INTERMEDIATE QUALITY REPORT

EU-SILC 2006 OPERATION POLAND

Warsaw, November 2007

CONTENTS   **Page**

1. OVERARCHING INDICATORS

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

	Indicator	Value
1	At-risk-of-poverty rate after social transfers - total	19
2	At-risk-of-poverty rate after social transfers - men total	20
3	At-risk-of-poverty rate after social transfers - women total	19
4	At-risk-of-poverty rate after social transfers - 0-17 years	26
5	At-risk-of-poverty rate after social transfers - 18-64 years	19
6	At-risk-of-poverty rate after social transfers - 65+ years	8
7	At-risk-of-poverty rate after social transfers - men 18-64 years	20
8	At-risk-of-poverty rate after social transfers - men 65+ years	6
9	At-risk-of-poverty rate after social transfers - women 18-64 years	18
10	At-risk-of-poverty rate after social transfers - women 65+ years	9
11	At-risk-of-poverty threshold - single	7510 PLN
12	At-risk-of-poverty threshold - 2 adults, 2 children	15772 PLN
13	Relative median at-risk-of-poverty gap - total	25
14	Relative median at-risk-of-poverty gap - men total	26
15	Relative median at-risk-of-poverty gap - women total	24
16	Relative median at-risk-of-poverty gap - 0-17 years	27
17	Relative median at-risk-of-poverty gap - 18-64 years	25
18	Relative median at-risk-of-poverty gap - 65+ years	14
19	Relative median at-risk-of-poverty gap - men, 18-64 years	26
20	Relative median at-risk-of-poverty gap - men, 65+ years	14
21	Relative median at-risk-of-poverty gap - women, 18-64 years	25
22	Relative median at-risk-of-poverty gap - women, 65+ years	14
23	Inequality of income distribution S80/S20 income quintile share ratio	5.6
24	In work at-risk-of-poverty rate - total	13
25	In work at-risk-of-poverty rate - men total	14
26	In work at-risk-of-poverty rate - women total	11
27	Relative median income ratio people aged 65+/0-64	1.07
28	Relative median income ratio people aged 65+/0-64 - men	1.17
29	Relative median income ratio people aged 65+/0-64 - women	1.02
30	Aggregate replacement ratio pensions 65-74/earnings 50-59	0.59
31	Aggregate replacement ratio pensions 65-74/earnings 50-59 - men	0.67
32	Aggregate replacement ratio pensions 65-74/earnings 50-59 - women	0.57
	Before social transfers except old-age and survivors' benefits	
33	At-risk-of-poverty rate before social transfers - total	29
34	At-risk-of-poverty rate before social transfers - men total	30
35	At-risk-of-poverty rate before social transfers - women total	28
36	At-risk-of-poverty rate before social transfers - 0-17 years	36
37	At-risk-of-poverty rate before social transfers - 18-64 years	30
38	At-risk-of-poverty rate before social transfers - 65+ years	12
39	At-risk-of-poverty rate before social transfers - men, 18-64 years	31
40	At-risk-of-poverty rate before social transfers - men, 65+ years	9
41	At-risk-of-poverty rate before social transfers - women, 18-64 years	28
42	At-risk-of-poverty rate before social transfers - women, 65+ years	13

	Indicator	Value
	Before social transfers including old-age and survivors' benefits	
43	At-risk-of-poverty rate before social transfers - total	**49**
44	At-risk-of-poverty rate before social transfers - men total	**47**
45	At-risk-of-poverty rate before social transfers - women total	**51**
46	At-risk-of-poverty rate before social transfers - 0-17 years	**43**
47	At-risk-of-poverty rate before social transfers - 18-64 years	**43**
48	At-risk-of-poverty rate before social transfers - 65+ years	**87**
49	At-risk-of-poverty rate before social transfers - men, 18-64 years	**43**
50	At-risk-of-poverty rate before social transfers - men, 65+ years	**87**
51	At-risk-of-poverty rate before social transfers - women, 18-64 years	**44**
52	At-risk-of-poverty rate before social transfers - women, 65+ years	**87**
53	Mean equivalised disposable income	**14902 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 first-stage sampling units (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 strata were the voivodships (NUTS2) and within voivodships primary sampling units were classified by class of locality. In urban areas 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 dwellings in the first year of the survey. 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 were selected from each PSU. Altogether

5912 census areas and 24044 dwellings were selected for the sample in the first year of the survey[1]. Subsample 5 selected for the survey in 2006 to replace subsample 1 consisted of 1476 census areas and 6002 dwellings.

