Final Quality Report SILC2006 - BELGIUM

0. Introduction

This report contains a description of the accuracy, precision and comparability of the Belgian SILC2004 to SILC2006-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 Sampling Design

2.1.1 Type of sampling (stratified, multi-stage, clustered)

The Belgian EU-SILC 2006 survey is a 2-stage sampling.
There is stratification of sampling units.
There is no clustering of sampling units.

2.1.2 Sampling units (one stage, two stages)

Primary units:
The *Primary Sampling Units* are the municipalities (or part thereof in the larger ones); in each of the 11 strata, they were drawn PPS, i.e. with repetitions allowed (for instance, Schaerbeek was drawn 6 times). In total, 275 draws were made in 2004, once forever (for the whole duration of EU-SILC).
Secondary units:
The *Final Sampling Units* are the (private) households.

Recall that, in 2004, 40 households had been selected in each PSU, numbered 1 to 40.

The first 10 (whether or not they responded irrelevant) vanished from the panel in 2005 (to be replaced by newly drawn households), the second 10 in 2006; the other 20 (including possible split-offs) were followed according to the tracing rules.

Hence, the (cross-sectional) sample of SILC 2006 consists of

- "old" (longitudinal (2004-)2005-2006, perhaps longer) households

and

- "new" households (drawn in 2006, staying until 2009).

In fact, it is only the selection of the new households that gave us some degree of freedom (see in particular 2.1.4)

In the D-file, three variables have been added:

- ✓ DB061 is the identification of the primary units (concatenation of 5 digits for the municipalities and one letter).
- ✓ DB063 is the 'multiplicity order', the number of times each letter was drawn in the sample.
- ✓ DB071 is the order of selection of the new households within each letter.

2.1.3 Stratification and sub-stratification criteria

The stratification criterion is the region (NUTS2 level). The 11 strata are the 10 provinces of Belgium and the Brussels Capital Region.

2.1.4 Sample size and allocation criteria

Although our initial intention was draw 10 new households in each PSU (whether or not they would respond unknown). However, the first wave (2005) participation was a bit disappointing:, so we drew many more households in 2005. In 2006 (and hopefully for the next years!) we managed to keep the number of responding households close to 6000, drawing 16 new hh in each PSU**.**

Table 1: sample size and achieved response by NUTS2-units

NUTS2	Name	Old (or strange) hh	New hh	Total hh	Accepted hh (DB135=1)
BE10	Brussels	1039	587	1626	800
BE21	Antwerpen	1108	410	1518	903
BE22	Limburg	427	185	612	402
BE23	Oost-Vlaanderen	783	341	1124	674
BE24	Vlaams-Brabant	618	270	888	515
BE25	West-Vlaanderen	547	295	842	631
BE31	Brabant Wallon	150	106	256	156
BE32	Hainaut	924	369	1293	788
BE33	Liège	645	310	955	592
BE34	Luxembourg	170	63	233	167
BE35	Namur	285	116	401	232
Total	**Belgium**	6696	3052	9748	5860

2.1.5 Sample selection schemes

Systematic sampling of secondary units (new households) in each primary unit selected, the households have been ordered according to the age of the reference person.

2.1.6 Sample distribution over time

2.1.7 Renewal of sample: Rotational groups

See above.

2.1.8 Weightings

Recall that, for the first year of the panel (=SILC 2004 in Belgium), the computation of weights involved three stages (described in 134-04)

(a) initial weights
(b) weights corrected for nonresponse
(c) final (calibrated) weights

In 2006, a distinction has to be made between

<u>"old" households</u> i.e. households that contain at least one sample person who took part in 2005, and had to be surveyed again in 2006 according to the rotation and tracing rules (excluding the outgoing fourth) (household composition may have changed, whence quotations marks)

<u>"new" households</u> i.e. households that were drawn for the first time in 2006, among those households not containing any sample person drawn in 2004 or 2005 (quotations marks superfluous)

<u>"strange" households</u> i.e. a small number of hh that were drawn in 2004, did not take part in 2005 but did in 2006

This distinction pertains to

Since the "old" households are selected indirectly from the 2005 sample, and household composition may have changed, some kind of "weight sharing" must be applied to determine the (2006) initial weights, or rather base weights. On the other hand, "new" households have their own inclusion probability, whose inverse gives the initial weights;

For the "old" households, (2006) nonresponse=attrition can be linked with (2005) SILC information. For the "new" households, all we can rely upon to correct for initial nonresponse is auxiliary information (household size, urban/rural character,..) from the Population Register. We chose to give strange households their initial weight, without any correction.

On the other hand,

Calibration can be done together for "old", strange and "new" households. With respect to our 2004 model, we decided to relax the constraints (basically, calibrating at NUTS1-level instead of NUTS2), in order to decrease the standard deviation of weights.

This introduces the following sections

2.1.8.1 Initial weights for the new households
2.1.8.2 Nonresponse correction for the new households
2.1.8.3 Base weights for the old households
2.1.8.4 Attrition correction for the old households
2.1.8.5 Calibration (all households)

2.1.8.1. Initial weights for the new households

Belgium chose to draw the Primary Sampling Units (= municipalities or parts thereof) "forever", and to rotate the Secondary Sampling Units (=households) within the selected PSU's.
The 2004 PPS two-stage sampling design was self-weighting within each stratum h: x denoting any households in municipality X), we had (in 2004)
P (x drawn) = P(x drawn|X drawn) . P(X drawn) = $n_h/N_X \,.\, N_X/N_h \,.\, g_h = n_h/N_H \,.\, g_h$, where
n_h denotes the number of households to be drawn in the (selected) PSU (viz. 40)
N_X the number of households in the PSU (in 2004)
N_h the number of households in the stratum (in 2004)
g_h the number of PSU's drawn in the stratum.
(This is an oversimplification, since PSU are drawn with repetition; the selection probability for a PSU should be replaced by the expectation of selection multiplicity, and the term 40 by a multiple depending on the selection multiplicity…but the idea is the same).

