Final Quality Report SILC2005 - BELGIUM

0. Introduction

This report contains a description of the accuracy, precision and comparability of the Belgian SILC2005-surveydata. It is structured following the guidelines in the commission regulation (EC) no. 28/2004. This results in three chapters:

1. Indicators
2. Accuracy
3. Comparability
4. Coherence

1. Indicators

The common longitudinal EU indicators based on the longitudinal sample of EU-SILC can't be computed after 2 waves and will be given from 2007 on.

2. Accuracy

For second and following waves of the longitudinal component the following information has to be provided

2.1.8.5 Attrition for the old households

Before "sharing" the 2004 weights (to define the 2005 base weights for the "old" households-, a correction for attrition should be introduced. It appears that the 2005 response rate of (components of) old households varies according to their (2004) tenure. The only other attrition predictor we looked at (rotational group) was hardly significant.

Main steps of computation:

Merge (by DB030 :)

- D-file
- last year's D-file (just for calibrated=final weights 2004, **calwei04**, see IV)
- last year's H-file (just for tenure 2004)

Compute (weighted) response rates by tenure, only for «principal components» [1].

Compute ATTwei (for all components)[2]

1

Code	HH020(2004)	Rate(2005)
1	owner	75,0
2	tenant	68,4
3	reduced	72,7
4	free	63,2
	Total	73,0

N.B. The accepted new households, weighted by ATTwei, represent approximately ¾ of population (one fourth left the panel!)

Weight sharing

We followed Eurostat's (ultimate = 20 June 2006) recommendation "EU-SILC weighting procedures: an outline" and shared the calibrated 2004 weights (instead of the initial weights, see Lavallée).
Fortunately, no respondent (2005) household was the result of a fusion (viz. DB110=10), so weight sharing amounted to defining …
LINK= quotient "2005 household members already in household in 2004 (or age=0)/2005 household members", this quotient is <= 1, and was 1 in most cases.
However, in 3 cases[3], **LINK** turned out to be 0, and –under time pressure- we could not determine whether the computation of **LINK** (merging the 2004 and 2005 R-files) had gone wrong, or these 3 households should not have been interviewed.
Finally, we found a compromise, defining **LINKPLUS**= ((2*__LINK__+1)/3) (if **LINK**=1 : **LINK** = **LINKPL**US = 1, if **LINK**=0: **LINKPLUS**=1/3).
The final step was: **SHAwei**=**ATTwei*LINKPLUS**

2.1.8.6 Calibration

In Belgium, 11 *sampling* strata were used (provinces= NUTS2)
In order to avoid a large std of calibrated weights, we use 3 *extrapolation* strata (the 3 NUTS1 regions BRUssels=BE1, VLAanderen=BE2 and WALlonia=BE3).
Extrapolation strata: 3

2139 new households,	sum of NRwei	4393146	(about whole population)
2998 old households,	sum of SHAwei	3371037	(about ¾ of whole population)

Goal: Final weights such that new represent 40%, old 60% →
define **WEwei** ('initial weight' in g-calib syntax) as 0,4*__NRwe__i+0,8*__SHAwei.__

Calibration model
VLA, WAL:
WAVE1+SIZE4+(AGE8XSEX2)+PROV5 →20 individual[4] + 5 household constraints
BRU:
WAVE1+SIZE4+(AGE8XSEX2) →16 individual + 5 household constraints

Prov = province where interviewed (differs from DB040 in two cases)

---

[2] IF (HH020=1) ATTwei = calwei04/0.750 .
IF (HH020=2) ATTwei = calwei04/0.684 .
IF (HH020=3) ATTwei = calwei04/0.727 .
IF (HH020=4) ATTwei = calwei04/0.632 .
[3]See 2.3.1 coverage errors

[4] Five provinces and 16 age*sex categories, but sum over provinces = sum over age*sex

Individual constraints 27=16+11 (age*sex + prov; note that each province belongs to one single region (extrapolation stratum), for the other two regions, the total is set to 0 and the condition is vacuous)

Household constraints 5=4+1 (size: "1", "2", "3 or "4 & more", the distinction between "old" and "new" households yields only one additional constraint since the total number of households is already known)

Wave: only at household level, only new vs. old (no distinction between old).

Calibration type (after some trials and errors…): truncated, 0.4 <= g-weight <= 2 (little diffrence with alternative methods).

2.1.8.7 Final longitudinal weights

Explanation (following "*EU-SILC weighting procedures: an outline*"), at hh level

- Coe04 We start from the 2004 calibrated weight of all (4120) hh due to stay in panel ("represent" about 75% of population 2004 ...or a bit more: split hh!, respecting the age -sex - hh size distribution)

- ATTwei We correct for attrition (multiply Coe04 by inverse response rate) according to tenure, in such a way that the 2998 hh that did stay "represent" about 75% of population 2004 (or a bit more: split hh!)

- SHAwei We "share" weights, i.e. multiply ATTwei by the proportion LINK of 2005 hh members that already belonged to the 2004 sample), in such a way that the 2998 hh that did stay "represent" about 75% of population 2004 (SHAwei <= ATTwei), keeping in mind that

* *persons who are born or have immigrated into the target population since time 1, and have remained in that population since that time.* was approximated by "persons whose age (2005) <1 year (relying on final calibration to account for other changes, we were not able to determine, for the whole population, who/how many stayed eligible throughout, i.e. the population 2004&2005)
* due to a few linkage problems, the proportion LINK had to be recomputed as LINKPLUS (in more than 90% of cases LINK=LINKPLUS=1 anyway!)

WEwei Just a rescaling of SHAwei (WEwei=4/5 SHAwei), in such a way that the 2998 hh that did stay "represent" about 60% of population 2004 (which is a bit less than population 2005, and a bit more than population 2004&2005)

2.1.8.8. Final cross-sectional weights

Statistics

	N	Minimum	Maximum	Mean	Std. Dev.
Final weights	5137	58,18	7878,79	871,64	325,86

Remark:

Although we relaxed constraints, the standard deviation of weights increased…

- 2004:(exponential calibration) 5275 households, range of final weights [135 → 5817], mean 842, std 293;
- 2005:(truncated, $0.4 \leq g \leq 2$) 5137 households, range of final weights [58 → 7879], mean 871, std 326.

Reason for the large standard deviation (wide range): “old” households, std 383 (vs. 245 for the new households); “snowball effect” or “entropy”: 2005 calibration

performed on weights that had already been calibrated in 2004 (then increased by the attrition correction, decreased for some households by weight sharing).

2.2 Sampling errors

The table is based on the results of EU-SILC 2005.

Income components	Mean	Number of observations before imputation	Number of observations after imputation	Standard error
HY010	39352.8	4515	5136	3058.2
HY020	28742.9	4709	5136	2952.8
HY022	25130.4	4602	4925	2966.2
HY023	20384.7	4522	4644	2961.8
Net income components at household level				
HY030N				
HY040N				
HY090N	1100.8	1074	3073	405.7
HY050N	2915.9	1809	1837	160.7
HY060N				
HY070N				
HY080N				
HY100N				
HY110N				
HY120N				
HY130N				
HY140N				
HY145N				
Gross income components at household level				
HY030G				
HY040G	9860.6	330	368	1600.3
HY090G	1100.8			405.7
HY050G	2942.8	1809	1837	163.2
HY060G	6112.6	114	117	965.8
HY070G	1541.2	33	41	924.7
HY080G	3389.7	322	346	618.9
HY100G	2813.5	1372	1559	146.1
HY110G	760.7	8	9	378.2
HY120G				
HY130G	2808.8	433	477	558.0
HY140G	12379.9	3612	4445	484.4
net income components at personal level				
PY010N	19939.8	4208	4492	3367.4
PY020N	1910.7	305	305	99.4
PY035N	0.0			
PY050N	23008.6	339	606	4288.2
PY070N				
PY080N	7092.5	13	13	7346.1

PY090N	7846.5	970	1248	436.9
PY100N	12291.5	1607	1896	361.1
PY110N	11244.8	68	77	1043.5
PY120N	4759.7	110	162	805.4
PY130N	9548.2	264	322	617.3
PY140N	507.8	153	211	130.7
gross income components at personal level				
PY010G	30051.4	3767	4492	3411.7
PY020G	1910.7	305	305	99.4
PY030G				
PY035G				
PY050G	31150.6	327	606	4994.5
PY070G				
PY080G	7092.5	13	13	7346.1
PY090G	8648.5	702	1248	661.8
PY100G	14138.9	1095	1896	511.0
PY110G	12343.2	48	77	1257.6
PY120G	4967.0	80	162	853.3
PY130G	10390.4	185	322	680.9
PY140G	507.8	153	211	130.7
PY200G	2337.3	3863	3863	54.1

