Statistics Finland

INTERMEDIATE QUALITY REPORT

RELATING TO

THE EU-SILC 2005 OPERATION

IN FINLAND

Pauli Ollila, Marie Reijo, Hannele Sauli, 30 December 2006

1 Common cross-sectional European Union indicators

1.1 Common cross-sectional EU indicators based on the cross-sectional component of EU-SILC

Primary Laeken indicators of social cohesion	
At-risk-of-poverty rate after social transfers	
At-risk-of-poverty threshold	
- 1-person household	10 447.20
- 2 adults, 2 dependent children	21 939.12
At-risk-of-poverty rate by age and gender	
- Total	11.71
- Males, total	10.61
- Females, total	12.76
- Total, 0-15	9.93
- Total, 16-24	21.51
- Males, 16-24	19.96
- Females, 16-24	23.12
- Total, 25-49	8.28
- Males, 25-49	8.65
- Females, 25-49	7.91
- Total, 50-64	8.52
- Males, 50-64	9.11
- Females, 50-64	7.94
- Total, 65+	18.47
- Males, 65+	11.02
- Females, 65+	23.49
- Total, 16+	12.12
- Males, 16+	10.79
- Females, 16+	13.37
- Total, 16-64	10.61
- Males, 16-64	10.74
- Females, 16-64	10.47
- Total, 0-64	10.45
- Males, 0-64	10.54
- Females, 0-64	10.36
At-risk-of-poverty rate by most frequent activity status and by gender and selected age group	
- Total, 16+	11.98
- Males, 16+	10.56
- Females, 16+	13.31
- Of which: ‘At work’, total, 16+	3.83
- Of which: ‘At work’, males, 16+	3.91
- Of which: ‘At work’, females, 16+	3.74
- Of which: ‘Not at work’, total, 16+	21.13
- Of which: ‘Not at work’, males, 16+	19.51
- Of which: ‘Not at work’, females, 16+	22.39
- Of which: ‘Unemployed’, total, 16+	35.47
- Of which: ‘Unemployed’, males, 16+	39.45
- Of which: ‘Unemployed’, females, 16+	30.57
- Of which: ‘Retired’, total, 16+	16.69
- Of which: ‘Retired’, males, 16+	11.40
- Of which: ‘Retired’, females, 16+	20.69
- Of which: ‘Other inactive’, total, 16+	23.42
- Of which: ‘Other inactive’, males, 16+	24.32
- Of which: ‘Other inactive’, females, 16+	22.82
At-risk-of-poverty rate by household type	
- All households, no dependent children, total	14.39
- All households, no dependent children, 1-person household, total	29.87
- All households, no dependent children, 1-person household, males	27.02

- All households, no dependent children,1-person household, females	31.94
- All households, no dependent children,1-person household, aged < 65 years	26.63
- All households, no dependent children,1-person household, aged 65+	36.15
- All households, 2 adults, no dependent children, both aged < 65 years	6.24
- All households, 2 adults, no dependent children, at least one aged 65+	7.87
- All households, other households, no dependent children	3.33
- All households with dependent children, total	8.77
- All households with dependent children, single parent, at least 1 dependent child	19.12
- All households with dependent children, 2 adults, 1 dependent child	6.97
- All households with dependent children, 2 adults, 2 dependent children	5.32
- All households with dependent children, 2 adults, 3+ dependent children	11.88
- All households with dependent children, other households with dependent children	5.80
At-risk-of-poverty rate by accommodation tenure status	
- Total	11.71
- Owner or rent-free, total	8.11
- Tenant, total	21.31
At-risk-of-poverty rate by work intensity of the household	
- All households without dependent children, WI=0	27.26
- All households without dependent children, 0<WI<1	10.92
- All households without dependent children, WI=1	2.89
- All households with dependent children, WI=0	56.04
- All households with dependent children, 0<WI<0.5	26.18
- All households with dependent children, 0.5<=WI<1	7.28
- All households with dependent children, WI=1	2.94
Inequality of income: S80/S20 income quintile share ratio	
- Total	3.60
Relative median at-risk-of-poverty gap by gender and selected age group	
- Total	13.53
- Males, total	14.73
- Females, total	12.83
- Total, 0-15	10.32
- Total, 16+	14.43
- Males, 16+	15.98
- Females, 16+	13.53
- Total, 16-64	17.47
- Males, 16-64	18.04
- Females, 16-64	16.55
- Total, 65+	10.27
- Males, 65+	8.87
- Females, 65+	10.81
Secondary indicators	
Dispersion around the at-risk-of-poverty threshold	
- 40% median, total	2.10
- 40% median, males, total	2.38
- 40% median, females, total	1.83
- 50% median, total	4.97
- 50% median, males, total	4.83
- 50% median, females, total	5.10
- 70% median, total	20.45
- 70% median, males, total	18.83
- 70% median, females, total	21.99
At-risk-of-poverty rate before social transfers	
At-risk-of-poverty rate before all social transfers except old-age/survivors' benefits by gender and selected age group	
- Total	28.10
- Males, total	27.22
- Females, total	28.93

- Total, 0-15	31.33
- Total, 16+	27.34
- Males, 16+	26.04
- Females, 16+	28.56
- Total, 16-64	27.38
- Males, 16-64	27.31
- Females, 16-64	27.46
- Total, 65+	27.16
- Males, 65+	19.35
- Females, 65+	32.42
At-risk-of-poverty rate before all social transfers including old-age/survivors' benefits by gender and selected age group	
- Total	40.48
- Males, total	38.04
- Females, total	42.82
- Total, 0-15	31.73
- Total, 16+	42.53
- Males, 16+	39.46
- Females, 16+	45.40
- Total, 16-64	30.72
- Males, 16-64	29.99
- Females, 16-64	31.45
- Total, 65+	92.04
- Males, 65+	89.10
- Females, 65+	94.02
Inequality of income distributions: Gini coefficient	
Inequality of income distributions: Gini coefficient, total	25.74

1.2 Other indicators

Other indicators	
The average of equivalised disposable income	
- Total	19 392.96

2 Accuracy

2.1 Sampling Design

The sampling design of the Finnish EU-SILC survey, the collection year 2005, (also parallel with the design of the Finnish Income Distribution Survey [IDS]) is a *two-phase sampling design*. The copy of the population register some weeks before the end of the study year included 4,185,517 non-institutional persons aged 16 years or over. The type of the frame was based on the *domicile code*, i.e. very exact identification of all the possible places where people can live. The first digits of this code include regional information (municipality code). Systematic sampling of persons was carried out from that frame in order to get the basis for a master sample (in 2005 exceptionally 100,000). After various checks and combinations we get 98,248 dwelling units with all their relevant members. The loss of 1,752 persons is due to the difference between the register which the selector of the master sample has and the final population register of the end of the study year. This final information (coming with the tax information to be connected to the master sample in order to create the strata, for example) is available after the master sample has been selected. At this point those who have died, moved permanently abroad or placed into an institution after the time point of the copy of the register and before the end of the year are excluded from the master sample. With this processing we correct the effect of the frame imperfection (not exactly describing the right time) in the sample.

This master sample of dwelling units is used for different sampling purposes, and one of them is the Income Distribution Survey. For that the master sample is stratified by socio-economic criteria, emphasising high-earners, farmers and entrepreneurs in the allocation. The sample size of the first wave is 7,500. The second wave of the IDS (5,869) is included in the set of households to be interviewed. The final definition of the structure of the household is done during the interview. The stratum is identified for these IDS waves separately in the variable DB050.

Referring to the description of the sampling design above it can be observed that
* **the Finnish cross-sectional data 2005 are based on a nationally representative probability sample of the population residing in private households** (non-institutionalised persons, two-phase sampling in both IDS waves),
* **all private households and all persons aged 16 and over within the household are eligible for the operation** (selection of persons, creation of dwelling units around persons and definition of households during the interviews),
* **representative probability samples are achieved both for households, which are the basic units of sampling, data collection and data analysis, and for individual persons in the target population** (selection of persons aged 16 and over from the register, creation of dwelling units around persons and definition of households during the interviews), and
* **the sampling frame and methods of sample selection ensure that every individual and household in the target population is assigned a known and non-zero probability of selection** (for every non-institutionalised person the probability of selection is identified and greater than zero).

2.1.1 Type of Sampling and Sampling Units

The Finnish sampling design includes only sampling of persons, thus there is only **one stage of sampling**. The **stratification** is constructed in the first-phase master sample, **not** in the population. Sampling is conducted **in two phases**: in the first phase **persons** are selected (first phase sampling unit), in the second phase the **target persons together with their dwelling units** are selected (second phase sampling unit). In a sense the second phase contains **clustering** (though constructed around the target person). However, the **sampling unit** can be still considered as a **person** (only he/she answers the personal questions).

2.1.2 Stratification Criteria

The SILC data selection follows parallelly the sampling design of the Income Distribution Survey. The IDS stratification is conducted in the first-phase master sample containing dwelling units. The strata are created by using a socio-economic categorisation based on the register information available for the members at the time of sample selection. The stratification takes the highest earning person as the categorising person, but the entrepreneur need not be the highest earning one to define the household in the class of entrepreneurs. The income class division is used to allocate the sample more to high-earners. The stratification variable is **DB050**, containing values 1-13 for the first IDS wave and 14-26 for the second IDS wave, based on the dwelling units created around the selected persons.

Table 2.1 Stratification Criteria for the IDS

IDS Wave 1 (CY 2005)			**IDS Wave 2 (CY 2004)**		
Socio-economic categorisation of the household	Income Class	Stratum code	Socio-economic categorisation of the target person	Income Class	Stratum code
Wage earners	Lowest	1	Wage earners	Lowest	14
	2nd lowest	2		2nd lowest	15
	3rd lowest	3		3rd lowest	16
	Highest	4		Highest	17
Entrepreneurs	Lower	5	Entrepreneurs	Lower	18
	Higher	6		Higher	19
Farmers	Lower	7	Farmers	Lower	20
	Higher	8		Higher	21
Pensioners	Lower	9	Pensioners	Lower	22
	Higher	10		Higher	23
Others	Lower	11	Others	Lower	24
	Higher	12		Higher	25
No tax information	-	13	No tax information	-	26

2.1.3 Sample Size and Allocation Criteria

The effective sample size and other relevant sample size information of the Finnish EU-SILC sampling design can be found in the following tables.

Table 2.2 Sampling Design Information of the Finnish EU-SILC

Cross-sectional sample 2004	Value	Definition
Minimum effective sample size	**6 750**	For household selection, not the case of Finland
Minimum effective sample size (sample of persons)	**5 063**	Finland uses registers for income and other data; thus a sample of persons (instead of a sample of households) is selected. *Regulation 1177/2003* Article 9 (paragraph 3) states that *"the minimum effective sample size in terms of the number of persons aged 16 or over to be interviewed in detail shall be taken as 75% of the figures shown in columns 3 and 4 of the table in Annex II, for the cross-sectional and longitudinal components respectively"*.
Minimum achieved sample size	**6 329**	The achieved sample size *"depends on the efficiency of the sample design used (i.e. on the 'design effect')"*. The design effect term ($deff^2$) is *"the ratio of variance of a certain statistics) under the actual design, to that variance under a simple random sample of the same size"*. The reference statistic to be used in the design effect calculations is *at-risk-of-poverty-rate at national level (after social transfers)*. This design effect term for Finland based on the calculations from the Finnish Income Distribution Survey 2001, i.e. here $deff^2$ = **1.25**.
Minimum sample to be selected	**8 328**	Taking the nonresponse into account, the sample to be selected must be larger in order to get the minimum achieved sample size. Currently the overall response rate *R* is approximately **0.76** in Finland.
Actual sample	**13 373**	Combined with the structure of the Income Distribution Survey of Finland, the Finnish EU-SILC provides the actual sample to be selected larger than the minimum sample to be selected. This includes 7,500 from the first wave and 5,869 from the second wave of the Income Distribution Survey.
Expected number of respondents	**11 104**	When excluding the nonresponse (24% for the first wave and 8% for the second wave)
Realised number of accepted respondents	**11 229**	This includes **5,797** for the first IDS wave and **5,432** for the second IDS wave. **Thus the requirement of the minimum sample to be selected is reached (11,229 > 8,328)**.

Table 2.3 Information Concerning Stratification

Stratum in the master sample		**Master sample**		**2nd phase sample**		**2nd phase sample excluding over-coverage**		**2nd phase accepted respondents**	
1st w.	2nd w.	1st w.	2nd w.	1st w.	2nd w.	1st w.	2nd w.	1st w.	2nd w.
1	14	19 521	9 848	**999**	**794**	992	**787**	749	737
2	15	17 946	9 087	**900**	**727**	896	**722**	707	682
3	16	15 890	7 965	**849**	**692**	836	**688**	679	655
4	17	7 659	3 731	**750**	**579**	744	**578**	578	544
5	18	3 807	1 887	**699**	**536**	688	**532**	531	503
6	19	1 879	955	**501**	**387**	495	**384**	385	347
7	20	2 976	1 604	**561**	**481**	554	**476**	472	457
8	21	1 749	971	**441**	**401**	438	**399**	390	376
9	22	12 845	6 310	**576**	**418**	537	**397**	420	355
10	23	8 529	4 165	**572**	**435**	554	**420**	449	398
11	24	4 536	2 268	**398**	**258**	394	**257**	276	233
12	25	636	276	**202**	**141**	201	**141**	142	127
13	26	275	138	**52**	**24**	50	**23**	19	18
All		**98 248**	**49 205**	**7 500**	**5 873**	**7 379**	**5 804**	**5 797**	**5 432**

Note that the strata were created only for those who were not dead or otherwise included in the over-coverage. Stratum variable **DB050**: 1-13 first wave, 14-26 second wave (i.e. stratum code + 13). The primary response probabilities for each stratum used before calibration can be calculated from this table by using *"number of respondents in the stratum" / "number of selected observations in the stratum "*.

2.1.4 Sample Selection Schemes

The master sample of persons (1st phase) is selected with **systematic sampling** from the population *sorted by the domicile code*. The SILC/IDS sample of the first wave with dwelling units constructed around the target persons is selected from the **stratified** master sample with **simple random sampling without replacement** within every stratum and using *non-proportional allocation* (see Table 2.3). The IDS second wave respondents from the previous year were selected at that time in the same way.

2.1.5 Sample Distribution over Time

The income reference period is not a moving one; it is the calendar year preceding the survey year for all households and persons. The reference population is defined as the population registered as resident in Finland on 31 December 2004. Household composition is also dated 31 December 2004. The fieldwork period took five months; it started in January and ended in May 2005.

The cross-sectional sample of the EU-SILC consists of two rotational groups. The "old panel" started fieldwork right after the training period in January and ended on 15 March. The "new panel" households were interviewed between February and May 15.