In 2005, the picture had become
P (x drawn) = P(x drawn|X drawn) . P(X drawn) = $m_h/M_X \,.\, N_X/N_h \,.\, g_h$, where
m_h is the number of households to be drawn in the (selected) PSU (depending on h)
M_X is the number of households in the PSU (in 2005)
The factor N_X/M_X indicates the increase-decrease in inclusion probabilities in PSU X (still assuming X has been drawn) between 2005 and 2004.
Now it would seem logical to replace N_X by N_X-(40+x), to account for the 40 households[1] already drawn in 2004 and the x (depending on the province) drawn in 2005, whence immunized from being drawn again in 2006.
However, the following argument shows that (assuming momentarily that X has been drawn) matters are not so easy:

P(x drawn in 2005) =
(P(x drawn in 2005|x drawn en 2004) . P(drawn in 2004)) +
(P(drawn in 2005|x not drawn in 2004) . P(not drawn en 2004)),

the first term vanishes and the second equals $n_h/(M_X\text{-}40)$. $(N_X\text{-}40)/N_h$; since both fraction terms are much larger than 40 (at least 900 in all selected PSU's, for both years), the ratio $(N_X\text{-}40)/(M_X\text{-}40)$ is very close to N_X/M_X. Since the term 40 is an

[1] Perhaps a bit less (households that vanished already subtracted) or a bit more (split households, both components of which stayed in PSU, should be subtracted twice)

approximation anyway[2], we chose to stick to m_h/M_X . N_X/N_h. g_h as inclusion probabilities, and its inverse for initial weights **INIwei=DB080**.

The same argument applies in 2006.

2.1.8.2.Nonresponse correction for the new households

Following Eurostat's suggestion, we replaced the homogeneous response groups (based on household size crossed with urbanity) ratio by a multiple regression model (based on the same dummy variables). By "responding", we mean only those households whose results were accepted (DB135=1). For technical reasons, we used linear regression instead of logistic; since the (predicted) response turned out to be close to 50% for all categories, this is harmless.
The variables used were

- DB100 = urbanity, taking three values [3]
- HOUSEHOLD size, recoded into the two values "single" and "more"[4]

The regression produced a new variable "expresp", allowing us to define
NRwei = INIwei/expresp

---

[2] See previous footnote; determining how many of the new households in the PSU are new in the population, how many immigrated from a selected PSU and how many from a non-selected PSU would be too complicated, similarly we will investigate what happened to all households that belonged to the PSU in 2004 but no longer in 2005

[3] categorical variable, but since the answer rate depends almost linearly on the coding, we elected to regard DB110 as a numerical variable.

[4] HOUSEHOLD size is a numerical variable, but it appears that the greatest response contrast separates on the one hand the 1-person households, on the other the 2+-person households (the impact of 6+-person households is marginal), whence the recoding into a dichotomic variable. Note that a few (nonresponding) households whose size remained unknown were excluded, explaining the slightly higher response rate.

Size	Rate
1	39,0
2	50,0
3	51,7
4	55,0
5	56,8
6	49,2
7	53,1
8	84,6
9	37,5
10	100,0
11	100,0
Total	47,8

2.1.8.3 Attrition for the old households

Before “sharing” the 2005 weights (to define the 2006 base weights for the “old” households, a correction for attrition should be introduced. This year, we elected to perform this correction at the level of individuals, since a 2005 sample person either stays in the panel or leaves it (rotated out, left population, noncontact, refusal or inability to respond, while the structure of a household can change. However, this distinction (between attrition predictors) is not essential

- individual characteristics (e.g. PE040) can be averaged (actually have to be, since it is collected only for adult members) to the household,

in the same fashion as

- household characteristics (e.g. HH020) could be distributed to the members.

The strongest attrition predictor turned out to be a combination of PB090 and PB100 (from the P-file, but almost always identical for all hh members), showing that “late” 2005 respondents were more likely to drop out:

Month of interview 2005	Attrition in 2006	Stayed in 2006
Sep	16,2	83,8
Oct	20,3	79,7
Nov	20,7	79,3
Dec	26,6	73,4
Total	19,7	80,3

The final (regression) model used nationality, educational attainment and tenure; we used linear regression, since the attrition rates were far from 0 and from 1.

2.1.8.4 Weight sharing

Like one year ago, we followed Eurostat’s recommendation "EU-SILC weighting procedures: an outline" and shared the calibrated 2005 weights (instead of the initial weights, see Lavallée).
Fortunately, no respondent (2006) household was the result of a fusion (viz. DB110=10), so weight sharing amounted to defining the quotient …**SHACOF** = "# 2006 household members already in household in 2005 (or age=0)/” # 2006 household members", this quotient is <= 1, and was 1 in most cases.
However, in quite a few cases **SHACOF** turned out to be 0, meaning these households should in fact not have been interviewed (it must be admitted that the tracing rules are ambiguous if “DB130=24”; we decided to keep these households in the database, giving them their initial weight, without nonresponse correction)

2.1.8.5 Calibration

We first put the pieces together:

- new = started in 2006 (initial weight, corrected for initial nonresponse)
- old = took part in 2005 (2005 weight, corrected for attrition)
- strange = did not take part in 2005 (initial weight, non correction)

In Belgium, 11 *sampling* strata were used (provinces= NUTS2).
In order to avoid a large std of calibrated weights, after reuniting the three sorts of households (

- we use 3 *extrapolation* strata (the 3 NUTS1 regions BRUssels=BE1, VLAanderen=BE2 and WALlonia=BE3)
- we performed a transformation (up to a coefficient, replacing the weights by their square root), analogous to truncation, but perhaps a bit less radical

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

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

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 4 (size: "1", "2", "3 or "4 & more",)

Calibration type (after some trials and errors…): linear (the square root transformation making truncation superfluous)

This produced the (final) 2006 cross-sectional weights.

2.1.8.7 Final longitudinal weights

Recall that the longitudinal (household or individual) files is obtained by merging two or three datasets, corresponding to different survey years (DB010 for households, RB010 for individuals). Instead of DB075, we found it a bit easier to identify the rotational groups by their last year in the panel, Year_end (this determines the first year: 2005 if Year_end = 2008, 2004 else).

We have eight basic blocks:

↓DB010 Year_end→	2006	2007	2008
2004	X	X	
2005	X	X	X
2006	X	X	X

Most longitudinal analyses are conducted at the personal level (indeed: household composition may change).
After cleaning the file (setting the weight RB060 to zero if RB110 took any of the values 3, 5, 6 or 7) we defined (by rescaling the various RB050) RB060 in such a way that each of the eight blocks "represents" the whole population (some ten million people), then defined (only for those still present in 2006) RB062 [RB063] in such a way that the sum of the three [two] rotational groups involved corresponds to the whole population.

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

2.1.8.8. Final cross-sectional weights

See above for the process (by the way, wouldn't it be more natural to switch 2.1.8.7 and 2.1.8.7?)

Statistics

	N	Minimum	Maximum	Mean	Std. Dev.
Final weights	5860	192,51	3292,18	771,67	246,75

Remark:

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

- 2004:(exponential calibration) 5275 households, range of final weights [135 → 5817], mean 842, std 293;
- 2005:(truncated, 0.4≤g≤2) 5137 households, range of final weights [58 → 7879], mean 871, std 326.
- 2006: linear calibration, after weight trimming 5860 households, range of final weights [192 → 3292], mean 772, std 245.

2.1.9 Substitutions

No substitution was applied in our survey.