Equivalised disposable income	Mean	Number of observations before imputation	Number of observations after imputation	Standard error
Subclasses by household size				
1 household member	18813.6	1049	1314	8703.2
2 household members	18091.8	3187	3445	683.6
3 household members	19632.5	2426	2487	1327.4
4 and more	17800.0	5467	5510	739.5
Population by age group				
<25	17244.4	4060	4096	670.1
25 to 34	19313.9	1525	1577	816.6
35 to 44	21752.4	1930	1982	8213.7
45 to 54	20032.3	1377	1785	948.9
55 to 64	19367.3	1346	1484	1366.1
65+	14477.7	1551	1832	533.4
Population by sex				
Male	19091.5	6032	6301	2565.1
Female	17705.5	6095	6452	522.2

2.3 Non-sampling errors

2.3.2 Measurement and processing errors

Mismatch in time between household composition and household income (see also §3.1)

A number of inconsistencies result from a mismatch between the composition of the household at the moment of the interview (between September and December of year x) and the income of the previous year (year x-1).

This mismatch can bias the measurement of poverty status in several ways. For example:

- ✓ Persons who were full-time students in year x-1 (and depending on their parents), but were employed at the time of the interview (and living independently in a one person household for example) will report an income equal to 0 in year x-1 and will be wrongly classified as a poor household.

Other examples can also occur for persons where the household composition changed:

- ✓ For a housewife who was married in year x-1, but divorced and is working at the time of the survey there will also be a mismatch
- ✓ For a household which received family allowances for a student in year x-1, but where the student is no longer part of the household in year x there will also be a mismatch
- ✓ For a household with a person working in year x-1, but retired at the moment of the survey (in year x) a mismatch will also occur. Take notice of the fact that, as the examples show the bias can go in both directions: under and over reporting of income. In each one of the examples, the choice to situate the income reference period in the past is the cause, however.

An extensive screening in the post data collection process revealed that a number of interviewers had filled in at least some questions about income, loans and (housing) costs in Belgian Franks instead of Euros. This occurred less than for the previous survey and mainly for the loans, despite the fact that the amount in Belgian Franks appeared on the screen as soon as the loan (in Euro) was filled. This control was added in 2004.

- ***Error in the routing wave 2004***

An error in the routing occurred for Questions H100 and H101 on the 'Revenus du patrimoine' (Interests, dividends, profit from capital investments in unincorporated business)(To be included in Variable HY090G). Only individuals responding precisely on Question H99 about 'Revenus des placements financiers' were asked to precise whether the amounts were profit or loss. For individuals responding the question H100 (not an amount but a scale value) H101 was never asked. For these cases, the incomes were considered as profit.
H 36 (HY040): if the person answered that he didn't let out a part of his house, we still asked how much the profit was.

- ***Error in the routing wave 2005***

There was one error in the routing in the household questionnaire for tenants. They skipped the question "Can you tell me what is the amount you pay monthly for your consumption of electricity and gas together? Give a rough estimation. If a part of your dwelling is professionally used, give the total only for the non-professional part.**"**

- ***Correspondence French/Dutch versions of Questionnaires wave 2004***

There was no mistake in the formulation of the French/Dutch versions in 2004.

- ***Correspondence French/Dutch/German versions of Questionnaires wave 2005***

For the question about the mode of contact, the French version was wrongly asking whether the **household** was contacted where the Dutch version asked whether the **address** was contacted.
In the German version, question I8. 'Retirement' is coded 8 as it is coded 7 in the other languages because 'Student' and 'Unpaid work experience' were unfortunately split in 2 codes (6 & 7). Other consequence: 'Permanently disabled' and 'Fulfilling domestic tasks' were collected on the same code (9). We estimate that 0,18% of the response on this question could have been influenced by this.

• ***Differently asked questions***

HH050: The question in 2004 did not point out that the inability to keep home adequately warm was the **inability to pay** to keep home adequately warm. We then changed the question in 2005 and the interviewee was then asked 'do you have financial difficulties to keep home warm?'.

Problem: in the French version, the question did not mention 'to keep home **adequately** warm', whereas the Dutch version did.

The answers in 2005 are thus barely comparable to those of 2004.

2004 :

N°	Question
H 1	**Pouvez-vous chauffer votre logement convenablement ?**
	Oui Non

2005 :

N°	Question	Codes	Routing	*EV*
H 11	**Avez-vous financièrement des difficultés pour chauffer votre logement ?**		H 12	
	Oui Non	1 2		*HH050*

2.3.2.2. Processing errors

Belgium used the CAPI–method to interview the persons. The questionnaire was programmed in Blaise. So processing errors due to data entry (from a written to an electronic format) were reduced to a minimum.
Statistics Belgium programmes several data entry and coding controls in the Blaise program. Those were identical for both waves.

Next to these controls, some warnings were implemented **in 2005** in order to ask the interviewer to verify the introduced data in the case of abnormally high or low

amounts. A warning is a simple text box with a message such as ‘This amount is very low, are you sure the amount is right?’ or ‘This amount is very high, are you sure the amount is right?’. The interviewer has then to confirm the value or to change it in case of error.

Household questionnaire	
H16	If lower than 500 or higher than 1000000
H22 (monthly)	If lower than 20 or higher than 2000
H22 (half-yearly)	If lower than 100 or higher than 10000
H22 (yearly)	If lower than 200 or higher than 20000
H23 (monthly)	If lower than 20 or higher than 2000
H23 (half-yearly)	If lower than 100 or higher than 10000
H23 (yearly)	If lower than 200 or higher than 20000
H26	If lower than 25 or higher than 5000
H33	If lower than 50 or higher than 10000
H34, H37, H41	If lower than 100 or higher than 5000
H43, H77, H84	If lower than 25 or higher than 1000
H66	If lower than 100 or higher than 25000
H71B	If lower than 25 or higher than 750
H79, H86	If lower than 300 or higher than 12000
H93	If lower than 100 or higher than 1500
Individual questionnaire	
I25, I27, I47, I50, I90, I91	If lower than 500 or higher than 5500
I53, I86, I93, I94	If lower than 6000 or higher than 66000
I58	If higher than 1200
I98B, I98C, I115B, I115C	If higher than 1350
I99, I102B, I102C	If higher than 5400

Some warnings concern other values than amounts. It’s the case for H17 when the value is higher than 30 years (‘A period of 30 years is really exceptional, are you sure it is right?’) and for H18 when the interest equals 0 or is higher than 15.

2.3.3. Non-response errors

2.3.3.1. Achieved sample size

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

2004	2005
4004	2997

- number of persons 16 years or older, number of sample persons and number of co-residents, members of households for which an interview is accepted in the database and who completed a personal interview:

	2004	2005

Persons 16 y and more	7755	5846
Sample persons	7755	5708
Co-residents with interview		138

2.3.3.2. Unit non-response

Response rate for households

- Wave response rate

$$\text{Wave response rate} = \frac{4112}{7327-112} = 57\%$$

$$\text{Refusal rate} = \frac{1666}{7327-112} = 23.09\%$$

$$\text{Non contacted and others rate} = \frac{1325}{7327-112} = 18.36\%$$

- Longitudinal follow-up rate

$$\text{Longitudinal follow - up rate} = \frac{2262}{2262+336} = 87.06\%$$

- Follow-up ratio:

$$\text{follow - up ratio} = \frac{2262+2212}{2262+336} = 1.72$$

- Achieved sample size ratio

$$\text{Achieved sample size ratio} = \frac{4112}{2600} = 1.58$$

		SAMPLE OUTCOME IN WAVE2											
		DB130=11											Total
		DB135=1 (A)	DB135=2 (B)	DB120=22 (C)	DB130=22 (D)	DB130=23 (E)	DB130=24 (F)	DB130=21 (G)	DB120=21 (H)	NC (I)	DB110=10 (J)	DB120=23 (K)	(T)
		4112	25	23	232	112	1006	1666	20	15	4	112	7327
		SAMPLE OUTCOME IN WAVE1											
DB130=11	DB135=1	1934	11	0	43	28	246	314	3	15	4	2	2600
2638	DB135=2	0	0										0
DB120=21 to 23													
DB130=21 to 24													
TOTAL													
		NEW HOUSEHOLDS IN WAVE 2											
DB110=8		38	1	1	7	0	12	7	1	NA	NA	1	68
DB110=9		2140	13	22	182	84	748	1345	16	NA	NA	109	4659

Personal interview response rates

Response rate for persons

- Wave response rate

Wave response rate of sample persons =

$$= \frac{9831}{11909} = 82.5\%$$

- Wave response rate of non sample persons:

Not computed for 2005 because there were no co-residents in 2004 (all 16+ were sample persons for the first wave).