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 proportionate to the number of dwellings. Then in each of the census areas belonging to the 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 a new one, selected independently as described above. For the 2006 survey subsample 5 was selected as a replacement of subsample 1.

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.

---

[1] In 2006 18 494 households should be contacted (according to DB110 variable). It should be pointed out, however, that following Eurostat's decision the sample of 12 000 households was adopted for the estimation of 2006 EU-SILC costs to be co-financed by the European Commission.

Non-response adjustments

DB080 weights were then adjusted with the use of household non-response rates estimated for each class of locality separately:

Code o f class of locality (p)	Class of locality	Completeness rate (cr$_p$)
	Poland	0.699
1	Warsaw	0.399
2	Towns 500 000 – 1 000 000 inhabitants	0.567
3	Towns 100 000 – 500 000 inhabitants	0.636
4	Towns 20 000 – 100 000 inhabitants	0.695
5	Towns less than 20 000 inhabitants	0.740
6	Rural areas	0.823

The adjusted weights were calculated according to the formula:

$$DB080_p^{corrected} = \frac{DB080_p}{Ra_p * Rh_p},$$

Weights DB080 and DB080^{corrected} were calculated for subsample 5. The next step consisted in calculating the weights DB090 and RB050 for the households of subsample 5 with the use of the integrated calibration method. For the subsamples 2, 3 and 4, surveyed for the second time, the base weights were determined by the correction of the base weights from the previous year. The base weight of 2005 is equal to RB050 multiplied by 4. This weight was then adjusted by non-response and households' and individuals' falling out of the population surveyed. The calculations were made on the subsamples of the so called sample persons i.e. those who were in the surveyed sample at the age of 14 and over in 2005 and who should be surveyed in 2006. The modifying factor was determined for each subsample (2, 3 and 4) separately according to the class of locality and took the form:

$$\frac{R(1)_p - M}{R(2)_p}$$

where:

$R(1)_p$ – estimated number of respondents belonging to the sample person group in the p-th class of locality in the subsample surveyed for the t-th time,

M – estimated number of sample persons who belonged to the surveyed population in the first year and in the next year were out of the survey scope.

The base weights of 2005 were used for the calculation of numerator and denominator. The above expression is the reciprocal of the empirical estimate of probability that a given person will be interviewed again in the second year of the survey. In the second stage of the base weight calculation for the second year of the survey children of "sample persons" received the weights of mothers and "co-residents' i.e. additional persons included in the household surveyed were ascribed zero weights. Then the respondents' weights were averaged and all the members of a given household were ascribed such a mean weight. Then for the base weights thus obtained the trimming of extreme weights was applied. The last stage of calculations consisted in combining the four independent subsamples, applying the integrated calibration as described below (for sample 5 repeatedly) and trimming. As a result, DB090 and RB050 weights are obtained for households and individuals from samples 2, 3, 4 and 5.

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 2006 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 separately in the groups according to gender and age in each voivodship according to urban and rural area,

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 $n_h - 1$ PSU's (primary sampling units) with replacement (McCarthy and Snowden method (1985)), where n_h denotes the sample size of PSU 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, we implemented the linearization method of variance estimation for the main poverty indicators, and the results of comparisons with those obtained by the bootstrap method showed they were very similar.

Indicator	Value	Standard error	Achieved sample size	Design effect	Effective sample size
At-risk-of-poverty rate after social transfer	19.09	0.402	45122	4.04	11169
S80/S20 income quintile share ratio	5.64	0.083	45122	3.11	14509
Relative median at-risk-poverty gap	25.00	0.671	45122	2.96	15244
Gini coefficient	33.26	0.273	45122	2.73	16528
Mean equivalised disposable income	14902	93.35	45122	2.57	17557

2.3. Non-sampling errors

Sampling frame and coverage errors

The samples for EU-SILC 2005 and EU-SILC 2006 were selected from the sampling frame based on the TERYT system, i.e. the *Domestic Territorial Division Register*. Two kinds of primary sampling units (PSU) were distinguished in the sampling frame:

- about 178 000 *CEA – census enumeration areas* with about 68 dwellings each,
- about 33 000 *ESD – enumeration statistical districts*, with about 377 dwellings each.