2.2 Sampling errors

The table is based on the results of EU-SILC 2006.

Income components	Mean	Number of observations before imputation	Number of observation after imputation	Standard error
HY010	41089.9539	2285	5851	2871.7
HY020	29905.9143	2248	5854	1637.2
HY022	26044.3338	2096	5658	1645.9
HY023	21047.9784	1979	5486	1642.3
Net income components at household level				
HY030N				
HY040N				
HY090N	2564.0	1360	3694	3167.4
HY050N	3332.6	2063	2124	75.8
HY060N				
HY070N				
HY080N				
HY100N				
HY110N				
HY120N				
HY130N				
HY140N				
HY145N				
Gross income components at household level				
HY030G				
HY040G	7109.6	402	429	1027.0

HY090G	2564.0	1360	3694	3167.4
HY050G	3360.3	2049	2124	77.5
HY060G	6578.9	94	94	772.1
HY070G	1559.5	46	61	652.0
HY080G	3416.8	416	437	327.3
HY100G	2814.6	1528	1795	86.7
HY110G	1718.5	13	14	895.2
HY120G				
HY130G	3824.8	471	481	704.8
HY140G	13104.3	3274	5192	1369.9
net income components at personal level				
PY010N	19501.6	4739	5327	297.2
PY020N	1883.4	253	387	61.3
PY035N				
PY050N	19866.0	533	758	1393.6
PY070N				
PY080N	2871.7	22	27	1226.7
PY090N	8031.8	1386	1453	386.2
PY100N	13062.0	1943	2056	244.1
PY110N	11672.7	90	92	809.6
PY120N	5193.5	174	192	544.7
PY130N	9222.6	361	398	375.0
PY140N	648.8	218	225	103.8
gross income components at personal level				
PY010G	29765.5	4082	5327	508.7
PY020G	1883.4	253	387	61.3
PY030G				
PY035G				
PY050G	23689.7	293	758	1534.0
PY070G				
PY080G	2871.7	22	27	1226.7
PY090G	8739.3	968	1453	483.0
PY100G	15098.9	1372	2056	367.3
PY110G	13095.1	59	92	1684.9
PY120G	5377.2	152	192	558.1
PY130G	9669.8	277	398	405.4
PY140G	648.8	218	225	103.8
PY200G				

Equivalised disposal income	Mean	Number of observations before imputation	Number of observation after imputation	Standard error
Subclasses by household size				
1 household member	18204.9	767	1637	4713.1
2 household members	18518.6	1400	3898	525.1
3 household members	19822.2	978	2778	480.3
4 and more	19225.8	2066	6002	523.8
Population by age group				
<25	18281.4788	1785	4732	396.7
25 to 34	20823.0994	580	1618	576.8
35 to 44	21813.6713	823	2206	4534.7

45 to 54	20451.4242	676	2059	571.6
55 to 64	19548.2897	604	1689	732.3
65+	14525.7095	743	2016	408.6
Population by sex				
Male	19681.6062	2512	6993	281.0
Female	18348.6294	2699	7327	523.8

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.

• 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.**”**

• ***Error in the routing wave 2006***

There was one error in the routing. In the household questionnaire, in the part concerning childcare, the selection was made on the base of actual age instead of age in the income reference period. So we missed information for some children born in 1993 or 1994.

• ***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.

• ***Correspondence French/Dutch/German versions of Questionnaires wave 2006***

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 about 0,2% 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 **longitudinal database 2004-2006**:

2004	2005	2006
2638	4112	3618

- 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 **longitudinal database 2004-2006** and who completed a personal interview:

	2004	2005	2006
Persons 16 y and more	5153	8075	7016
Sample persons	5153	7973	6744
Co-residents with interview		102	272

2.3.3.2. Unit non-response

Response rate for households

- Wave response rate

$$\text{Wave response rate} =$$
$$= \frac{5860}{9796 - 59} = 60\%$$

$$\text{Refusal rate} =$$
$$= \frac{2413}{9796 - 59} = 24.8\%$$
$$\text{Non contacted and others rate} =$$
$$= \frac{1319}{9796 - 59} = 13.6\%$$

- Longitudinal follow-up rate

$$\text{Longitudinal follow - up rate} =$$
$$= \frac{4387}{4387 + 899} = 83\%$$

- Follow-up ratio:

$$\text{follow - up ratio} =$$
$$= \frac{4387 + 2132}{4387 + 899} = 1.24$$

- Achieved sample size ratio

$$\text{Achieved sample size ratio} =$$
$$= \frac{5860}{5298} = 1.11$$

SAMPLE OUTCOME IN WAVE3

DB130=11										
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)	DB1 (K)
5860	0	18	475	145	775	2413	18	29	4	

SAMPLE OUTCOME IN WAVE2

DB130=11	DB135=1									
		3772	0	0	192	72	351	850	1	44	4
3772	DB135=2	0	0								
DB120=21 to 23											
DB130=21 to 24											
TOTAL											

NEW HOUSEHOLDS IN WAVE 3

DB110=8										
	78	0	0	18	1	25	53	2	NA	NA
DB110=9										
	2010	0	3	265	72	399	1510	15	NA	NA

Personal interview response rates

Response rate for persons

- Wave response rate

Wave response rate of sample persons =

$$= \frac{10983}{11117} = 98.8\%$$

- Wave response rate of non sample persons:

$$= \frac{244}{259} = 94\%$$

- Longitudinal follow-up rate:

$$= \frac{11206}{11340} = 98.8\%$$

Rate (RB250=21) $= \frac{17}{11340} = 0.1\%$

Rate (RB250=23) $= \frac{17}{11340} = 0.1\%$

Rate (RB250=31) $= \frac{37}{11340} = 0.3\%$

Rate (RB250=32) $= \frac{4}{11340} = 0.01\%$

Rate (RB250=33) $= \frac{1}{11340} = 0.01\%$

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

- Achieved sample size ratio

$$= \frac{11227}{9974} = 112\%$$

- Response rate for non-sample persons

$$= \frac{291}{309} = 94\%$$

Personal interview response rate in wave 2

		RB250=11,12,13	Not completed because of RB250=21	RB250=22	RB250=23	RB250=31	RB250=32	RB250=33	HHnc	Pn	Pl	TOTAL
Sample persons (RB100=1 and rb245=1-3) from the sample forwarded from last wave												
(1) RB110=1-2		7042	8	0	10	18	2	1				7081
(2) RB110=6												13
(3) RB110=-1												0
(4) RB120=2												6
(5) RB120=3												8
(6) RB120=4												61
(7) DB135=2 or -1 or DB110=7 or DB120=21-23 or DB130=21-24 or -1												0
(8) DB110=3-6												0
												0
New sample persons												0
(9) Reached age 16												0
(10) Sample additions		3941	6	0	7	19	2	0				3975
Non-sample persons 16+												
	From w 1	244	0	0	6	7	1	1				259
(11) this wave	Not in w1	47	0	0	2	1						50
(12) Earlier wave	From w 1											
	Not in w1											
Sample persons from sample not forwarded from last wave (excluded died or non eligible)		223										223

