

INTERMEDIATE QUALITY REPORT

EU-SILC 2005 OPERATION POLAND

Warsaw, December 2006

CONTENTS	Page
1. COMMON CROSS-SECTIONAL EUROPEAN UNION INDICATORS.....	3
2. ACCURACY	5
2.1. Sample design	5
2.2. Sampling errors	8
2.3. Non-sampling errors.....	8
2.4. Mode of data collection.....	15
2.5. Imputation procedure	16
3. COMPARABILITY	16
3.1. Basic concepts and definitions	16
3.2. Components of income.....	18
4. COHERENCE	19
4.1. Comparison of the selected Laeken indicators based on EU-SILC and HBS.....	20

1. COMMON CROSS-SECTIONAL EUROPEAN UNION INDICATORS

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

	Indicator	Value
1	At-risk-of-poverty rate after social transfers - total	21
2	At-risk-of-poverty rate after social transfers - men total	21
3	At-risk-of-poverty rate after social transfers - women total	20
4	At-risk-of-poverty rate after social transfers - 0-15 years	29
5	At-risk-of-poverty rate after social transfers - 16-24 years	26
6	At-risk-of-poverty rate after social transfers - 25-49 years	21
7	At-risk-of-poverty rate after social transfers - 50-64 years	16
8	At-risk-of-poverty rate after social transfers - 65+ years	7
9	At-risk-of-poverty rate after social transfers - 16+ years	19
10	At-risk-of-poverty rate after social transfers - 16-64 years	21
11	At-risk-of-poverty rate after social transfers - 0-64 years	23
12	At-risk-of-poverty rate after social transfers - men 16-24 years	25
13	At-risk-of-poverty rate after social transfers - men 25-49 years	22
14	At-risk-of-poverty rate after social transfers - men 50-64 years	19
15	At-risk-of-poverty rate after social transfers - men 65+ years	5
16	At-risk-of-poverty rate after social transfers - men 16+ years	20
17	At-risk-of-poverty rate after social transfers - men 16-64 years	22
18	At-risk-of-poverty rate after social transfers - men 0-64 years	23
19	At-risk-of-poverty rate after social transfers - women 16-24 years	27
20	At-risk-of-poverty rate after social transfers - women 25-49 years	21
21	At-risk-of-poverty rate after social transfers - women 50-64 years	14
22	At-risk-of-poverty rate after social transfers - women 65+ years	9
23	At-risk-of-poverty rate after social transfers - women 16+ years	18
24	At-risk-of-poverty rate after social transfers - women 16-64 years	20
25	At-risk-of-poverty rate after social transfers - women 0-64 years	22
26	At-risk-of-poverty rate after social transfers - employed	8
27	At-risk-of-poverty rate after social transfers - unemployed	46
28	At-risk-of-poverty rate after social transfers - retired	11
29	At-risk-of-poverty rate after social transfers - other inactive	27
30	At-risk-of-poverty rate after social transfers - men, employed	9
31	At-risk-of-poverty rate after social transfers - men, unemployed	48
32	At-risk-of-poverty rate after social transfers - men, retired	11
33	At-risk-of-poverty rate after social transfers - men, other inactive	26
34	At-risk-of-poverty rate after social transfers - women, employed	7
35	At-risk-of-poverty rate after social transfers - women, unemployed	43
36	At-risk-of-poverty rate after social transfers - women, retired	10
37	At-risk-of-poverty rate after social transfers - women, other inactive	27
38	At-risk-of-poverty rate after social transfers - single, < 65 years	26
39	At-risk-of-poverty rate after social transfers - single, 65+ years	7
40	At-risk-of-poverty rate after social transfers - single, male	25
41	At-risk-of-poverty rate after social transfers - single, female	12
42	At-risk-of-poverty rate after social transfers - single, total	17
43	At-risk-of-poverty rate after social transfers - 2 adults, no children, both < 65	14
44	At-risk-of-poverty rate after social transfers - 2 adults, no children, at least one 65+	6
45	At-risk-of-poverty rate after social transfers - other households without children	14