Table 2.4 Distribution of interviews through time in 2005

	Cross-section		Longitudinal panels				
			1st wave	2nd wave			All
	Total number	%	number	2-year number	3-year number	4-year number	number
January	2 683	24	-	928	897	858	2 683
February	3 454	31	288	870	870	872	2 900
March	2 216	20	694	33	51	53	831
April	1 806	16	562	-	-	-	562
May	1 070	10	368	-	-	-	368
Total	11 229	100	1 912	1 831	1 818	1 783	7 344

2.1.6 Renewal of Sample: Rotational Groups

The Finnish cross-sectional SILC data collection year 2005 contains two groups based on the Income Distribution Survey: one is a new rotation group (1st IDS wave) and another is a set of responded households of the IDS of the previous year (2nd IDS wave). In subsequent years the second wave of the IDS serves as the second wave of the SILC as well. Note that the Finnish SILC design is not purely integrative from 2006 on, only the first two SILC waves are included in the cross-sectional SILC data. The third and fourth waves appearing in the collection year 2006 and later are not connected to the Finnish Income Distribution Survey, but they are conducted separately.

2.1.7 Weightings

2.1.7.1 Master Sample

The age criterion of the IDS changed due to the EU-SILC connection for the collection year (CY) 2004. Separately calculated from the master samples CY 2003 (of size **100,000**) and 2004 (of size **50,000**) we got the population figures for the person selection, e.g., where $\pi_{a,person\,k}$ is the **inclusion probability of the selected person *k*** in the master sample. The **inclusion probabilities of the dwelling units** created around the

selected persons in the master sample were $\pi_{ak} = \pi_{a,person\,k} n_{16+,dwelling\ of\ k}$. *Note that in this year and subsequent years concerning the EU-SILC in Finland the principles of weighting at this stage are parallel to the principles which are recommended by Eurostat, i.e. the first phase weight includes the master sample information in full.*

2.1.7.2 Income Distribution Survey Sample

The **inclusion probabilities of two-phase sampling** (the effect of selecting the master sample and the IDS sample) were calculated, at the second phase based on the stratification (13 strata) of the master sample and the allocation used. *Note that the over-coverage is now included.* For those waves we separately calculated the inclusion probabilities $\pi_k^* = \pi_{ak}\pi_{k|s_a}$, where

$$\pi_{ak} = \pi_{a,person\,k} n_{16+,HH\ of\ k} = \frac{n_{s_a} n_{16+,HH\ of\ k}}{N}$$

and $\pi_{k|s_a} = n_h / N_{h,s_a}$ is the conditional inclusion probability at the second phase taking the stratification of the master sample into account. The Finnish SILC D file has the design weight variable **DB080** (the inverse of the inclusion probability), in which the original design weights were calculated *separately for the two IDS waves* and with a multiplication by 0.5 in order to get coherent information about the households.

PB070 (*personal design weight for selected respondents*) is an estimate of the **inverse of the inclusion probability of the person** (**DB080***$n_{16+,HH}$). *This weight was not needed in the weighting procedure of the IDS*. Again in this case these weights were calculated *separately for both waves* (resulting in about double number of 16+ population). In addition, the calculation was conducted for *all of the sample* (excluding over-coverage). However, the weight **PB070** is defined only for the households that have been accepted (P file), not all the sample (including non-response). In this case there should be a non-response correction included in the weight in order to get the figures right. We did *the simple adjustment* n_{sample} / $n_{respondents}$ *in every stratum. In addition,* to get the separate wave effect to disappear, *we multiplied the weights by 0.5.* The sum of the weights is N_{16+}.

As the basis of calibration **the unit non-response was corrected** by $n_{sample,h}$ / $n_{respondents,h}$ *in every stratum h* (interpreted as the inverse of the response probability in every stratum). The sum of these corrected weights calculated separately in the data of accepted 16+ persons in the HHs coincides with N_{16+}. The data set of accepted target persons is the same as in the P file (1st IDS wave **5,797**, 2nd IDS wave **5,432** --> **11,229** observations).

These weights containing a simple correction were used in **calibration (the raking method)** conducted with the macro CALMAR (applicable in SAS) for the accepted households. The calibration process was carried out *separately for both waves*. The calibration could be interpreted as **integrative**, i.e. both the household and the person levels were included in the process. The percentual marginal distributions and the statistics used in calibration are the following:

1) Households: province; type of municipality; HH size; sums of 15 different income variables. *The first three distributions of the households were obtained from the master sample, using weights for which a primary calibration (population register: 16+ persons and persons under 16/15 by region; gender*age class) was conducted. The income information comes from different registers.*

2) Persons: gender and age classes (0-4, 5-9, … , 80-84, 85+)

Table 2.5 Description of the Calibration Variables

Variable name	Description
Alue	Region (NUTS 3 level), Capital region separated
Ask8	Size of dwelling unit
Haastkur	Degree of urbanisation

Mibs01-Mibs18	Men 0-4, 5-9, 10-14, … , 80-84, 85-
Nibs01-Nibs18	Women 0-4, 5-9, 10-14, … , 80-84, 85-
Trplopti	Income 1: Cash or near cash employee income
Saipalk	Income 2: Income 1 > 0
Lelake	Income 3: Pensions
Tyotts	Income 4: Unemployment benefits 1
Perustur	Income 5: Unemployment benefits 2
Saityott	Income 6: Income 4 > 0
Elintul3	Income 7: Income from self-employment
Yhtytulo	Income 8: Capital income 1
Maattulo	Income 9: Income from agriculture
Omaitul2	Income 10: Income from property and forestry 1
Muupaao2	Income 11: Other capital income
Metstulo	Income 12: Income from forestry 2
Myvo	Income 13: Capital gains
Saielake	Income 14: Pensions > 0
Askorot	Mortgage interests

In addition, **2,415,000** was used as the **fixed number of households** in the process. The result of this calibration was the weight that produced exactly these margins when used in the summation of these variables in the data set containing accepted observations. **DB090** is this calibrated weight multiplied by 0.5 in order to adjust the effect of separate calculations.

When **DB090** is connected to the R file ("All persons currently living in households or temporarily absent"), these weights (in this context **RB050**) give the sum which coincides with the exact number of non-institutionalised persons at the end of 2004, i.e. **5,160,767**. Furthermore, when **DB090** is linked to the P file ("All eligible persons for whom the information could be completed"), these weights (in this context **PB040**) give the sum which equals the number of households defined (**2,415,000**). These operations are in line with the document *"Description of the Target Variables"*, page 38: *"We have **DB090** = **RB050** = **PB040**"*.

Finally, the personal cross-sectional weight for the selected respondent, i.e. **PB060** is **DB090** multiplied by $n_{16+,HH}$. The number of 16+ is fixed in this phase as well.

An additional weight for children aged 0 to 12, i.e. **RL070** (*Children cross-sectional weight for child care*) is calculated by multiplying **RB050** with the term *" number of non-institutionalised children in age class X from the register" / "number of children in age class X estimated with RB050"*, where *X* = 0 to 12.

2.1.8 Substitutions

Substitution: *refers to replacement of the original units selected in the sample, which do not supply the required information, either because the address cannot be located or is inaccessible, or because the household refuses to co-operate, the entire household is temporarily away, or the household is unable to respond, by other units.*

The Finnish IDS and SILC data contain **no substitutions**.

2.2 Sampling Errors

The Framework Regulation 1177/2003 states that

"The precision requirements concerning publication of the data collected in EU-SILC shall be expressed in terms of the number of sample observations on which the statistic is based and the level of item non-response (additional to total non-response at unit level). The Commission shall not publish an estimate if it is based on fewer than 20 sample observations, or if non-response for the item concerned exceeds 50%. The data shall be published by the Commission with a flag if the estimate is based on 20-49 sample observations, or if non-response for the item concerned exceeds 20% and is lower than or equal to 50%. The data shall be published by the Commission in the normal way when based on 50 or more sample observations and the item non-response does not exceed 20%.
All data publications shall include technical information for each Member State on the effective sample size as well as a general indication of standard error of at least the main estimates."

That is, the MSs have to calculate the **effective sample size** and the **standard errors of at least the main estimates**, which are defined as follows:

Effective sample size
The effective sample size used in the construction of each common cross-sectional EU indicator based on the cross-sectional component of EU-SILC, for the equivalised disposable income and for the unadjusted gender pay gap, will be provided.
Standard errors
The standard errors for the common cross-sectional EU indicators based on the cross-sectional component of EU-SILC, for the equivalised disposable income and for the unadjusted gender pay gap, will be provided.

Table 2.6 Effective Sample Sizes, Item Non-responses and Standard Errors of the Main Estimators

Estimator	Accepted observations in general	Item non-response	Effective sample size	Standard error
Equivalised disposable income	29 112	0	29 112	57.78
At-risk-of-poverty rate after social transfers	29 112	0	29 112	0.338
Inequality of income distribution S80/S20 income quintile share ratio	29 112	0	29 112	0.055
Relative median at-risk-of-poverty gap	29 112	0	29 112	0.601
Dispersion around the risk-of-poverty threshold	29 112	0	29 112	0.136
At-risk-of-poverty rate before social transfers except old-age and survivors' benefits	29 112	0	29 112	0.361
At-risk-of-poverty rate before transfers including old-age and survivors' benefits	29 112	0	29 112	0.375
Inequality of income distribution: Gini coefficient	29 112	0	29 112	0.369

The *"gender pay gap"* comes from another source, not utilising the SILC data. Note that this table contains the calculations in general; when these indicators are classified with some variables (e.g. *main activity status* and *work intensity*), some item non-response may appear due to the classification variables.

The sampling design of the Finnish EU-SILC and the Finnish Income Distribution Survey is a two-phase design, with simple random sampling without replacement (1st phase) and stratified simple random sampling with unequal allocation emphasising some groups (2nd phase). The standard error calculations are conducted with the bootstrap method (10,000 replications). The idea is to estimate the standard error of the second phase by separately carrying out simple random sampling with replacement in every stratum with the original sample size of the stratum. *Unlike with the 2004 data, now the calibration has been conducted in every replication, and the weights are an outcome of this process.* The variance to be used is simply the variance of the bootstrap estimator. In addition, in order to take the non-negligible sampling fraction into account the variance is multiplied by the finite population correction at the whole sample level, i.e.

approximately 0.77. The standard error is the square root of the variance. The standard error of the equivalised disposable income is calculated with the software CLAN.

The variance estimation process includes some aspects of uncertainty. The non-response effect is not taken into account in variance estimation. The with-replacement nature of selection differs from the original selection, and the use of the finite population correction at the general level does not take the non-proportional allocation into account.

2.3 Non-sampling Errors

2.3.1 Sampling Frame and Coverage Errors

The target population is the set of elements about which information is wanted and parameter estimates required. The Commission Regulation on sampling and tracing rules states that "*The target population of EU-SILC shall be all private households and their current members residing in the territory of the Member State at the time of data collection. Persons living in collective households and in institutions are generally excluded from the target population. Small parts of the national territory amounting to no more than 2% of the national population and the national territories listed in the Regulation may be excluded from EU-SILC, after agreement between the Member States concerned and the Commission (Eurostat).*" There is no register of households in Finland, so the selection is based on the population register and the creation of the households begins with the dwelling unit information available in the register.

2.3.1.1 Description of the Sampling Frame

The sample is drawn from the Population Information System maintained by the Population Register Centre of Finland. The register is a continuously updated population register based on domicile. It is updated daily with information on population changes: births, deaths, migration, immigration and emigration, marriages, divorces, adoptions and changes of names. The Population Information System is a compilation of local registers kept up by population register districts.

The Population Information System (PIS) includes information on Finnish citizens and aliens permanently resident in Finland. It includes persons living in private households, institutions, persons living temporarily abroad, and also homeless persons. Persons living in institutions, collective households or residential homes do not belong to the target population, but they are included in the PIS household population and have to be excluded from the master sample (see below).

Every person residing in Finland has a unique identification code and each dwelling has a domicile code. Each person is registered in the municipality where he/she has a permanent place of residence. The domicile code is the link between a person and his/her permanent dwelling. Persons without an address are registered in municipal registers as homeless persons. A person with a permanent address may also have registered a temporary address. The linkage between identification and domicile codes enables the pre-entry into the IDS-SILC questionnaire of all persons permanently registered in the dwelling unit-households before the interviewer contacts the household.

The copy of the population register some weeks before the end of the study year was the **sampling frame** for the selection of the new Income Distribution Survey (IDS) sample. After the separation of the persons placed in institutions and the homeless (a specific code identifies both cases), this frame included 4,185,517 persons aged 16 years or over. The sort of the frame was based on the domicile code, i.e. a very exact identification of all the possible places where persons can live. This code includes regional information at the beginning (municipality code). That frame is used for the **construction of the dwelling units for the master sample** as well. After various checks and combinations (e.g. excluding collective households, e.g. members of the same hall of residence as the target person) we get the dwelling units with all their relevant members for the selected master sample. Before the fieldwork begins the information of the second panel of the IDS and the changes after the selection of the sample are updated based on the register of the end of the year (then already available).

2.3.1.2 Information about the Frame: Reference Period, Updating Actions, Quality Review Actions

In general, the Population Information System of the Population Register Centre can be considered exhaustive and up-to-date as regards persons. Updating activities occur constantly. The system is maintained by notifications of changes. Maternity hospitals immediately report new-born children to local register offices. Deaths have to be reported at once either to a physician or to the police. They have to report the death to the Population Information System. The inhabitants are themselves responsible only of notification of changes of residence. Those who move or immigrate are expected to report to the local register office of the new place of residence on the change of address within one week of the move, specifying all the members of the family or household involved in the move.

Those emigrating should supply a notice of change of address in the country of entry. According to an agreement between the Nordic countries - which are the main destinations of migrants – the population register authorities of the country of entry inform the population register authorities of the country of exit. In the years when municipal elections are arranged (every 4th year), the population is corrected by around 1,000 persons, when emigrants whose emigration has been left unnoticed return notifications of voting.

A reliability survey on the Population Information System is conducted yearly by means of a sample survey of 10,000 persons. From the EU-SILC point of view, reliability of its address information is of special relevance. In the 2004 survey, the address was correct for 99.2 per cent of the respondents. The non-response rate of the reliability survey was 8.8 per cent. The addresses of the non-respondents were checked from other sources and found correct among 78.3 per cent, incorrect among 6.1 per cent and non-verifiable among 15.6 per cent. Assuming that all the unverifiable addresses were incorrect the final proportion of the correct addresses was 97.5 per cent.

The Population Information System has no under-coverage in any population groups. Asylum seekers and refugees are not included in the resident population until their permit of residence has been processed. The small over-coverage is a consequence of the necessity to draw the sample in good time before the actual date of defining the sample households (31 Dec.) and may also be related to register updates - delays in the notifications of emigration, moving to reside permanently in institutions or deaths.

