wave cannot be calculated since all co-residents selected in first wave have not yet reached the age of 16 years.

346 non-sample persons aged 16+ completed personal interviews (RB250 = 11,12,13) .There were 368 co-residents aged 16 and over listed in the households accepted for the database in 2007 (quests not included), 20 of which were out-of-scope. In addition, there were 30 co-residents aged 16 and over in non-responded households forwarded to 2007 from 2006 (according to last household interview), 2 of whom in households no longer in scope in 2006. **Response rate for non-sample persons** is thus: 346/(368-20+30-2)=92.0%

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

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

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

	Total	DB110									
		1	2	3	4	5	6	7	8	9	10
Total	4072	3455	330	4	11	23	0	85	138	0	26
%	100.0	84.8	8.1	0.1	0.3	0.6	0.0	2.1	3.4	0.0	0.6

RECORD OF CONTACT AT ADDRESS

	Total	DB120				
		11	21	22	23	Missing
Total (DB110 = 2,8,9)	468	430	34	2	2	0
%	100.0	91.9	7.3	0.4	0.4	0.0

HOUSEHOLD QUESTIONNAIRE RESULT

	Total	DB130					
		11	21	22	23	24	Missing
Total (DB110=1 or DB120=11)	3885	3624	135	20	19	87	0
%	100.0	93.3	3.5	0.5	0.5	2.2	0.0

HOUSEHOLD INTERVIEW ACCEPTANCE

	Total	DB135		
		1	2	Missing
Total (DB130=11)	3624	3620	4	0
%	100.0	99.9	0.1	0.0

2.3.3.4. Distribution of persons by membership status (RB100)

MEMBERSHIP STATUS

	Total	Current household members				No current household members			Missing
		RB110=1	=2	=3	=4	RB110=5	=6	=7	
Total	10472	9836	81	147	77	213	85	33	0
%	100.0	93.9	0.8	1.4	0.7	2.0	0.8	0.3	0.0

MOVED TO

	Total	RB120			
		1	2	3	4
Total	213	186	10	17	0
%	100.0	87.3	4.7	8.0	0.0

'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 2.3. Distribution of item non-response, household-level variables, 2007

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)	3610	99.7	28	0.8	1891	52.4
Total disposable household income (HY020)	3615	99.9	18	0.5	1754	48.5
Total disposable household income before social transfer other than old-age and survivors' benefits (HY022)	3581	99.0	56	1.6	1135	31.7
Total disposable household income before social transfers including old-age and survivors' benefits (HY023)	3378	92.4	154	4.6	1028	30.4
Net income components at household level						
Imputed rent (HY030N)	3500	96.7	3500	100.0	0	0.0
Income from rental of a property or land (HY040N)	50	1.4	0	0	0	0.0
Family/ children related allowances (HY050N)	1482	40.9	0	0.0	7	0.4
Social exclusion not elsewhere classified (HY060N)	22	0.6	0	0.0	0	0.0
Housing allowances (HY070N)	66	1.8	9	10.8	0	0.0
Regular inter-household cash transfers received (HY080N)	153	4.2	3	2.0	0	0.0
Alimonies received (compusory + voluntary (HY081N)	80	2.2	1	1.3	0	0
Interest, dividends, profit from capital investments in incorporated business (HY090N)	1373	37.9	1287	93.7	36	2.6
Interest repayments on mortgage (HY100N)	361	10.0	115	31.9	0	0.0