Sum of rows	10983	14	0	17	37	4	1	0	0	0	11117
	11206	14	0	17	37	4	1	0	0	0	11340
	11030	14	0	19	38	4	1	0	0	0	11167

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	4670	4188	248	18	8	19	0	3	182		4
%	100	89.7	5.3	0.4	0.2	0.4	0	0.1	3.9		0.1

Record of contact at address
DB120=

	Total	11	21	22	23	missing
Total (DB110=2,8,10)	434	425	3	0	2	4
%	100	97.9	0.7	0	0.5	0.9

Household questionnaire result
DB130=

	Total	11	21	22	23	24	missing
Total (DB120=11 or DB110=1)	4613	3618	585	142	40	224	4
%	100	78.4	12.7	3.1	0.9	4.9	0.1

Household interview acceptance
DB135=

	Total	1	2	missing
Total(DB130=11)	3618	3618	0	0
%	100	100	0	

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	8913	8414	100	247	72	63	13	4	
%	100	94.4	1.1	2.8	0.8	0.7	0.1	0.0	

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	204	0	130	9	8	57
%	100	0	63.7	4.4	3.9	27.9

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=14	RB250=21	RB250=23	RB250=31	RB250=32	RB250=33
Total	7016	6962	52	0	0	0	0	2
%	100	99.2	0.7	0	0	0	0	0.1

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

	Total	RB250=11	RB250=14	RB250=21	RB250=23	RB250=31	RB250=32	RB250=33
Total	6744	6706	36	0	0	0	0	2
%	100	99.4	0.5	0	0	0	0	0.1

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

	Total	RB250=11	RB250=14	RB250=21	RB250=23	RB250=31	RB250=32	RB250
Total	272	256	16	0	0	5	1	0
%	100	94.1	5.9	0	0	0	0	0

Distribution of household members aged 16 and over by RB260

(Household members 16 + RB250=11)

	Total	RB260=2	RB260=5
Total	6962	5929	1033
%	100	85.2	14.8

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

	Total	RB260=2	RB260=5
Total	6706	5747	959
%	100	85.7	14.3

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

	Total	RB260=2	RB260=5
Total	256	182	74
%	100	71.1	28.9

2.5 Imputation procedure

2.5.0 Preceding important remark

In contrast to 2004 and as 2005 – in 2006 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. 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.

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

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.)

- An integrated strategy.

As it was the case for SILC-2004 & 2005 we used for SILC-2006 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 2005 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, <u>the fit was very good and R-squares very high</u> (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	
Code	Description	Code	Description	(total number of observations)	Method
HY040	Income from rental of a property or land	H37	Rental of a part of the house	0 (21)	
HY040	Income from rental of a property or land	H74	Rental of property or land other than own house	2.2(446) 3.4 (446) 0.9 (446)	01: Hot deck (imputation of a randomly drawn given amount) 08: Intervals: imputation of the median of the given amounts falling in the same interval 09: deductive imputation: correction based on answer in 2005
HY040	Income from rental of a property or land			6.3 (429)	
HY050	Family/children related allowances	H91	Child allowance	1.1 (2108)	04: Regression with number of children and age of the oldest child as auxiliary variables
HY050	Family/children related allowances	H93	Birth grant	5.9 (170)	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	11.6 (112) 1.8 (112) 16.1 (112)	08: Imputation based on legal amounts 09: deductive imputation: correction amount given for the whole period but asked monthly Net-gross model

HY050	**Family/children related allowances**	**(I116B)**	**Parental leave benefit**	**27.1 (48)**	**08: Imputation of legal amounts**
HY050	**Family/children related allowances**			**3.5 (2124)**	
HY060	**Social assistance**	**H71A, H71B**		**0 (101)**	
HY070	**Housing allowance**	**H43**	**Allowance for housing (tenants)**	**12.1 (33)**	**05: Median**
HY070	**Housing allowance**	**H26**	**Intervention of authorities for repayments on mortgage**	**33.3 (27)**	**05: Median**
				7.4 (27)	**09: correction based on legal amounts**
HY070	**Housing allowance**			**24.6 (61)**	
HY080	**Regular inter-household cash transfer received**	**H86**	**Alimony and child support received**	**1 (309)**	**05: Median**
				0.6 (309)	**06: observation in 2005 carried forward**
HY080	**Regular inter-household cash transfer received**	**H88**	**Regular cash support**	**9.1 (165)**	**01: Hot deck**
				0.6 (165)	**06: observation in 2005 carried forward**
HY080	**Regular inter-household cash transfer received**			**4.8 (437)**	
HY090	**Interests, dividends, etc.**	**H99, H100**		**11.5 (3694)**	**02: Regression (auxiliary variables: sort assets (bank accounts, bonds,…), tenure status, subjective rent) + random term**
				51.7 (3694)	**02: Ranges of values: regression with bounds**
HY110	**Income received by people aged**	**H69**		**7.1 (14)**	**05: Median**

	< 16				
HY130	**Regular inter-household cash transfer paid**	**H79**	**Alimony and child support paid**	**0.4 (258)**	**05: Median**
HY130	**Regular inter-household cash transfer paid**	**H81**	**Regular cash support**	**3.1 (257)** **0.4 (257)**	**01: Hot deck** **09: deductive imputation based on answer in 2005**
HY130	**Regular inter-household cash transfer paid**			**2.1 (481)**	
HY140	**Tax on income and social contributions**	**I130**	**Repayments for tax adjustment**	**5.9 (2211)**	**08: other source was used: fiscal data**
HY140	**Tax on income and social contributions**	**I132**	**Receipts for tax adjustment**	**3.9 (3096)**	**08: other source was used: fiscal data**
HY140	**Tax on income and social contributions**			**37 (5192)**	**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**	**0.04 (5153)** **14.2 (5153)** **0.08 (5153)** **0.08 (5153)** **0.14 (5153)**	**1) Corrections** **2) Net income is given, imputation based on regression** **3) current income is given, imputation based on regression** **4) Imputation on basis of EU-SILC 2005** **6) other**
PY010	**Employee cash income**	**I52**	**Number of months I47-**	**0.1 (4851)**	**1) correction**