The whole territory of Poland is divided into enumeration statistical districts and census enumeration areas. In EU-SILC census enumeration areas are used as primary sampling units. The secondary sampling units are dwellings. For each census enumeration area a list of dwellings was made up to form the secondary sampling frame. All the households from the selected dwellings are supposed to enter the survey.
The TERYT system is updated annually with respect to the territorial division into statistical districts and census enumeration areas. The lists of dwellings, names of towns, villages and streets are updated. Other changes due to new construction, dismantle of buildings and administrative division modifications are also introduced.

The sample for EU-SILC 2005 was selected in September 2004 from the sampling frame updated as for January 1, 2004. In the sample selected some 6.8% of dwellings were found to be non-existing (cancelled, changed for non-residential units) as well as uninhabited or temporarily inhabited, while in sample 5, selected in 2005 for the 2006 survey about 6.2% of such dwellings were recorded.

Measurement and processing errors

As with any other statistical survey, EU-SILC may be burdened with non-sampling errors which occur at various stages of the survey and which cannot be eliminated completely. This mainly applies to interviewers' errors at the stage of collecting the information, errors due to the respondents' misunderstanding of questions and inaccurate or sometimes even false answers as well as the errors taking place at the stage of data recording.

After the household and individual interview completion the respondents were obliged to answer a few questions concerning interview performance. On the basis of this material it is possible to state that about three quarters of respondents (78% of those filling in the household questionnaire and 75% of those filling in the individual questionnaire) showed a favourable attitude towards the survey, while about 3% (both in the case of the household and individual interview) were unwilling towards it. In the interviewers' opinion, in about 88% 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. It should also be pointed out that, in our opinion, the quality of data concerning net income categories is much higher than in the case of gross income. The reason is that non-response to the highest degree affected the information on taxes and social and health insurance contributions.

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 indicate the measures taken to reduce different types of errors.

In Poland EU-SILC was carried out in May/June 2006.

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 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. The interviewers received written instructions concerning the survey performance.

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

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

Data recording from the questionnaire forms was carried out with the use of Microsoft Visual FoxPro version 9 operating under the WINDOWS system. The following two applications were designed:
- The so called interviewer's application – to be used by the interviewers to record and check the data from their areas with the use of Laptops and PCs. The data were recorded on the local disk in the VFP database. After the work was completed, the data were transmitted using Web services to the MS SQL server for the national database;
- The so called server application – to be used by the staff of Statistical Offices recording the data directly for the national database and for those supervising the regional data preparation; this application was published in the CITRIX server and made accessible with the customer's software.

Both applications shared a number of modules.

The server application had a module which allowed for works (such as checking, viewing, making statements) on the national data (from all the voivodships). The national file completeness was also checked with the use of Microsoft Visual FoxPro. Additional check-up was made with SAS checking programmes.

Tables of EU-SILC results were compiled with the use of: SAS, SPSS, Microsoft Visual FoxPro.

According to the regional survey coordinators and on the basis of the analysis of errors identified at the stage of data recording and during the additional check-up of the national data files, it can be judged that in comparison with the first EU-SILC wave performed in 2005 the quality of data from EU-SILC 2006 was improved. This is because:
- EU-SILC 2006 was performed almost exclusively by professional interviewers (because of the size of sample in 2005 interviews were also carried out by the staff members of statistical office who only received some training but were not professional interviewers);
- Additional questions concerning incomes were introduced to check the period in which a given type of income was received, which reduced errors consisting in filling in monthly instead of annual incomes, but also facilitated and improved the income data imputation;
- The instructions for the interviewers were improved;
- The data checking at the recording stage was made more precise;

Moreover, with regard to the possibility of comparison of certain data from panel subsamples (such as gender, age, education completed, the year of taking the first job, vocational experience etc.) the so called file was set up and some additional guidelines introduced to allow for comparison of these variables between subsequent years of the survey.

Non-response errors

Achieved sample size

Sample size	Rotational group				
	2	3	4	5	Total
A	3605	3594	3614	4101	14914
B	8463	8453	8525	9452	34893
C	10899	10955	10947	12321	45122

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 persons who are members of the households for which the interview is accepted for the database.