Income received by people aged under 16 (HY110N)	58	1.6	6	10.3	3	5.2
Regular taxes on wealth (HY120N)	2479	68.4	77	3.1	0	0.0
Regular inter-household cash transfers paid (HY130N)	176	4.9	2	1.1	0	0.0
Tax on income and social contributions, net (HY140N)	0	0.0	0	.	0	.
Repayments/ receipts for tax adjustment (HY145N)	1341	37.0	130	9.7	23	1.7
Gross income components at household level						
Imputed rent (HY030G)	3500	96.7	3500	100.0	0	0.0
Income from rental of a property or land (HY040G)	50	1.4	0	0	0	0.0
Family/ children related allowances (HY050G)	1482	40.9	0	0.0	7	0.4
Social exclusion not elsewhere classified (HY060G)	22	0.6	0	0.0	0	0.0
Housing allowances (HY070G)	66	1.8	9	10.8	0	0.0
Regular inter-household cash transfers received (HY080G)	153	4.2	3	2.0	0	0.0
Alimonies received (compusory + voluntary (HY081G)	80	2.2	1	1.3	0	0
Interest, dividends, profit from capital investments in incorporated business (HY090G)	1373	37.9	1287	93.7	36	2.6
Interest repayments on mortgage (HY100G)	361	10.0	115	31.9	0	0.0
Income received by people aged under 16 (HY110G)	58	1.6	6	10.3	3	5.2
Regular taxes on wealth (HY120G)	2479	68.4	77	3.1	0	0.0
Regular inter-household cash transfers paid (HY130G)	176	4.9	2	1.1	0	0.0
Tax on income and social contributions, gross (HY140G)	2663	73.6	2663	100.0	0	0.0

Table 2.4. Distribution of item non-response, person-level variables, 2007

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)	4693	55.0	343	7.3	30	0.6
Non-cash employee income (PY020N)	996	11.7	260	26.1	103	10.3
Company car (PY021N)	68	0.8	68	100.0	0	0.0
Contributions to individual private pension plans (PY035N)	485	5.7	57	11.8	1	0.2
Cash benefits or losses from self employment (PY050N)	575	6.7	73	12.7	2	0.3

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	%
Pension from individual private plans (PY080N)	5	0.1	0	0.0	0	0.0
Unemployment benefits (PY090N)	90	1.1	9	10.0	0	0.0
Old-age benefits (PY100N)	2093	24.5	33	1.6	1	0.0
Survivors' benefits (PY110N)	87	1.0	1	1.1	0	0.0
Sickness benefits (PY120N)	646	7.6	187	28.9	0	0.0
Disability benefits (PY130N)	604	7.1	13	2.2	0	0.0
Education-related benefits (PY140N)	199	2.3	8	4.0	0	0.0
Gross income components at personal level						
Employee cash or near cash income (PY010G)	4693	55.0	343	7.3	30	0.6
Non-cash employee income (PY020G)	996	11.7	260	26.1	103	10.3
Company car (PY021G)	68	0.8	68	100.0	0	0.0
Employer's social insurance contributions (PY030G)	4517	52.9	4517	100.0	0	0.0
Contributions to individual private pension plans (PY035G)	485	5.7	57	11.8	1	0.2
Cash benefits or losses from self employment (PY050G)	575	6.7	73	12.7	2	0.3
Pension from individual private plans (PY080G)	5	1.1	9	10.0	0	0.0
Unemployment benefits (PY090G)	90	24.5	33	1.6	1	0.0
Old-age benefits (PY100G)	2093	1.0	1	1.1	0	0.0
Survivors' benefits (PY110G)	87	7.6	187	28.9	0	0.0
Sickness benefits (PY120G)	646	7.1	13	2.2	0	0.0
Disability benefits (PY130G)	604	2.3	8	4.0	0	0.0
Education-related benefits (PY140G)	199	1.1	9	10.0	0	0.0

2.4. Mode of data collection

Table 2.5. Distribution of household members aged 16 and over in responded households by data status (RB250), 2007

HOUSEHOLD MEMBERS 16+ (RB245=1 to 3)

	Total	RB250=11	=12	=13	=14	=21	=22	=23	=31	=32	=33
Total	8532	8471	0	0	61	0	0	0	0	0	0
%	100.0	99.3	0.0	0.0	0.7	0.0	0.0	0.0	0.0	0.0	0.0

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

	Total	RB250=11	=12	=13	=14	=21	=22	=23	=31	=32	=33
Total	8184	8125	0	0	59	0	0	0	0	0	0
%	100.0	99.3	0.0	0.0	0.7	0.0	0.0	0.0	0.0	0.0	0.0