- Longitudinal follow-up rate:

$$= \frac{9831}{14935} = 65.8\%$$

$$\text{Rate (RB250=21)} = \frac{17}{14935} = 0.1\%$$

$$\text{Rate (RB250=23)} = \frac{10}{14935} = 0.1\%$$

$$\text{Rate (RB250=31)} = \frac{41}{14935} = 0.3\%$$

$$\text{Rate (RB250=32)} = \frac{9}{14935} = 0.1\%$$

$$\text{Rate (RB250=33)} = \frac{4}{14935} = 0.03\%$$

Note that these results are provisional. Some clarifications of the Eurostat technical document are necessary in order to get accurate results.

- Achieved sample size ratio

Achieved sample size ratio for sample persons (and co-residents)

$$= \frac{9974}{10146} = 98.3\%$$

No achieved sample size for co-residents can be computed because there were no co-residents in 2004.

- Response rate for non-sample persons

$$= \frac{143}{158} = 90.5\%$$

Personal interview response rate in wave 2												
		RB250=11,12,13	Not completed because of									TOTAL
			RB250=21	RB250=22	RB250=23	RB250=31	RB250=32	RB250=33	HHnc	Pn	PI	
Sample persons (RB100=1 and rb245=1-3) from the sample forwarded from last wave												
(1) RB110=1-2		5703	8	0	4	27	4	2				5748
(2) RB110=6												29
(3) RB110=-1												0
(4) RB120=2												4
(5) RB120=3												7
(6) RB120=4												44
(7) DB135=2 or -1 or DB110=7 or DB120=21-23 or DB130=21-24 or -1												1953
(8) DB110=3-6												31
												0
New sample persons												0
(9) Reached age 16												0
(10) Sample additions		4128	9	0	6	14	5	2				4164
Non-sample persons 16+												
(11) this wave	From w 1											
	Not in w1	143	0	0	4	6	1	4	0	1	0	159
(12) Earlier wave	From w 1											
	Not in w1											
Sample persons from sample not forwarded from last wave (excluded died or non eligible)												3026
Sum of rows		9831	17	0	10	41	9	4	0	0	0	11909
		9831	17	0	10	41	9	4	0	0	0	14935
		9974	17	0	14	47	10	8	0	1	0	12068

2.3.3.3 Distribution of households by household status, by record of contact at address, by household questionnaire result, by household acceptance

Household status
DB110=

	Total	1	2	3	4	5	6	7	8	9	10
Total	4150	3710	238	5	2	16	7	26	142		4
%	100	89.4	5.7	0.1	0.1	0.4	0.2	0.6	3.4		0.1

Record of contact at address
DB120=

	Total	11	21	22	23	missing
Total (DB110=2,8,10)	384	306	6	1	5	66
%	100	79.7	1.6	0.2	1.3	17.2

Household questionnaire result
DB130=

	Total	11	21	22	23	24	missing
Total (DB120=11 or DB110=1)	4016	3013	495	81	47	378	2
%	100	75	12.3	2	1.2	9.4	0.1

Household interview acceptance
DB135=

	Total	1	2	missing
Total(DB130=11)	3013	2997	16	0
%	100	99.5	0.5	

2.3.3.4 Distribution of persons for membership status (RB110)

	Total	Current HH member				No current HH member			missing
		RB110=1	RB110=2	RB110=3	RB110=4	RB120=2 to 4	RB110=6	RB110=7	
Total	7531	7099	68	177	90	63	29	5	
%	100	94.3	0.9	2.4	1.2	0.8	0.4	0.1	

Distribution of persons moving out by variable RB120

	Total	RB110=5				
		RB120=1		RB120=2	RB120=3	RB120=4
		This person is a current HH member	This person is not a current HH member			
Total	174	69	42	4	7	52
%	100	100	39.7	24.1	2.3	4.0

2.3.3.5 Item non-response

In the following table an overview of the item non-response for all income variables is presented. The percentage households having received an amount, the percentage of households with missing values and the percentage of households with partial information is calculated.
These percentages are calculated as follows:

- % of households having received an amount : number of households (or persons) who have received something (yes to a filter) / total
- % of households with missing values : number of households (or persons) who said that they have received something but did not give any amount (no partial information) / number of households (or persons) who have received something (yes to a filter)
- % of households with partial information: number of households (or persons) who said that they have received something but gave partial information (amounts were not given for all components) / number of households (or persons) who have received something (yes to a filter)

Overview of the non-response for the income variables - % households having received an amount, % of households with missing values and % of households with partial information.

Item non-response	% of households having received an amount	% of households with missing values	% of households with partial information
Total gross household income (HY010)	**100**	**12.1**	**52.9**
Total disposable household income (HY020)	**100**	**8.3**	**57.1**
Total disposable household income before social transfers except old-age and survivor's benefits (HY022)	**95.9**	**6.6**	**59.8**
Total disposable household income before social transfers including old-age and survivor's benefit (HY023)	**90.4**	**2.6**	**65.4**
Net income components at household level			
Family related allowances (HY050N)	**35.8**	**1.5**	**1.9**
Interests, dividends, etc. (HY090N)	**59.8**	**65.1**	**0**
Gross income components at household level			
Income from rental of a property or land (HY040G)	**7.2**	**10.3**	**0.8**

Family related allowances (HY050G)	**35.8**	**1.5**	**2.7**
Social exclusion not elsewhere classified (HY060G)	**2.3**	**2.6**	**0**
Housing allowance (HY070G)	**0.8**	**19.5**	**0**
Regular inter-household cash transfer received (HY080G)	**6.7**	**6.9**	**1.4**
Interest repayments on mortgage (HY100G)	**30.3**	**12**	**1.3**
Income received by people aged < 16 (HY110G)	**0.2**	**11.1**	**0**
Regular inter-household cash transfer paid (HY130G)	**8.7**	**3.1**	**0**
Tax on income and social contributions (HY140G)	**86.5**	**18.7**	**24.5**
Net income components at personal level			
Employee cash or near cash income (PY010N)	**45**	**6.3**	**11.2**
Cash benefits or losses from self-employment (PY050N)	**6.1**	**44.1**	**1.8**
Pension from individual private plans (PY080N)	**0.1**	**0**	**0**
Unemployment benefits (PY090N)	**12.5**	**22.3**	**0.5**
Old age benefits (PY100N)	**19**	**15.2**	**0.7**
Survivor' benefits (PY110N)	**0.8**	**11.7**	**0**
Sickness benefits (PY120N)	**1.6**	**32.1**	**0**
Disability benefits (PY130N)	**3.2**	**18**	**0.6**
Gross income components at personal level			
Employee cash or near cash income (PY010G)	**45**	**8.2**	**16.1**
Non cash employee income (PY020G)	**3.1**	**0**	**0**
Cash benefits or losses from self-employment (PY050G)	**6.1**	**46**	**2.1**

Pension from individual private plans (PY080G)	**0.1**	**0**	**0**
Unemployment benefits (PY090G)	**12.5**	**43.8**	**0.5**
Old age benefits (PY100G)	**19**	**42.2**	**1.3**
Survivor' benefits (PY110G)	**0.8**	**37.7**	**0**
Sickness benefits (PY120G)	**1.6**	**50.6**	**0**
Disability benefits (PY130G)	**3.2**	**42.5**	**0.6**
Education-related allowances (PY140G)	**2.1**	**27.5**	**0**
Gross monthly earnings for employees (PY200G)	**41.9**	**3.1**	**0**

2.4 Mode of data collection

Distribution of household members aged 16 and over by RB250

(Household members RB245=1)