Unit non-response

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

Ra = 0.997
Rh = 0.866

Ra – the address contact rate
Rh – the proportion of complete household interviews accepted for the database

NRh = 13.71

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

Rp = 0.954
NRp = 4.64

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 = 17.71

- DB120 - Contact at address

DB120	Rotational group				
	2	3	4	5	**Total**
Address contacted (11)	3916	3950	3949	5409	17224
Address cannot be located (21)	1	1	0	55	57
Address impossible to access (22)	0	0	0	2	2
Address does not exist or is non-residential or is unoccupied or not the principal residence (23)	36	36	41	766	879
Total	3953	3987	3990	6232	18162

- DB130 - Household questionnaire result

DB130	Rotational group				
	2	3	4	5	**Total**
Household questionnaire completed (11)	3605	3595	3614	4105	14919
Refusal to co-operate (21)	232	258	254	1107	1851
Entire household temporarily away for duration of fieldwork (22)	52	55	52	81	240
Household unable to respond (illness, incapacity,…) (23)	20	31	21	94	166
Other reasons (24)	7	11	8	22	48
Total	3916	3950	3949	5409	17224

- DB135 - Household interview acceptance

DB135	Rotational group				
	2	3	4	5	**Total**
Interview accepted for database (1)	3605	3594	3614	4101	14914
Interview rejected (2)	0	1	0	4	5
Total	3605	3595	3614	4105	14919

Item non-response (income variables)

Item non-response	(A) % of households having received an amount	(B) % of households with missing values	(C) % of households with partial information
Total household gross income	42.02	5.25	52.64
Total disposable household income	73.03	4.51	22.41
Total disposable household income before social transfers other than old-age and survivor's benefits	72.60	6.16	19.69
Total disposable household income before social transfers including old-age and survivor's benefits	65.93	8.28	15.31
Net income components at household level			
HY040N	1.05	0.20	0.12
HY050N	23.16	0.26	0.24
HY060N	5.30	0.07	0.03
HY070N	5.36	0.19	0.02
HY080N	6.20	0.52	0.00
HY090N	0.76	0.52	0.00
HY110N	3.69	0.06	0.00
HY120N	45.83	4.24	0.00
HY130N	5.15	0.27	0.00
HY140N	41.46	31.14	24.88
HY145N	43.97	2.64	0.02
Gross income components at household level			
HY040G	1.17	0.20	0.00
HY050G	22.26	0.26	1.14
HY090G	0.42	0.52	0.34
HY110G	3.35	0.06	0.35
HY140G	41.34	30.90	25.41
	% 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	30.73	7.03	0.07
PY020N	0.10	0.22	0.00
PY035N	2.63	0.70	0.00
PY050N	5.69	2.83	0.34
PY080N	0.01	0.01	0.00
PY090N	3.40	0.28	0.00
PY100N	22.34	1.71	0.24
PY110N	1.44	0.11	0.00
PY120N	0.40	0.04	0.00
PY130N	6.25	0.48	0.03
PY140N	1.26	0.06	0.00

	% 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	15.83	7.03	14.97
PY050G	5.06	1.78	3.14
PY080G	0.00	0.01	0.01
PY090G	2.09	0.28	1.31
PY100G	15.85	1.71	6.73
PY110G	0.90	0.11	0.55
PY120G	0.25	0.04	0.15
PY130G	4.41	0.48	1.87
PY200G	27.84	6.55	0.00

Income variables' imputation methods applied

Imputation is aimed at obtaining complete records at the level of target variables. Target variables do not simply reflect questionnaire variables and their calculation algorithm is often complicated, although it mainly consists in aggregation. So it is necessary to take a decision at what aggregation level the imputation should take place. There are three possible options:
- the level of questionnaire variables,
- the level of partly aggregated components,
- the level of ready calculated target variables.

Since the only formal requirement is to obtain imputed target variables, all the above options are permissible and practicable depending on the specific character of variables. However, the most frequent practice is the imputation at the level of questionnaire variables, because there are certain arguments for this approach, if on condition that the quantity of data and calculation algorithm details allows for it without much complication.

First of all, imputation at the lowest aggregation level possible may be desirable for the principal reasons related to the quality of imputation when:
- a target variable implies components of different character (e.g. taking different but rather predictable values, e.g. various social benefits, or dependent on a number of explanatory variables and thus easier to be modelled separately);
- target variables include many components and it is often the case that some of them have the missing items, while others – the correct ones which would be missed during the imputation of an aggregated variable.

Secondly, there are practical arguments for the imputation of disaggregated variables, as the same data serve as a basis for calculating national variables differing from the Eurostat's target variables. Thus the imputation of disaggregated components may be required so as to ensure the imputed data needed for other calculations.

