

EU-SILC 2006 in Estonia: Final Quality Report

Tallinn 2008

CONTENTS

INTRODUCTION	3
1. COMMON LONGITUDINAL EUROPEAN UNION INDICATORS BASED ON THE LONGITUDINAL COMPONENT OF EU-SILC	3
2. ACCURACY	3
2.1. Sample design.....	3
2.1.1. Type of sampling design.....	3
2.1.2. Sampling units	3
2.1.3. Stratification and sub-stratification criteria	3
2.1.4. Sample size and allocation criteria	3
2.1.5. Sample selection schemes	3
2.1.6. Sample distribution over time.....	3
2.1.7. Renewal of sample: Rotational groups	3
2.1.8. Weightings.....	3
2.1.9. Substitution.....	4
2.2. Sampling errors	4
2.3. Non-sampling errors	7
2.3.1. Sampling frame and coverage errors	7
2.3.2. Measurement and processing errors	7
2.3.3. Non-response errors.....	11
2.4. Mode of data collection	16
2.5. Imputation procedure.....	16
Interest payments received from demand deposit or time deposit account in the previous calendar year.....	21
2.6. Imputed rent	23
2.7. Company cars	23
2.8. Within-household non-response inflation factor	24
3. COMPARABILITY	24
3.1. Basic concepts and definitions	24
3.2. Components of income.....	24
3.2.1. Differences between the national definitions and standard EU-SILC definitions	24
3.2.2. The source or procedure used for the collection of income variables.....	24
3.2.3. The form in which income variables at component level have been obtained	24
3.2.4. The method used for obtaining income target variables in the required form	24
3.3. Tracing rules	25
4. COHERENCE	25
4.1. Comparison of income target variables and number of persons who receive income from each 'income component', with external sources	25
4.2. Comparison of other target variables with external sources	26

INTRODUCTION

The EU-SILC survey in Estonia started in 2004. In the first year, a sample of 6000 households was selected for the survey. Households were randomly divided into four rotational groups. According to original rotational scheme, one of these 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 also introduced into the survey in 2005 and 2006 to ensure cross-sectional representability. Thus, in 2006 the sample consists of six rotational groups (four started in 2004, one started in 2005 and one started in 2006).

The present report concerns mostly the longitudinal part of the survey, i.e. with five rotational groups, four of which retained from 2004 and one from 2005. Unless specially mentioned, all tables in the report use data from these five sub-samples only.

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

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

Longitudinal indicators are not available, as no rotational group has yet been in the survey for four years.

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

Longitudinal database of 2006 contains two kinds of households:

S3	Households introduced into the survey in 2004 and their split-offs. Randomly divided into four rotational groups. Year 2006 is their third year in the survey.
S2	Households introduced into the survey in 2005 and their split-offs. Form one rotational group. Year 2006 is their second year in the survey. .

Thus, there are two longitudinal sets of interest in year 2006:

- Longitudinal set of two year duration, involving data from year 2005 and 2006. Both sub-samples S2 and S3 contribute to this set. Longitudinal weight to be used for this set is RB062.
- Longitudinal set of three year duration, involving data from year 2004, 2005 and 2006. Only the sub-sample S3 contributes to this set. Longitudinal weight to be used for this set is RB063.

Three-year longitudinal weight RB063

This weight is the 2006 base weight for the sub-sample S3, with the exception that children born between interviews of 2005 and 2006 receive zero weight. As S3 contains in fact four independent rotational groups (DB075=1,2,3,4), weighted independently, then computationally RB063 is the base weight divided by four.

Two-year longitudinal weight RB062

The basis for this weight is again the base weight of 2006. To combine S2 and S3 into one set, base weights were multiplied by a factor according to the size of a sub-sample. No immigrant-households were present in S3 and thus no correction was needed for that. Children born between interviews of 2005 and 2006 received, again, zero weight.

Sub-sample S3 contained also some co-residents which responded in both 2005 and 2006 (so-called longitudinal co-residents). Weight RB062 was extended to cover them also (in households containing longitudinal residents, weight RB062 was averaged over all eligible persons).

Personal base weight PB050

Weight PB050 is a base weight for responded household members aged 16 and over. This was calculated from the base weight RB060. Non-response was corrected by post-stratification in groups by sex and exact age in years, so that dataset of responded persons provides the same age-sex structure as personal register. Post-stratification was performed independently in each of rotational groups.

Cross-sectional household weight DB090 for longitudinal database

Cross-sectional households weights DB090 for year 2006 were recalculated in a way to correspond only to sub-samples S2 and S3. That is, without taking new households of 2006 into account. Thus, weight DB090 is different from similar weight in cross-sectional database of 2006.