	Total	RB250=11	RB250=21	RB250=23	RB250=31	RB250=32	RB250=33
Total	5902	5846	8	8	33	5	2
%	100	99.1	0.1	0.1	0.6	0.1	0.0

(Sample persons 16+ RB245=1 and RB100=1)

	Total	RB250=11	RB250=21	RB250=23	RB250=31	RB250=32	RB250=33
Total	5754	5708	8	4	28	4	2
%	100	99.2	0.1	0.1	0.5	0.1	0.0

(Co-residents 16+ RB245=1 and RB100=2)

	Total	RB250=11	RB250=21	RB250=23	RB250=31	RB250=32	RB250=33
Total	148	138	0	23	5	1	0
%	100	93.2	0.0	15.5	3.4	0.7	0.0

Distribution of household members aged 16 and over by RB260

(Household members 16 + RB250=11)

	Total	RB260=2	RB260=5
Total	5846	4953	893
%	100	84.7	15.3

(Sample persons 16 + RB100=1 and RB250=11)

	Total	RB260=2	RB260=5
Total	5708	4872	836
%	100	85.4	14.6

(Co-residents 16 + RB100=2 and RB250=11)

	Total	RB260=2	RB260=5
Total	138	81	57
%	100	58.7	41.3

2.5 Imputation procedure

2.5.0 Preceding important remark

In contrast to 2004 – in 2005 the calendar question (i40 in the questionnaire) was presented to every respondent rather the only those who indicated that had been a change in their social-economic position. This was a very simple but nevertheless very successful change as it enabled us to assess and check much thoroughly the link between the social-economic position and the income variables. Notably for the self-employed this resulted in a substantive number of cases (being identified as being self-employed) who would be otherwise (and who were to some extent in 2004) not identified as being self-employed. These cases mainly concern people in jobs 'somewhere on the bridge' between being self-employed and employee but who nevertheless indicated in the calendar that they were self-employed.

2.5.1 Overall strategy: Emphasis on internal information and integration of outlier detection- , imputation- and control-phases.

- Emphasis on internal information.

We can't emphasise enough that to correct and impute our data (for any variable) we relied:

1) **as much as possible on internal information present in the data itself**
2) on formal and legal sources of information and
3) only as final resort turned to statistical procedures (random imputations for ex.)

Some examples to make it more clear what we mean:

→ for the self-employed almost half of the imputations were done on basis of the recorded household income of the household to which the self-employed person belonged to (22.9% - 48.3% is imputed); for another 6% information (on the missing income) was gathered from question I86 questioning the amount of money they retrieved from their business. Less than 10% of the cases were imputed by use of purely statistical procedures (9.2% - imputation of medians and regression with characteristics of the job as independent variables).

→ For the employed – the vast majority of the imputations concerned recorded net wages which needed to be converted in to gross.

- An integrated strategy.

As it was the case for SILC-2004 we used for SILC-2005 again an 'integrated approach' to organise the detection of outliers and the imputations. Crucial to the understanding of our way of working are the concepts of what we call 'vertical' and 'horizontal integration'.

By 'vertical integration' we mean that the phases of outlier detection and imputation were done together for each variable separately (1) rather than that both phases were done separately for all variables together (2). The differences between (1) – the way we did things for SILC 2004 - and (2) the way it was done for SILC 2003 – are subtle but nevertheless more than semantics, especially when combined with horizontal integration.

By horizontal integration we mean that information for each respondent on one variable was checked against information on another variable or another source. Information on the monthly gross income for example was – if both possible and applicable- checked with information on the net income, the yearly income, the current income (if no changes had occurred), the household income, other 'proxi'-variables to income (status etc…) and very important external sources of information like legislation.

The interplay between what we call vertical and horizontal integration leads to a dynamic strategy: variables are checked for outliers and inconsistencies, variables are compared to each other and corrected, (corrected) variables are immediately imputed consistently to the information in other (also corrected) variables – and this several times repeated.

We believe that the emphasis of this strategy on consistency of internal information for respondents throughout the survey and the use of external sources of information (legislation) is a far more successful way of detecting outliers and imputing missing values compared to methods of screening for outliers entirely based on (univariate) distributional features of variables (box-plot methods for example) and imputation methods mainly based on statistical probability models (IVE for example).

Outlier detection: The shift in strategy also implies – of course - a shift in the techniques that are used. As far as the outlier detection concerns there is far less emphasis on univariate - purely distributional related methods like box-plots but more emphasis on inconsistency checks. For the income variables these checks were done in 2 ways: i. comparison of ratio's between variables and ii. comparison of the relative position of a respondent's answer on one variable to its position on another variable.

i. Comparison of ratio's between variables:

Comparison of the ratio between two inputs on comparable income variables is a straightforward way to detect outliers. Atypical large or small ratios between gross and net variants of income variables are obviously an indication of 'something being wrong'.

ii. Comparison of relative positions on income variables:

The central issue in this procedure is the comparison of two income variables by comparison of the normal scores calculated for each case on both variables, after log-transformation. The log-transformation is necessary to normalize the otherwise poisson-distributed income variables.
The inputs of both comparable incomes are considered to be consistent if both normal scores are within predefined boundaries (for example -1,96 and 1,96) and/or the difference between the normal scores is limited (less than 1,96).
There is an indication of bias if the input of one of the incomes for a case is situated within 'normal boundaries' (-1.96 – 1.96) but the other input is not and/or if the difference between the two normal scores differ substantially (>1.96). In fact, the entire procedure consist out of 4 steps:

1. Identification of the variables to be compared.
2. Log-transformations, normality checks, calculation of means and standard deviations.
3. Calculation of normal scores.
4. Consistency control and identification of inconsistencies.

iii. Other techniques :

There was explicitly more emphasis on the above techniques but this does not imply that the 'conventional' box-plot method was not used at all. In this method input outside the interval below were considered to be outliers:
[First Quartile – 1,5 * (Third Quartile – First Quartile) ; Third Quartile + 1,5 * (Third Quartile – First Quartile)]

Furthermore and as already mentioned, where applicable and usable legal maximums and minimums were also used to some extent.

Finally, we also checked for outliers via controls on a 'case to case' base in which we maximally used information of proxi-variables like professional status and other variables. In this process manifest errors in proxi- and/or other variables associated with the income variables were also removed/corrected (for example 'the number of months').

Imputation: We did no longer make use of IVE. Instead we i. corrected (not imputed – in fact) a greater number of cases and if correction was not desirable or possible, but information on a directly comparable variable was present anyway (see section on internal information above), we ii. resorted to direct imputation, via a regression model.

i. Corrections.

Corrections were also mainly done on basis of information in other comparable variables. Gross-net ratio of around 40 - 1 Euro = +/- 40Belgian Francs - or 12 - yearly income entered as monthly or vice versa - lead to simple corrections of the gross or the net, for example.

ii. Regressions.

If correction was not desirable or possible but information on a directly comparable variable was present anyway, we resorted to direct imputation, via a regression model, of the variable for which input was missing. Below we describe how this was done for net –gross imputation, which were the most prevalent instances of that sort. The method was extended, however, to other imputations (imputations of the 2003 income based on the current income, for example).

Missing values on gross income variables (PY010G, PY020G, … and components) were, if collected, imputed on the basis of the corresponding net variables (PY010N, PY020N, … and components). The implementation of this imputation procedure was quasi-similar for almost all (income) variables on which it was applied. The procedure implied 6-steps:

1. Identification of the 'reference cases' (both gross and net collected) and identification of the cases to be imputed (net collected – gross missing).
2. Calculation of the gross/net ratio for the reference cases. Cases with an extreme value on this ratio were excluded from further use in the procedure.
3. Curve estimation of the relation (regression model) between gross and net income. The best fitting model (linear, logarithmic, quadratic, exponential) was being implemented.
4. Implementation of the regression model for the reference cases to identify outliers.

5. Re-implementation of the regression model for the reference cases after removal of the outliers.
6. Actual imputation step: missing (gross) values are imputed on the basis of
    a) net values and
    b) the estimates for the relation between gross and net income assessed in the steps above.

In step 1 the cases of which both gross and net income were collected are identified. We refer to these cases as 'reference cases' (step 1). The relationship between their net and gross income serves as reference for the imputation of the gross incomes for the cases where only the net was collected (cases to be imputed).