The presence of the members of the households are checked in the interview. Persons who recently changed place of residence and/or household, new-borns, recently moved to institutions or died are the usual sources of non-correct register-based pre-entries in the IDS-SILC questionnaire (1.4% of the sample).

2.3.2 Measurement and Processing Errors

First of all, since income information is almost exclusively linked from the registers, this Section 2.3.2 mostly concerns with collection and processing of the interviewed data. The interviews were carried out by CATI or CAPI. The quality of register data will not be reported here, only a short description of the register data processing is included in this report.

Table 2.7 Mode of data collection (%), SILC 2005

	Cross-section	Longitudinal				
	total	all	Start year / panel duration			
			2005 / R4	2004 / R2	2004 / R3	2004 / R4
CATI	97	97	94	94	98	99
CAPI	3	3	6	6	2	1

2.3.2.1 Questionnaire build-up, the testing procedures, interviewer training

Questionnaire build-up has its starting point in the previous year's questionnaire, feedback from the field interviewers and feedback from the data editing process and users. The leading principle in the questionnaire build-up is a gradual integration process of the SILC to the IDS, and to avoid too many changes in the national IDS.

During the process of BLAISE programming, the questionnaire was table-tested by the team responsible for the IDS and EU-SILC. Eight persons were involved. In weekly meetings logical checks, signals, routings, question formulations and other questionnaire refinements were discussed. In the end, a group of professional interviewers checked the questionnaire against their experience. Finally, the tools were tested in the interviewer organisation before they were sent to the field. A major problem with the questionnaire build-up is the testing: a complex routing system, several checks and forced entries make the testing rather difficult.

Cognitive laboratory studies

The fieldwork of 2004 was thoroughly evaluated by Statistics Finland's Survey Laboratory. This was done during and after the actual fieldwork period and the results were exploited in the development of the questionnaire for the next wave. The evaluation consisted of three different methods:

- a table evaluation of the questionnaire by experienced survey experts
- careful and detailed feedback given by some 30 interviewers based on fieldwork experience
- a study of the fieldwork based on behavioural coding of tape-recorded actual interviews.

Some 180 pages of analysis were produced. The results are not easily summarised. Most of the feedback received consisted of detailed observations on the understandability of question formulations, on the ability of respondents to grasp the question and give adequate answers on the telephone, or questionnaire logic. These documents consist of abundant material that has been only partly exploited.

During the fieldwork period, 20 randomly selected interviewers gave detailed, written feedback about their field experience and the fieldwork tools on a standardised query.

Behavioural coding of the tape-recorded interviews produced some impressive observations about how the interviewers followed the standardised interviewing method and, if they did not follow the rules, why not. Some questions were too long for a telephone interview. The questionnaire did not support probing. The ordering of questions was sometimes illogical and occasionally repetitious. It was evident that the questionnaire did not function well enough, especially in the complex section of yearly activities and current working life.

All this feedback led to a renewal of the questionnaire as already reported below in this section. There is a lot of material that cannot be taken into account at once; gradual development of the questionnaire is still ahead.

Interviewer training

Statistics Finland's interviewer organisation employs about 160 field interviewers on a permanent work contract. They work mostly part-time. They are given basic training on interviewing and questionnaire standards and codes of practices when they start working. They collect most of Statistics Finland's survey data, for the Labour Force Survey, Household Budget Survey, Time Use Survey and Adult Literacy Survey, for example. In other words, they are experienced. Of them 127 were involved with EU-SILC interviews in spring 2005.

The questionnaire changes for 2005 were introduced to the interviewers in a separate written report and, of course, in the instructions book. The instructions book is rewritten every year and it is also under constant development. This year, new charts about the domain order on the questionnaire and a directory were added. The interviewers are paid to get acquainted with the material and practice with it.

The changes were also presented and demonstrated in the interviewers' training courses that took place in January 2005. Before this training, they had training material in the form of the CATI questionnaire and interviewer instructions and they were paid to study the material. One-and-a-half-hour training on the changes of the questionnaire was given to all interviewers in connection with some other training arranged at the same gatherings (one training day for each of five different areas of Finland).

Besides that, newly recruited interviewers were trained separately. They had one day's training about the SILC, which was a shortened version of the training given to all interviewers in 2004. The training programme included a lecture on the planning of the survey, including a description of Eurostat's process, legislation and future uses of the data, and Eurostat guidelines on data protection. Concern over international comparability was underlined. Instructions on the fundamental rules of central data collection were given and discussed, such as the definition of target population, household definition and its implementation in practice, different concepts and classifications of activity, especially labour market activities, child care questions, housing costs and mortgages. A major part of the training time was used on going through the videoed BLAISE questionnaire with the aid of three lecturers. The panel design and the future modules were described. The last part of training consisted of data transferring, data protection and other practicalities.

During the whole fieldwork period, interviewers' information desk is open for them. They can ask for support from the IDS-SILC team. The interviewers, who are distributed all over the country, also have organised district meetings with each other to discuss professional matters.

2.3.2.2 Possible sources of measurement errors

Measurement errors stemming from
- difficulties in understanding complex questions on the telephone,
- difficulties in remembering complex life course events like the year's activities, day care changes, payments of many sorts, and
- difficulties in knowing/reporting another household member's activities

are not systematically surveyed, but the questionnaire was also evaluated in the cognitive laboratory from the above-mentioned points of view.

The potentials for error prevention were used extensively in BLAISE programming. Most relevant instructions were added to the screen with the questions. Routing questions to proper respondents (on the grounds of information on age, sex, activity, etc.) relieves the interviewer from difficult decisions. Advance fillings from the Population Register were used to help household construction. Occupational and housing data already collected last year were made use of in rotational group 2 if there was no change of job or dwelling. Coherence was maintained by introducing logical checks to interconnected questions. Questions presuming numerical answers were given upper and lower limits where possible. Signals were pre-programmed to possible incoherent answers, to possible crossings of given numerical limits or missing answers. The questionnaire was also programmed to accommodate the mode of addressing the respondent depending on whether the selected person him/herself or another member of the household was responding (talking about the respondent: Did you… ; talking about another member: Did N.N. …), which helped the interviewer and respondent to keep track of the member-specific data collection.

Of the many possible sources of measurement errors, the focus in this section is on errors due to *integration problems*, *questionnaire techniques* and *fieldwork problems*. The problems are presented as *possible sources of error*. The exact nature and size of error, if any, can only rarely be detected. In this section, changes made in the 2005 questionnaire are also reported.

Proxy interviews

In the chosen design where only one member of the household is interviewed, all other members of the household are, of course, represented by the respondent. The general mode of collection was a personal interview of a selected respondent or a household respondent. The household respondent was chosen by the interviewer as the one who had the best knowledge of the household's affairs. The interviewers were

instructed to prefer interviewing the selected person whenever possible. In the following, a proxy respondent is defined as the respondent who is not the selected respondent. Of the 11,299 household interviews, 2,668 interviews (24%) were given by someone else than the selected respondent.

Table 2.8 Proxy interviews in the cross-sectional and longitudinal components, SILC 2005

Selected person represented by	Cross-section		Longitudinal component: Start year/panel duration							
			2005 / R4		2004 / R2		2004 / R3		2004 / R4	
	n	%	n	%	n	%	n	%	n	%
Proxy	2 668	23.8	441	23.1	415	22.6	446	24.5	398	22.3
No proxy	8 561	76.2	1 471	76.9	1 416	77.3	1 372	75.5	1 385	77.7
Total	11 229	100.0	1 912	100.0	1 831	100.0	1 818	100.0	1 783	100.0
Proxy=person responsible for the accommodation	2 377	89.1	388	88.0	370	89.2	392	87.9	358	89.9

The interviewers have been instructed to negotiate with the selected person whether he or she is the best person to give information about the household economy and housing and about the other household members, too. According to an estimate of the interviewers, about 85 per cent of their informants are those who have the best knowledge of the household's affairs. In case the selected person is aged less than 18 years, the contact letter is also sent to his/her parents or guardians. The interviewers have been trained to find a household respondent in the earlier years when collecting the IDS data and they have been continuing this procedure. The high percentage of proxy interviews guarantees a higher quality of the household information. Proxies are mostly (89%) 1st or 2nd persons responsible for the accommodation (Table 2.7). As can be seen in Table 2.8, the youngest selected persons under the age of 18 have most often been represented by a proxy respondent. Most of the proxy respondents are parents or spouses.

Table 2.9 Distribution of proxy interviews by their relationship to the selected person and age, SILC 2005.

All selected pers.		Proxy resp.		The proxy's relationship to the selected person					Age of proxy respondent			
Age	Number	n	%	parent	spouse	child	sibling	other	<24	25-44	45-64	65+
-18	297	260	88	85	0	0	1	2	1	36	50	0
18-24	1 224	450	37	33	3	0	1	0	2	7	27	1
25-44	3 513	666	19	2	16	0	0	0	1	14	4	0
45-64	4 429	876	20	0	19	0	0	0	0	2	16	2
65+	1 766	386	22	0	18	3	0	1	0	1	5	16
All	11 229	2 668	23	7	16	1	0	0	0	7	13	3

In 77 per cent of the interviews, the selected respondent was interviewed. Other members are allowed to be consulted during the interview if they are available. This option is often used. Problematic situations occur if the members disagree on participation in the survey.

Fieldwork problems

Mode of data collection (CATI). According to interviewers' estimate, about half of the interviews are conducted through mobile phones and about 6 per cent of them outside home. The respondent can answer while driving a car, sitting in a restaurant, travelling abroad, shopping, or anywhere. Batteries may run down, the signal may be weak, etc. Secret numbers are becoming more and more common. Telephone interviews in general are plagued by a sense of rush. This certainly has an effect on remembering and reporting income items, housing costs, months spent in different activities, etc. In large households, the interview is too long for telephone.

Since 2005, *the interviewers' feedback survey* has been routinely collected from all interviewers at the end of the project. It showed that 29 per cent of them felt that the duration of the interview was too long and half of

those who felt so, also thought that it had an effect on the refusal rate and weakened the quality of responses.

Refusals. The share of sampled households who refuse co-operation with the interviewer slowly rises each year. The interviewers feel that especially the *name of the survey* hampers their success in contacting the households – it is 'income' in the name that is the reason for rejection.

Fieldwork tools. According to the feedback from the interviewers, some of them were not ready for the changes and felt the remodelled questionnaire was even more difficult to manage than the previous year's questionnaire. The interviewers' feedback survey shows that 22 per cent of them felt that the questionnaire did not work well technically and 26 per cent also found it unsatisfactory as to the substance. The survey produced a long list of problems that can be useful in developing the fieldwork tools. Many of the problems have been the same from year to year, even before the SILC was started.

Integration of the questionnaires of the national IDS and the EU-SILC
A major part of the 2005 questionnaire contents was shared with the national IDS and EU-SILC, but there were differences, too. A serious concern in the integration process is to preserve the national time-series without violating demands made to EU-SILC comparability. A stepwise integration strategy aims to achieve full integration in 2007. The questionnaire for the second wave of the EU-SILC operation was reshaped. The motive for the reshaping was, firstly, the feedback from the field interviewers who found the questionnaire repetitive and burdensome[1], and secondly, the test results provided by the Cognitive Laboratory which showed that the first questionnaire did not support sufficiently uniform interviewing techniques. Reshaping concerned more the ordering and routings - and less new formulations of questions.

Labour information in the IDS and EU-SILC
In the IDS, the reference period for the labour information is the income reference year. In the SILC, a lot of labour information refer to the current situation. As a consequence, respondents should tell about all their monthly activities during the IRP, main job and second job during the IRP and the corresponding occupation and NACE, and at the same time the current situation should also be recorded about the same variables. This "overlap" concerns SILC variables such as PL030, PL040, PL050, PL110, PL130, PL140, PL150. Also, SILC variables PL070 - PL090 overlap with the IDS variables but with a different definition: in the IDS overlaps are permitted, in the SILC, one should define one's main activity for each month.

The reference periods for the activities and job-taking in the IDS and EU-SILC are not easily reconciled. The respondent with several job (or activity) changes cannot be burdened with a complicated question structure that would pick up all the required episodes during the IRP and the time of the interview. The first solution was to reduce the number of reference periods. That was achieved in defining "current" to be included in the IRP.

2.10 Examples of labour information with different requirements in the IDS and EU-SILC

Concepts / Variables	Requirements		Solution
	IDS	EU-SILC	Integrated
			Current = December 2003
Main job	Longest period of employment during the year or highest income	Current	If main job is different from current job, both are collected
Second job	The second longest period of employment during the year or second highest	Current	If second job is different from current second job, both are collected

[1] From the 2004 final report: According to the feedback from the interviewers, household formation, the calendar of activities and child care were considered as the most problematic areas of the questionnaire. Especially these questions were described as repetitive and particularly burdensome for large households. Silent moments caused by a difficult questionnaire can be intolerable in a telephone interview. The consequent measurement failures are by their nature incoherence or missing values.

	income		
PL020	---	Current - 4 weeks	December
PL025	---	Current + 2 weeks	December
PL030	---	Current	December
PL040	Status in main job	Current	If main job is different from current job, both are collected
PL050	Occup. in main job	Current	If main job is different from current job, both are collected
PL070, PL072, PL080, PL085, PL087, PL090	Number of months for each activity - 12 categories - overlaps allowed	Number of months for each main activity - no overlaps allowed	Number of months and calendar of activities collected for all members 16+
PL110	NACE in main job	Current	If main job is different from current job, both are collected
PL140	Contract in main job	Current	If main job is different from current job, both are collected
PL180	---	since last interview	constructed through calendar of activities

Reshaping the activities questionnaire: December is the basis for the variable PL030 - self-defined current economic status in 2004 and 2005. In 2004, this variable was collected from the calendar of activities which was constructed for all members of the household aged 16 or older. In 2005, the variable was collected as a starting question in the labour market questions section. The wordings were harmonised with the LFS questionnaire. The calendar of activities was only constructed for the target person. Months of labour and other activities (PL060 - PL090) were integrated with the IDS questions: the IDS questions on different activities during the IRP were changed from a freely filled table to a standardised sequence of questions. Detailed follow-up questions were inserted in the context of each activity. Questions on the main and second job of the year and the current main and second, third,… jobs were routed effectively. All labour questions were first asked about the selected respondent and after that about the other household members. In the data processing phase, working-age persons receiving disability pension (and no other pensions) who had defined themselves as 'retired' were moved to the 'permanently disabled'. This editing was based on register information on pensions.