2.1.9. Substitution

No substitution was 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, 2006

Income components	Mean ¹	Number of observations		Standard error
		Before imputation ²	After imputation	
Total household gross income (HY010)	129091	2613	3985	2271
Total disposable household income (HY020)	107551	2548	3985	1718
Total disposable household income before social transfer other than old-age and survivors' benefits (HY022)	100777	3396	3985	1713
Total disposable household income before social transfers including old-age and survivors' benefits (HY023)	83863	3401	3985	1792
Net income components at household level				
Imputed rent (HY030N)	29141	136	3985	426
Income from rental of a property or land (HY040N)	144	3984	3985	50
Family/ children related allowances (HY050N)	3546	3985	3985	164
Social exclusion not elsewhere classified (HY060N)	9	3985	3985	3
Housing allowances (HY070N)	135	3978	3985	20
Regular inter-household cash transfers received (HY080N)	547	3981	3985	103
Interest, dividends, profit from capital investments in incorporated business (HY090N)	454	2982	3985	154
Interest repayments on mortgage (HY100N)	1077	3896	3985	110
Income received by people aged under 16 (HY110N)	22	3985	3985	5
Regular taxes on wealth (HY120N)	263	3926	3985	11
Regular inter-household cash transfers paid (HY130N)	646	3984	3985	77
Tax on income and social contributions, net (HY140N)	0	3985	3985	0
Repayments/ receipts for tax adjustment (HY145N)	-820	3975	3985	41
Gross income components at household level				
Imputed rent (HY030G)	29141	136	3985	426
Income from rental of a property or land (HY040G)	190	3984	3985	65
Family/ children related allowances (HY050G)	3832	3985	3985	199
Social exclusion not elsewhere classified	9	3985	3985	3

¹ Zeros are included in calculations.

² 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 ¹	Number of observations		Standard error
		Before imputation ²	After imputation	
(HY060G)				
Housing allowances (HY070G)	135	3978	3985	20
Regular inter-household cash transfers received (HY080G)	547	3981	3985	103
Interest, dividends, profit from capital investments in incorporated business (HY090G)	572	2982	3985	202
Interest repayments on mortgage (HY100G)	1077	3896	3985	110
Income received by people aged under 16 (HY110G)	23	3985	3985	5
Regular taxes on wealth (HY120G)	263	3926	3985	11
Regular inter-household cash transfers paid (HY130G)	646	3984	3985	77
Tax on income and social contributions, gross (HY140G)	20632	1046	3985	569
Net income components at personal level				
Employee cash or near cash income (PY010N)	40790	8991	9292	813
Non-cash employee income (PY020N)	341	9122	9292	38
Contributions to individual private pension plans (PY035N)	359	9230	9292	27
Cash benefits or losses from self employment (PY050N)	577	9179	9292	60
Pension from individual private plans (PY080N)	13	9292	9292	11
Unemployment benefits (PY090N)	111	9285	9292	18
Old-age benefits (PY100N)	8437	9284	9292	181
Survivors' benefits (PY110N)	85	9290	9292	12
Sickness benefits (PY120N)	269	9177	9292	33
Disability benefits (PY130N)	1003	9287	9292	61
Education-related benefits (PY140N)	170	9288	9292	56
Gross income components at personal level				
Employee cash or near cash income (PY010G)	50866	8991	9292	1064
Non-cash employee income (PY020G)	449	9122	9292	49
Employer's social insurance contributions (PY030G)	16648	4437	9292	355
Contributions to individual private pension plans (PY035G)	359	9230	9292	27
Cash benefits or losses from self employment (PY050G)	790	9179	9292	81
Pension from individual private plans (PY080G)	14	9292	9292	13
Unemployment benefits (PY090G)	132	9285	9292	23

Income components	Mean ¹	Number of observations		Standard error
		Before imputation ²	After imputation	
Old-age benefits (PY100G)	8551	9284	9292	200
Survivors' benefits (PY110G)	85	9290	9292	12
Sickness benefits (PY120G)	316	9177	9292	41
Disability benefits (PY130G)	1003	9287	9292	61
Education-related benefits (PY140G)	170	9288	9292	56

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, 2006

	Mean	Number of observations		Standard error
		Before imputation ³	After imputation	
Subclasses by household size				
1 household member	51270	615	807	2361
2 household member	67444	1463	2212	1519
3 household members	74784	1306	2221	1550
4 and more	70291	2240	4046	783
Population by age group				
<25	67429	1204	2065	1384
25-34	86197	633	1079	2939
35-44	70661	813	1471	1600
45-54	68226	939	1614	1276
55-64	67473	778	1279	1546
65+	49381	1257	1778	691
Population by sex				
Male	71100	2500	4222	1190
Female	64771	3124	5064	855

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

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 these types of errors completely, steps were taken to reduce them as much as possible.

The 2006 questionnaires were modified using the experience from the first and second waves of the survey. 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 concerned the following variables:

³ Imputation includes both fully and partially missing values of national components of the income variable. Net/gross conversion is not considered as imputation. Households for which within-household inflation factor (HY025) could not be calculated are excluded.

- 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 use of CAPI continued in 2006 allowing for a further reduction in measurement errors stemming from the data collection method. In CAPI primary consistency checks were in place that allowed for an immediate correction of logical inconsistencies on the spot during the interview. 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.

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

2.3.2.2. Processing errors

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

As in 2005, the data for the 2006 operation was collected using CAPI. The data-entry program was written in Blaise and contained most of the consistency checks. In case of each consistency check the interviewer must check if the situation is correct, if it is not, correct it, if it is, then make a remark. This way, most of the errors could already be corrected during the interview. All remarks and suppressed consistency errors were manually checked during the secondary in-office data editing procedure.