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

	Total	RB250=11	=12	=13	=14	=21	=22	=23	=31	=32	=33
Total	348	346	0	0	2	0	0	0	0	0	0
%	100.0	99.4	0.0	0.0	0.6	0.0	0.0	0.0	0.0	0.0	0.0

Table 2.6. Distribution of household members aged 16 and over in responded households by type of Interview (RB260), 2007

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

	Total	RB260=1	=2	=3	=4	=5	Missing
Total	8532	124	7351	14	1	981	61
%	100.0	1.5	86.2	0.2	0.0	11.5	0.7

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

	Total	RB260=1	=2	=3	=4	=5	Missing
Total	8184	119	7065	13	1	927	59
%	100.0	1.5	86.3	0.2	0.0	11.3	0.7

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

	Total	RB260=1	=2	=3	=4	=5	Missing
Total	348	5	286	1	0	54	2
%	100.0	1.4	82.2	0.3	0.0	15.5	0.6

2.5. Imputation procedure

As 2007 was the third survey year, it was possible, for some households and persons, to use values of previous year to impute missing values. Data of 2006 was used only if household or person received particular kind of income in 2006 and analysis showed that these two incomes are sufficiently closely related. If analysis indicated no correlation between the incomes of 2006 and 2007, values were not used in imputation.

Details on the number of values forwarded from 2006 to 2007 are given in Table 2.7.

If missing value could not be imputed with data from previous year, the following methods were used:

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

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

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

- Random regression with IveWare;

For some income variables having highly skewed distribution, imputation was conducted on the log-scale. In general, empirical bounds of values present in the dataset were used in IVEware to bound imputed values. For some income components, amount per month was imputed and then converted into amount per year.

If an income component was collected only net (PY020, PY080, PY090, PY100, PY110, PY120, HY050, HY140, HY145), then missing net values were imputed and then converted to gross using net/gross conversion algorithm, where necessary. Respectively, if an income component was collected only gross (PY035, PY130, PY140, HY060, HY070, HY080, HY090, HY120, HY130), then a gross value was imputed and then converted to net.

For income components, which were collected both net and gross (PY010, PY050, HY040, HY110), the procedure was as follows. If only gross value was obtained, it was first converted to net using gross/net conversion algorithm. If both net and gross value were obtained, the net value was used, since it is believed that people know this value better. Missing net values were imputed using IVEware. Gross components of EU-SILC variables were obtained with net/gross conversion algorithm. In this way, when only gross value was obtained, a value recorded in gross component was equal to the collected gross value, since net/gross and gross/net algorithm are in accordance with each other. Also, it allows basing both net and gross recorded values on the same collected value.

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

Following table provides numbers of values imputed for each income component by method of imputation. Numbers are given for the full sample of 2007, i.e. rotational group 7 is included in calculations.

Table 2.7. Percentage of imputed cases by income component in national questionnaire, 2007

Code	Description	National code	Description	Total number of persons/ households having received an amount	No of values from previous year	No of imputed values and method of imputation	Comments
NET INCOME COMPONENTS ON PERSONAL LEVEL							
PY010N	Employee cash or near cash income	H01N	Net wages in a year	6639	68	392 (IveWare)	Amount per month,
		H07B	Total amount of additinal payments that had not been taken into consideration in net wages	1275	0	45 (IveWare)	
PY020N	Non-cash employee income	H21A2	Approximate value of motor fuel compensated by employer	169	2	18 (IveWare)	
		H21B2	Approximate value of use of public transportation compensated by employer	133	0	21 (IveWare)	
		H21C2	Approximate value of food at work received free or at a reduced price	367	23	70 (IveWare)	
		H21D2	Approximate value of using company housing free of charge or on favorable terms	52	5	16 (IveWare)	
		H21N2	Approximate value of housing costs compensated by employer	9	0	9 (IveWare)	
		H21E2	Approximate value of (mobile) phone or postal services compensated by employer	412	12	42 (IveWare)	
		H21F2	Approximate value of health services compensated by employer	174	8	54 (IveWare)	
		H21G2	Approximate value of training not related to work but compensated by employer	19	0	4 (IveWare)	
		H21H2	Approximate value of sporting possibilities free of charge or at reduced prices	255	9	48 (IveWare)	
		H21I2	Approximate value of holiday trip paid by employer	60	0	12 (IveWare)	
		H21J2	Approximate value of foodstuffs	74	0	12 (IveWare)	