Reshaping the child care questionnaire: Integration with the IDS and SILC meet again the problem of different reference periods in the child care questionnaire. In the IDS, child care has been studied for all the income reference period - i.e. for all children for the whole year. In 2005, just as in 2004, the reference period for SILC child care questions was defined as December of the IRP. The questions on hours of care relate to normal hours in the form of care that the child was attending in December. In 2005, these questions were re-programmed - not the formulations, but the order of the questions, routings and permitted value ranges were tightened so as to prevent wrong and missing entries.

Reshaping the health questionnaire: in the 2004 experience, it was evident that the questions on the use of health care services did not function in a valid way. Because of the dominance of public services in Finland's health care system, many respondents did not even think about health services that are available in the private sector. In 2005, some questions were added to consider private health care as well.

Testing the changes: The changes were table-tested by members of the IDS-SILC team and five experienced interviewers. Since there are no systematic testing tools available for complex BLAISE programs, the testing method in this phase is simply a set of different households - some representative typicalities, some utterly exotic circumstances - chosen on the basis of long experience with the IDS.

Measurement failures due to questionnaire techniques

Variable-specific problems
Income information or income-related information collected by interviews

Since the income data are mainly collected from registers, the questionnaire covers only those types of income that are not registered at all or on individual level (transfers between households, income from abroad, interest received, grants, non-taxed insurance compensations, strike pay). These variables cover a very small part of the total household income. The nature of questionnaire-collected income data is supplementary, important for some special groups such as single parents or students. They usually form only small fractions of target variables. These income items are collected similarly for the IDS and EU-SILC as to the income definitions, reference periods and units of observation. Problems of the respondent's willingness to answer, perception and remembering constitute the source of measurement errors and, as a consequence, possible under-coverage of non-registered income data.

There is not much information available for comparisons because of the obvious reason that these variables are not registered elsewhere on an individual level. As an exception, however, municipalities report the sum of maintenance allowances paid to single parents if the non-resident parent fails to pay. Of that sum, only 64 per cent were covered by our survey in 2004 (the IRP). This item is the less than 2 per cent subcomponent of target variable HY050G. Similarly, income from interest received collected through interviews covered only 58 per cent of the comparable income total reported in the national accounts.

HY080, HY130 Regular inter-household cash transfers. Regularity is not well emphasised in the questionnaire. The transfers are separated into maintenance allowances, bills, education support, gifts and monetary donations paid or received. Gifts and occasional money donations are excluded from the above-mentioned target variables.

Non-monetary deprivation

HS130 Lowest monthly income to make ends meet. The difficulty of this question for the respondent is well illustrated by the 1,327 (11.8%) cases of item non-response in the cross-section data. Unconvincingly, very low and very high figures were also given. According to the interviewer's code of action, questions of opinion should not be helped in any way, the question can only be repeated. The wording of the question is essential. In the longitudinal data, the number of missing answers varies between 11 and 14 per cent of the cases.

Other variables

PH010 - PH070 Health questions: item response rate is somewhat lower than the overall response rate since the health questions were not allowed to be answered by a proxy respondent. As a consequence, 1.4 per cent of the values are missing due to item non-response (=318 selected respondents were not reached). It has appeared only in 2006 in connection with a cognitive laboratory study that the questions are not formulated according to the regulation. The scale used in PH010 is not comparable, and the formulation of PH030 needs to be corrected (possible only from 2007 onwards).

HB100, PB120 - Household and personal interview duration - measurement failed in both 2004 and 2005.

PE030 Year when the highest level of education was attained - a large number of missing values due to register imperfection. (See Final report 2004 for details.)

PL040 Status in employment, PL050 Occupation, PL140 Type of contract, PL 150 Managerial position: were not asked from all persons who were currently inactive due to a routing error.

2.3.2.3 Processing errors

Description of the process

Fieldwork management and data reception. The interviewers collect the data and transmit them to the central unit. At Statistics Finland, there is a separate organisation, the interviewers' central unit, to control, monitor and supervise the fieldwork. The central unit transmits the fieldwork tools to the field and organises interviewer training at the beginning of the project, follows the fieldwork progress, and receives the output from the field, checks that all the sampled units are adequately processed and transmits the data to the IDS-

SILC team. It also collects feedback from the interviewers with a standardised questionnaire. All data contents processing takes place in the IDS-SILC team, either using the BLAISE system or SAS. Mainly the IDS and SILC data processing is integrated.

Checking and editing of the interview data

The BLAISE programming system already described above (Section 2.3.2.2) is a major data entry controller. However, there is still much processing to be done in the central unit. The checking process starts with the interviewers' remarks saved on the questionnaires. They comment whenever they feel that the coded answer does not reflect the individual real world. All comments are read and the need to edit the data is evaluated. This work starts during the fieldwork period, in 2005 it was begun in mid-February. All comments were processed before the end of May.

After the fieldwork period, the IDS-SILC team looks through incomplete interviews and makes a decision on the achieved sample. Some of the received incomplete interviews are rejected. Since the register income data are nearly perfect, the decision criteria are the sufficiency of the activities and housing information. In the 2005 operation, seven interviews were excluded from the received sample as incomplete (DB120=21). Missing identification codes are found out with the help of the Population Information System and added to the database.

After this phase, checking against the register data is started as soon as the relevant register information is available. Occupation and NACE are processed through automatic coding. Some of the cases will remain open, and they are processed manually.

Activity months, occupation, NACE and housing costs are checked with special intensity. The checks include error lists generated by comparisons of interview and register data. In a register country such as Finland, statistical offices have access to administrative data on an individual level, which makes this data process especially useful. Great differences between different sources of information are processed one by one. All variables, except variables where opinions are expressed, are checked: missing answers, denials and don't knows are checked against other information. Clear mistakes are corrected. Missing values are completed whenever possible (e.g. missing dwelling rents are corrected with average rents per m^2 in the area, other missing housing cost information is completed with supporting information collected on the questionnaire). Illogical answers are straightened if possible. Outliers (considerably small or high values in numerical variables, e.g. inter-household transfers, housing costs) are detected and checked against other information.

Processing inconsistency in the integrated project. The 12 IDS variables on months of activity are heavily edited to comply with register data. That can be done, since in the IDS there are not too many connections between months of activity and other interview variables. As a result, some of the respondents' own answers are rejected and replaced with answers in coherence with their earnings. Corresponding editing is <u>not</u> executed on the SILC variables concerning categories of activity or inactivity, since that would destroy the coherence of the large set of other interviewed variables interconnected with activities. In other words, as a result of different editing, activity information in the IDS and SILC differs from each other. Months of activity (PL070, PL072, PL080, PL085, PL087, PL090) in the EU-SILC are, thus, subjective responses given by respondents, as defined in the EU-SILC document 065/04.

Database construction. Simultaneously with the checking process, a database is opened and variable formation begins. Interview-based and register-based variables construction is started. Most of the interview-based variables are transferred from the questionnaires in their original form. Variables that need constructing are added gradually into the database after all the checks have been made. In 2004, the SILC data files for Eurostat were compiled from the database by SAS after the IDS data were completed.

Processing register data. Register data - that have been subscribed from the register authorities with a special procedure - arrive in electronic form to the Statistics Finland's data processing unit. In 2004, use was made of eleven registers. The incoming data are checked technically and contentually. Possible defects are notified to the authority in charge. They then transmit the corrected data. The registers cover all units - population, dwelling units, income receivers, etc. The data are linked to the sample persons and transmitted into the database of the IDS-SILC. The data are compared with available external data, i.e. those of the tax authority, pensions authority and other statistics. In this phase, the data are in their elementary form. Imputations are made using the hot-deck method (interest income) or modelling (imputed rent). The cross-sectional SILC target variables are constructed only after their elements have been checked.

Comparison of aggregates. Routines have been developed to compare the results on a variable level with external sources such as the Labour Force Survey, National Accounts, wage statistics and statistics on different social transfers and taxation produced by the National Pensions Institute, National Board of Taxes and National Research and Development Centre for Welfare and Health. These comparisons can also have an effect on error detection and possible needs to develop the instrument.

2.3.3 Non-response Errors

This section concentrates on non-response errors in the cross-sectional SILC data. **Many of the subsequent tables include the *rotational breakdown* as a requirement.** The second Finnish SILC data include the rotational group variable **DB075**, which is coded as follows:

1: Households included in the second wave of the Income Distribution Survey, the Collection Year 2005, the Study Year 2004.
2: Households included in the first wave of the Income Distribution Survey (CY 2005, SY 2004).

This year the 2004 groups for the longitudinal study (2-year, 3-year, 4-year panels) are included in the second wave of the IDS. Also the new IDS sample includes a smaller panel group (of size 2,500) for the longitudinal study, the others continue only for the cross-sectional purposes. The Finnish SILC design can be interpreted as *semi-rotational*, i.e. only a part of longitudinal rotational groups are included in the cross-sectional data. However, the forthcoming tables include both the cross-sectional rotational breakdown as well as the longitudinal study groups.

2.3.3.1 Achieved Sample Size

Table 2.11 Interview Information

Rotational group	Number of households for which an interview is accepted for the database (DB135 = 1) .	Number of persons aged 16 or older who are members of the households for which the interview is accepted for the database (DB135 = 1) and who completed a personal interview (RB250 = 11 to 13).	Number of selected respondents who are members of the households for which the interview is accepted for the database (DB135 = 1) and who completed a personal interview (RB250=11 to 13).
Total	11 229	22 961	11 229
1	5 797	11 909	5 797
2	5 432	11 052	5 432
Longitudinal 2004			
2-year	1 831	3 696	1 831
3-year	1 818	3 718	1 818
4-year	1 783	3 638	1 783
Longitudinal 2005	1 912	3 890	1 912

2.3.3.2 Unit Non-response

For Member States using a rotational design, information on unit non-response will be provided for the new replications in accordance with the formulas described below. For the total sample, the unit non-response will be calculated by removing, from the numerator and the denominator of the formulas described below, those units that according to the tracing rules are out of scope.

** **Household non-response rates*** *NRh = (1-(Ra * Rh)) *100*

Ra *(address contact rate) = addresses successfully contacted / valid addresses selected = sum(DB120=11) / [sum(DB120=all) - sum(DB120=23)]*
Rh *(proportion of complete household interviews accepted for the database) = number of HH interviews completed and accepted for the database / number of eligible households at contacted addresses) = sum(DB135=1) / sum(DB130=all)*

DB120 is the record of contact at the addresses

DB130 is the household questionnaire result
DB135 is the household interview acceptance result

**** Individual non-response rates*** *NRp = (1-Rp)*100*

Rp *(proportion of complete personal interviews within the households accepted for the database) = Number of personal interviews completed / number of eligible individuals in the households whose interviews were completed and accepted for the database = sum(RB250=11+12+13) / sum(RB245 = 1+2+3)*

RB245 is the respondent status
RB250 is the data status

** **Overall individual non-response rates*** *******NRp = (1-(Ra*Rh*Rp))*100*

For those MSs where a sample of persons rather than a sample of HHs (addresses) was selected, the individual non-response rates will be calculated for the selected respondent (RB245=2), for all individuals aged 16 or older (RB245=2+3) and for the non-selected respondent (RB245=3).

Table 2.12 Non-response Rates

Rotational group	Household non-response rate	Individual non-response rate			Overall individual non-response rate		
		Selected respondent	All individuals 16 or older	Non-selected respondent	Selected respondent	All individuals 16 or older	Non-selected respondent
Total	14.6408	0	0	0	14.6408	14.6408	14.6408
1	21.1614	0	0	0	21.1614	21.1614	21.1614
2	6.3771	0	0	0	6.3771	6.3771	6.3771
Longitudinal 2004							
2-year	5.7643	0	0	0	5.7643	5.7643	5.7643
3-year	6.6256	0	0	0	6.6256	6.6256	6.6256
4-year	6.7469	0	0	0	6.7469	6.7469	6.7469
Longitudinal 2005	22.2132	0	0	0	22.2132	22.2132	22.2132

2.3.3.3 Distribution of households (original units) by 'record of contact at address' (DB120), by 'household questionnaire result' (DB130), and by 'household interview acceptance' (DB35), for each rotational group (if applicable) and for the total

In the CY 2004 the first rotation is the old panel of the Income Distribution Survey, the other three are forthcoming SILC waves of two, three and four years. See above for further details of the rotation system.

Table 2.13 Distribution of Households by DB120, DB130 and DB135

Description	Total		Rotation 1		Rotation 2		L 2004 R2		L2004 R3		L2004 R4		L2005 R4	
	number	%	number	%	number	%								
Total	13 373	100	7 500	100	5 873	100	1 965	100	1 966	100	1 942	100	2 500	100
Address contacted	13 155	98.36	7 353	98.04	5 802	98.79	1 943	98.88	1 947	99.03	1 912	98.45	2 458	98.32
Address non-contacted	218	1.63	147	1.96	71	1.2	22	1.11	19	0.96	30	1.54	42	1.68
Total address non-contacted	218	100	147	100	71	100	22	100	19	100	30	100	42	100
Address cannot be located	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Address unable to access	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Address does not exist, etc.	218	100	147	100	71	100	22	100	19	100	30	100	42	100

Description	Total		Rotation 1		Rotation 2		L 2004 R2		L2004 R3		L2004 R4		L2005 R4	
	number	%	number	%	number	%								
Total	13 155	100	7 353	100	5 802	100	1 943	100	1 947	100	1 912	100	2 458	100
Household questionnaire completed	11 229	85.35	5 797	78.83	5 432	93.62	1 831	94.23	1 818	93.37	1 783	93.25	1 912	77.78
Interview not completed	1 926	14.64	1 556	21.16	370	6.37	112	5.76	129	6.62	129	6.74	546	22.21
Total interview not completed	1 926	100	1 556	100	370	100	112	100	129	100	129	100	546	100
Refusal to co-operate	1 147	59.55	968	62.21	179	48.37	56	50	60	46.51	63	48.83	338	61.9
Entire household temporarily away for duration of fieldwork	213	11.05	131	8.41	82	22.16	19	16.96	35	27.13	28	21.7	54	9.89
Household unable to respond	152	7.89	112	7.19	40	10.81	11	9.82	16	12.4	13	10.07	41	7.5
Other reasons	414	21.49	345	22.17	69	18.64	26	23.21	18	13.95	25	19.37	113	20.69
Household questionnaire completed	11 229	100	5 797	100	5 432	100	1 831	100	1 818	100	1 783	100	1 912	100
Interview accepted for database	11 229	100	5 797	100	5 432	100	1 831	100	1 818	100	1 783	100	1 912	100
Interview rejected	0	0	0	0	0	0	0	0	0	0	0	0	0	0

2.3.3.4 Distribution of substituted units

The Finnish IDS and SILC data contain **no substitutions**.