The imputation at the target variable level is carried out only when the above circumstances do not occur or when the practical difficulties can be overcome more easily than the imputation of disaggregated data.

There are several methods of component imputation. They can be classified as deterministic and stochastic methods. In case of deterministic methods the selected method and the set of explanatory variables (algorithm) clearly determine the imputation values for each record. In

stochastic methods the imputation value is determined using a random component and that is why 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 brings about variable variance reduction in the file and random error underestimation; it also distorts more the correlation structure (increasing correlations with explanatory variables). According to item 2.7 of Regulation 1981/2003 it is recommended that for EU-SILC imputation the methods retaining distribution characteristics should be applied, 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 categorizing variables occur in the hot-deck method, a random representative is selected out of the records showing adequate values of auxiliary variables. If it is not possible to find a donor with the equivalent values for all the auxiliary variables, the so called sequence approach is applied. The categorising variables were ranked from the most to the least significant ones. If there are no donors, categorization is carried out with the subsequent explanatory variables being left out, starting from the least significant ones so as to obtain a subset containing donors.
-   Stochastic regression imputation
Auxiliary variables are the explanatory variables of the regression model. The model takes the linear form or the logarithmic transformation is used. It is fitted on the basis of the correct records. The imputed value (or its logarithm in the case of transformed models) is a sum of the theoretical value derived from the model and randomly selected model residual. The set of records of which the residual is selected is restricted to those which are nearest to the record imputed for the theoretical value derived from the model.

Out of the deterministic methods the following are applied:
-   Regression deterministic imputation
The theoretical value from the model is adopted 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 the most significant components of target variables, stochastic imputation is applied in order to retain the variable properties distribution as required by Regulation 1981/2003.

The application of stochastic regression imputation 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 dataset. That is why in the cases where in accordance with the assumption referred to above, stochastic imputation is required, the hot-deck method is applied rather than regression imputation. 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.

As in the case of data for 2005 stochastic regression imputation is most widely used for incomes from hired employment, as:

- it is an important category of income, declared by a significant rate of respondents which, if present, has 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.

As compared with 2005 the application of the stochastic regression imputation method has increased due to imputation based on the panel data. It is applied to many other income categories than the income from hired employment only if the incomes of a given person from the previous year are known. In such a case this method is treated as basic, however, the hot-deck method is also used when a suitable model fitting is difficult. In the case of applying a quantitative categorizing variable in the hot-deck method, a breakdown into decile groups is used as a categorization criterion.

Considering a wider application of the stochastic regression imputation, a supplementary protection against the effects of potential insufficient model adequacy was introduced. The residuals are not generated from the distribution of residuals for the whole sample but they are selected from a restricted subset. Although in an ideal model residuals should be in the form of white noise, showing no trend whatsoever, in reality some trends may be observed in the distribution of residuals, which are not detected by the model, e.g. related to non-linearity of relationships which cannot be removed by known transformations. In such a case the use of residuals from the restricted range reduces a risk of generating values diverging from the real variable distribution by combining the theoretical value and the residual which would be quite improbable (in combination with this theoretical value).

Deterministic imputation is applied where missing data concern less significant components of target variables (taxes, burdens to the main component, additions, etc.) in the situation when 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 accessibility 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 have been retained, even if their impact on the imputed variable has not been statistically confirmed, if they expressed an economically important relationship or provided a grouping condition (interpretation criterion) in the calculation algorithm.

For the persons and households not surveyed in 2005 (a new sample, new household members, persons who could not be interviewed) or for those who did not gain a particular type of income in 2005, explanatory variables derived from the current data file are applied. Wherever the same type of income is found in the data for 2005, its value is treated as the main explanatory (categorizing) variable, both in the case of variables subjected to regression imputation and the hot-deck method. The current variables may be treated as additional explanatory variables.

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 are applicable.

Distribution of RB250 and RB260

- RB250 – Data status

RB250	Rotational group				
	2	3	4	5	Total
Information completed only from interview (11)	8463	8453	8525	9452	34893
Individual unable to respond (illness, incapacity, etc) (21)	25	28	28	43	124
Refusal to co-operate (23)	199	203	200	264	866
Person temporarily away and no proxy possible (31)	139	172	142	162	615
No contact for another reason (32)	25	16	19	29	89
Total	8851	8872	8916	9950	36589

- RB260 – Type of interview

RB260	Rotational group				
	2	3	4	5	Total
Face to face (1)	6784	6908	6940	7771	28403
Proxy interview (2)	1679	1545	1585	1681	6490
Total	8463	8453	8525	9452	34893

As for individual interviews, in 2006 a relatively high share (18.6%) 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 34 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 93 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. Problem of the interview duration was already pointed out in the previous Intermediate Quality Report for EU-SILC 2005.