The primary data-entry consistency 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/should not have received a certain type of income reported it or not (e.g. whether households with

- children received family benefits, retired people (or people below retirement age) received pensions, employed persons received wages and so on);
 - b. Whether benefits reported to have been received were logical in the age and gender dimensions. For instance student benefits for over 50 year-olds, income taxes for under 15 year-olds, maternity leave and childbirth allowances for men etc.
 - c. Whether an educational level attained was possible below a certain age.
 - d. whether answers provided to different non-monetary deprivation items agreed with each other;
 - e. whether the relationships in the household matrix were consistent with each other as well as with the age and sex of the household members;
 - f. Whether the difference between the starting and finishing time of the interview was too short or too long and so on.
 - g. whether reported taxes or medical benefits received were consistent with income
 - h. Membership in pension plans checked by year of birth to see if legally bound to have joined pension pillar.
 - i. Checks for correct survey area, interviewer code and personal numbers matching household numbers.
- 2) Lower and upper bounds of income variables (incl. benefits). These checks were developed with regard to data collected in the previous wave as well as administrative information.
 - 3) Tracing checks. These controls were implemented to ensure that all split-off households and new household members were assigned correct split numbers and person numbers respectively.
 - 4) Checks for correct survey area, interviewer code and household and personal numbers matching.
 - 5) Checks not allowing for occupations to be written on too general a scale for coding. (e.g. salesperson, cleaner)
 - 6) Checks with information from the previous year. These controls concerned demographic data, information on educational level and labour status as well as the calendar of activities.

The in-office staff promptly checked the questionnaires that were electronically transmitted to the central office. This stage included 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 category 'other' answers were gone through to see if they could be classified under one of the given options.
- 5) Additionally paid income tax was checked in-household to check for double-reporting.
- 6) Errors in coding.
- 7) Study benefits were checked by possibility of obtaining them in the school the respondent attended and legally set amounts.
- 8) Consistency between time reported working under socio-economic status and months that salary was received. Also time spent in prison.
- 9) Reported amounts of family benefits were checked compared with eligibility based on the structure of the family and benefit levels set out in legislation.
- 10) Demographic information in the interviewers' reports was compared to the data recorded in the electronic questionnaires.

All mistakes found 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.

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.

On average there were 1,2 mistakes per household questionnaire. The maximum number of mistakes in one questionnaire was six, 88% of imperfect questionnaires had one mistake in them, and every tenth had 2 or 3 mistakes. In personal questionnaires the average mistake rate was 1,4. . The maximum number of mistakes in one questionnaire was 13. 74% had one mistake; over a fifth of the questionnaires had 2-3 mistakes.

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.

In 2006, 1934 mistakes or 35% of all mistakes were sent to be clarified. Out of the 4943 interviewers' mistakes 1308 mistakes or less than a third (26%) were sent to be clarified. The others could be solved without this. 90% of all mistakes were clarified in one round, but the rest had to be respecified 2-3 times before final clarity was reached.

The largest mistake categories coming up during interview as consistency checks.

The questions that posed the most problems were related to education obtained, job classification and social benefits. Mistakes to do with education levels amounted to 35% (1335 people) of all consistency checks: the same person was reported to have different education levels than in previous waves. A fifth of the mistakes had to do with years spent in employment and age when started first job. The answers provided did not match previous waves' data. These mistakes could have resulted both from the interviewer or the interviewee making a mistake either in 2004, 2005 or 2006. The solution implemented was pre-filling education variables for people in the longitudinal sample while leaving them changeable so that the interviewer would double-check the data the next year.

The third biggest problem area was family benefits—in itself a complex field because there are many different kinds of benefits as well as diverse criteria for eligibility. The interviewers' training dedicated a special session to going through these issues but still 169 mistakes came out in these questions to which the interviewer did not respond when the initial consistency error popped up.

The biggest mistake categories in the in-office secondary data cleaning

Secondary checks revealed 2138 mistakes, of which 16% (339 mistakes) were sent to be clarified. The largest mistake category covering almost a fourth of all mistakes (23%) concerned family benefits. During secondary checks in these questions the following is checked for:

- Checking for the children's age, was a family entitled to certain benefits. Two kinds of mistakes came up: a family should have legally received certain benefits but they were not reported, or a family was not entitled to certain benefits but they were reported. All adoption benefits were reviewed. Over 500 mistakes concerned family benefits.
- Many errors occurred in putting down the socio-economic status. Some mistakes were typos; for instance the interviewer typed in 1 when the number should have been 11. 12% of the mistakes revealed by secondary checks had to do with socio-economic status.
- 1/10 of the mistakes came from comparing a person's socio-economic status and the months they had reported to have been receiving salary. Mostly people forgot to count their month off work on holiday.
- Another tenth of the mistakes had to do with levels of education. Secondary checks revealed that the reported education could not have been obtained in the reported school in the reported year.

- Study benefits caused 5% of all mistakes. They are well regulated by law and the interviewers' training discussed the topic at length but many mistakes were still made.
- Extreme income values accounted for over 80 mistakes.

The third and final stage of data checks involved later in-office data cleaning. The controls implemented at this stage involved further checks of data consistency, comparison with values from previous years, 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 or employment. Also extreme values for all income components as well as total income were checked.