Indicator		Value
46	At-risk-of-poverty rate after social transfers - single parent, at least one child	39
47	At-risk-of-poverty rate after social transfers - 2 adults, 1 child	17
48	At-risk-of-poverty rate after social transfers - 2 adults, 2 children	23
49	At-risk-of-poverty rate after social transfers - 2 adults, 3+ children	45
50	At-risk-of-poverty rate after social transfers - other households with children	22
51	At-risk-of-poverty rate after social transfers - households without children	13
52	At-risk-of-poverty rate after social transfers - households with children	25
53	At-risk-of-poverty rate after social transfers - owner or rent-free	20
54	At-risk-of-poverty rate after social transfers – tenant	25
55	At-risk-of-poverty rate after social transfers - households without children, w = 0	24
56	At-risk-of-poverty rate after social transfers - households without children, 0 < w < 1	14
57	At-risk-of-poverty rate after social transfers - households without children, w = 1	8
58	At-risk-of-poverty rate after social transfers - households with children, w = 0	62
59	At-risk-of-poverty rate after social transfers - households with children, 0 < w < 0.5	43
60	At-risk-of-poverty rate after social transfers - households with children, 0.5 < w < 1	22
61	At-risk-of-poverty rate after social transfers - households with children, w = 1	15
62	Median of the equivalised disposable household income	11467 PLN
63	At-risk-of-poverty threshold - single	6880 PLN
64	At-risk-of-poverty threshold - 2 adults, 2 children	14448 PLN
65	Inequality of income distribution S80/S20 income quintile share ratio	6,6
66	Relative median at-risk-of-poverty gap – total	30
67	Relative median at-risk-of-poverty gap - men total	31
68	Relative median at-risk-of-poverty gap - women total	30
69	Relative median at-risk-of-poverty gap - 0-15 years	32
70	Relative median at-risk-of-poverty gap - 16-64 years	30
71	Relative median at-risk-of-poverty gap - 65+ years	17
72	Relative median at-risk-of-poverty gap - 16+ years	29
73	Relative median at-risk-of-poverty gap - men, 16-64 years	31
74	Relative median at-risk-of-poverty gap - men, 65+ years	19
75	Relative median at-risk-of-poverty gap - men, 16+ years	30
76	Relative median at-risk-of-poverty gap - women, 16-64 years	30
77	Relative median at-risk-of-poverty gap - women, 65+ years	16
78	Relative median at-risk-of-poverty gap - women, 16+ years	29
79	Dispersion around the risk-of-poverty threshold - 40%	9
80	Dispersion around the risk-of-poverty threshold - 50%	15
81	Dispersion around the risk-of-poverty threshold - 70%	28
Before social transfers except old-age and survivors' benefits		
82	At-risk-of-poverty rate before social transfers – total	30
83	At-risk-of-poverty rate before social transfers - men total	31
84	At-risk-of-poverty rate before social transfers - women total	29
85	At-risk-of-poverty rate before social transfers - 0-15 years	39
86	At-risk-of-poverty rate before social transfers - 16-64 years	31
87	At-risk-of-poverty rate before social transfers - 65+ years	11
88	At-risk-of-poverty rate before social transfers - 16+ years	28
89	At-risk-of-poverty rate before social transfers - men, 16-64 years	32
90	At-risk-of-poverty rate before social transfers - men, 65+ years	8
91	At-risk-of-poverty rate before social transfers - men, 16+ years	29
92	At-risk-of-poverty rate before social transfers - women, 16-64 years	30
93	At-risk-of-poverty rate before social transfers - women, 65+ years	13
94	At-risk-of-poverty rate before social transfers - women, 16+ years	27

Indicator		Value
Before social transfers including old-age and survivors' benefits		
95	At-risk-of-poverty rate before social transfers – total	51
96	At-risk-of-poverty rate before social transfers - men total	49
97	At-risk-of-poverty rate before social transfers - women total	52
98	At-risk-of-poverty rate before social transfers - 0-15 years	45
99	At-risk-of-poverty rate before social transfers - 16-64 years	45
100	At-risk-of-poverty rate before social transfers - 65+ years	88
101	At-risk-of-poverty rate before social transfers - 16+ years	52
102	At-risk-of-poverty rate before social transfers - men, 16-64 years	45
103	At-risk-of-poverty rate before social transfers - men, 65+ years	88
104	At-risk-of-poverty rate before social transfers - men, 16+ years	50
105	At-risk-of-poverty rate before social transfers - women, 16-64 years	45
106	At-risk-of-poverty rate before social transfers - women, 65+ years	88
107	At-risk-of-poverty rate before social transfers - women, 16+ years	53
108	Gini coefficient	36
109	Mean equivalised disposable income	13761 PLN

2. ACCURACY

2.1. Sample design

Type of sampling design

The two-stage sampling scheme with different selection probabilities at the first stage was used. Prior to selection, sampling units were stratified.

Sampling units

The primary sampling units (PSU) were enumeration census areas, while at the second stage dwellings were selected. All the households from the selected dwellings are supposed to enter the survey.