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, 2005.

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 should 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 (2005).

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 2005. The account clearance with the Treasury Office (including payments and returns) effected in 2005 refers to the income for 2004.

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 (2005) 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 19 2006.

Basic information on activity status during the income reference period

In EU-SILC 2006 the definition of retired person (now this definition is in accordance with international recommendations) was changed. In EU-SILC 2005 people obtaining disability pensions were included in PL085 and in variable PL210 in category 6 while in EU-SILC 2006 they are included in variable PL090 and in variable PL210 in category 8.

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

Dwelling conditions and material deprivation items

The analysis of questions and explanatory notes from the guideline for interviewers concerning dwelling conditions and material deprivation items showed that some records differed from those included in document 065/04:

Arrears on mortgage payment – it was not clarified that only arrears on mortgage should be taken into account, so other dwelling related credits might have been included.

Arrears on rent payment – there was some inaccuracy in the questionnaire form – the question referred to arrears on housing-related bills (other than utility bills) not only to arrears on rent. That situation was clarified as the guideline was updated later on; however, the inaccurate expression remained in the questionnaire form which might have caused wrong answers.

Arrears on hire purchase instalments other than loan payments – this question included arrears on hire purchase and credits other than dwelling-related ones.

Capacity to afford paying for one week annual holiday away from home – first of all the question included the expression "if the household wants"; secondly, family as such was concerned and it was not pointed out that the question referred to the household as a whole.

Leaking roof, damp walls/ floors/foundation, or rot in window frames or floor – the question was formulated in a different way, namely: "Do you think your dwelling requires renovation because of…?"

Indoor flushing toilet for sole use of the household – the toilet could have been shared with other households.

Additionally, for the variables from HS010 to HS050 no information was given that paying through borrowing meant that household was not in arrears.

There were no other major 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

4.1. Comparison of EU-SILC and HBS results

The objective of this section is to compare HBS (Household Budget Survey) and EU-SILC results.
Up to 2004 the HBS provided the main source of data on the living conditions of the Polish population, among others on incomes, dwelling conditions and households' equipment.
The HBS has been regularly conducted every year since 1993 up to now with the use of the rotational method. The households are surveyed in the two year panel.
In the HBS the main source of data is the so called diary. Two additional questionnaires are also filled in.

When comparing these two sources we must take into account the discrepancies. The differences are to great extent brought about by the methodological diversity. Here are the main diverging points:
  - Different reference periods for income variables – in HBS the reference period is 1 month and, following Eurostat's recommendation, the annual income is the monthly income multiplied by 12, which in the case of irregular income, like that from farming, can bring about considerable distortions. In EU-SILC the reference period is a calendar year preceding the survey;
  - Different types of income are taken into account i.e. in HBS the information is collected both about the income in cash and in kind, while in EU-SILC – only about the income in cash (with a few exceptions), which may be important for the income from farming and social benefits other than retirement pay and pension. Moreover, EU-SILC does not take into account the so called lump sums which is the case in HBS;

- Different way of data collection – in HBS the respondents make records in the so called diary. They have to determine the data sources themselves and do not have them listed in the diary. This may cause omissions. In EU-SILC each respondent is asked detailed questions. In EU-SILC all the income missing data are imputed, while there is no imputation in HBS;
- Different way of sample selection – in HBS households which refused to participate in the survey are replaced with those from the so called reserve list. No replacement is applied in EU-SILC;
- Slightly different weighting of results.

In some tables below socio-economic groups' breakdown is used. The household survey results are traditionally prepared by CSO according to the so called socio-economic groups of households. The main criterion for socio-economic group classification is the prevailing source of income.

In tables below only weighted data are presented.