The following actions have been implemented to reduce the errors discussed above in future:

- 1) Giving each interviewer personal feedback during the start of the new survey year about the mistakes they had made the most in the previous year
- 2) Training interviewers to react to consistency checks
- 3) Focusing on the most common errors in the interviewers' training
- 4) Training interviewers during the fieldwork period when necessary
- 5) Training the fieldwork managers in EU-SILC specifics
- 6) Further developing the problematic questions concerning education and family benefit variables.

2.3.3. Non-response errors

2.3.3.1. Achieved sample size in longitudinal component

Number of households for which an interview is accepted for the database (DB135=1): 3985

Number of persons 16 years or older in these households, who completed a personal interview (RB250=11 to 13): 9292

of which sample persons (RB100=1): 8973

co-residents (RB100=2): 319

2.3.3.2. Unit non-response

In total, 4282 households were passed on to 2006 from 2005. Another 165 households were added to the sample as a result of split-off of original households (DB110=8). Among them, 42 households were out-of-scope or non-existent in 2006 (DB110=3,4,5,6 or DB120=23). Interview of 3985 households was accepted to the database (DB135=1).

Wave response rate is thus: $3985 / (4282 + 165 - 42) = 90.5\%$.

Rotational groups 1 and 2 will be dropped from the sample in 2007 due to rotation. In rotational groups 3, 4 and 5, 2437 households were passed on to 2006 from 2005. Among them, 25 households were out-of-scope or non-existent in 2006 (DB110=3,4,5,6 or DB120=23). Among those 2437 households, 2235 households will be passed on to 2007.

Longitudinal follow-up rate is thus: $2235 / (2437 - 25) = 92.7\%$.

In addition to 2235 households passing on to 2007 among those passed on to 2006 from 2005 in rotational groups 3, 4 and 5, there are 53 households among split-off households that will be passed on to 2007.

Follow-up ratio is thus: $(2235 + 53) / (2437 - 25) = 94.9\%$.

In 2005 longitudinal component, interview of 3511 households was accepted for the database.

Achieved sample size ratio is thus: $3985 / 3511 = 1.14$.

In total, 10184 sample persons aged 16 and over were passed on to 2006 from 2005. Among them, 56 persons belonged to households no longer in scope (DB110 = 3,4,5,6) and 139 were no longer in scope in existing households (RB110=6 or RB120=2,3). Among those 10184 sample persons, 8965 persons completed personal interview.

Wave response rate for persons is thus: $8965/(10184-56-139)=89.7\%$.

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 2006 from 2005, **longitudinal follow-up rate for persons** coincides with wave response rate for persons: $8965/(10184-56-139)=89.7\%$.

Non-response reasons among 9989 (=10184-56-139) above mentioned persons were the following:

- 925 or 9.3% belonged to households that did not respond;
- 7 or 0.07% were not able to respond due to illness or incapacity
- 64 or 0.64% refused to cooperate;
- 20 or 0.21% were temporarily away;
- 8 or 0.08% could not be contacted for other reasons;

In 2005, 8028 sample persons and 176 co-residents completed personal interview in longitudinal components. In 2006, 319 co-residents completed personal interview.

Achieved sample size ratio for sample persons is thus: $8965/8028 = 1.12$

Achieved sample size ratio for sample persons and co-residents is thus: $(8965+319)/(8028+176) = 1.13$

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.

There were 345 co-residents aged 16 and over in households accepted to database in 2006 (quests not included), 7 of which were out-of-scope. In addition, there were 22 co-residents aged 16 and over in non-responded households forwarded to 2006 from 2005 (according to last household interview), 2 of whom in households no longer in scope in 2006.

Response rate for non-sample persons is thus: $319/(345-7+22-2)=89.1\%$

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 Intermediate Quality Report):

DB120=21	DB120=22
<ul style="list-style-type: none"> ▪ 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 	<ul style="list-style-type: none"> ▪ The house given is located but given address can not be accessed (due to locked doors or gates, etc) ▪ Address of address-person can not be 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).

HOUSEHOLD STATUS

	Total	DB110									
		1	2	3	4	5	6	7	8	9	10
Total	4447	4010	146	4	14	24	0	80	165	0	4
%	100.0	90.2	3.3	0.1	0.3	0.5	0.0	1.8	3.7	0.0	0.1

RECORD OF CONTACT AT ADDRESS

	Total	DB120				
		11	21	22	23	Missing
Total (DB110 = 2,8,9)	311	229	81	0	1	0
%	100.0	73.6	26.0	0.0	0.3	0.0

HOUSEHOLD QUESTIONNAIRE RESULT

	Total	DB130					
		11	21	22	23	24	Missing
Total (DB110=1 or DB120=11)	4239	3990	173	26	37	13	0
%	100.0	94.1	4.1	0.6	0.9	0.3	0.0

HOUSEHOLD INTERVIEW ACCEPTANCE

	Total	DB135		
		1	2	Missing
Total (DB130=11)	3990	3985	5	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	11627	10750	138	224	83	297	92	43	0
%	100.0	92.5	1.2	1.9	0.7	2.6	0.8	0.4	0.0