Stratification and substratification

The stratification was based on the administrative division into voivodships (NUTS2) and then within voivodships primary sampling units were classified by class of locality. In urban areas enumeration census areas were grouped by size of town, but in the five largest cities districts were treated as strata. In rural areas strata were represented by rural gminas (NUTS5) of a subregion (NUTS3) or of a few neighbouring poviats (NUTS4). Altogether 211 strata were distinguished.

Sample size and allocation criteria

It was decided that the sample should include about 24 000 dwellings¹. Proportional allocation of dwellings to particular strata was applied. The number of dwellings selected from a particular stratum was in proportion to the number of dwellings in the stratum. Furthermore, the number of the first-stage units selected from the strata was obtained by dividing the number of dwellings in the sample by the number of dwellings determined for a given class of locality to be selected from the first-stage unit. In towns with over 100 000 population 3 dwellings per PSU were selected, in towns with 20-100 thousand population – 4 dwellings per PSU, in towns with less than 20 000 population – 5 dwellings per PSU, respectively. In rural areas 6 dwellings from each PSU were selected. Altogether 5912 enumeration census areas and 24044 dwellings were selected for the sample.

Sample selection schemes

Census areas were selected according to the Hartley-Rao scheme. Prior to selection, census areas were put in random order, for each stratum separately and then the determined number of PSU was selected with probabilities proportional to the number of dwellings. Then in each of the census areas belonging to a given PSU sample dwellings were selected using the simple random selection procedure.

Renewal of sample: rotational groups

The selected sample of first-stage units was divided into four subsamples, equal in size. Starting from 2006, one of the subsamples is eliminated and replaced with another one, selected independently as described above.

Weightings

Design factor

Design factor – DB080 is equal to the dwelling sampling fraction reciprocal in the h-th stratum i.e.

$$f_h = \frac{n_h * m'_h}{M_h},$$

$$DB080 = \frac{1}{f_h}$$

where:

n_h - number of PSU selected from the h-th stratum,

m'_h - number of dwellings selected from PSU in the h-th stratum,

M_h – number of dwellings in the h-th stratum.

¹In 2005 the real gross sample size amounted to 24 thousand dwellings. It should be pointed out, however, that following Eurostat's decision the sample of 15000 households was adopted for the estimation of 2005 EU-SILC costs to be co-financed by the European Commission.

Non-response adjustments

DB080 weights were then adjusted with the use of completeness indicator, estimated for each class of locality separately:

Code of class of locality (p)	Class of locality	Completeness rate (cr _p)
	Poland	0.706
1	Warsaw	0.490
2	Towns 500 000 – 1 000 000 inhabitants	0.524
3	Towns 100 000 – 500 000 inhabitants	0.679
4	Towns 20 000 – 100 000 inhabitants	0.702
5	Towns less than 20 000 inhabitants	0.723
6	Rural areas	0.816

The adjusted weights were calculated according to the formula

$$DB080_p^{corrected} = \frac{DB080_p}{cr_p},$$

Adjustments to external data

Using the integrated calibration method (in hyperbolic sinus version) weights were calculated for individuals and for households simultaneously. To do this, the information about households was used (4 size categories: 1-person, 2-person, 3-person and 4- and more person households) and number of persons by age and gender (15 age groups: under 16, 16-19 years, then eleven 5-year groups, 75 years and over). This information at the level of NUTS2, additionally classified by urban/rural areas, was derived from the 2002 Census and current demographic estimates.

Final cross-sectional weight

In EU-SILC 2005 the following cross-sectional weights were calculated:

DB090 – weight for households,

RB050 – weight for all household members,

$$RB050_{ij} = DB090_i$$

where:

i – household number,

j – person number in the i-th household.

PB040 – weight for respondents at the age of 16 and over who had individual interview. This weight is obtained by the adjustment of RB050 weight separately in the gender and age groups in each voivodship according to urban and rural area,

PM005 – weight for people at the age of 25–65 years. PM005 = RB050,

RL070 – weight for children at the age of 0–12 years. It is obtained by the adjustment of RB050 weight in 26 groups, i.e. 13 years of birth and gender.

Substitutions

No substitution was applied if the household did not enter the survey.

2.2. Sampling errors

Standard error and effective sample size

Estimation of standard errors was based on a resampling approach. We used a bootstrap method which resamples 200 times from each stratum $n_h - 1$ PSUs (primary sampling units) with replacement (McCarthy and Snowden method (1985)), where n_h denotes the sample size of PSUs in the h th stratum. After resampling the original weights were properly rescaled and bootstrap variance estimate of the corresponding indicator was obtained by the usual Monte Carlo approximation based on the independent bootstrap replicates. Computations were carried out using SAS software. Additionally, the linearisation method of variance estimation for the main poverty indicators was applied and the results were very similar to those obtained by the bootstrap method.