			I48		
PY010	Employee cash income	(i60_a_ne)	Pay for overtime	0.6 (162)	1) correction
PY010	Employee cash income	(i60_b_ne)	Commissions	0.0 (46)	No imputation
PY010	Employee cash income	(i60_c_ne)	Tips	0.0 (28)	No imputation
PY010	Employee cash income	(i60_d_ne)	Additional payments based on productivity	0.0 (91)	No imputation
PY010	Employee cash income	(i60_e_ne)	End of the year payments	0.03 (3163)	1) correction
PY010	Employee cash income	(i60_f_ne)	Thirteenth month payment	0.0 (612)	No imputation
PY010	Employee cash income	(i60_g_ne)	Fourteenth month payment	0.0 (48)	No imputation
PY010	Employee cash income	(i60_h_ne)	Holiday payments	0.2 (4007)	1) correction
PY010	Employee cash income	(i60_i_ne)	Profit sharing	0.0 (106)	No imputation
PY010	Employee cash income	(i60_j_ne)	Shares	0.0 (17)	No imputation
PY010	Employee cash income	(i60_k_ne)	Allowances paid for working in remote locations	2.5 (39)	1) correction
PY010	Employee cash income	(i60_l_ne)	Other additional payments	0.0 (152)	No imputation
PY010	Employee cash income	I53	Income from irregular jobs : wages and salaries	2.35 (213) 0.9 (213)	1) corrections 2) Imputation fixed amount
PY010	Employee cash income	I93	Income from jobs other than main job : wages and salaries	4.82 (83) 2.41 (83)	1) Household income is source 2) Gross/net
PY010	Employee cash income	I92	Number of months	0.0 (83)	No imputation

			Income from jobs other than main job : wages and salaries		
PY010G	Employee cash income			22.8 (5327)	
PY010N	Employee cash income			10.4 (5327)	
PY050	cash benefits or losses from self-employment	I93	Income for jobs other than main job : self-employed	15.56 (45) 4.44 (45)	1) Household income is source 2) Gross/net
PY050G	cash benefits or losses from self-employment			42.1 (758)	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.
PY050N	cash benefits or losses from self-employment			29.7 (758)	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)	18 (17) 35 (17)	01: Hot deck 09: One-shot amount converted into annuity
PY080	Pension from Individual private plans	I112	Life insurance (Assurance-vie)	81 (11)	09: One-shot amount converted into annuity
PY080	Pension from Individual private plans			18 (27)	
PY090	Unemployment benefits	I98_a	Subsistence income for persons entering the labour market	0 (14)	

PY090	**Unemployment benefits**	**(i98_b)**	**Full unemployment benefits**	**4.5 (988)**	**09: Number of months modified or imputed based on the calendar**
				28.6 (988)	**04: Net income is given, imputation based on regression**
				1.2 (988)	**06: imputation based on previous 2005**
				0.8 (988)	**09: deductive imputation based on current income or on total income of the household**
				0.3 (988)	**05: Median of the given amounts (in classes based on type of households)**
				1.5 (988)	**08: Imputation of legal amounts**
PY090	**Unemployment benefits**	**I98_c**	**Partial unemployment benefits**	**25.7 (81)**	**04: Net income is given, imputation based on regression**
PY090	**Unemployment benefits**	**I98_d**	**Other financial assistance (Allocation de garantie de revenus)**	**29.4 (17)**	**04: Net income is given, imputation based on regression**
PY090	**Unemployment benefits**	**(I98_e)**	**Other financial assistance (Allocation du fonds de sécurité d'existence)**	**22.7 (22)**	**04: Net income is given, imputation based on regression**
				13.6 (22)	**05: Median**
				22.7 (22)	**09: Number of months modified or imputed based on the calendar**
PY090	**Unemployment benefits**	**(I98_f)**	**Vocational training allowance**	**4.7 (21)**	**05: Median**
PY090	**Unemployment benefits**	**(I98_h)**	**Other cash benefits**	**16 (25)**	**05: Median**
				4 (25)	**09: Number of months modified or imputed based on**

					the calendar
PY090	**Unemployment benefits**	**I99_b**	**Early retirement benefits**	**3.7 (269**	**09: Number of months modified or imputed based on the calendar**
				25.3 (269) (234)	**net is given (regresion net-gross)**
				1.8 (269)	**04: last month value is given and used in regression**
				1.1 (269)	**06: observation in 2005 carried forward**
PY090	**Unemployment benefits**			**31.6 (1415)**	
PY100	**Old age benefits**	**I104**	**Pension Fund (Fonds de pension)**	**1.5 (64)**	**Conversion lump sum to annuity**
				12.5 (64)	**01: Hot deck**
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**	**3.6 (1813)**	**09: Number of months modified or imputed based on the calendar**
				27.3 (1813)	**04: Net pension is source**
				2 (1813)	**06: observation in 2005 carried forward**
				1 (1813)	**04: last month value is given and used in regression**
				0.3 (1813)	**09: deductive imputation based on total income given by the respondent**
				0.7 (1813)	**01: Hot deck**
PY100	**Old age benefits**	**(I_102_C)**	**Other financial assistance to old aged people[6]**	**14.6 (41)**	**04: Net pension is source**
				2.4 (41)	**06: 2005 observation carried forward**
				7.3 (41)	**09: Number of months modified or imputed based on the calendar**

[6] Revenus garantis aux personnes âgées

PY100	**Old age benefits**	**(I_102_D)**	**Other financial assistance to old aged people[7]**	**17.5 (40)**	**04: Net pension is source**
				2.5 (40)	**06: 2005 observation carried forward**
				2.5 (40)	**09: Number of months modified or imputed based on the calendar**
PY100	**Old age benefits**	**(I_102_E)**	**Type of old age benefits not given**	**26.9 (78)**	**04: Net pension is source**
				3.8 (78)	**06: 2005 observation carried forward**
				2.5 (78)	**04: current pension is source**
				1.3 (78)	**01: hot deck**
				1.3 (78)	**08: Imputation of legal amounts**
				7.7 (78)	**09: Number of months modified or imputed based on the calendar**
PY100	**Old age benefits**			**33.3 (2056)**	
PY110	**Survivor's benefits[8]**	**(I102_A)**		**28.5 (382)**	**04: Net pension is source**
				0.5 (382)	**04: current pension is source**
				1.6 (382)	**06: 2005 observation carried forward**
				0.5 (382)	**01: hot deck**
				0.8 (382)	**09: deductive imputation based on total income given by the respondent (h66)**
				3.1 (382)	**09: Number of months modified or imputed based on the calendar**
PY120	**Sickness benefits**	**(I115_c)**	**Paid sick leave (temporary inability to work due to sickness)**	**16.8 (143)**	**04: Net income is given, imputation based on regression**
				1.4 (143)	**06: 2005 observation carried**

[7] Complément au revenu garanti aux personnes âgées

[8] 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.