MOVED TO

	Total	RB120			
		1	2	3	4
Total (RB110=5)	297	244	4	49	0
%	100.0	82.2	1.3	16.5	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, 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)	3979	99.8	32	0.8	1340	33.7
Total disposable household income (HY020)	3981	99.9	10	0.3	1427	35.8
Total disposable household income before social transfer other than old-age and survivors' benefits (HY022)	3946	99.0	22	0.6	567	14.4
Total disposable household income before social transfers including old-age and survivors' benefits (HY023)	3683	92.4	95	2.6	489	13.3
<i>Net income components at household level</i>						
Imputed rent (HY030N)	3849	96.6	3849	100.0	0	0.0
Income from rental of a property or land (HY040N)	64	1.6	1	1.6	0	0.0
Family/ children related allowances (HY050N)	1627	40.8	0	0.0	0	0.0
Social exclusion not elsewhere classified (HY060N)	28	0.7	0	0.0	0	0.0
Housing allowances (HY070N)	98	2.5	7	7.1	0	0.0
Regular inter-household cash transfers received (HY080N)	140	3.5	4	2.9	0	0.0
Interest, dividends, profit from capital investments in incorporated business (HY090N)	1140	28.6	944	82.8	59	5.2
Interest repayments on mortgage (HY100N)	324	8.1	89	27.5	0	0.0
Income received by people aged under 16 (HY110N)	64	1.6	8	12.5	3	4.7
Regular taxes on wealth (HY120N)	2567	64.4	59	2.3	0	0.0
Regular inter-household cash transfers paid (HY130N)	189	4.7	1	0.5	0	0.0
Tax on income and social contributions, net (HY140N)	0	0.0	0	.	0	.
Repayments/ receipts for tax adjustment (HY145N)	1379	34.6	92	6.7	18	1.3
<i>Gross income components at household level</i>						
Imputed rent (HY030G)	3849	96.6	3849	100.0	0	0.0
Income from rental of a property or land (HY040G)	64	1.6	1	1.6	0	0.0
Family/ children related allowances (HY050G)	1627	40.8	0	0.0	0	0.0
Social exclusion not elsewhere classified (HY060G)	28	0.7	0	0.0	0	0.0
Housing allowances (HY070G)	98	2.5	7	7.1	0	0.0
Regular inter-household cash transfers received (HY080G)	140	3.5	4	2.9	0	0.0
Interest, dividends, profit from capital	1140	28.6	944	82.8	59	5.2

investments in incorporated business (HY090G)						
Interest repayments on mortgage (HY100G)	324	8.1	89	27.5	0	0.0
Income received by people aged under 16 (HY110G)	64	1.6	8	12.5	3	4.7
Regular taxes on wealth (HY120G)	2567	64.4	59	2.3	0	0.0
Regular inter-household cash transfers paid (HY130G)	189	4.7	1	0.5	0	0.0
Tax on income and social contributions, gross (HY140G)	2939	73.8	2939	100.0	0	0.0

Table 2.4. 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	%
<i>Net income components at personal level</i>						
Employee cash or near cash income (PY010N)	5042	54.3	282	5.6	19	0.4
Non-cash employee income (PY020N)	170	1.8	170	100.0	0	0.0
Contributions to individual private pension plans (PY035N)	487	5.2	57	11.7	5	1.0
Cash benefits or losses from self employment (PY050N)	722	7.8	104	14.4	9	1.2
Pension from individual private plans (PY080N)	4	0.0	0	0.0	0	0.0
Unemployment benefits (PY090N)	133	1.4	1	0.8	6	4.5
Old-age benefits (PY100N)	2384	25.7	7	0.3	1	0.0
Survivors' benefits (PY110N)	88	0.9	2	2.3	0	0.0
Sickness benefits (PY120N)	604	6.5	117	19.4	0	0.0
Disability benefits (PY130N)	605	6.5	5	0.8	0	0.0
Education-related benefits (PY140N)	185	2.0	4	2.2	0	0.0
<i>Gross income components at personal level</i>						
Employee cash or near cash income (PY010G)	5042	54.3	282	5.6	19	0.4
Non-cash employee income (PY020G)	170	1.8	170	100.0	0	0.0
Employer's social insurance contributions (PY030G)	4855	52.2	4855	100.0	0	0.0
Contributions to individual private pension plans (PY035G)	487	5.2	57	11.7	5	1.0
Cash benefits or losses from self employment (PY050G)	726	7.8	109	15.0	4	0.6
Pension from individual private plans (PY080G)	4	0.0	0	0.0	0	0.0
Unemployment benefits (PY090G)	133	1.4	1	0.8	6	4.5
Old-age benefits (PY100G)	2384	25.7	7	0.3	1	0.0
Survivors' benefits (PY110G)	88	0.9	2	2.3	0	0.0

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	%
Sickness benefits (PY120G)	604	6.5	117	19.4	0	0.0
Disability benefits (PY130G)	605	6.5	5	0.8	0	0.0
Education-related benefits (PY140G)	185	2.0	4	2.2	0	0.0

2.4. Mode of data collection

Table 2.5. Distribution of household members aged 16 and over by Data (RB250), 2006
HOUSEHOLD MEMBERS 16+ (RB245= 1 to 3)