Indicator	Value	Standard error	Achieved sample size	Design effect	Effective sample size
At-risk-of-poverty rate after social transfer	20.53	0.39	49044	3.72	13184
S80/S20 income quintile share ratio	6.64	0.12	49044	3.69	13291
Relative median at-risk-poverty gap	30.06	0.73	49044	2.82	17391
Gini coefficient	35.59	0.36	49044	3.50	14013
Mean equivalised disposable income	13761	1.08	49044	3.39	14467

2.3. Non-sampling errors

Sampling frame and coverage errors

The sample for EU-SILC 2005 was selected from the sampling frame based on the TERYT system, i.e. *Domestic Territorial Division Register*.

Two types of primary sampling units were distinguished:

- about 178 00 census enumeration areas (CEA), containing on average 68 dwellings each;
- about 33 00 enumeration statistical districts (ESD), containing on average 377 dwellings each.

The whole territory of Poland was divided into enumeration statistical districts and enumeration census areas. In EU-SILC enumeration census areas are used as primary sampling units. The secondary sampling units are dwellings. For each CEA the list of dwellings was set up to provide the secondary sampling frame. All the households in the selected dwellings were surveyed.

The TERYT system is updated every year. The updating concerns the territorial division of the country into districts and areas. The lists of dwellings, names of the towns/villages and streets are updated too. Some other changes resulting from the construction of new buildings, dismantle of the old ones and changes in the territorial division are also introduced.

The sample for EU-SILC 2005 was selected in September 2004 from the sampling frame which had been updated as of January 1, 2004. About 6.8% of the selected dwellings were found to be non-existing (cancelled, changed into non-residential places), uninhabited or temporarily inhabited.

Measurement and processing errors

As with any other statistical survey, EU-SILC can be burdened with non-sampling errors which occur at various stages of the survey and which cannot be totally eliminated.

According to the interviewers, who after the household and individual interview completion were obliged to answer a few questions concerning interview performance, over 70% of the respondents showed a favourable attitude towards the survey, while about 5% were unwilling towards it. In the interviewers' opinion, in about 86% of questionnaires (both household and individual ones) the quality of non-income data collected could be recognised as good or very good and in 1% - as doubtful. The quality of income data was evaluated as slightly worse, mainly because of item non-response.

Measurement and processing errors will be subject to a more detailed analysis in the final report. Below we only give a synthetic review of the survey organisation and measures taken to reduce different types of errors.

In Poland EU-SILC was carried out in May/June 2005.

EU-SILC is a non-obligatory, representative survey of individual households, performed by a face-to-face interview technique with the use of paper form questionnaires (the so called PAPI method). Two types of questionnaire: individual and household questionnaire were applicable. The main survey was preceded by the pilot survey carried out in 2004 and the questionnaire testing (November/December 2003).

The organisation and performance of the survey in the field was within the responsibility of regional statistical offices. Most of the interviewers were regular employees of the statistical offices having experience in other social surveys. Survey performance in the field was preceded by a series of trainings. Regional survey coordinators were instructed by CSO Social Statistics Division staff members and then the regional survey coordinators trained interviewers at the regional statistical offices.

Interviewers' visits to households were preceded by the introductory letter of the CSO President.

The interviewers received written instructions concerning the survey performance.

Small gifts were given to the families participating in the survey. Each statistical office chose the type of gift for its respondents.

Data recording and check-up took place in regional statistical offices and was done with the use of Microsoft Visual FoxPro. After all the questionnaires for a given household had been recorded (the identifiers being voivodship number, dwelling number and household number), it was possible to make the household screening which consisted of logical and calculation check-up at the section, inter-section and inter-questionnaire levels. The regional files were then transferred to the CSO Computing Centre and combined together to make up the general files at the national level. The national file completeness was also checked with the use of Microsoft Visual FoxPro. Additional check-up was made with SAS checking programmes. On the basis of overall data files it was possible to create files for Eurostat. Some of the primary target variables could be found directly in the questionnaires, others had to be calculated with the algorithms especially prepared for this purpose. Tables of EU-SILC results were compiled with the use of: SAS, SPSS, Microsoft Visual FoxPro.

Non-response errors

Achieved sample size

Sample size	Rotational group				
	1	2	3	4	Total
A	4116	4026	4053	4068	16263
B	10124	9810	9907	9825	39666
C	12519	12134	12274	12117	49044

A - number of households for which an interview is accepted for the database

B - number of persons at the age of 16 years or more who are members of the households for which the interview is accepted for the database, and who completed an individual interview.