To avoid bias in this imputation model atypical reference cases (both outliers and errors) were identified and removed at several steps in the procedure (step 2 and 4).

In step 2 (reference)cases for whom the ratio between gross and net income exceeded what can be considered typical for the taxation regime applicable to the income concerned, were excluded.
In the case of almost all variables the boundary value of this ratio was set at 2,5. This boundary was arbitrary chosen.
Scrutiny of the excluded cases, however, validates this value's potential to discriminate between incomes which were subjected to real(istic) taxation and outliers or errors.
The latter category seldom counted more than a few percent of the total population in the survey and their gross/net ratio often exceeded the 2,5 considerably.
Further exploration also revealed that the exclusion of these cases from the procedure results in a dramatic increase of the fit of the regression model on which the imputation is based.

In step 4 outliers in the regression model were identified and removed using default regression diagnostics.

The underlying probability model of the net-gross relation was assessed with SPSS' 'curve-estimation' procedure (step 3). It can be hypothesised that in most taxation schemes this relation will not be linear as higher revenues will be subjected to disproportionate higher taxes. The concern therefore is that application of a linear regression model may lead to biased result. Step 3 is an answer to that concern, which turned out to be unfounded, however. In fact, for most variables the linear model fitted the data well. For a few variables the fit of the quadratic model was slightly better, however. Overall, and we underline this, the fit was very good and R-squares very high (always > 0.85)

The estimates of this regression model (step 5) served as direct input for the implementation of the actual imputation (step 6).

iii. Other techniques.

Although we preferred the techniques above we were in some instances forced to resort to other techniques (due to lack of information – for example).

For some cases we imputed median values calculated after categorising using relevant variables. Most of the median values imputed, were for example, calculated after categorisation for status.

2.5.2 Description on imputation per target variable

In the following table is shown which imputation method we used for each target variable (and also for each component within the Belgian questionnaire). The percentage of imputed cases and the total number of observations is added.

Table 11: Percentage of imputation over the total number of observations per (target) variable

Income Component		Question in the Belgian questionnaire		Percentage imputed cases (total number of observations)	Method
Code	**Description**	**Code**	**Description**		
HY040	Income from rental of a property or land	H37	Rental of a part of the house	7.4 (27)	05: Median of the given amounts
HY040	Income from rental of a property or land	H74	Rental of property or land other than own house	3.7 (352)	01: Hot deck (imputation of a randomly drawn given amount)
				3.5 (375)	08: Intervals: imputation of the median of the given amounts falling in the same interval0
				4 (375)	09: deductive imputation: correction amount given for one year but asked monthly
HY040	Income from rental of a property or land			11.1 (368)	
HY050	Family/children related allowances	H91	Child allowance	1.5 (1825)	04: Regression with number of children and age of the oldest child as auxiliary variables

HY050	**Family/children related allowances**	**H93**	**Birth grant**	**5.3 (133)**	**05: Median of the given amounts (in classes based on number of children)**
HY050	**Family/children related allowances**	**(I116A)**	**Income maintenance benefit in the event of childbirth**	**13.1 (61)**	**08: Imputation based on legal amounts**
				13.1 (61)	**09: deductive imputation: correction amount given for the whole period but asked monthly**
				16.4 (61)	**09: Number of months modified based on the calendar**
HY050	**Family/children related allowances**	**(I116B)**	**Parental leave benefit**	**14 (50)**	**08: Imputation of legal amounts**
				6 (50)	**09: Number of months modified based on the calendar**
				12 (50)	**09: deductive imputation: correction amount given for the whole period but asked monthly**
HY050	**Family/children related allowances**			**4.2 (1837)**	
HY060	**Social assistance**	**H71A, H71B**		**2.6 (117)**	**08: Imputation based on legal amounts**
HY070	**Housing allowance**	**H43**	**Allowance for housing (tenants)**	**8.7 (23)**	**05: Median**
HY070	**Housing allowance**	**H26**	**Intervention of authorities for repayments on mortgage**	**22.2 (18)**	**05: Median**

				11.1 (18)	09: correction based on legal amounts
HY070	Housing allowance			19.5 (41)	
HY080	Regular inter-household cash transfer received	H86	Alimony and child support received	0.9 (234) 0.4 (234)	05: Median 09: deductive imputation based on other question (H84: amount that has to be received)
HY080	Regular inter-household cash transfer received	H88	Regular cash support	18.3 (131) 1.5 (131)	01: Hot deck 09: correction of the periodicity
HY080	Regular inter-household cash transfer received			8.4 (346)	
HY090	Interests, dividends, etc.	H99, H100		13 (3073) 52 (3073)	02: Regression (auxiliary variables: sort assets (bank accounts, bonds,…), tenure status, subjective rent) + random term 02: Ranges of values: regression with bounds
HY110	Income received by people aged < 16	H69		11.1 (9)	05: Median
HY130	Regular inter-household cash transfer paid	H79	Alimony and child support paid	0.5 (218)	09: deductive imputation based on other question (H77: amount that has to be paid)
HY130	Regular inter-household cash	H81	Regular cash support	5.2 (249)	01: Hot deck

	transfer paid				
HY130	**Regular inter-household cash transfer paid**			**3.1 (447)**	
HY140	**Tax on income and social contributions**	**I130**	**Repayments for tax adjustment**	**6.7 (1313)**	**08: other source was used: fiscal data**
HY140	**Tax on income and social contributions**	**I132**	**Receipts for tax adjustment**	**4.3 (1213)**	**08: other source was used: fiscal data**
HY140	**Tax on income and social contributions**			**43.2 (4445)**	**Tax was computed as the sum of all differences between gross and net in income variables, corrected by tax adjustment. In case a gross-net model or a net-gross regression was used, the difference (tax) was considered as imputed.**
PY010	**Employee cash income**	**I47-I48**	**Monthly Wages and salaries**	**0.2 (4492)** **10.0 (4492)** **1.0 (4492)** **0.2 (4492)** **2.3 (4492)** **0.4 (4492)**	**1) Corrections** **2) Net income is given, imputation based on regression** **3) current income is given, imputation based on regression** **4) yearly income is given, imputation based on regression** **5) Imputation on basis of EU-SILC 2004** **6) other**

PY010	**Employee cash income**	**I52**	**Nombur of months I47-I48**	**0.4 (4492)**	**1) correction**
PY010	**Employee cash income**	**(i60_a_ne)**	**Pay for overtime**	**31.7 (281)**	**Median % of Employee income - <u>calculated as follows:</u> (1) for all individuals answering i60_a_ne and reporting a net salary the % of i60_a_ne in relation to the net income was calculated (2) the median of this percentage was used (in relation tot the net income) for the cases with non-response on i60_a_ne**
PY010	**Employee cash income**	**(i60_b_ne)**	**Commissions**	**15.4 (52)**	**Median % of Employee income (calculation see PY010 – i60_a_ne)**
				3.8 (52)	**Correction (for ex. Bfrancs → Euro's)**
PY010	**Employee cash income**	**(i60_c_ne)**	**Tips**	**12.5 (16)**	**Median % of Employee income (calculation see PY010 – i60_a_ne)**
PY010	**Employee cash income**	**(i60_d_ne)**	**Additional payments based on productivity**	**18.2 (110)**	**Median % of Employee income (calculation see PY010 – i60_a_ne)**
PY010	**Employee cash income**	**(i60_e_ne)**	**End of the year payments**	**7.8 (3033)**	**Median % of Employee income (calculation see PY010 – i60_a_ne)**
				0.2 (3033)	**Correction (for ex. Bfrancs → Euro's)**
PY010	**Employee cash income**	**(i60_f_ne)**	**Thirteenth month payment**	**9.8 (644)**	**Median % of Employee income (calculation see**