	Total	RB250=11	=12	=13	=21	=22	=23	=31	=32	=33
Total	9398	9292	0	0	7	0	69	21	9	9398
%	100	98.9	0.0	0.0	0.1	0.0	0.7	0.2	0.1	100

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

	Total	RB250=11	=12	=13	=21	=22	=23	=31	=32	=33
Total	9072	8973	0	0	7	0	64	20	8	9072
%	100	98.9	0.0	0.0	0.1	0.0	0.7	0.2	0.1	100

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

	Total	RB250=11	=12	=13	=21	=22	=23	=31	=32	=33
Total	326	319	0	0	0	0	5	1	1	326
%	100	97.9	0.0	0.0	0.0	0.0	1.5	0.3	0.3	100

Table 2.6. Distribution of household members aged 16 and over by Type of Interview (RB260), 2006
HOUSEHOLD MEMBERS 16+ (RB245= 1 to 3) and RB250= 11 or 13

	Total	RB260=1	=2	=3	=4	=5	Missing
Total	9292	149	8537	18	5	568	15
%	100	1.6	91.9	0.2	0.1	6.1	0.2

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

	Total	RB260=1	=2	=3	=4	=5	Missing
Total	8973	136	8260	18	5	546	8
%	100	1.5	92.1	0.2	0.1	6.1	0.1

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

	Total	RB260=1	=2	=3	=4	=5	Missing
Total	319	13	277	0	0	22	7
%	100	4.1	86.8	0.0	0.0	6.9	2.2

2.5. Imputation procedure

As 2006 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 2005 was used only if household or person received particular kind of income in 2005 and analysis showed that these two incomes are sufficiently closely related. If analysis indicated no correlation between the incomes of 2005 and 2006, values were not used in imputation. For some variables, values of previous year were corrected to take into account trends present in the data. Details on the number of values forwarded from 2005 to 2006 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;
- Hot-deck imputation (random donor), when proportion of missing values was very small;
- 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 2006, i.e. rotational group 6 is included in calculations.

Table 2.7. Percentage of imputed cases by income component in national questionnaire, 2006

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 annual wages	7162	98	261 (hot-deck in interval) 90 (IveWare)	Amount per month, log-scale
		H07B	Total amount of additional payments that had not been taken into consideration in net wages	1150	3	45 (IveWare)	
PY020N	Non-cash employee income	H18	Possibility to use company car	7162	0	8 (logical deduction)	
		H20	Number of months a company car was used	233	0	2 (logical deduction)	
PY035N	Contributions to individual private pension plans	HK1	Joining the 3rd pillar of pension insurance	13406	2	0	
		HK2	Payments made into the 3rd pillar of pension insurance	615	8	76 (IveWare)	
		HK4	Joining any other collecting insurance scheme	13406	0	0	
		HK5	Payments made to the collecting insurance	145	0	9 (hot-deck)	
PY050N	Cash benefits or losses from self employment	H27N	Amount of loss from self-employment	142	1	24 (IveWare)	Log-scale
		H28N	Net profit from self-employment	387	1	113 (hot-deck in interval) 17 (IveWare)	Log-scale
		H35N	Net amount of royalties, remuneration or payment under contract for creative or scientific work	42	0	4 (IveWare)	Amount per month
		H46A2	Income from private provision of fee-charging services to other persons or households	323	3	9 (hot-deck)	

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
		H46B2	Income from the sale of own-produced consumer goods (e.g. handicrafts, souvenirs, etc.)	26	0	1 (hot-deck)	
		H46C2	Income from the sale of own-produced foodstuffs (e.g. pies, waffles, shashlik, etc.)	6	0	0	
		H46D2	Income from intermediate commercial transactions	17	0	2 (hot-deck)	
		H46E2	Income from agricultural or forestry activities	204	2	6 (hot-deck)	
		H46F2	Income from other unregistered self-employment	3	0	0	
PY080N	Pension from individual private plans	HK3	Whether received any payments from the 3rd pillar of pension insurance	13406	1	0	
		HK3A	Payments received from the 3rd pillar of pension insurance	4	0	1 (average)	
		HK6	Whether received payments from collecting insurance in the previous calendar year	13406	0	0	
		HK6A	Payments from collecting insurance in the previous calendar year	2	0	0	
PY090N	Unemployment benefits	H55A	Amount of unemployment benefit or any other benefits relating to unemployment	109	0	2 (median)	
		HF2A	Amount of redundancy benefit used up in the previous calendar year	67	0	5 (logical deduction)	