C - number of selected persons who are members of the households for which the interview is accepted for the database, and who completed an individual interview.

Unit non-response

- Household non-response rates $NRh = [1 - (Ra * Rh)] * 100$,

$Ra = 0,992$

$Rh = 0,706$

Ra – the address contact rate

Rh – the proportion of complete household interviews accepted for the database

$NRh = 29,96$

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

Rp = 0,950
NRp = 5,03

Rp – the proportion of complete personal interviews within the households accepted for the database

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

*NRp = 33,47.

Distribution of households

- DB120 - Contact at address

DB120	Rotational group				
	1	2	3	4	Total
Address contacted (11)	5567	5530	5561	5580	22238
Address cannot be located (21)	42	48	40	31	161
Address impossible to access (22)	1	2	1	6	10
Address does not exist or is non-residential or is unoccupied or not the principal residence (23)	401	432	409	394	1635
Total	6011	6011	6011	6011	24044

- DB130 - Household questionnaire result

DB130	Rotational group				
	1	2	3	4	Total
Household questionnaire completed (11)	4147	4060	4088	4100	16395
Refusal to co-operate (21)	1138	1150	1135	1177	4600
Entire household temporarily away for duration of fieldwork (22)	298	341	321	279	1239
Household unable to respond (illness, incapacity,...) (23)	125	166	145	174	610
Other reasons (24)	37	55	49	57	198
Total	5745	5772	5738	5787	23042

- DB135 - Household interview acceptance

DB135	Rotational group				
	1	2	3	4	Total
Interview accepted for database (1)	4116	4026	4053	4068	16263
Interview rejected (2)	31	34	35	32	132
Total	4147	4060	4088	4100	16395

Item non-response (income variables)

Item non-response	(A)	(B)	(C)
	% of households having received an amount	% of households with missing values	% of households with partial information
Total household gross income	40.3	6.5	52.8
Total disposable household income	67.7	5.8	26.2
Total disposable household income before social transfers other than old-age and survivor's benefits	67.4	8.5	22.1
Total disposable household income before social transfers including old-age and survivor's benefits	60.6	11.3	16.4
Net income components at household level			
HY040N	1.2	0.1	0.0
HY050N	23.4	0.3	0.5
HY060N	4.4	0.1	0.0
HY070N	6.0	0.2	0.0
HY080N	5.1	0.6	0.0
HY090N	1.1	0.7	0.0
HY110N	2.3	0.1	0.0
HY120N	42.8	4.7	0.0
HY130N	5.0	0.3	0.0
HY140N	40.5	33.4	22.8
HY145N	44.3	3.5	0.0
Gross income components at household level			
HY040G	1.2	0.1	0.0
HY050G	22.9	0.3	1.0
HY090G	0.6	0.7	0.5
HY110G	2.0	0.1	0.3
HY140G	40.5	33.3	23.4
	% of persons 16+ having received an amount	% of persons 16+ with missing values	% of persons 16+ with partial information
Net income components at personal level			
PY010N	29.3	6.8	0.1
PY020N	0.1	0.4	0.0
PY035N	3.0	0.6	0.0
PY050N	4.6	3.1	0.2
PY080N	0.0	0.0	0.0
PY090N	3.6	0.3	0.0
PY100N	22.4	1.4	0.4
PY110N	1.4	0.1	0.0
PY120N	0.4	0.1	0.0
PY130N	6.3	0.3	0.0
PY140N	0.8	0.0	0.0

	% of persons 16+ having received an amount	% of persons 16+ with missing values	% of persons 16+ with partial information
Gross income components at personal level			
PY010G	16.1	6.8	13.3
PY050G	4.7	1.4	3.3
PY080G	0.0	0.0	0.0
PY090G	2.1	0.3	1.5
PY100G	15.5	1.4	7.3
PY110G	0.9	0.1	0.5
PY120G	0.1	0.1	0.3
PY130G	4.3	0.3	2.1
PY200G	28.0	3.7	0.0

Income variables' imputation methods applied

Imputation is aimed at obtaining complete records at the level of target variables. Thus the target variable level is the highest level of aggregation at which imputation can be made. This approach is applied wherever it does not cause loss of significant information from the file.

In the situation where:

- a target variable includes components of different character (e.g. taking different but highly predictable values, like benefits, or dependent on explanatory variables and thus easier to be modelled separately),
- there are many components of a target variable and it is often the case that in some of them there are missing data, while in others there are correct ones which could be lost during the imputation of the aggregated variable,

imputation is carried out at the level of particular components of target variables, frequently at the level of questionnaire variables. In some cases the target variables are identical with the questionnaire variables.