					PY010 – i60_a_ne)
PY010	Employee cash income	(i60_g_ne)	Fourteenth month payment	7.3 (55)	Median % of Employee income (calculation see PY010 – i60_a_ne)
PY010	Employee cash income	(i60_h_ne)	Holiday payments	7.9 (3830)	Median % of Employee income (calculation see PY010 – i60_a_ne)
				0.4 (3830)	Correction (for ex. Bfrancs → Euro's)
PY010	Employee cash income	(i60_i_ne)	Profit sharing	15.7 (108)	Median % of Employee income (calculation see PY010 – i60_a_ne)
				0.9 (108)	Correction (for ex. Bfrancs → Euro's)
PY010	Employee cash income	(i60_j_ne)	Shares	56.2 (73)	Median % of Employee income (calculation see PY010 – i60_a_ne)
				8.2 (73)	Correction (for ex. Bfrancs → Euro's)
PY010	Employee cash income	(i60_k_ne)	Allowances paid for working in remote locations	25.9 (58)	Median % of Employee income (calculation see PY010 – i60_a_ne)
PY010	Employee cash income	(i60_l_ne)	Other additional payments	2.0 (150)	Median % of Employee income (calculation see PY010 – i60_a_ne)
PY010	Employee cash income	I53	Income from irregular jobs : wages and salaries	3.1 (226)	1) corrections
				46.5 (226)	2) net income is given and regression net-gross
PY010	Employee cash income	I91	Income from jobs other than main job : wages and salaries	1.6 (62)	1) corrections

PY010	**Employee cash income**	**I92**	**Number of months Income from jobs other than main job : wages and salaries**	**0.0 (62)**	**No imputations/all values given**
PY010	**Employee cash income**			**24.4 (4492)**	
PY050	**cash benefits or losses from self-employment**	**I78A, I78B**	**Income for main job : self-employed**	**22.9 (547)**	**1) imputation on basis of the household income**
				6.0 (547)	**2) imputation on basis of question I86 – money retrieved from business**
				3.5 (547)	**3) imputation of median values on basis of type of business**
				2.7 (547)	**4) values 2004**
				5.7 (547)	**5)imputation via regression on characteristics of the job**
				4.4 (547)	**6) correction**
				4.0 (547)	**7) net recorded – brut via regression**
				1.8 (547)	**8) minor elements of income were missing and imputed**
PY050	**cash benefits or losses from self-employment**	**I91**	**Income for jobs other than main job : self-**	**23.7 (59)**	**Correction**

			employed		
PY050	**cash benefits or losses from self-employment**			**48.3 (606)**	**Please take notice of the important remarks in 2.6.0 and 2.6.1 to assess the nature of the imputations for the self-employed.**
PY080	**Pension from Individual private plans**	**I109**	**Savings for ones old day (Epargne-pension)**	**0 (13)**	
PY080	**Pension from Individual private plans**	**I112**	**Life insurance (Assurance-vie)**	**0 (13)**	
PY080	**Pension from Individual private plans**			**0 (13)**	
PY090	**Unemployment benefits**	**I98_a**	**Subsistence income for persons entering the labour market**	**15.1 (33)** **18.2 (33)** **39.4 (33)**	**Net-gross model** **04: current income is given, imputation based on regression** **08: Imputation based on legal amounts**

PY090	**Unemployment benefits**	**(i98_b)**	**Full unemployment benefits**	**22.3 (913)**	**09: Number of months modified or imputed based on the calendar**
				24.5 (913)	**04: Net income is given, imputation based on regression**
				1.5 (913)	**06: imputation based on previous 2004**
				13.8 (913)	**09: deductive imputation based on current income or on total income of the household**
PY090	**Unemployment benefits**	**I98_c**	**Partial unemployment benefits**	**31.5 (54)**	**04: Net income is given, imputation based on regression**
PY090	**Unemployment benefits**	**I98_d**	**Other financial assistance (Allocation de garantie de revenus)**	**0 (8)**	
PY090	**Unemployment benefits**	**(I98_e)**	**Other financial assistance (Allocation du fonds de sécurité d'existence)**	**20 (15)**	**04: Net income is given, imputation based on regression**

PY090	**Unemployment benefits**	**(I98_f)**	**Vocational training allowance**	**42.8 (14)**	**04: Net income is given, imputation based on regression**
PY090	**Unemployment benefits**	**(I98_g)**	**Allowance for child care/ mobility payments**	**0 (3)**	
PY090	**Unemployment benefits**	**(I98_h)**	**Other cash benefits**	**28.6 (21)**	**04: Net income is given, imputation based on regression**
				4.8 (21)	**09: correction of specific value**
PY090	**Unemployment benefits**	**I99_b**	**Early retirement benefits**	**14.9 (234)**	**09: Number of months modified or imputed based on the calendar**
				22.3 (234)	**net is given (regresion net-gross)**
				6.8 (234)	**04: last month value is given and used in regression**
				2.1 (234)	**00:no imputation: amount was not given in i99 but in i98_b**
PY090	**Unemployment benefits**			**44.2 (1248)**	

PY100	**Old age benefits**	**I104**	**Pension Fund (Fonds de pension)**	**24.1 (54)**	**09: Monthly interest was given as annuity.**
PY100	**Old age benefits**	**I106**	**Group insurance (Assurance-groupe)**	**11 (9)**	**Conversion lump sum to annuity**
PY100	**Old age benefits**	**(I_102_B)**	**Old age pensions**	**14.2 (1655)**	**09: Number of months modified or imputed based on the calendar**
				29.5 (1655)	**04: Net pension is source**
				2.9 (1655)	**06: observation in 2004 carried forward**
				9.5 (1655)	**04: last month value is given and used in regression**
				2.5 (1655)	**09: deductive imputation based on total income given by the respondent**
PY100	**Old age benefits**	**(I_102_C)**	**Other financial assistance to old aged people[5]**	**9.1 (22)**	**04: Net pension is source**
PY100	**Old age benefits**	**(I_102_D)**	**Other financial assistance to old aged people[6]**	**64.7 (17)**	**04: Net pension is source**

[5] Revenus garantis aux personnes âgées
[6] Complément au revenu garanti aux personnes âgées

PY100	**Old age benefits**	**(I_102_E)**	**Type of old age benefits not given**	**30.4 (46)**	**04: Net pension is source**
PY100	**Old age benefits**			**43.6 (1896)**	
PY110	**Survivor's benefits**[7]	**(I102_A)**		**30.6 (314)**	**04: Net pension is source**
				1.6 (314)	**04: current pension is source**
				3.2 (314)	**06: 2004 observation carried forward**
				0.6 (314)	**05: imputation median**
				0.3 (314)	**09: deductive imputation based on total income given by the respondent (h66)**
PY120	**Sickness benefits**	**(I115_c)**	**Paid sick leave (temporary inability to work due to sickness)**	**16 (119)**	**04: Net income is given, imputation based on regression**
				4.2 (119)	**04: current income is given, imputation based on regression**
				33.6 (119)	**09: corrections based on the calendar and the total income given by the respondent (h66)**
PY120	**Sickness benefits**	**(I115_d)**	**Paid sick leave (temporary inability to**	**50 (6)**	**04: Net income is given, imputation based on**

[7] Individuals could answer 'yes' to the filter of question I102_a and be more than 65 years. After imputation, the values of the benefits were classified as old-age benefits.

			work due to professional sickness or injury)		regression
PY120	Sickness benefits	(I115_e)	Other sickness benefits	47 (17)	04: Net income is given, imputation based on regression
PY120	Sickness benefits			50.6 (162)	
PY130	Disability benefits	I115_a	Disability pension	28.3 (261)	04: Net income is given, imputation based on regression
				8.4 (261)	09: corrections based on last month household income, on the calendar or current income
PY130	Disability benefits	(I115_b)	Integration income for the handicapped	38.2 (34) 8.8 (34)	net is given (gross-net model) 09: correction based on the calendar
PY130	Disability benefits			43.2 (322)	
PY140	Education-related allowances	H95	Grants, scholarship and other educational help to pupils (of secondary schools)	28.7 (150)	01: Hot deck
PY140	Education-related allowances	H97	Grants, scholarship and other educational help to students (of colleges)	24.6 (61)	01: Hot deck

PY140	**Education-related allowances**			**27.5 (211)**	
PY200	**Employee cash income - CURRENT**	**I25-I26**	**Wages and salaries**	**0.0 (3863)**	**No imputations**

Additional remarks on imputations.

- **Gross/Net imputations.**

For a limited number of monetary variables a limited number of respondents had given only a value for the gross variant of the variable (the opposite – only net is given - occurred much more). For these cases a net value was imputed on basis of the gross using the Belgian rules of taxation. A small number of net- pensions and unemployment benefits were imputed in this way.