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
		HF6A	Net amount of unemployment benefits paid by Estonian Unemployment Insurance Fund	64	0	4 (IveWare)	
PY100N	Old-age benefits	H51A	Amount of old-age benefits	3246	1	10 (IveWare)	Amount per month, log-scale
		H52A	Amount of pension for incapacity for work or any other benefits relating to disability	1102	1	7 (IveWare)	Amount per month, log-scale
		HF5A	Amount spent last year of lump-sum benefits upon retirement from employer	6	0	0	
PY110N	Survivors' benefits	H53A	Amount of survivors' pension or any other benefits relating to the loss of a provider	129	2	0	
PY120N	Sickness benefits	H54A	Amount of sickness benefits or any other benefits relating to health	861	24	143 (IveWare)	Amount per day
PY130N	Disability benefits	H52A	Amount of pension for incapacity for work or any other benefits relating to disability	1102	1	7 (IveWare)	Amount per month, log-scale
PY140N	Education-related benefits	H57A2	Amount of state stipend	16	0	2 (median)	
		H57G2	Education allowance	279	0	6 (median by age and education)	
NET INCOME COMPONENTS ON HOUSEHOLD LEVEL							
HY040N	Income from rental of a property or land	D09N	Net income from rental of property	93	1	1 (median)	
HY050N	Family/ children related allowances	D11B	Parental benefit received in the previous calendar year	2376	0	0	
		D11C	Total amount of other benefits received in the previous calendar year	2376	0	0	

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
		D11D	Amount of pregnancy and maternity leave received in the previous calendar year	2376	0	0	
		D11E	Adoption allowance received in the previous calendar year	2376	0	0	
HY060N	Social inclusion not elsewhere classified	H58A2	Amount of other support/benefit/pension not mentioned above	47	0	0	
HY070N	Housing allowances	D03A	Amount of subsistence benefit	129	5	4 (median by hh size)	
HY080N	Regular inter-household cash transfers received	D16A	Amount of regular payments from other households	196	0	5 (IveWare)	
HY090N	Interest, dividends, profit from capital investments in incorporated business	H49A1	Interest income from deposits in a foreign bank	13406	0	1 (logical deduction)	
		H49A2	Interest income from deposits in a foreign bank - amount	26	0	8 (hot-deck)	
		H49B2	Interest income from securities (shares, bonds)	30	0	9 (hot-deck)	
		H49C2	Dividend income from securities (i.e. shares, bonds)	57	0	9 (hot-deck)	
		HE1	Whether received interest payments from demand deposit or time deposit account from a bank situated in Estonia	13406	0	1 (logical deduction)	
		HE1A	Interest payments received from demand deposit or time deposit account in the previous calendar year	2576	0	2214 (hot-deck in interval) 81 (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
HY100N		D08A_A	Mortgage interests paid in the previous calendar year	511	130	When interest is not reported, details about mortgage are requested and interest is deducted analytically	
HY110N	Income received by people aged under 16	D19A	Income received by children aged 16 or less	0	0	0	
HY120N	Regular taxes on wealth	D10	Tax on land or any other property tax	5679	2	1 (logical deduction)	
		D10A	Amount of tax on land or any other property tax paid	3633	38	51 (IveWare)	Log-scale
HY130N	Regular inter-household cash transfers paid	D14A	Amount of regular payments to other households	276	0	2 (median)	
HY145N	Repayments/receipts for tax adjustment	H64A	Income tax return for the income received in the previous year	118	0	11 (IveWare)	Log-scale
		H63	Additional amount of tax paid on income in the previous calendar year	2377	33	120 (IveWare)	log-scale

2.6. Imputed rent

Imputed rent must be computed for those households that are either the owners of their living space or use it without a rental fee.

The calculation of imputed rent was done in three steps.

1. *Finding the best model of square-meter price of the accommodations.* Square-meter price model was found separately for detached houses (including semidetached and terraced houses) and apartments. Model was fitted on the database of real estate sell-transactions in Estonia in 2006, containing 4682 transaction records.
Square-meter price of a detached house was modelled, using the following explanatory variables: county group (4 groups), level of urbanisation (capital city, big towns, small towns, rural areas), quality of accommodations (poor, satisfactory, good), area, interaction of county group and urbanisation level, interaction of quality and urbanisation level.
Square-meter price of an apartment was modelled, using the following explanatory variables: county group, level of urbanisation, quality of accommodations, number of rooms (1,2,3,4+), area, interaction of county group and urbanisation level.
2. *Applying the model to EU-SILC data.* The square-meter price models were applied to all households. The distribution of the estimated square-meter price is characterized in the following table:

Table 2.8. Characteristics of the square-meter price (in thousand kroons)

<i>Square-meter price of a detached house</i>		
Mean and standard error	12,06	0,01
Median	10,52	
Standard deviation	5,96	
Minimum	1,00	
Maximum	27,73	
<i>Square-meter price of an apartment</i>		
Mean and standard error	17,5	0,01
Median	18,05	
Standard deviation	7,57	
Minimum	1,36	
Maximum	37,69	

3. *Computation of imputed rent, based on the total price of accommodations.* The calculation formula for imputed rent is the following:

$$(\text{annual rent}) = (\text{price of the accommodations}) \times (\text{smoothed Euribor})$$

Price of the accommodations is multiplied by the mean value of Euribor from 01.01.1999 – 31.12.2006, which is equal to 3,239 (per cent). In the future we intend to use the mean value over ten years; at this point it is not possible, since earlier values are unavailable.

Applying this formula on all the households in EU-SILC, we obtained the following mean rental estimations:

Table 2.9. Mean estimated rent (in thousand kroons)

	Price	Annual rent	Monthly rent
Detached house	1067,42	34,574	2,881
Apartment	971,12	31,455	2,621
Total	999,71	32,381	2,698

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.