There are several methods of component imputation. They can be classified as deterministic and stochastic methods. In case of deterministic methods the method and the set of explanatory variables (algorithm) determines the imputation value for each record. In stochastic methods the imputation value is determined randomly so that with the same algorithm and the same data file each algorithm realisation may give slightly different imputation values. Although the stochastic methods slightly increase estimator variance (introducing an additional random error component), they do not distort variance or original data distribution characteristics, allowing for the correct estimation of random error. Deterministic imputation causes variable variance reduction in the file and underestimation of random error; it also distorts the correlation structure (increasing correlations with explanatory variables). According to item 2.7 of Decision 1981/2003 it is recommended that for EU-SILC imputation the methods retaining distribution characteristics should be used, which means the preference for the stochastic methods.

Out of the stochastic methods the following were used in the task presented here:

- Hot-deck method

Random selection of a representative (donor) out of the correct records.

If auxiliary categorising variables occur in the hot-deck method, a random representative is selected out of the records showing adequate values of auxiliary variables. If it proved impossible to find a donor of the equivalent values for all the auxiliary variables, the so called sequence approach was applied. The categorising variables were ranked from the most to the least significant ones. If there are no donors in the group, grouping is carried out with the subsequent explanatory variables left out, starting from the least significant ones so as to obtain a subset containing donors.

- Regression imputation with simulated residuals

Auxiliary variables are the explanatory variables of the regression model. The model takes a logarithmic form because of the income variable distribution. It is fitted on the basis of the correct records. The imputed value (its logarithm) is a sum of the theoretical value derived from the model and the pseudo-random number of the normal distribution with variance corresponding to the estimated variance of an error term in the model.

Out of the deterministic methods the following are applied:

- Regression deterministic imputation

A theoretical value from the model is taken as the imputation value.

- Deduction imputation

The imputation value is directly determined on the basis of the relationships between variables.

In the case of imputation at the target variable level or imputation of their most significant components, stochastic imputation is applied in order to retain the variable properties distribution as required by Decision 1981/2003.

The employment of regression imputation with simulated residuals requires a model which describes well the formation of a variable with relatively small variance of an error term and good statistical qualities. With high variance of an error term, there is a danger of getting accidental values which are not typical of the correct part of the data set. That is why in the majority of cases, where in accordance with the assumption referred to above stochastic imputation is required, the hot-deck method is applied. This is particularly justified when the number of records for imputation is rather low, or when the number of correct records is too small for a suitable model fitting. Regression imputation with randomly generated residuals is applied to incomes from hired employment, as:

- it is an important category of income, declared by a significant percentage of respondents and, if present, having a significant share in the total household's income,
- this category can be successfully modelled with the use of the variables included in the questionnaire,
- there is a large (absolute) number of missing data, the percentage, however, being rather small; a large number of correct records makes it possible to design a well-fitted model.

Deterministic imputation is applied where missing data concern less significant components of target variables (taxes, burdens to the main component, additions, etc.) and the main component is known. In such cases deterministic regression imputation is usually applied. Gross/net conversion is carried out with the use of the deterministic regression method. Deduction imputation is employed in rare cases of obvious relationships and can be treated as a supplementary stage of data editing.

The explanatory variables in the models and the grouping ones in the case of hot-deck method have been selected so as to represent the relationships which, according to logics and knowledge about the phenomena studied, should occur in the data set, taking into account availability of the potential variables in the questionnaire. The relationships have been tested on the file of correct data and in the majority of cases they proved to be significant. Some of the explanatory variables, when expressing an economically important relationship or providing a grouping condition (interpretation criterion) in the calculation algorithm, have been retained, even if their effect on the imputed variable has not been statistically significant.

2.4. Mode of data collection

EU-SILC is a non-obligatory, representative survey of individual households, performed by a face-to-face interview technique with the use of paper form questionnaires (the so called PAPI method). Two types of questionnaire: individual and household questionnaire were applicable.

Distribution of RB250 and RB260

- RB250 – Data status

DB250	Rotational group				
	1	2	3	4	Total
Information completed only from interview (11)	9607	9357	9397	9310	37671
Individual unable to respond (illness, incapacity, etc) (21)	38	37	36	46	157
Refusal to co-operate (23)	238	218	213	256	925
Person temporarily away and no proxy possible (31)	209	165	235	181	790
No contact for another reason (32)	32	33	26	32	123
Total	10124	9810	9907	9825	39666

- RB260 – Type of interview

RB260	Rotational group				
	1	2	3	4	Total
Face to face (1)	7750	7519	7619	7503	30391
Proxy interview (2)	1857	1838	1778	1807	7280
Total	9607	9357	9397	9310	37671

As for individual interviews, in 2005 a relatively high share (19%) of proxy interviews was noted. This was thoroughly discussed with the survey coordinators in the field.