2.3.3.5 Item non-response

Item non-response before imputing follows from the interviewed item on interest income taxed at source, and cause partial non-response in variable HY090G on interest, dividends, profit from capital investments in unincorporated businesses and in the total household income variables HY010, HY020, HY022 and HY023. Based on the sample observations, a partial item non-response rate was 6.7 per cent in these variables. The proportion of income recipients of imputed rent (HY030G) of all households is presented, but it has not been counted to total household income. The figures on item-non responses are not relevant to the income component.

Table 2.14 Distribution of item non-response in the EU-SILC 2005 survey

Components of income	% of households having received an amount (<0, >0)	% of households with missing values (before imputation)	% of households with partial information (before imputation)	% of households with collected values (before imputation) of the households having received the income	% of households with partial information (before imputation) of the households having received the income
HY010	100.0	0.0	6.7	93.3	6.7
HY020	100.0	0.0	6.7	93.3	6.7
HY022	98.3	0.0	6.7	93.1	6.9
HY023	96.6	0.0	6.7	93.0	7.0
HY030G	76.9	.	.	.	.
HY040G	10.9	0.0	0.0	100.0	0.0
HY050G	33.4	0.0	0.0	100.0	0.0
HY060G	7.0	0.0	0.0	100.0	0.0
HY070G	17.0	0.0	0.0	100.0	0.0
HY080G	8.3	0.0	0.0	100.0	0.0
HY090G	71.4	0.0	6.7	90.6	9.4
HY100G	35.2	0.0	0.0	100.0	0.0
HY110G	3.6	0.0	0.0	100.0	0.0
HY120G	55.3	0.0	0.0	100.0	0.0
HY130G	15.1	0.0	0.0	100.0	0.0
HY140G	98.7	0.0	0.0	100.0	0.0
HY135G	..	..	..	..	..
Components of income	**% of persons 16+ having received the amount (<0, >0)**	**% of persons 16+ with missing values (before imputation)**	**% of persons 16+ with partial information (before imputation)**	**% of persons 16+ with collected values (before imputation) of the persons 16+ having received the income**	**% of persons 16+ with partial information (before imputation) of the persons 16+ having received the income**
PY010G	63.5	0.0	0.0	100.0	0.0
PY020G	2.2	0.0	0.0	100.0	0.0
PY030G	..	..	..	..	..
PY035G	10.5	0.0	0.0	100.0	0.0
PY050G	21.9	0.0	0.0	100.0	0.0
PY070G	..	..	..	..	..
PY080G	..	..	..	..	..
PY090G	14.1	0.0	0.0	100.0	0.0
PY100G	17.5	0.0	0.0	100.0	0.0
PY110G	1.4	0.0	0.0	100.0	0.0
PY120G	6.2	0.0	0.0	100.0	0.0
PY130G	8.3	0.0	0.0	100.0	0.0
PY140G	10.1	0.0	0.0	100.0	0.0
PY200G	..	..	..	..	..

.. Information is not available
. Information is not logical

2.3.3.6 Total item non-response and number of observations in the sample at unit level of the common cross-sectional EU indicators based on the cross-sectional component of the EU-SILC, for equivalised disposable income and for the unadjusted gender pay gap (if applicable)

The *total item non-response* means *"the non-response for an item' + 'the non-response due to the total non-response unit' (household and/or person)".* Correspondingly, non-response for an item for the cross-sectional EU common indicators based on the EU-SILC cross-sectional component means:

- No information at household level for variable HY020 or HY025, or HY020 equals zero or negative
- No information at household level on equivalised sample size
- No information on age/sex, when applicable to the indicator, because these variables are missing.
- No information on household type, when applicable to the indicator
- No information on accommodation tenure status, when applicable to the indicator
- No information on HY022 or HY023, when applicable
- No information on PY200G, when applicable

Non-response due to the total non-response unit means:

- Non-response at individual level (i.e. an individual questionnaire is missing, when applicable)
- Non-response at household level (i.e. interview rejected for the database; address cannot be located or address unable to access).

The interviewed information is collected mostly from the household respondent because use of the registers is the major data source for income in Finland, and therefore the non-response at individual level is not applicable here. The following tables are based on the indicator programs in SAS provided by Eurostat in order to obtain the required indicators. The first table includes the item non-response information based on persons, and the second table contains the classification variables concerning households, and the information is based on households there. The latter table also includes the unit non-response and the total item non-response. The unit non-response information is only available at the household level, not containing any information about the structure of the household (e.g. tenure, household type, work intensity). Note that the unit non-response does not include the over-coverage.

The estimation of the common cross-sectional indicators qualifies for the sample observations criteria provided by the European Commission for the publishing. All estimates are based at least over 50 observations in the sample, and the item non-response rates are lower than 20 per cent.

The missing items appear in the estimates of at-risk-of-poverty rates by work intensity and by most frequent status. With regard to the both variables, missing items are mostly due to the definitions for these derived variables in the SAS indicator program. Values were responded to in the original target variables, but were excluded from the derived variables in the program. Information about this has been written to the footnotes of the figures in Tables 2.14 and 2.15.

Table 2.15 Item Non-response Concerning Indicators with Different Classifications, Persons

At-risk-of-poverty rate after social transfers	Number of sample observations	Number of sample observations not taken into account due to item non-response	Item non-response in calculation of indicator	Item non-response of classification variable
Gender, total	**29 112**	**0**	**0**	**0**
- female	14 373	0	0	
- male	14 739	0	0	

Age, total	**29 112**	**0**	**0**	**0**
- under 16 years	6 151	0	0	
- 16 - 24 years	3 867	0	0	
- 25 - 49 years	9 258	0	0	
- 50 - 64 years	6 765	0	0	
- 65 years and over	3 071	0	0	
- 0 - 64 years	26 041	0	0	
- 16 years and over	22 961	0	0	
- 16 - 64 years	19 890	0	0	
Age and gender, total	**29 112**	**0**	**0**	**0**
- under 16 years, female	3 028	0	0	
- under 16 years, male	3 123	0	0	
- 16 - 24 years, female	1 785	0	0	
- 16 - 24 years, male	2 082	0	0	
- 25 - 49 years, female	4 657	0	0	
- 25 - 49 years, male	4 601	0	0	
- 50 - 64 years, female	3 299	0	0	
- 50 - 64 years, male	3 466	0	0	
- 65 years and over, female	1 604	0	0	
- 65 years and over, male	1 467	0	0	
- 0 - 64 years, female	12 769	0	0	
- 0 - 64 years, male	13 272	0	0	
- 16 years and over, female	11 345	*0*	*0*	
- 16 years and over, male	11 616	*0*	*0*	
-16 - 64 years, female	9 741	0	0	
-16 - 64 years, male	10 149	0	0	
Most frequent activity and gender, total	**22 328**	**0**[2]	**0**	**0**
- Employed, female	5 878	0	0	
- Employed, male	6 848	0	0	
- Unemployed, female	598	0	0	
- Unemployed, male	610	0	0	
- Retired, female	2 386	0	0	
- Retired, male	2 239	0	0	
- Other inactive, female	2139	0	0	
- Other inactive, male	1 630	0	0	
Tenure status, total	**29 112**	**0**	**0**	**0**
- Owner or rent free	23 633	0	0	
- Tenant	5 479	0	0	
Household type	**29 112**	**0**	**0**	**0**
- One person, under 65 years	1 780	0	0	
- One person, 65 years and over	610	0	0	
- One person, female	1 282	0	0	
- One person, male	1 108	0	0	
- One-person households, total	2 390	0	0	
-Two adults, both under 65 years old, no dependent children	5 520	0	0	

[2] No non-responded items in the original target variables
Missing items in the derived variables are as follows:
n = 6 151: age < 16 years;
n = 633: age >= 16: total months = sum(PL070, PL072, PL080, PL085, PL087, PL090) <= 6 or (sum(PL070,PL072,0) / total months <= 0.5) and (sum(PL080,0) / total months <= 0.5) and (sum(PL085,0) / total months <= 0.5) and (sum(PL087,PL090,0) / total months <= 0.5)

- Two adults, at least one 65 years old, no dependent children	2 466	0	0	
- Other households without dependent children	2 049	0	0	
- Single parent households with dependent children	1 063	0	0	
- Two adults, one dependent child	3 792	0	0	
- Two adults, two dependent children	5 376	0	0	
- Two adults, three or more dependent children	4 681	0	0	
- Other households with dependent children	1775	0	0	
- Households without dependent children	12 425	0	0	
- Households with dependent children	16 687	0	0	
- Others (not possible to determine type)	0	0	0	
Work intensity	**26 615**	**1[3]**	**0**	**1**
Household without dependent children, W = 0	1 528	0	0	
Household without dependent children, 0 < W < 1	4 287	0	0	
Household without dependent children, W = 1	4 118	0	0	
Household with dependent children, W = 0	564	0	0	
Household with dependent children, 0 < W <= 0.5	955	0	0	
Household with dependent children, 0.5 <= W < 1	6 469	0	0	
Household with dependent children, W = 1	8 694	0	0	

Table 2.16 Item Non-response Concerning Indicators with Different Classifications, Households

	Number of sample observations	Number of sample observations not taken into account due to item non-response	Unit non-response	Total item non-response
Household type	**11 229**	**0**	**1 926**	**1 926**
- One-person households, total	2 390	0		
-Two adults, both under 65 years old, no dependent children	2 760	0		
- Two adults, at least one 65 years old, no dependent children	1 233	0		
- Other households without dependent children	643	0		
- Single parent households with dependent children	399	0		
- Two adults, one dependent child	1 264	0		
- Two adults, two dependent children	1 344	0		
- Two adults, three or more dependent children	837	0		
- Other households with dependent children	359	0		
- Households without dependent children	7 026	0		
- Households with dependent children	4 203	0		
- Others (not possible to determine type)	0	0		
Work intensity	**9 590**	**1[4]**	**1 926**	**1 927**
- HH without dep. ch., W = 0	945	0		
- HH without dep. ch., 0<W<1	2 096	0		
- HH without dep. ch., W = 1	2 348	0		

[3] Non-responded items in PL030: n=11; if age >=18 & age <=24 then n=1.
Missing items are due to the definition of the variable derived from the original target variables.
[4] Households with non-responded items in PL030: n=11; if age >=18 & age <=24 then n=1.
Missing items are due to the definition of the variable derived from the original target variables.

- HH with dep. ch., W = 0	167	0		
- HH with dep. ch., 0<W<= 0.5	244	0		
- HH with dep. ch., 0.5<=W<1	1 556	0		
- HH with dep. ch., W = 1	2 234	0		
Tenure status	**11 229**	**0**	**1 926**	**1 926**
- Owner or rent free	8 532	0		
- Tenant	2 697	0		

2.4 Mode of Data Collection

Table 2.17 Distribution of Household Members aged 16 and over by 'RB250' and 'RB245'

Rotational group	Total	RB250=11	RB250=12	RB250=13	RB250=21	RB250=22	RB250=23	RB250=31	RB250=32	RB250=33
Household members 16+ and RB245 = 1 to 3										
.	22 961	0	0	22 961	0	0	0	0	0	0
.	100	0	0	100	0	0	0	0	0	0
1	11 909	0	0	11 909	0	0	0	0	0	0
	100	0	0	100	0	0	0	0	0	0
2	11 052	0	0	11 052	0	0	0	0	0	0
	100	0	0	100	0	0	0	0	0	0
Longit. 2004										
2-year	3 696	0	0	3 696	0	0	0	0	0	0
	100	0	0	100	0	0	0	0	0	0
3-year	3 718	0	0	3 718	0	0	0	0	0	0
	100	0	0	100	0	0	0	0	0	0
4-year	3 638	0	0	3 638	0	0	0	0	0	0
	100	0	0	100	0	0	0	0	0	0
Longit. 2005	3 890	0	0	3 890	0	0	0	0	0	0
	100	0	0	100	0	0	0	0	0	0
Household members 16+ and RB245 = 2										
.	11 229	0	0	11 229	0	0	0	0	0	0
.	100	0	0	100	0	0	0	0	0	0
1	5 797	0	0	5 797	0	0	0	0	0	0
	100	0	0	100	0	0	0	0	0	0
2	5 432	0	0	5 432	0	0	0	0	0	0
	100	0	0	100	0	0	0	0	0	0
Longit 2004.										
2-year	1 831	0	0	1 831	0	0	0	0	0	0
	100	0	0	100	0	0	0	0	0	0
3-year	1 818	0	0	1 818	0	0	0	0	0	0
	100	0	0	100	0	0	0	0	0	0
4-year	1 783	0	0	1 783	0	0	0	0	0	0
	100	0	0	100	0	0	0	0	0	0
Longit.	1 912	0	0	1 912	0	0	0	0	0	0
2005	100	0	0	100	0	0	0	0	0	0
Household members 16+ and RB245 = 3										
.	11 732	0	0	11 732	0	0	0	0	0	0
.	100	0	0	100	0	0	0	0	0	0
1	6 112	0	0	6 112	0	0	0	0	0	0
	100	0	0	100	0	0	0	0	0	0
2	5 620	0	0	5 620	0	0	0	0	0	0
	100	0	0	100	0	0	0	0	0	0

Longit. 2004										
2-year	1 865	0	0	1 865	0	0	0	0	0	0
	100	0	0	100	0	0	0	0	0	0
3-year	1 900	0	0	1 900	0	0	0	0	0	0
	100	0	0	100	0	0	0	0	0	0
4-year	1 855	0	0	1 855	0	0	0	0	0	0
	100	0	0	100	0	0	0	0	0	0
Longit.	1 978	0	0	1 978	0	0	0	0	0	0
2005	100	0	0	100	0	0	0	0	0	0

Table 2.18 Distribution of Household Members aged 16 and over by 'RB260' and 'RB245'

Rotational group	Total	RB260=1	RB260=2	RB260=3	RB260=4	RB260=5	RB260= missing
Household members 16+ and RB245 = 1 to 3							
.	22 961	0	357	10 871	0	11 733	0
.	100	0	1.55	47.35	0	51.10	0
1	11 909	0	259	5 538	0	6 112	0
	100	0	2.17	46.50	0	51.32	0
2	11 052	0	98	5 333	0	5 621	0
	100	0	0.89	48.25	0	50.86	0
Longit. 2004							
2-year	3 696	0	39	1 791	0	1 866	0
	100	0	1.06	48.46	0	50.48	0
3-year	3 718	0	33	1 785	0	1 900	0
	100	0	0.89	48.01	0	51.10	0
4-year	3 638	0	26	1 757	0	1 855	0
	100	0	0.71	48.30	0	50.99	0
Longit.	3 890	0	106	1 806	0	1 978	0
2005	100	0	2.72	46.43	0	50.84	0
Household members 16+ and RB245 = 2							
.	11 229	0	293	8 268	0	2 668	0
.	100	0	2.61	73.63	0	23.76	0
1	5 797	0	214	4 174	0	1 409	0
	100	0	3.69	72.00	0	24.31	0
2	5 432	0	79	4 094	0	1 259	0
	100	0	1.45	75.37	0	23.18	0
Longit. 2004							
2-year	1 831	0	31	1 385	0	415	0
	100	0	1.69	75.64	0	22.67	0
3-year	1 818	0	28	1 344	0	446	0
	100	0	1.54	73.93	0	24.53	0
4-year	1 783	0	20	1 365	0	398	0
	100	0	1.12	76.56	0	22.32	0
Longit.	1 912	0	74	1 381	0	441	0
2005	100	0	3.87	72.23	0	23.06	0
Household members 16+ and RB245 = 3							
.	11 732	0	64	2 603	0	9 065	0
.	100	0	0.55	22.19	0	77.26	0
1	6 112	0	45	1 364	0	4 703	0
	100	0	0.74	22.32	0	76.94	0
2	5 620	0	13	1 239	0	4 362	0
	100	0	0.23	22.05	0	77.62	0
Longit. 2004							
2-year	1 865	0	8	406	0	1 451	0

	100	0	0.43	21.77	0	77.80	0
3-year	1 900	0	5	441	0	1 454	0
	100	0	0.26	23.21	0	76.53	0
4-year	1 855	0	6	392	0	1 457	0
	100	0	0.32	21.13	0	78.54	0
Longit.	1 978	0	16	425	0	1 537	0
2005	100	0	0.80	21.49	0	77.71	0

2.5 Interview Duration

HB100, PB120, Lengths of the household and personal interviews. Measuring the length of the interview separately for the household and person interviews failed because the household and person interview questions were interlaced, which makes it difficult to measure the time.