- **Imputation of 'total housing cost'**

For the calculation of the total housing cost, we examined the current costs for small, average and large usage and used these amounts for both outlier detection and imputation, while taking into account other variables such as the number of household members and the household income. The cost for the water usage for example can be subdivided in subscriber money (fixed) and costs for the actual usage (variable). The cost for the usage of electricity depends largely whether the heating is electric or not: Singles in an apartment without electric heating consume approximately 600 kWh per year (~ 7 euro), while large consumers with accumulation warmth have an annual usage of approximately 20.000 kWh (~ 240 euro).

2.6.3 Imputation of partial unit non-response

The method chosen for Belgium was imputation of an income for each member of the household who did not answer the questionnaire. Imputation is based on the variable RB210 (basic activity status) of the individual given in the R-file. When the answer is missing or 4 (other inactive person), it is chosen not to impute any income. The method for imputation differs with the categories: imputation based on a regression for the wages (no difference between employee and employer, independent variables are age and gender), imputation of a sub-category median for the unemployment and retirement incomes. Net incomes were computed with a gross to net model, based on the imputed gross incomes.
HY025 is calculated as total net disposable income including these individual imputed incomes divided by HY020.

2.6 Imputed rent: only from 2007 on.

2.7 Collection variable company Car

For 2004, the benefit for individuals of using a company car for private goals was not directly assessed at the interview but afterwards calculated by applying the applicable taxation rules.
The fiscal benefit of all nature that a person has - due to disposition of a company car for private goals - is calculated by multiplying the number of kilometres driven for private use by a coefficient. To calculate the latest we need the cylinder capacity of the car. Unfortunately we only asked the mark, type and registration year of the car. Following Eurostat, with these information we could use the method of valuation on the basis of accrued savings. The benefit equals the sum of :

1. depreciation over the reference period in the capital value of the car, plus
2. coverage by the employer of other costs which would normally fall on the user of his/her own car.

The problem for Belgium with this method is that we should calculate the coverage by the employer of other costs, and this is very difficult to estimate.

Finally we decided to work with **the national rules of the tax authorities**. As mentioned before, we need the cylinder capacity.
To calculate the cylinder capacity, we did the following. We assumed that a company car is mostly diesel driven. We looked up for each mark, type and diesel engine what the corresponding cylinder capacity is. If we had several cylinder capacities for the type of the mark, we calculated the weighted mean of the cylinder capacity. If there is not diesel version for a type of car, we did the same logic but than for petrol.

Once we had that we could easily find the corresponding fiscal coefficient. As mentioned before we didn't ask in the questionnaire how many km a respondent does for private goals with his company car. We assumed - using the logic of the national rules of tax authorities- that a person does 5000 km a year for private goals.

After we calculated the fiscal benefit of all nature for a whole year, we weighted it for respondents who didn't dispose for a whole year of the company car. **The fiscal benefit of all nature is a gross non-cash employee income**.
In 2005, we decided to work with **the national rules of the tax authorities**. The benefit for individuals of using a company car for private goals was not directly assessed at the interview but afterwards calculated by applying the applicable taxation rules.
The fiscal benefit of all nature that a person has - due to disposition of a company car for private goals - is calculated by multiplying a fixed amount of kilometres driven for private use by a coefficient. To calculate the latest we need the fiscal cylinder capacity of the car. This fixed amount of kilometres driven for private use is for the tax authorities 5000 km if the distance home-work is less than 25 km, and 7500 if it's more than 25 km.

In 2005, we asked directly the fiscal cylinder capacity and the distance between work and home. In case of non response of the cylinder capacity, we asked the mark, type and registration year of the car. Than we had to use an imputation method.
Imputation: To calculate the cylinder capacity, we did the following. We assumed that a company car is mostly diesel driven. We looked up for each mark, type and diesel

engine what the corresponding cylinder capacity is. If we had several cylinder capacities for the type of the mark, we calculated the weighted mean of the cylinder capacity. If there is not diesel version for a type of car, we did the same logic but than for petrol.

Once we had that we could easily find the corresponding fiscal coefficient. Than we only had to multiply it by the fixed amount of kilometres driven for private use to obtain the fiscal benefit of all nature

Example:

Type of car	Fiscal cylinder capacity	Forfait	Distance home work	Fixed amount	Fiscal benefit of all nature
Smart fortwo	5	0,1864	< 25 km	5000	931 €
Smart fortwo	5	0,1864	> 25 km	7500	1396 €

After we calculated the fiscal benefit of all nature for a whole year, we weighted it for respondents who didn't dispose for a whole year of the company car. **The fiscal benefit of all nature is a gross non-cash employee income**.

3.Comparability

All household members of 16 year and older **at the time of the interview**, are selected for a personal interview. From 2006 on the age of 16 will be calculated at the end of the income reference period.

3.1 Basic concepts and definition

Only changes from first wave are reported.

Basic information on activity status during the income reference period
Basic information on activity status during the income reference period was mainly obtained via the calendar question (I40) in contrast to 2004 where it was obtained by combining the answer for question I8 (PL030) with the answer(s) for question(s) I38 (PL200) and for those with a change I40 (calendar question)). ALSO SEE REMARK 2.5.0.

3.2 Components of income

3.2.1 Differences between the national definitions and standard EU-SILC definitions, and an assessment, if available, of the consequences of the differences mentioned will be reported for the following target variables.

Total household gross income
HY010 = PY010 + PY020G + PY050G + PY090G + PY100G + PY110G + PY120G + PY130G + PY140G + HY040G + HY050G + HY060G + HY070G + HY080G + HY090G + HY110 G.

PY020G was not part of HY010 for 2004.
For 2005 PY020G only contains the value of company cars.

Family/children related allowances

For the SILC 2004 Belgium asked allowances received from the federal government. In 2005 it also includes birth grants given by some local authorities and medical organizations.

Income received by people aged under 16: in 2004 we asked the amount for last month (current) but the reference period for the variable is income reference period (year 2003). This was corrected for 2005 and the question aimed at the total income received last year by people aged fewer than 16.

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

No change from the first wave.

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

No change from the first wave.

3.2.4 The method used for obtaining income target variables in the required form (i.e. gross values)

See above for information on control, correction, imputation and creation of the gross target variables.

Tracing rules
Although the 'tracing rules' from Eurostat say that sample households non enumerated the first year of the panel 'may be dropped', some households who did not participate in 2004 were contacted in 2005. These cases concern households who were not interviewed in 2004 because they were temporarily away, unable to respond due to illness or due to other reason (DB130=22 to 24).

4 Coherence

Below we present the results of a number of analyses we have done to check the coherence of the EU-SILC 2004 and EU-SILC 2005 data. We make a difference between 'internal' coherence and 'external' coherence. With the first we mean the coherence between the two available EU-SILC waves. With the later we mean the coherence between EU-SILC and other datasources. Coherence is one of our main concerns so we are working very hard to expand our insights on this point. The results we show below are therefore to be seen as merely a start.

4.1 EXTERNAL COHERENCE

We compared both EU-SILC waves on a number of critical dimensions with comparable information coming from the Labour Force Survey (LFS) and the HouseholdBudget Survey (HBS). These results we show below and other results form are elaborately discussed in a separate document on coherence. This document is for now however only available in Dutch. Our apologies for that. The document is public however and our experts have full insight in it f.ex..

COHERENCE BETWEEN EU-SILC and LFS – SELF-DEFINIED SOCIAL-ECONOMIC STATUS.

males

	werkend	werkloos	pensioen	nt. actief
LFS2003	2276554	269851	955194	643824
	54,9	6,5	23,0	15,5
LFS2004	2312246	260494	946633	644675
	55,5	6,3	22,7	15,5
LFS2005	2342481	257918	976181	632167
	55,7	6,1	23,2	15,0
LFS2006	2350703	256465	994314	642357
	55,4	6,0	23,4	15,1
HBS2005	2242242	229145	977353	885418
	51,7	5,3	22,6	20,4
SILC2004	2270129	287924	956442	550357
	55,8	7,1	23,5	13,5
SILC2005	2213854	279771	916696	596901
	55,2	7,0	22,9	14,9

females

	werkend	werkloos	pensioen	nt. actief
LFS2003	1713957	315502	1092267	1285996
	38,9	7,2	24,8	29,2
LFS2004	1737427	332656	1098738	1265777
	39,2	7,5	24,8	28,5
LFS2005	1806896	312967	1122029	1229850
	40,4	7,0	25,1	27,5
LFS2006	1837515	301739	1106693	1259475
	40,8	6,7	24,6	28,0
HBS2005	1915198	324055	791937	1581246
	41,5	7,0	17,2	34,3