The interviewers decided on proxy interviews only if the substitute respondents were well informed about the situation in the household and there was no other possibility to get the information. Proxy interviews were performed in the following situations:

- no contact with the respondent because of long-term absence (e.g. work in another town or abroad);
- respondent's disability, illness or pathology (such as alcoholism);

- according to other members of the household, the respondent was only available late at night and was not willing to participate in such a long interview, while at the same time the proxy could provide detailed information, even based on the documents, such as tax statements.

2.5. Interview duration

The average household interview duration was about 36 minutes, while the average individual interview duration was about 25 minutes. In total the average time needed to carry out a household interview and individual interviews with persons at the age of 16 years and over was 94 minutes.

This value exceeded significantly that assumed in the regulation, which results from the fact that in the Polish SILC all the information is collected during the interview. The questionnaire parts covering social benefits and self-employment (in and outside farming) have been expanded by many auxiliary questions which help to answer but, on the other hand, prolong the interview. We have already pointed out to the interview duration after the tests and pilot survey.

3. COMPARABILITY

3.1. Basic concepts and definitions

The reference population

There were no essential differences between the national concepts and standard EU-SILC concepts.

The survey unit was a household and all the household members who had completed 16 years of age by December 31, 2004.

The survey did not cover collective accommodation households (such as boarding house, workers' hostel, pensioners' house or monastery), except for the households of the staff members of these institutions living in these buildings in order to do their job (e.g. hotel manager, tender etc.).

The households of foreign citizens could participate in the survey.

The private household definition

There were no essential differences between the national concepts and standard EU-SILC concepts.

Household is a group of persons related to each other by kinship or not, living together and sharing their income and expenditure (multi-person household) or a single person, not sharing his/her income or expenditure with any other person, whether living alone or with other persons (one-person household).

Family members living together but not sharing their income and expenditure with other family members make up separate households.

The household size is determined by the number of persons comprised by the household.

The household membership

There were no essential differences between the national concepts and standard EU-SILC concepts.

The household composition accounted for:

- persons living together and sharing their income and expenditure who have been in the household for at least 6 months (either the real or the intended time of staying in the household should be considered),
- persons absent from the household because of their occupation, if their earnings are allocated to the household's expenditure,
- persons at the age of up to 15 years (inclusive), absent from the household for education purposes, living in boarding houses or private dwellings,
- persons absent from the household at the time of the survey, staying at education centres, welfare houses or hospitals, if their real or intended stay outside the household is less than 6 months.

The household composition did not account for:

- persons at the age of over 15 years, absent from the household for education purposes, living in boarding houses, students' hostels or private dwellings,
- men in military service (those performing substitute military service working in companies and living at home are included in the household),
- persons in prison,
- persons absent from the household at the time of the survey, staying at education centres, welfare houses or hospitals, if their real or intended stay outside the household is more than 6 months,
- persons (household's guests) staying in the household at the time of the survey who have been or intended to be there for less than 6 months,
- persons renting a room, including students (unless they are treated as household members),
- persons renting a room or bed for the time of work in a given place (including such works as land melioration, geodetic measurements, forest cut-down or building constructions),
- persons living in the household and employed as au pairs, helping personnel on the farm, craft apprentices or trainees.

The income reference period(s) used

There were no differences between the national concepts and standard EU-SILC concepts. The income reference period was last calendar year (2004).

Reference period for taxes on income and social insurance contributions

The reference period for income tax prepayment and compulsory social insurance contributions is the year 2004. The account clearance with the Treasury Office (including payments and returns) effected in 2004 refers to the income for 2003.

The reference period for taxes on wealth

There were no differences between the national concepts and standard EU-SILC concepts. Taxes on wealth paid during the income reference period (2004) were recorded.

The lag between the income reference period and current variables

The lag between the income reference period and current variables is about 5 months.

The total duration of the data collection of the sample

EU-SILC was performed on the territory of the whole country between May 2 and June 17 2005.

Basic information on activity status during the income reference period

Considering the fact that the questionnaire form for 2005 had been prepared before the methodological changes of variables were notified to us by Eurostat, the variables: Change of job since last year (PL160), Reason for change (PL170), Most recent change in the individual's activity status (PL180) - were not recorded. Starting from EU-SILC 2006, they have been taken into account. All the remaining variables were in accordance with Eurostat's requirements.