Code	Description	National code	Description	Total number of persons/ households having received an amount	No of values from previous year	No of imputed values and method of imputation	Comments
		H21K2	Approximate value of leasing or loan at reduced interest rate	7	0	2 (IveWare)	
		H21L2	Approximate value of use of equipment and/or other tools (e.g. power saw, lawnmower, etc.)	116	0	15 (IveWare)	
		H21M2	Approximate value of other non-cash income from labour	28	0	1 (logical deduction)	
PY021N	Company car	H18	Possibility to use company car	6639	1	0	
		H20	Number of months a company car was used	268	0	3 (logical deduction)	
PY035N	Contributions to individual private pension plans	HK1	Joining the 3rd pillar of pension insurance	12137	1	1(logical deduction)	
		HK2	Payments made into the 3rd pillar of pension insurance	631	22	53 (IveWare)	
		HK5	Payments made to the collecting insurance	127	0	3 (median)	
PY050N	Cash benefits or losses from self employment	H27N	Amount of loss from self-employment		0	10 (IveWare)	
		H28N	Net profit from self-employment	300	0	58 (IweWare)	
		H35B	Net amount of royalties, remuneration or payment under contract for creative or scientific work	47	0	4 (IveWare)	Amount per month
		H46A2	Income from private provision of fee-charging services to other persons or households	308	2	23 (IveWare)	
		H46B2	Income from the sale of own-produced consumer goods (e.g. handicrafts, souvenirs, etc.)	24	0	1 (median)	

Code	Description	National code	Description	Total number of persons/ households having received an amount	No of values from previous year	No of imputed values and method of imputation	Comments
		H46C2	Income from the sale of own-produced foodstuffs (e.g. pies, waffles, shashlik, etc.)	7	0	0	
		H46D2	Income from intermediate commercial transactions	12	0	1 (median)	
		H46E2	Income from agricultural or forestry activities	190	1	8 (IweWare)	
		H46F2	Income from other unregistered self-employment	6	0	1 (logical deduction)	
PY080N	Pension from individual private plans	HK3	Whether received any payments from the 3rd pillar of pension insurance	2	1	0	
PY090N	Unemployment benefits	H55A	Amount of unemployment benefit or any other benefits relating to unemployment	75	2	3 (IweWare)	
		HF6A		36		6(IveWare)	
PY100N	Old-age benefits	H51A	Amount of old-age benefits	2880	21	11 (IveWare)	Amount per month, log-scale
		H52A	Amount of pension for incapacity for work or any other benefits relating to disability	1055	6	4	
PY110N	Survivors' benefits	H53A	Amount of survivors' pension or any other benefits relating to the loss of a provider	132	0	1 (median)	Amount per month
PY120N	Sickness benefits	H54A	Amount of sickness benefits or any other benefits relating to health	963	60	234 (IveWare)	Amount per day