2.8. Within-household non-response inflation factor

Ca 1.4% of households, those data was accepted for the database, are affected by within household non-response. Ignoring this fact may cause substantial bias in main indicators. To correct the effect of non-responding individuals within a household, the inflation factor (variable HY025) is calculated by which it is necessary to multiply total household gross income, total household disposable income and the total household disposable income before social transfers.

Calculation procedure was as follows.

1. Missing total individual gross income was imputed with the unweighted median of imputation class. Imputation classes were cross-sections of 9 age groups (16-17, 18-19, 20-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75+), sex, employment status (employed/other), and quintile of equivalised total household gross income. Where possible, total individual income of previous wave was used for imputation (if it was greater than zero).
2. Quintile of equivalised total household gross income was recalculated using imputed values of previous step.
3. Steps 1 and 2 were repeated until there were no more shifts in the income quintile (5 repetitions were sufficient in 2006).
4. Within-household non-response inflation factor was calculated as the ratio of total household gross income including imputed individual income to initial household gross income. For households with negative or zero household gross income, inflation factor was set to 1 (i.e. no inflation was done). If resulting value was greater than 5, inflation factor was set to missing.

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 2005.

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

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

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

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

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

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

3.3. Tracing rules

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

4. COHERENCE

This section will compare the longitudinal EU-SILC data to various external sources, including the National Accounts (NA), the Household Budget Survey (HBS), the Labour Force Survey (LFS), 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

The total annual income received from gross wages and salaries was 57,624.6 millions of kroons according to the EU-SILC cross-sectional component in 2005. (In Estonia, the income reference year is one year behind the survey year so all comparisons for 2006 data will be done with 2005 incomes) According to national register accounts, the same figure was somewhat lower at 50 766 million kroons (gross) in 2005. Our analyses of the comparability of register and survey data show that this is a recurring tendency. In 2003, the average gross income was reported to be 20,7% higher in EU-SILC than the records of the Estonian tax and customs board indicate. In 2004, EU-SILC's amount was 9,3% higher than register data. This can partly be due to unreported incomes and also from PY010 containing a wider set of income components than the definition from the Estonian tax and customs board. The latter calculates its amount from all incomes taxable by the social insurance tax, including pensions and scholarships paid in relation to an employment connection and other remunerations.

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

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

Income component	EU-SILC	Administrative sources*
Cash or near-cash employee income (PY010G)	57.624.600.000	47.486.489.539
Non-cash employee income (PY020G)	578.000.000	-
Unemployment benefits (PY090G)	148.000.000	173.280.000
Old-age benefits (PY100G)	9.460.000.000	9.142.230.000
Disability benefits (PY130G)	1.129.343.155	1.773.970.000

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

When looking at PY010G the difference with wage statistics is to be expected, given that the latter refer to the full-time equivalent and exclude part-time work contracts, and the unofficial work relationships are not included. In wage statistics, PY020G is included in PY010G and could not be

separated from it for individual analysis. The data concerning wage statistics comes from in-house sources, not administrative registers.

In the case of the unemployment benefits, the definitions that can be used differ between the sources and thus only the numbers can be only tentatively compared. The definition in administrative sources for the unemployed is more restrictive, but the number of people receiving unemployment benefits in EU-SILC is still substantially lower, suggesting under coverage in the survey.

Variable PY100G in EU-SILC also includes pension benefits received from abroad, which tend to be higher than national benefits. 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. The EU-SILC survey also includes other old-age benefits that are not taken into account in the national administrative sources (such as local benefits provided by the local government to pensioners residing in their municipality). This is reflected in the somewhat higher reporting of old-age pensions in the survey than in national sources. 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.

Table 4.2. Number of recipients of income components by source of information, income year 2005

Income component	EU-SILC	Other sources*
Cash or near-cash employee income (PY010G)	653,971	496,277
Non-cash employee income (PY020G)	26,099	
Unemployment benefits (PY090G)	16,004	31 347
Old-age benefits (PY100G)	288,343	296,082
Survivor's benefits (PY110G)	7,941	9,312
Disability benefits (PY130G)	65,049	

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

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.

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. The differences are minor, up to 2% 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.

Table 4.3. Distribution of population aged 16-74 by ISCED level, based on the cross-sectional EU-SILC and the LFS, 2006

ISCED level	EU-SILC	LFS
0 Pre-primary education	0.6	0.5
1 Primary education	4.8	4.1
2 Lower secondary education	18.7	18.2
3 (Upper) secondary education	43.0	44.5
4 Post-secondary non tertiary education	7.6	5.5
5 First stage of tertiary education	24.9	26.8
6 Second stage of tertiary education	0.3	0.4
Total	100.0	100.0

Finally, Table 4.4 presents the comparison of population aged 16 or over by most frequent 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 4.4. Distribution of population aged 16 and over by self-defined activity status based on longitudinal EU-SILC and the HBS, 2005

Activity status	EU-SILC	HBS
Working full-time	52.9	50.6
Working part-time	3.6	3,3
Unemployed	4.3	3.9
Pupil, student	8.8	10.7
In retirement	21.8	22.0
Permanently disabled	3.7	4.2
Fulfilling domestic tasks and care responsibilities	4.7	1.5
Other inactive	...*	3.6
Total	100.0	100.0

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