				1.4 (143) 0.7 (143) 1.4 (143) 13.3 (143)	forward 09: corrections based on the calendar and the total income given by the respondent (h66) 01: hot deck in classes 08: Imputation of legal amounts 09: Number of months modified or imputed based on the calendar
PY120	Sickness benefits	(I115_d)	Paid sick leave (temporary inability to work due to professional sickness or injury)	8.9 (45) 2.2 (45)	09: Number of months modified or imputed based on the calendar 04: current pension is source
PY120	Sickness benefits	(I115_e)	Other sickness benefits	3 (66)	09: Number of months modified or imputed based on the calendar
PY120	Sickness benefits			20.8 (192)	
PY130	Disability benefits	I115_a	Disability pension	24.1 (344) 3.5 (344) 0.3 (344) 0.3 (344) 0.3 (344) 5.2 (344)	04: Net income is given, imputation based on regression 06: 2005 observation carried forward 09: corrections based on the calendar and the total income given by the respondent (h66) 08: Imputation of legal amounts 04: current pension is source 09: Number of months modified or imputed based on the calendar
PY130	Disability benefits	(I115_b)	Integration income for the handicapped	5.2 (96) 5.2 (96)	06: 2005 observation carried forward 09: Number of months modified or imputed based on the calendar

PY130	**Disability benefits**			**9.3 (398)**	
PY140	**Education-related allowances**	**H95**	**Grants, scholarship and other educational help to pupils (of secondary schools)**	**3.5 (85)**	**01: Hot deck**
PY140	**Education-related allowances**	**H97**	**Grants, scholarship and other educational help to students (of colleges)**	**2.7 (374)**	**01: Hot deck**
PY140	**Education-related allowances**			**3.2 (219)**	**Note that in the P-file all grants received by someone in the household are given to the reference person of the household as they can concern persons aged under 16 who are not present in the P-file.**

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.

All the gross-net imputation for PY100 and PY110 was done following the Belgian taxing rules. We first (1) had to determine the status of the person (isolated or married, with or without dependant children, …), then (2) we applied all the taxing rules including reductions of taxes for e.g. dependant child. (3) Once this model has been applied to gross-net transformation, we could use it for the net to gross (very more useful in fact). To do that, we applied the model on each possible amount as fictive gross amount. As result, we got each possible net amount. We then only had to do the correspondence between net and gross amount.

- **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).

2.5.3 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. When available, we preferably used the longitudinal information's from 2005 for imputation. For the other cases the chosen method for imputation was imputation of a sub-category median based on age and sex. 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

From 2007 on.

2.7 Collection variable company Car

Since 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.

Since 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 and 2006 PY020G only contains the value of company cars.

Family/children related allowances

For the SILC 2004 Belgium asked allowances received from the federal government. From 2005 on 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 previous wave.

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

No change from the previouswave.

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

INTRODUCTION.

Although there is in our opinion an overall acceptable degree of coherence between the results of the Belgian EU-SILC 2005 and EU-SILC 2006 data there are admittedly also a number of marked differences which need explanation. The most eye catching differences are observed for the population under the age of 16 (a decrease in the population at risk of poverty of 2,9%) and the old (65+ - an increase of 2,4%). We have assessed, as far as the very tight time schedule permitted it, several possible explanations for these differences. Below we report on our findings.

Four explanations can be put forward to account for the observed differences between the two waves: (1) systematic error, (2) sampling error, (3) selection bias and (4) true effects.
We will not take the first two into consideration here.
There was a systematic problem with the child allowances in the 2004 and 2005 data. This problem has been fixed for the 2005 data and will be fixed in the very near future for the 2004 data. The experience with the 2005 data learned that the impact of this problem was limited anyway.
We certainly do not want to wipe sampling error as explanation of the table. In some subpopulations the sampling errors are definitely not small and some of the differences which seem big at first are in fact within the boundaries of the confidence intervals. It is therefore undoubtedly a contributing factor to the differences we observe and it is important to keep in mind that everything we present below is subject to sampling error.

Below we will however focus on to what extent selection and/or true effects explain the observed differences. The 'and/or' is necessary because both may co-exist (additively) or even interact (we elaborate on this below). This makes distinguishing between them and assessing their separate (net)impact complex.

To address this problem we have adopted a strategy which is mainly based on comparing the results of several subpopulations in the panel for each wave (year) and between the waves (years). The subpopulations are identified on their status in the panel: new to wave Y, present in wave Y-1, present in wave Y+1 and so on. We also distinguish between (what we call) the input-side of a wave and the output-side. In the former the identification is made on basis of the status in the previous wave. Cases are new to the survey (replacing cases lost due to the rotation or due to non-respons) or cases were present in wave Y-1. On the output side households are either present in wave Y+1 or they are lost due to the rotation or due to non-respons (whatever the reason: refusal, death, and …). This way of thinking results in the scheme below in which each household or better each single observation of a household is considered twice within a given year, once at the input side and once at the output side.

OBSERVATIONS EU-SILC 2004		OBSERVATIONS EU-SILC 2005		OBSERVATIONS EU-SILC 2006	
IN	OUT	IN	OUT	IN	OUT

NEW		NEW		NEW	
-	PANEL PRESENT 2005	PANEL PRESENT 2004	PANEL PRESENT 2006	PANEL PRESENT 2005	PANEL PRESENT 2007
	LOST		LOST		LOST

- NEW: New households/individuals to replace households/individuals lost due to rotation and other reasons.
- PANEL: Households/individuals participating longitudinally
- LOST: Households/individuals missing due to rotation and other reasons (refusal, death, …).

The values in this scheme can relatively easy be calculated for several variables and characteristics and within different breakdowns as we will do below.

RESULTS.

- **Overview**

In the table below – the results for the median equivalised income are shown for the subpopulation as outlined above and by age group.

		VALUES 2004		VALUES 2005		VALUES 2006	
		IN	OUT	IN	OUT	IN	OUT
MEDIAN EQ_INC							
ALL	OVERALL	15.540		16.570		17.318	
	NEW	15.540	-	16.100	-	16.984	-
	PANEL	-	15.740	16.974	16.876	17.432	
	LOST	-	15.378	-	15.983	-	
15 - YRS.	OVERALL	15.048		16.146		17.829	
	NEW	15.048	-	16.282	-	17.292	-
	PANEL	-	15.363	16.096	16.927	18.178	
	LOST	-	14.549	-	14.611	-	
16 - 64 YRS.	OVERALL	16.733		17.651		18.493	
	NEW	16.733	-	17.051	-	18.093	-
	PANEL	-	16.915	18.099	17.800	18.589	
	LOST	-	16.420	-	17.452	-	
65 + YRS.	OVERALL	12.448		12.667		12.933	
	NEW	12.448	-	12.707	-	12.699	-
	PANEL	-	12.417	12.650	12.891	12.979	
	LOST	-	12.468	-	12.323	-	

We come to the following overall conclusions:

1. Between the waves (2004-2005-2006) there is a steep increase in the income of the population participating longitudinally (PANEL)
2. Idem ditto there is a steep raise in income if the new households in each wave are compared (NEW)
3. There is an important difference in the income of the households lost between Y and Y+1 (LOST) and the new households in wave Y+1 (NEW)
4. These conclusions seem only to apply to the young and middle aged. They are absent or less marked in the older population.