Code	Description	National code	Description	Total number of persons/ households having received an amount	No of values from previous year	No of imputed values and method of imputation	Comments
PY130N	Disability benefits	H52A	Amount of pension for incapacity for work or any other benefits relating to disability	1055	6	4 (IveWare)	Amount per month, log-scale
PY140N	Education-related benefits	H57A2	Amount of state stipendium	8	0	1 (median)	
		H57G2	Education allowance	280	3	8 (IveWare)	
NET INCOME COMPONENTS ON HOUSEHOLD LEVEL							
HY040N	Income from rental of a property or land	D09N	Net income from rental of property	79	1	1 (IveWare)	
HY050N	Family/ children related allowances	D11B	Parental benefit received in the previous calendar year	236	0	0	
		D11C	Total amount of other benefits received in the previous calendar year	2133	0	0	
		D11D	Maternity leave benefits	76	0	8 (IveWare)	log-scale
HY060N	Social inclusion not elsewhere classified	H57B2	Amount of scholarship or grant awarded by a fund or organisation locating in Estonia	12	0	0	
		H57C2	Amount of scholarship or grant awarded by a foreign state	7	0	2 (logical imputation)	
		H58A2	Amount of other support/benefit/pension not mentioned above	36	0	0	
HY070N	Housing allowances	D03A	Amount of subsistence benefit	102	0	13 (IveWare)	
HY080N	Regular inter-household cash	D16A	Amount of regular payments from other households	107	0	4 (IveWare)	

Code	Description	National code	Description	Total number of persons/ households having received an amount	No of values from previous year	No of imputed values and method of imputation	Comments
	transfers received						
HY090N	Interest, dividends, profit from capital investments in incorporated business	H49B2	Interest income from securities (shares, bonds)	8	0	4 (IveWare)	
		H49C2	Dividend income from securities (i.e. shares, bonds)	64	3	14 (IveWare)	
		HE2	Interest income from deposits in a bank - interval	3395	0	29 (IveWare)	
		HE1A	Interest income from deposits in a bank - amount	3424	8	3318 (hot-deck within a given interval)	
HY100N	Interest repayments on mortgage	D08E_A	Mortgage interest paid last year	560		176	When interest is not reported, details about mortgage are requested and interest is deducted analyticall
HY110N	Income received by people aged under 16	D19A	Income received by children aged 16 or less	39	1	2 (IveWare)	
HY120N	Regular taxes on wealth	D10	Tax on land or any other property tax	5154		2 (logical deduction)	
		D10A	Amount of tax on land or any other property tax paid	3542	56	70 (IveWare)	Log-scale
HY130N	Regular inter-household cash transfers paid	D14A	Amount of regular payments to other households	181	0	3 (IveWare)	
HY145N	Repayments/receipts fro tax	H64A	Income tax return for the income received in the previous year	2292	63	162(IveWare)	Log-scale

Code	Description	National code	Description	Total number of persons/ households having received an amount	No of values from previous year	No of imputed values and method of imputation	Comments
	adjustment	H63	Additional amount of tax paid on income in the previous calendar year	119	0	15 (IveWare)	log-scale

2.6. Imputed rent

For calculating imputed rent the following method was used., worked out by E.-M.-Tiit in 2006. Based on registered prices of dwellings in the Real Estate data-base (REDB) regression models predicting the price of m^2 of dwelling by its location, type, size and quality were built. For 2007 EU-SILC the Real Estate data of year 2006 was used. For houses and flats two different regression models were built:

Price of m ^{2} of flat = 12126,1– 1949,0*county + 4898,5 *village+ 3259,0 *quality —-1297,5*rooms + 18,8 *size — 760,4*county*village

Price of m^2 of house = 12169,7 – 1505,1*county + 6752,5*village + 3494,7*quality – 28,5*size – 1246,7 *county*village – 963,7*quality*village

For calculating imputed rents the estimated price of the dwelling and the rate of 12-month EURIBOR (average of the 8 last years), which was 3.239% in 2006 was used:

rent per year = price of m ^{2}∗size of dwelling * 0.03239

2.7. Company cars

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

3. COMPARABILITY

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

3.2. Components of income

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

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

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

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

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

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

3.3. Tracing rules

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

4. COHERENCE

This section will compare the data of the 2006 and 2007 waves of EU-SILC with one another and the longitudinal EU-SILC data to various external sources, including the National Accounts (NA), the Household Budget Survey (HBS), the Labour Force Survey (LFS), wage statistics and social protection statistics.

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

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

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

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

In the following paragraphs, EU-SILC income data is compared component by component between the waves 2006 and 2007 and to income data from administrative sources for income years 2005 and 2006. Table 4.1 presents the comparisons by total amounts received/spent and Table 4.2 by number of recipients.