The average overall duration of the interview was 29 minutes.

Table 2.19 Distribution (%) of total duration of interview in cross-section and panels, SILC 2005

	Minutes						
	1-25	26-35	36-60	61+	missing	Total	Mean
Cross-section total	46	30	21	2	1	100	29.1
Longitudinal	48	29	20	2	1	100	28.7
Panel started 2005	43	31	23	2	1	100	30.1
Panels started 2004, second wave							
3-year panel	49	29	19	2	1	100	28.2
4-year panel	51	27	19	2	1	100	27.9

Note: In 163 interviews the measurement of interview duration failed.

3. Comparability

3.1 Basic Concepts and Definitions

Basic concepts and their definitions are in accordance with the Commission Regulation (EC) No 1980/2003 provided for the community statistics on income and living conditions as regards definitions and updated definitions. To some extent, adaptation of the definitions used in the national statistical system is allowed for the EU-SILC. In Finland, private household and household membership in particular are the ones that have been defined nationally (e.g. IDS) with less detailed information than stated in the regulations, but within the framework.

The reference population consists of the members of the private households permanently resident in Finland on 31 December 2004. Persons living in institutions, in collective households or in residential homes[5] are excluded.

The private household was constructed to include a person residing alone, or all the persons, related or not, who reside and have their meals together or otherwise use their income together. If a person was temporarily absent from the household's main dwelling and from home, no time duration was set for the absence provided that the above-mentioned criteria of household formation and membership were fulfilled. Such persons have close ties to the household. Therefore, the following persons are also counted in household members:

- Persons conducting military service or conscript service
- Persons residing and working in another locality or abroad if they are involved in the acquisition and use of household income
- Persons residing and studying in another locality if they use income received mostly from their parents
- Persons temporarily in institutions, on holiday or travelling.

The following persons form a household of their own:
- Subtenants
- Domestic staff
- Students living on their own if they live mostly on their own income or on a student loan
- Students residing in dormitories, unless they are married or officially cohabiting.

Other definitional solutions done are due to the collection of the information both from registers and by interviews. These are related to **reference times**. First, current as a reference time refers to several calendar time points. Information collected solely by interviews (e.g. non-monetary deprivation indicators, education, health) refers to the interview time point in the survey year (2005). Information collected by interviews, but used for the target variables as combined with the information from registers and other information interviewed on themes close to income is related to the income reference period, which is the fixed 12-month period before the survey year, i.e. the whole calendar year (2004). Then, the current is either the last day (dwelling, characteristics of dwelling for the imputed rent, housing environment, housing costs) or the last month (economic activity) of the income reference year. In particular, information on housing arrears is consistent with information on housing costs from the income reference period, not from the last twelve months preceding the interview time point as provided.

[5] Residential homes are situated either in residential or institutional care buildings and do not meet the definition of dwelling. They do not include a kitchen or cooking facilities, a water closet or cleaning facilities (shower, bathroom or sauna). Students dormitories which are counted in the private household definition above include these facilities.

Table 3.1 Finland's definitions for the reference periods in the EU-SILC 2005 survey.

Current, time point of interview for the respondent in the survey year 2005:
- Non-monetary household deprivation indicators
- Housing (amenities in the dwelling)
- Education
- Health

Current, last day (31 Dec.) of the income reference period (2004):
- Basic data
- Physical and social environment
- Housing (dwelling type, tenure status and housing conditions, housing costs)

Current, last month (December) of the income reference period (2004):
- Child care
- Labour information

Last 12 months preceding the time point of interview:
- Health (access to health care)

Income reference period (a fixed 12-month period), i.e. 2004:
- Income
- Labour information (basic information on activity status)
- Housing and non-housing related arrears.

The income reference period is the preceding calendar year of the survey year, i.e. a fixed 12-month period. Income taxed by the Bookkeeping Act received from the completed accounting periods in the income reference period is included. These are business income, income from dividends and interest, and avoir fiscal tax credit from the accounting periods 1993-1995, 1996-1999 and 2000-2004.

The reference period for taxes on income and social contributions is the years when taxes are paid from the income received during the income reference period. The taxes are paid in the income reference period (t) and the following years (t+1, t+2). The social contributions are mostly paid in the income reference period (t).[6]

The reference period of taxes on wealth is the years when taxes are paid from the wealth owned in the income reference period. Most of the payments are done during the income reference year.

The time lag between the income reference period and current variables is in its maximum when current information is from the interview time point. The last interview was conducted on 23 May in the survey year. The time lag is then 4.7 months. However, most of the current information is from the end of the income reference period and then the time lag does not exist.

[6] Most of the taxes (incl. taxes on net wealth owned) and social contributions are actually done during the income reference year (t) as withholdings by a payer or advance payments by a person, i.e. 90 per cent of enforced taxes in 2004 (National Board of Taxes 2005). Some of these payments can be done up till March of the year after the income reference period (t+1). As a result of the enforced taxation by tax authorities, 5.9 per cent of the enforced taxes were received as tax refunds in the year after the income reference period (t+1), 3.9 per cent of the enforced taxes were paid as residual taxes in the year after the income reference period (t+1) and further in the beginning of the following year (t+2). If demands of rectification and petition of appeals were proceeded, in a few cases, taxes are paid later (t+3,…,n).

The consistency is highest among employees and pensioners. 92 per cent of the total withholdings and advance payments for employees and 94 per cent of the total withholdings and advance payments for pensioners were in accordance with the enforced taxes in 2004. The consistency was not as high among self-employed persons, 75 per cent of total advance payments done by farmers and 70 per cent done by other self-employed persons were in accordance with the enforced taxes. (National Board of Taxes 2005).

Interviews were conducted from 7 January to 23 May in the survey year 2005. **The duration of interviewed data collection** was 4.5 months. Of all household interviews, 25 per cent were collected by 1 February, 50 per cent by 23 February, 75 per cent were collected by 5 April, and 90 per cent by 2 May.

For the register database, the last information was collected on 24 November in the survey year 2005. When data collection from registers is included in the measurement, **the duration of the whole data collection both from interviews and registers** was 10.5 months.

The basic information on activity status during the income reference period was derived from information on a person's main activity in each month by summing the activities over the months (twelve in total, see interviewed groups below). The information on a person's main activity was interviewed from the household respondent. For answering to a question, the respondent was instructed to give priority to employment over non-economic activity and inactivity if that person had had several activities during the month. Full-time and part-time work was separated by working hours. Work was full-time if a person worked at least 30 hours per week. Otherwise, work was part-time if a person worked under 30 hours per week. In economically inactive statuses, the answer is based on the respondent's assessment about his/her main activity during the month.

The target variables on a person's activity status during the income reference period and the detailed subgroups interviewed are as follows :

PL070, Number of months at full-time work :
- Employee working full-time (at least 30 hours per week)
- Entrepreuner or unpaid worker in family enterprise working full-time (at least 30 hours per week)

PL072, Number of months at part-time work:
- Employee working part-time (under 30 hours per week)
- Entrepreuner or unpaid worker in family enterprise working part-time (under 30 hours per week)

PL080, Number of months in unemployment:
- Unemployed

PL085, Number of months in retirement:
- Retiree

PL087, Number of months in studying:
- Pupil, student

PL090, Number of months in inactivity:
- On unpaid sickness leave, etc.
- Others
- In military service or conscript service

3.2 Components of Income

3.2.1 Differences between the National Definitions and Standard EU-SILC Definitions

Total household gross income and disposable household income

The target variables on gross income components, on **total gross household income, HY010**, and on **disposable household income HY020 and total disposable household income before social transfers other than old-age and survivors' benefits HY022 and including old-age and survivors' benefits HY023** are well in accordance with the requirements.

HY010 is the sum of gross income components at the household level. HY020 is HY010 after current transfers paid are deducted. HY010 is a positive value (incl. 0 values). Negative values of the net income variables HY020, HY022 and HY023 on total disposable household income are due to such current transfers paid which are not related to the total household gross income HY010. These are regular taxes on wealth HY120G, which may exceed the amount of the total household gross income by the EU-SILC definition. The number of the sample households with negative values was nine in HY020, 120 in HY022 and 538 in HY023. For calculating the common European Union indicators, the negative values were set for zero values. The conversion has an effect on the HY020 mean equivalised income and Gini coefficient estimates.

Tax on income and social insurance contributions HY140G and regular inter-household transfers paid HY130G were done from the total household gross income received during the income reference year. They did not cause negative values to total household income components.

The target variable **HY030G on imputed rent** is required as a compulsory variable from 2007 onwards. Information on the variable is according to the national Income Distribution Statistics in the 2005 EU-SILC survey. The rental equivalence method used follows the EU-SILC regulation (Table 3.3). Information on imputed rent was not counted for total household income components (HY010, HY020, HY022, HY023),.

The target variable HY030G covers imputed rent for all households occupying own dwellings and the households renting a dwelling from other households at a lower price than the market price, or got it free. Differing from what the regulation states about the coverage, imputed rent was not yet calculated for dwellings rented from a public, municipal, voluntary or non-profit agency in the 2005 survey.

For constructing the target variables **HY022 and HY023, total disposable household income before social transfers**, social transfers in gross amounts were converted to net amounts, and deducted from total disposable household income (HY020). Detailed income information from the Personal Tax Register was used. The phases in deriving HY022 and HY023 were as follows:

1. Deductions which are focused on social transfers subject to taxation were counted in order to obtain taxable social transfers. Deductions of the state and municipal taxation were done separately.
2. Taxes paid on taxable social transfers in state and municipal taxation were deducted. These are the actual taxes paid defined by the rate of the taxed social transfers and taxed earned income (incl. social transfers in the Finnish taxation).
3. The gross to net converted social transfers subject to taxation and social transfers not subject to taxation excluding and including old-age benefits and survivors' benefits were deducted from HY020, resulting in HY022 and HY023.

Income received

The variables on gross income components were obtained by summing the detailed gross items to the person and household unit level. Especially when register income is available as very detailed items, the aggregating of the items for the target variables is closely in accordance with the regulations. **Compared with the Regulation definitions on the EU-SILC gross income components, the following differences,** however, appear due to using register information within the Personal Tax Register frame:

- Payments (PY080G) received by persons from their voluntary personal insurances, which are in the same register item as income received from statutory, voluntary insurance taken by employers for their employees. This income could not be excluded exactly. It was counted in interests, dividends, profits from capital investment in an unincorporated business (HY090G).
- Income received from statutory, voluntary insurance taken by employers or employed persons (entrepreneurs) themselves to supplement compulsory social security was counted in interests, dividends, profits from capital investment in an unincorporated business (HY090G).
- Earned income from agriculture, forestry and business activities was counted in self-employment income (PY050G). In addition, income from business activities and earned income from dividends was included for shareholders in corporations. Income is in gross amounts after expenses except interest on loans for acquisition of income. They were counted as deductions for taxable income and result as lower taxes paid (HY140G).
- Loan interests diminish the taxable income as deductions, and result as lower taxes paid (HY140G).
- Self-employed income (PY050G) is positive (incl. 0 income). Losses were considered for lower taxes paid from other type of income in the income reference period, or in the spouse's taxes paid. If no taxable income was received at all, the confirmed losses are considered in taxes that will be paid from the income received in the following years. Therefore, confirmed losses both from the income of the income reference period and from previous periods as well can both have an effect on taxes paid from the reference period's income (HY140G).[7]
- Both received social benefits and social benefits obliged to be returned to payers were included in the certain target variables on social benefits (PY090G, PY100G, PY110G, PY120G, PY140G, HY050G , HY060G, HY070G). The statistical units have negative values on these variables if the social benefits were solely returned back, or the returned amount exceeded the amount received during the income reference period. The social benefits are obliged to be returned if the income or living conditions have changed and they are not valid in relation to the allowed criteria any more.
- Income received personally by people aged under 16 was counted in the target variable HY110G. The variable consists of the following type of income: employee income and self-employment income, education related allowances, survivors' benefits, disability benefits and part of family/children-related allowances. Other social benefits within the ESSPROS system are paid for children's carers, and were counted in family benefits (HY050G). Income received from interest, dividends, profit from capital investments in unincorporated businesses is a relatively significant income source for people aged under 16. It was counted in HY090G.

[7] In the sample, 17.8 per cent of self-employed persons (PL030 =1,2, & PL040 = 1,2) had 0 income on PY050G (n = 650 / 3644). Most of them had other income sources, 72.6 per cent received income on PY010G, 71.4 per cent received personal income on HY090G, 88.9 per cent got income on either PY010G or HY090G, and 5.8 per cent on only other type of income during the reference year. Persons who were temporarily away from work are counted in the numbers. Losses were in 6.1 per cent for all self-employed persons (n= 224), for 30.8 per cent of whom the losses (incl. losses from HY090G) were considered as deductions for taxes paid (incl. all income to which deductions focused), and for 73.2 per cent the confirmed losses (the rest of the losses or all) can be considered in the taxes paid from income received after the income reference year. In addition, a small number of losses were counted ín the spouse's taxation.