3.2. Components of income

Imputed rent

Variable was not recorded.

Interest paid on mortgages

Variable was not recorded.

Cash or near-cash employee income

This variable does not account for:

- assistance for foster families; since granting the benefit is not connected with quitting the job, this benefit has been qualified to the category of „Family related allowances' (HY050),
- benefit granted to the families when the only person providing income for the family is called up to the active military service; since this benefit is only granted when the only family supporter has been called to the military service, it has been included in the category of „Family related allowances' (HY050).

Non-cash employee income

The information collected only refers to the income gained from the use of the company car for private purposes.

Employers' social insurance contributions

Variable was not recorded.

Cash profits or losses from self-employment (including royalties)

The data on income from self-employment were collected in two different ways: the respondents were asked about the company's costs and profits and also about the amount of money gained from self-employment which was allocated to the household's expenditure. After a detailed analysis of data it was decided that the income from self-employment would be equal to the amount allocated to the household's needs.

Value of goods produced for own consumption

Variable was not recorded.

Survivors' benefits

Death grants are not included in the income because the whole sum is used to cover the cost of the funeral.

Sickness benefits

Sickness and childcare benefits are not included (a childcare benefit is granted to the working parent of a sick child), because they are paid by the employer and cannot be detached from the income from hired employment. Therefore, they are accounted for in the income from hired employment.

All the other variables not listed above

There were no divergences from common definitions.

The source or procedure used for the collection of income variables

The income data were collected during the interviews with respondents. The target income variables were split into components corresponding to particular benefits applicable in the Polish conditions.

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

The respondents were asked to give the net incomes and contributions (income tax prepayments and compulsory social insurance). Only in the case of income from rental of a property (HY040) the respondents were asked to give the gross income and the amount of tax paid.

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

The gross income was obtained by summing up net value, income tax prepayments and compulsory social insurance contributions. If the information on tax and insurance contributions was missing, the amounts were imputed on the basis of the results obtained. Only in the case of income from rental of property, the tax paid was subtracted from the gross income.

4. COHERENCE

A detailed analysis of the coherence between the results of EU-SILC 2005 and those of HBS, LFS and other sources will be presented in the final quality report. In this report a comparison was only made between the main social cohesion indicators calculated on the basis of EU-SILC 2005 and HBS. This type of indicators published so far by CSO and Eurostat were based on HBS results

4.1. Comparison of the selected Laeken indicators based on EU-SILC and HBS

The comparison of Laeken indicators calculated on the basis of EU-SILC and HBS is given below.

Indicators	HBS 2003 including income in kind	HBS 2003 SILC income definition	HBS 2004 including income in kind	HBS 2004 SILC income definition	SILC 2005
At-risk-of-poverty rate by age:					
Total (0+)	17	18	17	19	21
0-15	23	25	24	26	29
16-24	21	22	22	18	26
25-49	17	19	18	19	21
50-64	11	13	12	13	16
65+	6	8	6	7	7
16+	15	17	16	17	19
16-64	16	18	17	18	21
0-64	18	20	19	20	23
At-risk-of-poverty rate by frequent activity status:					
Total (age 16+)	15	17	16	17	19
At work	12	14	12	14	14
Unemployed	38	38	40	41	46
Retired	7	9	8	8	11
Other inactive	21	22	22	23	27
Dispersion around at-risk-of-poverty threshold:					
40% of median income	6	7	6	7	9
50% of median income	11	12	11	12	15
70% of median income	25	26	25	26	28
Income quintile ratio (S80/S20)	5.0	5.4	5.2	5.6	6.6

The data included in this table show that the change of data source from HBS to EU-SILC has a significant effect on the indicators, particularly when compared over time. As a rule, EU-SILC based indicators of income differentiation and relative poverty risk are higher. The differences between the results of the two surveys result among others from:

- different reference periods for income variables – in HBS the reference period was 1 month and, following Eurostat’s recommendation, the annual income was the monthly income multiplied by 12, which in the case of irregular income, like that from farming, could bring about considerable distortions;
- a different way of data collection – in HBS the respondents made records in the so called diary, while in EU-SILC each respondent was asked detailed questions about the income and if any items were missing, data imputation was applied;
- a different way of sample selection – in HBS households which refused to participate in the survey were replaced;

- slightly different weighting of results;
- a different reference period for labour force-related variables – in HBS the question was asked about the current situation, while in EU-SILC – about the prevailing situation during the year.