As illustrated in table 4.1, the total cash incomes from employment indicate a rise of about 20% in incomes from 2005 to 2006. (In Estonia, the income reference year is one year behind the survey year so the 2007 survey gives us annual incomes for 2006, and the 2006 survey gave us the incomes of 2005) The wage statistics figures show a similar increase although the incomes reported through wage statistics by companies indicate systematically lower total wage incomes. The under-representation is around 17%, aka wage statistics incomes make up 83% of the total wage incomes reported by wage-earners themselves. This can partly be due to unreported wages paid to temporary employees, or failing to report other cash benefits by companies and also from PY010 containing a wider set of income components than the definition for wage statistics. Wage statistics also refer to the full-time equivalent of income and exclude part-time work contracts; unofficial work relationships are also excluded. In EU-SILC data both are included. The data concerning wage statistics comes from in-house sources, not administrative registers.
The numbers of recipients for PY010G also have a comparability problem because in wage statistics the number of recipients is calculated in full time units and not actual persons.

Table 4.1. Total amounts of income components by source of information (in kroons), income years 2005 and 2006

Income component	2005		2006	
	EU-SILC	**Administrative**	**EU-SILC**	**Administrative**

		sources*		sources*
Cash or near-cash employee income (PY010N)	46,201,187,883	38,295,556,080	56,063,250,586	47,124,543,520
Unemployment benefits (PY090G)	148.000.000	173.280.000	137,000,000	129,360,000
Old-age benefits (PY100G)	9.460.000.000	9.142.230.000	10,800,000,000	10,908,510,000
Gross survivor's benefits (PY110G)	94,333,736	214,270,000	108,053,845	173,320,000
Disability benefits (PY130G)	1.129.343.155	1.773.970.000	1,321,145,473	1,969,140,000

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

Unemployment benefits went through a drop from 2005 to 2006, both in total amounts paid and number of recipients. This was in line with the trend of increasing employment rates during times of economic growth. The number of people receiving unemployment benefits in EU-SILC is substantially lower from the figure in administrative sources, amounts paid/received too differ by 5-15%, revealing undercoverage in the survey. This is further underlined by the fact that the types of benefits included in PY090 are wider than the definitions in administrative sources, including for instance redundancy payments which are excluded from the ESSPROS classification.

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

Another condition to be considered is that the administrative data includes the institutionalised population whereas the EU-SILC survey does not include institutionalised people in its sample—e.g. hospitalized, in retirement homes and imprisoned people. This difference influences old-age and disability benefits especially.

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

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

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

Table 4.2. Number of recipients of income components by source of information, income years 2005 and 2006

Income component	2005		2006	
	EU-SILC	Other sources*	EU-SILC	Other sources*
Cash or near-cash employee income (PY010N)	653,971	496,277	666,494	518,593
Unemployment benefits (PY090G)	16,004	31,667	10,836	20,375
Old-age benefits (PY100G)	288,343	296,082	287,568	291,580
Survivor's benefits (PY110G)	7,941	9,312	8,482	9,537
Disability benefits (PY130G)	65,049	64,623	66,379	65,477

* 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 2007 to the estimates from 2006 operation. Changes that emerge are, in general, in line with what could be expected. Much like the previous year, 2007 was a year of considerable economic growth and increase in real salaries in Estonia. While the average salary increased by 19%, the increase in wage receivers was more modest – 2%. Comparably, there was a decrease in people receiving unemployment benefits as many discouraged workers found employment and unemployment levels reached record lows. Administrative data confirms the survey results.