Current transfers paid

The target variable on **tax on income and social insurance contributions (HY140G)** includes taxes paid for the state taxation and for the municipal taxation. For the state taxation, taxes from earned income (incl. social benefits) are paid progressively by the person's income level, taxes from capital income are paid uniformly (29 per cent of capital income in 2004). For municipal taxation, taxes from earned income are paid by the tax rate of the place of domicile that a person hold at the end (31 Dec.) of the year preceding the income reference year.

The social contributions include the following items: compulsory sickness contributions, unemployment contributions and pension contributions.

The target variable on **regular taxes on wealth (HY120G)** includes taxes on real property owned and taxes paid on net wealth owned at the end (31 Dec.) of the income reference period. Besides, taxes on real property owned are paid indirectly in utility costs of dwellings by shareholders in housing corporations. The tax was not included in HY120G, but it was counted in housing costs (HH070) and consequently, as a part of the housing costs component it diminishes the gross rent value in the imputed rent (HY030G).

Table 3.2 Components of income. Finland's definitions and assessed consequences resulting from differences compared with the EU-SILC definition in the 2005 survey.

Components of income	Variable name	The definition	Consequences to comparability 1 = comparable 0 = not comparable
Total household gross income	HY010		1 See notes below
Total disposable household income	HY020		1 See notes below
Total disposable household income, before social transfers other than old-age and survivors' benefits	HY022		1 See notes below
Total disposable household income, before social transfers including old-age and survivors' benefits	HY023		1 See notes below
Imputed rent	HY030G	Imputed gross rentals for all households that do not report paying full rent, either because they are owner-occupiers or they live in accommodation rented from another household at a lower price than the market price or rent-free minus housing costs actually paid (incl. subsidies received from government, i.e. housing benefits and tax deductions for mortgage interests). Imputed for the dwellings which are used as the main residence of the household.	0 Note: Consistent with the Finnish IDS. Dwellings rented at a lower price than the market price from a public, municipal, voluntary or non-profit agency are not included. The information on imputed gross rentals (equivalent to market rent) is based on the rent including other utility costs done besides the "space rent". After deducting consistent housing costs paid by the household, the definition is comparable. Imputed rent has not been included in the gross household income variable (HY010) and the total disposable household income variables (HY020, HY022, HY023).
Income from rental of property or land	HY040G	Income received, during the income reference period, from renting a property less expenses except interest payments.	1 Note: Interest payments on loans for acquisition of income are considered as deductions from taxable income in taxation, and thus diminish the amount of taxes paid on the income (HY140G).
Family/children-related allowances	HY050G	Financial support to households for bringing up children and financial assistance to people who support relatives other than children: income maintenance benefit in the event of childbirth, birth grant, parental leave benefit, family or child allowance, other cash benefits.	1
Social exclusion payments not elsewhere classified	HY060G	Social benefits to the socially excluded or to those at risk of social exclusion: income support to people with insufficient resources, and other cash benefits as support for destitute and vulnerable persons to help alleviate poverty or assist in difficult situations.	1 Note: A register-based item on income support also includes a minor part of means-tested housing allowance. Parts are not separable from each other.
Housing allowances	HY070G	Rent benefit or benefit to owner-occupiers, means-tested	1
Regular inter-household cash transfers received	HY080G	Regular monetary amounts or monetary amounts over the certain minimum amount (EUR 100) received during the income reference period, from other households or persons:	1

		compulsory child support, voluntary support to education, voluntary payments for housing costs and utility bills.	
Interest, dividends, profit from capital investments in unincorporated businesses	HY090G	The amount of interest from assets, dividends and profits from capital investment in an unincorporated business in which the person does not work, received during the income reference period, less expenses incurred. Interests on loans for acquisition of income are considered as expenses for certain income items, but not for all income items.	1 Note: Interest payments on loans for acquisition of income are subtracted as deductions from taxable income in taxation, and thus diminish the taxes paid on income. (HY140G). The component includes income from statutory benefits (incl. pensions) undertaken voluntarily by an employer, an employed person (entrepreneur), or a person individually in addition to the compulsory scheme of social benefits. These are a few register items which cannot be subdivided.
Interest paid on mortgages	HY100G	Total gross amount, before deducting any tax credit or tax allowance, of mortgage interest on the main residence of the household during the income reference period.	1
Income received by people aged under 16	HY110G	Gross income received by all household members aged under 16 during the income reference period.	1
Regular taxes on wealth	HY120G	Taxes that are payable annually on the ownership or use of land and buildings paid during the income reference period (t). Taxes that are payable on the net wealth owned at the end (31 Dec.) of the income reference period (t) paid in the income reference year (t) and in the following years $_{(t+1,\ t+2)}$.	1 Taxes paid on the ownership and use of land or buildings of shareholders in housing companies are included indirectly in service charges as part of housing costs, and thus they diminish the gross value of imputed rent.
Regular inter-household transfers paid	HY130G	Regular monetary amounts or monetary amounts over the certain minimum amount (EUR 100) paid during the income reference period, to other households or persons: compulsory child support, voluntary support to education, voluntary payments for housing costs and utility bills.	1
Tax on income and social insurance contributions	HY140G	Taxes on income, profits and capital gains: taxes on individual, household or tax-unit income (income from employment, property, entrepreneurship, pensions, etc.) including taxes deducted by employers (i.e. withholdings), other taxes at source and taxes on the income of owners of unincorporated enterprises paid **from the income received in the income reference year**. Social insurance contributions paid during the income reference period.	1 Note: Interests charged on arrears of taxes due and any fines imposed by tax authorities are not included.
Repayments/receipts for tax adjustments	HY145G	-	-
Cash or near-cash employee income	PY010G	Monetary component of the compensation of employees in cash payable by an employer to an employee: value of any social contributions and income taxes payable by an employee or by the employer on behalf of the employee to social insurance schemes or tax authorities.	1 Note: Tips and bonuses, and benefits based on profit sharing from stock options (excl. the ones converted into cash) are included in this component according to the Finnish IDS.
Non-cash employee income	PY020G	Non-monetary income components which may be provided free or at a reduced price to an employee as part of the employment package by an employer: company car and associated costs, free or subsidised meals, luncheon vouchers, reimbursement or payment of housing-related expenses, other goods and services provided free or at a reduced price by their employer to their employees.	1 Note: Company car is included.
Employers' social insurance contributions	PY030G	-	-
Cash profits or losses from self-employment (including royalties)	PY050G	The income received, during the income reference period, by individuals, for themselves or in respect of their family members, as a result of their current or former involvement in self-employment jobs: operating profit accruing to working owners or partners of an unincorporated enterprise, royalties earned on writing, inventions and so on, not included in the profit/loss of unincorporated enterprises, rentals from business buildings, vehicles, equipment, etc., not included in the profit/loss of unincorporated enterprises, after deduction of related costs. Interests on loans for acquisition of income are considered as costs for a few income items, but not for all income items.	1 Note: Interest payments on loans for acquisition of income are subtracted as deductions for taxable income in taxation, and thus diminish the taxes paid on income (HY120G). Positive values (incl. 0 values). Losses are considered as deductions for taxable income and diminished taxes paid from other type of income in the income reference year, or in the spouse's taxes paid. If such taxable income that deductions concern has not been received at all, losses are considered as taxes which will be paid from the income received in the following years
Value of goods produced for own consumption	PY070G	-	Note: Value is not significant at national level, or to particular groups of households. The information is not collected.
Unemployment benefits	PY090G	Benefits that replace income lost by a worker due to the loss of gainful employment, provide subsistence income to persons	1

		entering or re-entering the labour market, provide subsistence income to unemployed persons not members in unemployment funds, provide subsistence income to persons in long-term unemployment, and to elderly persons who retire after long-term unemployment before the legal retirement age, contribute to the cost of training or re-training people looking for employment. The costs of travelling or relocating to obtain employment are included as deductions for taxes paid of unemployed benefits.	
Old-age benefits	PY100G	Benefits that provide replacement income when an aged person retires from the labour market, or guarantee certain income when a person has reached the prescribed age. Old-age pensions, early old-age pensions, deferred old-age pensions and part-time pensions are counted in old-age benefits. The statutory retirement age for old-age pension under the national and employment scheme is 65. Persons secured under the employment scheme are in certain professions entitled to start old-age pensions earlier. Early old-age pensions are awarded after the age of 58 in public sector contracts and 60 in private sector contracts under the employment scheme. Part-time pensions are awarded to persons aged 58 to 64 under the employment scheme.	1
Survivors' benefits	PY110G	Benefits that provide temporary or permanent income to people below the retirement age after the death of their spouse, partner or next-of-kin, usually when the latter represented the main breadwinner for the beneficiary. Survivors' pension to the deceased person's children, to a surviving spouse and under the employment pension scheme to a former spouse are counted in survivors' benefits.	1
Sickness benefits	PY120G	Benefits that replace in whole or in part loss of earnings during temporary inability to work due to sickness or injury.	1
Disability benefits	PY130G	Benefits that provide an income to persons below the standard retirement age whose ability to work and earn is impaired beyond the minimum level laid down by legislation by physical or mental disability. Income for the disabled persons entering or returning to work.	1
Education-related allowances	PY140G	Grants, scholarships and other education assistance received by students.	1
Gross monthly earnings for employees	PY200G		Note: The gender pay gap is calculated by the Wages and Salaries Statistics unit, Statistics Finland

3.2.2 The Source or Procedure Used for the Collection of Income Variables

Income information is primarily register information, which was linked to the EU-SILC sample persons from the register database, i.e. the Total Income Database (TIDB) maintained by Statistics Finland. The TIDB is compiled from register sources maintained by several administrative authorities[8], who are also in charge of the data quality. The sources cover the whole population of Finland. For the TIDB, information is further checked in order to ensure the consistency of the data from several sources.

Items which were not available from registers were collected by interviews (1.2 per cent from all gross income and 1.7 per cent from all paid transfers weighted at total households were interviewed). Interviewed items on income were as follows:

- Wages and salaries for persons who have no taxable income in Finland (incl. in PY010G)
- Income from agriculture received by a party to an estate (incl. in PY050G)
- Income from forestry after expenses (incl. in PY050G)
- Interest income taxed at source (incl. in HY090G)

[8] Administrative registers are the Personal Tax Register of National Board of Taxes, the Pension Register of the Finnish Centre for Pension, the Pension Register, Social Insurance Register, Rehabilitation Register, Study Aid Register, Housing Allowance Register of the Social Insurance Institution; the Registers of the Education Fund, the Farm Register of the Information Service Centre of the Ministry of Agriculture and Forestry, the Social Assistance Register of the National Research and Development Centre for Welfare and Health (STAKES), the Tax Database of the military injury benefits system of the State Treasury. The main frame for income information is the Personal Tax Register to which other registers give more detailed information, or supplement it by tax-free income information.

- Pensions from abroad to persons who have no taxable income in Finland (incl. in PY100G)
- Tax-free care allowances and convalescent's grants, unspecified tax-free pensions (incl. in PY130G)
- Small subsidies for studying (incl. in PY140G)
- Maintenance support for children (incl. in HY050G)
- Strike assistance (incl. in HY060G)
- Regular inter-household transfers received (HY080G)
- Regular inter-household transfers paid (HY130G)

Interviewed items were automatically checked and corrected in relation to acceptable values in the Blaise questionnaire on the basis of information received in the course of the interview and further, after the information collection, the checking was continued in order to detect and correct erroneous values (Section 2.3.3 Processing errors). The hot-deck method was used to impute item non-responses of interest income taxed at source in the component HY090G interest, dividends, profit from capital investments in unincorporated business to the households.

3.2.3 The Form in Which Income Variables at Component Level Have Been Obtained

Except for the target variables HY020, HY022 and HY023, the target variables on income are in gross amounts.

Table 3.3 Components of income. Finland's sources or procedures used for collection of income components, the form and the methods used for obtaining the target variables in the 2005 survey.

	Variable name	Source or procedure used for collection	The form	The method used for obtaining the target variable
Total household gross income	HY010	The register database, the IDS/EU-SILC interview	Gross value	The sum for all household members of gross personal income components (PY010G, PY020G (company cars), PY030G, PY50G, PY070G, PY090G, PY100G, PY110G, PY120G, PY130G, PY140G) plus gross income components at household level (HY040G, HY050G, HY060G, HY070G, HY080G, HY090G, HY110G)
Total disposable household income	HY020	The register database, the IDS/EU-SILC interview	Net value	The sum for all household members of gross personal income components (PY010G, PY020G (company cars), PY030G, PY50G, PY070G, PY090G, PY100G, PY110G, PY120G, PY130G, PY140G) plus gross income components at household level (HY040G, HY050G, HY060G, HY070G, HY080G, HY090G, HY110G) minus regular taxes on wealth (HY120G), regular inter-household cash transfers paid (HY130G), tax on income and social insurance contributions (HY140G)
Total disposable household income, before social transfers other than old-age and survivors' benefits	HY022	The register database, the IDS/EU-SILC interview	Net value	The total disposable income (HY020) minus **total gross to net converted transfers** of unemployment benefits (PY090G), sickness benefits (PY120G), disability benefits (PY130G), education-related allowances (PY140G), family/children-related allowances (HY050G), social exclusion not elsewhere classified (HY060G) and housing allowances (HY070G)
Total disposable household income, before social transfers including old-age and survivors' benefits	HY023	The register database, the IDS/EU-SILC interview	Net value	The total disposable income (HY020) minus **total gross to net converted transfers** of unemployment benefits (PY090G), old-age benefits (PY100G), survivors' benefits (PY110G), sickness benefits (PY120G), disability benefits (PY130G), education-related allowances (PY140G), family/children-related allowances (HY050G), social exclusion not elsewhere classified (HY060G) and housing allowances (HY070G)
Imputed rent	HY030G	The external data source is Rent statistics for which information is collected by monthly Labour Force Survey interviews (the whole sample size is 12,000), and from register sources maintained by Statistics Finland. Rent statistics are compiled by a conventional method based on classification and regression analysis (hedonic method). The available data from the statistics include mean rents/m2 for dwellings in different sizes, types, and areas. Source for repurchase prices: Federation of Finnish Insurance Companies, Finland's Tax Act The IDS/EU-SILC interviewed data. The HBS interviewed data (for estimating	Gross value	Information about mean rent / m2 (incl. utility costs which is not separable from "the space rent", incl. new and old contracts) of privately financed rented dwellings was imputed from the Rent statistics for the floor area of the sample households' main dwelling by using the following strata: - Statistical grouping of municipalities (urban / other) - Number of rooms (1, 2, 3, 4+) - Type of building (detached houses with 1-2 dwellings and other type of buildings, semi-detached or terraced house, block of flats) - Construction or renovation year (-60, 61-70, 71-80, 81-90, 91-) Since the base year, the mean rent (i.e. a price index) of the Rent Statistics by statistical grouping of municipalities has been annually extrapolated to the