- **These conclusions quantified and in more detail**

1. increase between 2004-2005-2006 = True effect

PANEL						
	VALUES 2004 - OUT	VALUES 2005 - IN	% 2005-2004	VALUES 2005 - OUT	VALUES 2006 - IN	% 2006-2005
ALL	15.740	16.974	107,8	16.876	17.432	103,3
15 - YRS.	15.363	16.096	104,8	16.927	18.178	107,4
16 – 64 YRS.	16.915	18.099	107,0	17.800	18.589	104,4
65 + YRS.	12.417	12.650	101,9	12.891	12.979	100,7

There is a substantial raise in income for the population that is followed between 2004 and 2005 on the one side and 2005 and 2006 on the other side. Overall this increase is 7,8%[9] for the transition 2004-2005 and 3,3% for the transition 2005-2006. This increase is however especially marked among children and adults (< 65 yrs.) – up to 7,4% between 2005 and 2006 for children for example – and almost non existing for the older age group (65+), with only an increase of 0.7% between 2005 and 2006.

[9] This percentage is biased upwardly. As we already explained the observations for child allowances were systematic biased in both 2004 and 2005 (not 2006). This has for the time being only been corrected for 2005. The same correction will in 2004 will result in a higher median equivalised income for families with children and (indirectly) in a higher median equivalised income overall.

We see a similar pattern – even slightly more marked – for the new selected households in each wave:

NEW						
	VALUES 2004 - IN	VALUES 2005 - IN	% 2005-2004	VALUES 2005 - IN	VALUES 2006 - IN	% 2006-2005
ALL	15.540	16.100	103,6	16.100	16.984	105,5
15 - YRS.	15.048	16.282	108,2	16.282	17.292	106,2
16 – 64 YRS.	16.733	17.051	101,9	17.051	18.093	106,1
65 + YRS.	12.448	12.707	102,1	12.707	12.699	99,9

This implies that the increase in income experienced by the longitudinal group (PANEL) is not different to the increase in income experienced by a random selection out of the general population (NEW). This indicates two things:

- the increase is a real effect[10]. Incomes are on the raise – especially for the households with young children.
- although the population that continues to participate might well be selected, there is no proof that the evolution in their income (a stabilization for the old and a firm increase for the rest) is biased.

2. Selection into the group that continues to participate

It's clear that the population that goes lost – or continues to participate on the other hand – is to some extent selected:

MEDIAN EQ_INC			
2004			
	VALUES 2004 - OUT		
	PANEL	LOST	% LOST/PANEL
ALL	15.740	15.378	97,7
15 - YRS.	15.363	14.549	94,7
16 - 64 YRS.	16.915	16.420	97,1
65 + YRS.	12.417	12.468	100,4

MEDIAN EQ_INC			
2005			
	VALUES 2005 - OUT		
	PANEL	LOST	% LOST/PANEL
ALL	16.876	15.983	94,7
15 - YRS.	16.927	14.611	86,3
16 - 64 YRS.	17.800	17.452	98,0
65 + YRS.	12.891	12.323	95,6

[10] There actually are other indications for this. We measured for example the increase in salary depending of whether or not a document was used to provide the information to the interviewer (tax declaration, pay check …). In both groups (with and without document) a substantial raise in the mean and median salary was observed. As the reliability of the information from the group with document can be assumed to be quite high it seems reasonable to assert that increase is for real. We were not able yet to incorporate this (and other) information in this paper, however.

Both in 2004 (transition 2004 – 2005) and 2005 (transition 2005 – 2006) the households continuing to participate in Y+1 are definite richer than those who are lost – up to 5,3% in 2005 (transition to 2006) for example. This is especially the case for the households with young children: based on the observations for 2005 the difference between those participating in 2006 and those lost between 2005-2006 is for the young age category 13,7%.

This selection effect is however only troublesome to the extent that it is not corrected by the rotational principle in EU-SILC and the entry of new households to compensate for the households that go lost. To check that is, however, less straightforward. One way to get an idea of this is to compare the observations of the households that go lost in Y with the observations in Y+1 of the new households :

MEDIAN EQ_INC			
2004- 2005			
	LOST 2004	NEW 2005	% LOST/NEW Y+1
ALL	15.378	16.100	95,5
15 - YRS.	14.549	16.282	89,4
16 - 64 YRS.	16.420	17.051	96,3
65 + YRS.	12.468	12.707	98,1

MEDIAN EQ_INC			
2005 - 2006			
	LOST 2005	NEW 2006	% LOST/NEW Y+1
ALL	15.983	16.984	94,1
15 - YRS.	14.611	17.292	84,5
16 - 64 YRS.	17.452	18.093	96,5
65 + YRS.	12.323	12.699	97,0

In both comparisons (2004-2005 and 2005-2006) the new households in Y+1 have a higher income than the households lost in Y – up to 13,5% for the children in 2005-2006.

This comparison is however upwardly biased because it can be expected that also the households who are lost between Y and Y+1 will have experienced an increase in income. It is however (intra-SILC) impossible to assess to what extent this was the case as we have obviously no observation for Y +1 for these cases. The other way around we have of course no insight in the income for Y for the new households entering in Y+1. One way around this is simply comparing the households lost in Y with the new households in Y:

MEDIAN EQ_INC			
2004- 2005			
	LOST 2004	NEW 2004	% LOST/NEW Y
ALL	15.378	15.540	101,1
15 - YRS.	14.549	15.048	103,4
16 - 64 YRS.	16.420	16.733	101,9
65 + YRS.	12.468	12.448	99,8

MEDIAN EQ_INC			
2005 - 2006			
	LOST 2005	NEW 2005	% LOST/NEW Y
ALL	15.983	16.100	99,3
15 - YRS.	14.611	16.282	89,7
16 - 64 YRS.	17.452	17.051	102,4
65 + YRS.	12.323	12.707	97,0

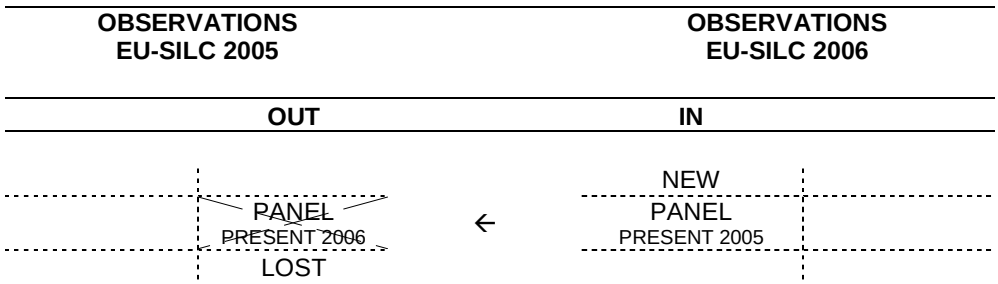
The differences are smaller now but still persist especially for the young.