Table 4.3. Mean and number of recipients of income components in EU-SILC 2006 and 2007 (income years 2005 and 2006)

	Mean		Number of recipients	
	2006	2007	2006	2007
Individual level components				
PY010N	70,768	84,116	653,971	666,494
PY020N	16,843	8,676	26,099	162,593
PY035N	5,207	5,111	78,039	84,401
PY050N	9,587	36,000	73,750	60,224
PY090N	7,693	10,566	16,004	10,836
PY100N	32,398	37,409	296,346	287,568
PY110N	12,533	12,739	10,964	8,482
PY120N	3,580	3,261	80,244	102,774
PY130N	17,362	19,903	65,049	66,379
PY140N	10,659	10,092	18,782	22,284
Household level components				
HY040N	12,067	11,290	6,321	5,726
HY050N	10,724	12,498	190,720	185,502
HY070N	6,066	4,295	12,604	11,790
HY080N	14,156	15,138	21,187	28,870
HY090N	1,420	1,519	183,962	238,307
HY110N	2,428	4,473	5,770	6,092
HY120N	485	430	308,450	356,151
HY130N	12,390	14,728	28,999	33,360
HY145N	-2,390	-2,455	201,894	213,965
HY010	128,581	155,991		
HY020	107,329	130,759		
HY022	102,080	125,883		
HY023	92,554	130,453		

The large differences in PY020N from 2006 to 2007 result from different calculation rules that came to force in 2007. In 2006 PY020N only included company car, but from 2007 onwards other kinds of non-cash income were added to the amount.
Cash benefits from self employment have grown in part due to favourable economic conditions and partly due to a decrease in income tax. The number of entrepreneurs seems to fluctuate between survey years which hints to a relatively big pool of short-lived businesses.
The increase in the mean of PY090N has to do with a new unemployment benefit attached to one's earlier salary which is considerably lower than the fixed unemployment allowance. There was also a small increase in the latter in 2007.

The amounts paid to private pension schemes and the 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.

Old-age benefits have increased due to a small increase of 300 kroons a month to the national pension, while the number of receivers has not changed much.
The survivors' and sickness benefits continued the trend of the previous year with the number of recipients increasing, whereas the actual amounts have not changed much. This is because neither allowance was increased by the government. Average disability benefits have increased while the number of recipients has not changed much, which is in line with what could be expected. Education-related benefits have on the other hand stayed the same while the numbers of recipient have gone up. This is also because the allowance levels are centrally fixed, and allowances did not increase that year. The increase in number of recipients is likely to be caused by a large birth cohort reaching the age of 18 and becoming eligible for university and vocational school student support.

Household level variables reflect changes in line with personal level variables. The average amounts have not changed much for HY145N, which should not fluctuate much regardless of the state of the economy; all other characteristics have had some changes.
For HY040N the decrease in the average amount is probably due a change in the questionnaire making it less likely for the respondent to mix up net or gross amounts. The small decrease in the number of households receiving income from rental of a property or land probably has to do with sample fluctuations.
Family allowances have increased, in compliance with increases of national benefit levels. 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.
A rapid increase in income collected through HY090 reflects a positive situation on the financial markets, with more people investing and large returns that have to do with overall economic growth. More people also had to pay taxes on wealth, but the amounts are fairly small and have not changed substantially. The decrease has to do with more people having to pay tax but on smaller sums.

The general economic picture should also account for the increase in income received by people aged under 16—more short-term and summer jobs were available for young people because in 2006, there was a big shortage of workers.

Total household gross income and net income increased by approximately 21% each in 2006. The increases stem from the higher wages and other income components, most of which have gone up considerably. HY022 and HY023 follow the same pattern.

4.2. Comparison of other target variables with external sources

In Table 4.3 the distribution of population aged 16-74 derived from EU-SILC and LFS is compared between survey years 2005-2007. The differences between the two data sources are minor, and can mostly be found in the category of post-secondary non tertiary education, which is one of the most difficult ISCED levels to capture accurately. The coherence here is very good. Longitudinal comparison indicates a stable educational structure in society with small variations attributable to sample fluctuations.