		insurance for detached houses)		base year rents by the strata, and imputed to the equivalent sample dwellings. To obtain the value of imputed rent, costs on housing the household actually paid (rents, maintenance and repair of the dwelling, electricity, gas and other fuels, incl. subsidies received for them) were subtracted from the value. Further, depreciation of detached houses was imputed for the equivalent dwellings by stratifying, and subtracted from the value. Depreciation was imputed to detached houses according to the following strata: - Statistical grouping of municipalities (urban / other) - Floor area m2 available to households (<60, 60-89, 90-124, 124-) - Construction or renovation year (-50, 50-64, 64-) - Construction material (wood, other)
Income from rental of property or land	HY040G	Register database	Gross value	
Family/children-related allowances	HY050G	Items either from the Register database or from the IDS/EU-SILC interview	Gross value	
Social exclusion payments not elsewhere classified	HY060G	Items either from the Register database or from the IDS/EU-SILC interview	Gross value	
Housing allowances	HY070G	Items either from the Register database or from the IDS/EU-SILC interview	Gross value	
Regular inter-household cash transfers received	HY080G	The IDS/FI-SILC interview	Gross value	
Interest, dividends, profit from capital investments in unincorporated businesses	HY090G	Items either from the Register database or from the IDS/EU-SILC interview	Gross value	Item non-responses of interest income taxed at source were imputed for the households who responded in the interview that they had received the income during the income reference year, but did not specify the amount. The hot-deck method was used for imputing. Grouping variables were the socio-economic group of the reference person and the number of members in the household.
Interest paid on mortgages	HY100G	Register database	Gross value	
Income received by people aged under 16	HY110G	Register database	Gross value	
Regular taxes on wealth	HY120G	Register database	Gross value	The item of tax on real property was edited by using data on the taxes paid on the real property owned in the year (t-1) before the income reference period (t) and data on change of taxable value of the real property from the year t-1 to the income reference period t.
Regular inter-household transfers paid	HY130G	The IDS/EU-SILC interview	Gross value	
Tax on income and social insurance contributions	HY140G	Register database	Gross value	
Repayments/receipts for tax adjustments	HY135G	-	-	-
Cash or near-cash employee income	PY010G	Register database	Gross value	
Non-cash employee income	PY020G	Register database	Gross value	
Employers' social insurance contributions	PY030G	-	-	-
Cash profits or losses from self-employment (including royalties)	PY050G	Register database, the IDS/EU-SILC interview, the external data source HBS	Gross value	The component includes the gross item of timber selling as earned and capital forestry income, for which expenses were imputed by using the parameter estimates of the regression model of the expenses based on the external data source (HBS). For imputing the parameters of the expenses values, the following variables were used in a linear scale: - Forestry income from timber selling - Subsidies for forest improvement - Forestry levy - Forest area
Value of goods produced for own consumption	PY070G	-	-	-
Unemployment benefits	PY090G	Register database	Gross value	
Old-age benefits	PY100G	Register database and the IDS/EU-SILC interview data	Gross value	Survivors' benefits and disability benefits which were received simultaneously with old-age benefits were regrouped into old-age benefits by using the retirement age of the national pension scheme, which is 65. This was not done if a person was on part-time pension, and still at work.
Survivors' benefits	PY110G	Register database	Gross value	
Sickness benefits	PY120G	Register database	Gross value	
Disability benefits	PY130G	Register database	Gross value	
Education-related	PY140G	Register database and the IDS/EU-SILC interview	Gross	

allowances			value	
Gross monthly earnings for employees	PY200G	-	-	-

4 Coherence

4.1 Comparison of Income Target Variables and Number of Persons Who Receive Income from Each Income Component with External Sources

Table 4.1 shows results from income comparisons with relevant data sources. They are the Income Distribution Statistics (IDS), Total Statistics on Income Distribution (TSID) and National Accounts (NA) by Statistics Finland. The IDS is the primary national source for the household income statistics. The TSID is compiled from the Total Income Database (TIDB) which is used as a register income source both for the IDS and EU-SILC. The EU-SILC comparisons with these two statistics (IDS, TIDB) have been done in more detail in the following tables (Tables 4.1, 4.2, 4.3).

Social transfers received are compared with the social expenditure on cash benefits by main group from the European System of Integrated Social Protection Statistics (ESSPROS) compiled by the National Research and Development Centre for Welfare and Health (STAKES), Finland.

The differences on total income amounts across the statistics are mostly due to differences in items defined to the components. These are described in Table 4.1. Almost all of the income information was collected from the TIDB to the EU-SILC sample units. Further, the EU-SILC data were estimated to the private households by using information on crucial demographic and income variables from the TIDB in the sampling and the weightings (Section 2.1). Therefore, inconsistencies between the estimated EU-SILC and TSID income are primarily resulting from the unit-non responses among the units having received certain type of register-based income not used in the weightings (see below). Interviewed information completes the register information on income, and as a result from this part, the income is slightly more complete in the EU-SILC than in the TSID.

The EU-SILC and IDS income data were compiled in the integrated statistical survey. The sample and the frame households were the same. All differences between these two statistics are due to income definitions.

Compared with the ESSPROS (Table 4.1) and with the TSID in more detail (Table 4.2), the under-coverage of the EU-SILC income components is related to certain social transfers received. These are housing allowances (HY070G) and social exclusion not elsewhere classified (HY060G), both focused on low income households as means-tested income, and therefore, having relevancy to the indicators on poverty and social exclusion. The income components were not considered in the calibration.

The amount on register-based housing allowances is 8.3 per cent lower in the EU-SILC than in the TSID. Of housing allowances, general housing allowance as the largest item is 6.4 per cent lower, and of other items, students' housing supplements is 5.9 per cent lower and pensioners' housing allowances 13.9 per cent lower in the EU-SILC than in the TSID.

Social assistance is the main item consisting of 94 per cent of income on social exclusion not elsewhere classified. Compared with the TSID, the EU-SILC amount is 14.2 per cent lower.

The differences from comparing income recipients by main income components in Table 4.3 follows from the same factors as the differences in total income sums. Further, the EU-SILC and IDS household keeping unit definition differs from the TSID household dwelling unit definition. This also has an effect on the figures.

Table 4.1 The total gross income of private households in the income reference year 2004 according to different data sources

	EU-SILC	IDS	Difference	Difference	Notes
Income components	Sum (EUR 1 000)	Sum (EUR 1 000)	Sum (EUR 1 000)	%	
2.1. Gross employee income	57 717 076	58 005 105	-288 028	-0.5	The IDS includes employee income received by those aged under 16, and other non-cash employee income than company car. Additionally, small differences in focusing costs of expenses of taxable earned income on income components.
2.2. Self-employment income	5 419 840	5 372 680	47 160	0.9	Small differences in forestry income definitions.
2.3. Imputed rent	..	..			Not a separate income component in the IDS.
2.4. Property income	7 102 987	15 279 513	.	.	The IDS includes imputed rent (incl. interest payments), profits from sales, and an item counted as social benefits (e.g. criminal liability substitution paid by the State Treasury) in the EU-SILC
excl. imputed rent	7 102 987	8 921 975	-1 818 988	-20.4	
2.5. Current transfers received	24 614 069	25 268 196	-654 126	-2.6	The IDS includes imputed rent from other household. Inter-household transfers received have been defined as more widely in the IDS than the EU-SILC.
excl. imputed rent	24 614 069	25 058 644	-444 575	-1.8	
2.6. Other income received	111 931	..	.	.	The income is included in other IDS income component.
2.7. Interest payments	1 188 389	..	.	.	
2.8. Current transfers paid	25 536 448	25 148 509	387 939	1.5	The IDS does not include inter-household transfers paid except compulsory child support. Taxes paid on profits from sales and deduction due to voluntary payments done by persons themselves to personal private pensions plans are included in the IDS
Total disposable household income (incl. imputed rent)	..	78 777 045	.	.	
Total disposable household income (excl. imputed rent, positive values)	69 517 400	72 210 471	-2 693 071	-3.7	
	EU-SILC	TSID	Difference	Difference	
Income components	Sum (EUR 1 000)	Sum (EUR 1 000)	Sum (EUR 1 000)	%	
2.1. Gross employee income	57 717 076	58 220 301	-503 224	-0.9	The TSID includes employee income received by those aged under 16, and other non-cash employee income than company car.
2.2. Self-employment income	5 419 840	5 249 414	170 426	3.2	Small differences in forestry income definitions.
2.3. Imputed rent	..	..			
2.4. Property income	7 102 987	8 784 608	-1 681 621	.	The TSID does not include certain interests income.
2.5. Current transfers received	24 614 069	23 955 888	658 182	2.7	The TSID does not include all inter-household transfers received.
2.6. Other income received	111 931	..	.	.	The income is included in other income components TSID.
2.7. Interest payments	1 188 389	..	.	.	
2.8. Current transfers paid	25 536 448	25 024 254	512 194	2.0	The TSID does not include inter-household transfers paid. Tax paid on profits from sales is counted in the TSID.
Total disposable household income (incl. imputed rent)	..	..			
Total disposable household income (excl. imputed rent, positive values)	69 517 400	71 196 893	-1 679 493	-2.4	The TSID does not include the EU-SILC and IDS interviewed income.
	EU-SILC	NA	Difference	Difference	
Income components	Sum (EUR 1 000)	Sum (EUR 1 000)	Sum (EUR 1 000)	%	
2.1. Gross employee income	57 717 076	58 989 029	-1 271 953	-0.9	Differences, e.g. employee income received by those aged under 16, and other non-cash employee income than company car is included in the NA.
2.2. Self-employment income	5 419 840	6 402 483	-982 644	3.2	
2.3. Imputed rent	..				
2.4. Property income	7 102 987	9 408 827	.	.	
excl. imputed rent	7 102 987	5 481 526	1 621 461	29.6	The NA does not include avoir fiscal tax credit and all interest items. Income from voluntarily taken insurances is counted in current transfers received.
2.5. Current transfers received	24 614 069	26 797 920	-2 183 851	-8.1	The NA does not include inter-household transfers received.
2.6. Other income received	111 931		111 931	.	The income is included in other income components in the NA.
2.7. Interest payments	1 188 389		1 188 389	.	
2.8. Current transfers paid	25 536 448	25 885 000	-348 552	-1.3	The NA does not include inter-household transfers paid.

	EU-SILC	ESSPROS	Difference	Difference	
Income components	Sum (EUR 1 000)	Sum (EUR 1 000)	Sum (EUR 1 000)	%	
PY090G. Unemployment benefits	3 532 417	3 483 570	48 847	1.4	
PY100G. Old-age benefits	12 531 961	11 604 384	927 577	8.0	The ESSPROS includes pensioners' housing allowances, it does not include income received from PY110G and PY130G for the persons who are on old-age pensions after the standard age
PY110G. Survivors' benefits	389 972	1 441 662	-1 051 690	-72.9	See PY100G
PY120G. Sickness benefits	422 707	1 890 945	-1 468 238	-77.6	The ESSPROS includes sickness benefits paid as employee income
PY130G. Disability benefits	3 110 190	3 821 364	-711 174	-18.6	See PY100G
PY140G. Education-related allowances	547 903	..	.	.	
HY050G. Family/children-related allowances	2 511 613	2 499 220	12 393	0.5	
HY060G. Social exclusion payments not elsewhere classified	408 519	454 971	-46 452	-10.2	
HY070G. Housing allowances	849 188	436 000	413 188	94.8	See PY100G. The ESSPROS does not include students' housing supplements.
Total, excl. education-related allowances	23 756 569	25 632 116	-1 875 547	-7.9	
Same definitions in accordance with ESSPROS:					
HY070G. Housing allowances	407 994	454 971	-46 977	-10.3	
PY100G,PY110G,PY130G	16 262 513	16 867 410	-604 897	-3.6	
Total, excl. all education-related allowances	23 545 765	25 632 116	-2 086 351	-8.1	

.. Information is not available
. Information is not logical

Table 4.2 Income items of social exclusion payment not elsewhere classified (HY060G) and housing allowances (HY070G) in the income reference period 2004 in the EU-SILC and TSID

HY060G	EU-SILC		TSID	
Income item	Mean	Sum (EUR 1 000)	Mean	Sum (EUR 1 000)
Social assistance	153	370 673	177	432 159
Conscript's allowance	4	9 087	6	14 218
Special assistance for immigrants	12	27 935	6	14 542
Daily allowance from liability insurance	0	255	0	1 088
Strike assistance	0	0	0	0
Strike assistance (interviewed)	0	317	.	.
HY070G	EU-SILC		TSID	
Income item	Mean	Sum (EUR 1 000)	Mean	Sum (EUR 1 000)
General housing allowance	168	405 860	178	433 777
Students' housing supplements	87	210 804	92	224 097
Pensioners' housing allowance	94	227 456	108	264 873
Spouse pensioners' housing allowance	1	2 934	1	2 842
Other (interviewed)	1	2 134	.	.

.. Information is not available
. Information is not logical

Table 4.3 The number of income recipients in the income reference period 2004 according to different data sources

	EU-SILC	IDS	Difference	EU-SILC	IDS	Difference
Income components	Households (1 000)	Households (1 000)	%	Persons (1 000)	Persons (1 000)	%
2.1. Gross employee income	1 666	1 668	-0.1	2 646	2 687	-1.5
2.2. Self-employment income	392	375	4.5	477	448	6.1
2.3. Imputed rent	..	..		..	..	..
2.4. Property income	1 533	1 892	.	.	.	.
excl. imputed rent	1 533	1 558	-1.6	.	.	.
2.5. Current transfers received	2 025	2 108	-4.1	.	.	.
excl. imputed rent	2 025	2 105	-3.9	.	.	.
2.6. Other income received	57	..	.	.	.	.
2.7. Interest payments	720	..	.	.	.	.
2.8. Current transfers paid	2 368	2 367	0.1	.	.	.
	EU-SILC	TSID	Difference	EU-SILC	TSID	Difference
Income components	Households (1 000)	Households (1 000)	%	Persons (1 000)	Persons (1 000)	%
2.1. Gross employee income	1 666	1 679	-0.8	2 650	2 685	-1.3
2.2. Self-employment income	392	366	6.7	491	461	6.1
2.3. Imputed rent	..	..		..	..	..
2.4. Property income	1 533	1 355	11.6	.	.	.
2.5. Current transfers received	2 025	1 991	1.7	.	.	.
2.6. Other income received	57	..	.	.	.	.

2.7. Interest payments	720	..	.	.	.	.
2.8. Current transfers paid	2 368	2 360	0.4	.	.	.

.. Information is not available
. Information is not logical