SILC2004	1776202	348165	976618	1179634
	41,5	8,1	22,8	27,6
SILC2005	1761316	374341	980481	1124612
	41,5	8,8	23,1	26,5

COHERENCE BETWEEN EU-SILC and LFS – HAVING WORKED LAST WEEK.

males

	gewerkt	niet gewerkt
LFS2003	2064660	2080763
	49,8	50,2
LFS2004	2062546	2101502
	49,5	50,5
LFS2005	2066132	2143324
	49,1	50,9
LFS2006	2090783	2153056
	49,3	50,7
HBS2005	2237110	2104708
	51,5	48,5
SILC2004	1870853	2194217
	46,0	54,0
SILC2005	1878568	2129172
	46,9	53,1

females

	gewerkt	niet gewerkt
LFS2003	2917541	1490181
	66,2	33,8
LFS2004	2937906	1496691
	66,2	33,8
LFS2005	2911756	1560225
	65,1	34,9
LFS2006	2910029	1595495
	64,6	35,4
HBS2005	2866424	1749195
	62,1	37,9
SILC2004	2608653	1673006
	60,9	39,1
SILC2005	2572819	1667931
	60,7	39,3

COHERENCE BETWEEN EU-SILC and LFS – STATUS OF OCCUPATION.

males

	1	2	3	4	5
LFS2003	809516	640435	378452	100610	387993
	34,9	27,6	16,3	4,3	16,7
LFS2004	801989	660348	380096	108003	403830
	34,1	28,0	16,1	4,6	17,2
LFS2005	800089	686425	373302	108968	418527
	33,5	28,8	15,6	4,6	17,5
LFS2006	830320	690047	345657	103915	421745
	34,7	28,9	14,5	4,3	17,6
SILC2004	745464	662195	401395	105181	357367
	32,8	29,2	17,7	4,6	15,7
SILC2005	653893	711785	402162	97405	342052
	29,6	32,2	18,2	4,4	15,5

females

	1	2	3	4	5
LFS2003	297477	710360	324260	199542	221774
	17,0	40,5	18,5	11,4	12,6
LFS2004	291766	716523	335768	226521	214329
	16,3	40,1	18,8	12,7	12,0
LFS2005	291781	764216	347046	218873	226626
	15,8	41,3	18,8	11,8	12,3
LFS2006	316449	792534	340734	201398	221178
	16,9	42,3	18,2	10,8	11,8
SILC2004	255361	777996	394650	161538	148298
	14,7	44,8	22,7	9,3	8,5
SILC2005	241899	770725	401090	166975	136858
	14,1	44,9	23,4	9,7	8,0

Categories:
1 = bleu collar – private sector
2 = white collar – private sector
3 = public sector
4 = public sector – contractual
5 = self-employed

COHERENCE BETWEEN EU-SILC, HBS and LFS – WORKREGIME.

males		
	fulltime	parttime
LFS2003	2168741	141503
	93,9	6,1
LFS2004	2193328	153909
	93,4	6,6
LFS2005	2205119	182193
	92,4	7,6
LFS2006	2215761	175922
	92,6	7,4
HBS2003	1760370	89610
	95,2	4,8
HBS2004	1559387	133713
	92,1	7,9
HBS2005	1864810	159866
	92,1	7,9
SILC2004	2131770	132247
	94,2	5,8
SILC2005	2050219	159865
	92,8	7,2

females		
	Full -time	parttime
LFS2003	1067831	656223
	61,9	38,1
LFS2004	1061632	690309
	60,6	39,4
LFS2005	1099035	749406
	59,5	40,5
LFS2006	1103209	769084
	58,9	41,1
HBS2003	960212	607000
	61,3	38,7
HBS2004	956179	573168
	62,5	37,5
HBS2005	1115554	802871
	58,1	41,9
SILC2004	1027534	705473
	59,3	40,7
SILC2005	1041340	684653
	60,3	39,7

COHERENCE BETWEEN EU-SILC, HBS and LFS : TYPE OF CONTRACT.

males		
	indefinite duration	temporary
LFS2003	1808979	120035
	93,8	6,2
LFS2004	1826404	124032
	93,6	6,4
LFS2005	1834977	133807
	93,2	6,8
LFS2006	1833373	136658
	93,1	6,9
HBS2003	1716910	112468
	93,9	6,1
HBS2004	1563208	103730
	93,8	6,2
SILC2004	1753028	159604
	91,7	8,3
SILC2005	1723682	142762
	92,4	7,6

females		
	indefinite duration	temporary
LFS2003	1361580	170059
	88,9	11,1
LFS2004	1387489	183088
	88,3	11,7
LFS2005	1437509	184408
	88,6	11,4
LFS2006	1471766	179351
	89,1	10,9
HBS2003	1379428	161809
	89,5	10,5
HBS2004	1324352	177075
	88,2	11,8
SILC2004	1392773	196572
	87,6	12,4
SILC2005	1401029	185888
	88,3	11,7

COHERENCE BETWEEN EU-SILC, HBS an LFS – MEDIAN MONTHLY SALARY.

	MEDIAN
LFS2003	1383,3333
LFS2004	1415
LFS2005	1466,6667
LFS2006	1500
HBS2003	1454,42
HBS2004	1445,94
HBS2005	1507,95
SILC2004	1535,7172
SILC2005	1557,6667

COHERENCE BETWEEN EU-SILC, HBS – MEDIAN DISPOSABLE HOUSEHOLD INCOME.AT RISK OF POVERTY THRESHOLD

EQ_INC

	MEDIAN	ARPT
HBS2003	16598,3	9959,0
HBS2004	16824,0	10094,4
HBS2005	17432,1	10459,2
SILC2004	15540,0	9324,0
SILC2005	16437,9	9862,7

COHERENCE BETWEEN EU-SILC, HBS – AT RISK OF POVERTY RATES

% BELOW ARPT

	ALL	MALES	FEMALES	<=15yrs.	15-64 yrs.	>=65 yrs
HBS2003	14,3	14,1	14,4	12,1	13,3	22,1
HBS2004	14,0	13,6	14,5	11,9	13,6	21,5
HBS2005	14,8	14,2	15,5	14,1	13,5	22,5
SILC2004	14,9	13,9	15,8	17,5	12,8	20,6
SILC2005	14,8	14,1	15,5	18,6	12,4	20,6

4.2 INTERNAL COHERENCE

To visualize evolutions between two waves we have calculated for every monetary variable several matrixes. The example below is for the median equivalised income (weighted).

	WEIGHTED	VALUES 2004				VALUES 2005		
	MEDIAN EQ_INC	2004-ALL	2004-ROT.GR.1	PRESENT 2004 - MISSING 2005	PRESENT 2004 - PRESENT 2005	PRESENT 2004 - PRESENT 2005	NEW - 2005	2005-ALL
VALUES 2004	2004-ALL	15540,0						
	2004-ROT.GR.1	98,5	15301,4					
	PRESENT 2004 - MISSING 2005	99,0	100,6	15388,7				
	PRESENT 2004 - PRESENT 2005	101,3	102,9	102,3	15742,4			
VALUES 2005	PRESENT 2004 - PRESENT 2005	107,8	109,5	108,9	106,4	16756,7		
	NEW - 2005	102,7	104,3	103,7	101,4	95,2	15957,6	
	2005-ALL	105,8	107,4	106,8	104,4	98,1	103,0	16437,9

In this matrix the median equivalised income for several parts of the panel:

- For the observations in 2004 :
  1 the population belonging to the rotational group to be removed in Y+1,
  2 the population refusing to further participate in Y+1
  3 the population participating in both Y and Y +1
- For the observations in 2005
  4 the population participating in both Y and Y+1
  5 the new population

The observations for population 3 and 4 are direct comparable (respectively coming from 2004 - 15742.4 Eur. and 2005 – 16756,7 Eur.). The 106,4 implies that there was for the population that was observed both in 2004 and 2005 an increase in median equivalised income between 2004 and 2005 of 6,4%. The other results are interpreted in a similar way.

We repeat that we have calculated these matrices for all monetary variables (both P and H) and a number of other variables. For each variable the matrix is calculated for the mean and the median in both a weighted and an unweighted version. We decided not to copy them all here. But if asked we would gladly provide you the excel-file with these matrices in it.