Tab. 1. Structure of population by age

Specification	EU-SILC 2006	HBS 2006
	in %	
Total	100,0	100,0
0-14	16,5	18,1
15-24	15,7	16,6
25-54	44,1	41,6
55-64	10,5	11,3
65+	13,3	12,4

Tab. 2. Structure of population by level of education

Specification	EU-SILC 2006	HBS 2006
	in %	
Total	100,0	100,0
No school education	2,5	0,9
Completed primary	19,2	20,0
Lower secondary	4,9	6,5
Elementary vocational	26,6	26,8
Secondary	33,5	33,5
Higher	13,2	12,2

Tab. 3. Structure of households and persons in households by socio-economic group

Households	Households		Persons in households	
	EU-SILC 2006	HBS 2005	EU-SILC 2006	HBS 2005
Total	13318760	13332605	37794851	37744302
Total = 100				
Employees	47,9	43,1	57,4	51,3
Farmers	2,5	4,7	3,7	7,0
Self-employed	5,0	5,9	5,6	7,0
Retirees	27,8	27,0	19,7	18,8
Pensioners	9,8	11,8	7,1	8,8
Maintained from non-earned sources	6,9	7,5	6,5	7,0

Tab. 4. Average yearly equivalent income in PLN by socio-economic group

Households	Disposable income		Income from hired work	
	EU-SILC 2006	HBS 2005	EU-SILC 2006	HBS 2005
Total	14902	13444	8701	6636
Employees	16872	14325	14117	11695
Farmers	9224	12229	689	1216
Self-employed	18271	18271	2970	2565
Retirees	13700	13309	1270	1266
Pensioners	9649	9622	1004	902
Maintained from non-earned sources	7213	8552	1387	727

Tab. 5. Average yearly equivalent income in PLN by number of persons

Households	Disposable income		Income from hired work	
	EU-SILC 2006	HBS 2005	EU-SILC 2006	HBS 2005
Total	14902	13444	8701	6636
1-person	14372	13704	4439	3673
2-persons	17611	15961	7247	5589
3-persons	17378	14954	11691	8839
4-persons	14928	13582	10713	8382
5-persons	12349	11318	7569	5888
6-persons and more	10959	9964	6113	3970

Tab. 6. Households provided with selected durables

Specification	EU-SILC 2006	HBS 2006
	in %	
Fixed telephone	74,0	71,9
Mobile telephone	70,7	73,1
Television set	96,9	98,5
Computer	44,4	43,7
Printer	31,8	29,5
Internet connection	28,5	28,4
Microwave oven	34,1	38,0
Dishwasher	8,0	6,2
Refrigerator	96,9	98,6
Washing machine	96,2	96,8
Passenger car	50,7	49,5

4.2. Comparison of Laeken Indicators based on EU-SILC 2005 and EU-SILC 2006

The results of the EU-SILC 2006[2] clearly indicate an improvement of the income condition of the Polish households between 2004 and 2005 as well as lower income differences. The preliminary analysis of Laeken indicators measured based on the EU-SILC 2006 shows that in 2005 as compared with 2004 the at-risk-of-poverty rate dropped by 2 percentage points (from 21% to 19%). The highest reduction – by 3 percentage points - was noted for children (at the age of 0-17 years), particularly in large families (2 adults and 3 or more children) and incomplete families (single parent with dependent children) – by 7 and 8 percentage points, respectively. This poverty rate reduction was accompanied by a significant increase of the real value of the poverty threshold (at 60% of the disposable income median). At the same time, the relative median at-risk-of-poverty gap showed a significant decrease (by 5 percentage points). The income quintile share ratio dropped from 6.6 in 2004 to 5.6 in 2005 (S80/S20), while Gini coefficient – from 36 to 33, respectively.

Generally, the positive trends in the income condition of households noted on the basis of EU-SILC are reflected by the macroeconomic figures (real increase of disposable income, reduction of unemployment which is one of the main factors increasing poverty risk). The improvement of the situation of families with children results from the revision of the respective regulations and higher benefits granted to this group.

Relative poverty reduction is also confirmed by the HBS-based estimates. It should be noted, however, that both the macroeconomic figures and HBS results indicate much lower dynamics of incomes than that measured on the basis of EU-SILC. Perhaps much greater changes in the income condition of households observed between the first (2005) and the second (2006) wave of EU-SILC could be explained among others by underestimation of income variables in the first wave and improvement of the quality of data in EU-SILC 2006. This was also referred to in the section of this report discussing non-sampling errors.

---

[2] It should be remembered that the reference period for income in EU-SILC 2005 is the year 2004, while in EU-SILC 2006 – the year 2005, respectively.