- **Assessing the impact of these effects.**

To asses the impact of the above effects we have created hypothetical populations by combining the observations – over the waves – of several subpopulations. We have done two trials with this principle:

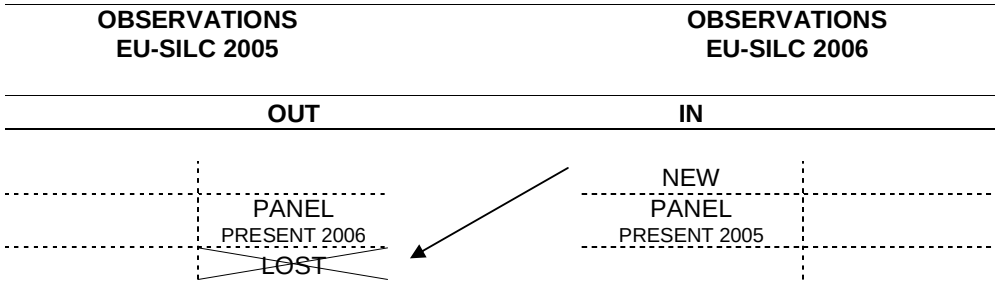
a. In the 2005 data we replaced the observations for the population participating in both 2005 and 2006 with their observed values for 2006. That gives an insight in the impact of the increase in income between 2005 and 2006 already stemming only from the part of the population that continues to participate.

Scenario A:



b. We replaced all observations for the entire population that was lost between 2005 and 2006 by the observation for the new households in 2006. This should allow to some degree an assessment of the impact the selection effect.

Scenario B:



1. results for scenario A:

SCENARIO A: LONGITUDINAL 2005 REPLACED BY VALUES 2006

ALL

	2005	2005 - SCENARIO 1	2006	% 2006/2005	% scen. A / 2005	% 2006/ scen. A
MEDIAN EQ_INC	16.570	16.932	17.318	105	102,2	102,3
% BELOW ARPT	14,8	14,8	14,9	-	-	-

YOUNG

	2005	2005 - SCENARIO 1	2006	% 2006/2005	% scen. A / 2005	% 2006/ scen. A
MEDIAN EQ_INC	16.146	16.973	17.829	110	105,1	105,0
% BELOW ARPT	17,9	16,4	15,0	-	-	-

ADULTS

	2005	2005 - SCENARIO 1	2006	% 2006/2005	% scen. A / 2005	% 2006/ scen. A
MEDIAN EQ_INC	17.651	18.209	18.493	105	103,2	101,6
% BELOW ARPT	12,3	12,3	12,6	-	-	-

SCENARIO A – continued

OLD

	2005	2005 - SCENARIO 1	2006	% 2006/2005	% scen. A / 2005	% 2006/ scen. A
MEDIAN EQ_INC	12.667	12.675	12.933	102	100,1	102,0
% BELOW ARPT	21,4	22,9	23,8	-	-	-

The major conclusions in the above table are the following:

- replacing the observations for 2005 with the observations 2006 for the households participating in both years does not change the overall poverty rate
- For the young however, the poverty rate decrease with 1,5%
- For the old, the rate increases with 1,5%.

The decrease by 1,5% of in the poverty risk for the young implies that one half the observed difference between 2005 and 2006 for that category is explained by the raise in their income. The increase by 1,5% of the risk for the old explains more than the half of the observed difference between 2005 and 2006.

2. results for scenario B:

SCENARIO B: LOST 2005 REPLACED BY NEW 2006

ALL

	2005	2005 - SCENARIO 1	2006	% 2006/2005	% scen. A / 2005	% 2006/ scen. A
MEDIAN EQ_INC	16.570	16.916	17.318	105	102,1	102,4
% BELOW ARPT	14,8	14,8	14,9	-	-	-

YOUNG

	2005	2005 - SCENARIO 1	2006	% 2006/2005	% scen. A / 2005	% 2006/ scen. A
MEDIAN EQ_INC	16.146	17.000	17.829	110	105,3	104,9
% BELOW ARPT	17,9	16,1	15,0	-	-	-

ADULTS

	2005	2005 - SCENARIO 1	2006	% 2006/2005	% scen. A / 2005	% 2006/ scen. A
MEDIAN EQ_INC	17.651	17.940	18.493	105	101,6	103,1
% BELOW ARPT	12,3	12,7	12,6	-	-	-

OLD

	2005	2005 - SCENARIO 1	2006	% 2006/2005	% scen. A / 2005	% 2006/ scen. A
MEDIAN EQ_INC	12.667	12.832	12.933	102	101,3	100,8
% BELOW ARPT	21,4	21,9	23,8	-	-	-

From this table the major conclusions are:

- Once again the overall poverty rate does not change.
- For the young however a decrease in the rate by 1,9% is observed
- For the old, the rate increases with 0,5%.

The above table gives an assessment of the impact of the selection effect but will likely overestimate that impact. The reason for that is that probably also the income of the households leaving the survey in 2005 will have risen between 2005 and 2006, as this is the general trend.
For the old the above result is remarkable. It implies that the increase in their poverty rate is too much larger extent caused by the increase in the income of the others than by the selection effect.

CONCLUSIONS.

We have indications for 5 major conclusions:

1. **There is a relatively strong increase in the incomes, especially for families with young children (< 16 yrs.). We observe this increase not only for household in the panel but (although not directly comparable) also for new households entering a wave and even households leaving the panel.**
2. **There is a selection effect. Somewhat poorer households seem to leave the panel more quickly than richer households. The loss of these poorer households is not compensated by the entry of new households.**
3. **Conclusion 1 and 2 do not concern the old population (65+) however. For the old there is neither evolution nor selection or at least not at a significant level.**
4. **The impact of the increase of income explains about half of the decrease of the poverty rate of the young and the biggest part of the increase in poverty among the old.**
5. **The selection effect explains probably the rest of the difference for the young. It explains only a smaller part of the difference in risk between 2005 and 2006 for the old.**

Index