Table 4.3. Distribution of population aged 16-74 by ISCED level, based on the cross-sectional EU-SILC and the LFS, 2005-2007

ISCED level	2005		2006		2007	
	EU-SILC	LFS	EU-SILC	LFS	EU-SILC	LFS
0 Pre-primary education	0.3	0.4	0.6	0.5	0,6	0,5
1 Primary education	5.4	3.1	4.8	4.1	4,2	3,5
2 Lower secondary education	18.3	17.5	18.7	18.2	18,6	17,9
3 (Upper) secondary education	43.0	44.8	43.0	44.5	44,1	45,8
4 Post-secondary non tertiary education	7.6	8.7	7.6	5.5	7,1	5,2

5 First stage of tertiary education	24.6	25.1	24.9	26.8	24,9	26,8
6 Second stage of tertiary education	…*	0.6	0,3	0.4	(0,4)**	(0,3)**
Total	100.0	100.0	100.0	100.0	99,9	100.0

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

Table 4.4 presents the longitudinal comparison of population aged 16 or over by most frequent current activity status in EU-SILC and HBS. The differences that can be observed between the two data sources are relatively minor with the exceptions of students, domestic workers and people in the "other inactive" category. This indicates that domestic workers are frequently reported as "other inactive" in HBS.

Table 4.4. Distribution of population aged 16 and over by self-defined activity status based on longitudinal EU-SILC and the HBS, 2005-2007

Activity status	2005		2006		2007	
	EU-SILC	HBS	EU-SILC	HBS	EU-SILC	HBS
Working full-time	51.3	51.1	52.9	50.6	54.2	50.8
Working part-time	3.9	3.2	3.6	3,3	3.9	3.2
Unemployed	5.8	5.2	4.3	3.9	3.2	2.8
Pupil, student	9.0	9.7	8.8	10.7	8.8	12.3
In retirement	22.6	21.5	21.8	22.0	21.7	21.5
Permanently disabled	3.5	4.3	3.7	4.2	3.7	4.1
Fulfilling domestic tasks and care responsibilities	3.9	3.4	4.7	1.5	4.4	1.5
Other inactive	…*	1.6	…*	3.6	(0.0)*	3.4
Total	100.0	100.0	100.0	100.0	99.9	99.6

* Extremely unreliable estimate, based on less than 20 sample observations

The differences in the share of household possessing various consumer durables are mostly negligible. In most cases, the difference between EU-SILC and HBS data is less than 1% and within the standard error of estimate. The differences are more noticeable with computers. This is not unexpected, as computer ownership is steadily increasing among the households and EU-SILC data reflects the situation in the first half of the year, whereas HBS estimate is an average for a whole year. The lower levels in HBS in ownership of a personal computer also stem from more narrow response categories—only a stationary computer at one's house is included in the question, laptops are not.

Table 4.5. Share of households in possession of various consumer durables based on EU-SILC and the HBS, 2005-2007

	2005		2006		2007	
Consumer durable	EU-SILC	HBS	EU-SILC	HBS	EU-SILC	HBS
Telephone, including mobile phone	91.7	92.9	94.5	95.1	95.6	97,8
TV	97.3	96.5	97.7	95.4	97.7	97.2
Washing machine	82.4	83.1	86.3	82.6	86.2	87.5
Car	45.7	45.2	48.2	48.1	50.4	49.9
Personal computer	40.4	42.1	48.1	45.1	53.3	51.7

Table 4.6 presents the distribution of households by dwelling type. In 2005, the differences between data from the two surveys were small, the greatest divergence appearing in the case of detached houses. In 2006 and 2007, the differences were non-existent regarding this variable.

Table 4.6. Households by the type of dwelling based on EU-SILC and the HBS, 2005-2007

	2005		2006		2007	
Type of dwelling	EU-SILC	HBS	EU-SILC	EU-SILC	EU-SILC	HBS
Detached house	27.2	25.4	26.2	25.7	26.5	26.5
Semi-detached or terraced house	3.9	4.1	3.5	3.7	4.1	4.1

Apartment or flat	67.8	68.2	69.3	68.7	68.7	68.7
Some other kind of accommodation	(1.1)*	2.3	(1.1)*	1.8	(0.8)*	(0,8)*
Total	100.0	100.0	100.0	100.0	100.0	100.0

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