CENTRUM VOOR SOCIAAL BELEID
HERMAN DELEECK

Measuring change with the Belgian Survey on Income and Living Conditions (SILC): taking account of the sampling variance

Report Financed by FPS Social Security Belgium

(Ref.: 2013/BOEP/UA-EU2020)

December 2013

Tim Goedemé

Herman Deleeck Centre for Social Policy – University of Antwerp

St. jacobstraat 2

2000 Antwerp

Belgium

tim.goedeme@uantwerpen.be

+32 3 265 55 55

Universiteit Antwerpen

Abstract

Early 2013 Eurostat published a paper which contains some estimations for the standard error of recent changes in the Europe 2020 poverty reduction indicator (Osier et al., 2013). For Belgium, the published standard errors were contested by national experts. At the same time, Osier et al. (2013) warned for limitations to the accuracy and consistency of the sample design variables, which could be the cause of some counterintuitive results. In this report, I review the Belgian SILC sample design and describe the available sample design variables. In addition, I document the importance of a consistent coding of the sample design variables on the basis of a dataset with consistent sample design variables prepared by Statistics Belgium. In order to provide some idea of the size of the standard error of changes over time, this dataset is used to report a number of estimations of the standard error of recent changes in the at-risk-of-poverty indicator for several vulnerable groups, both at the national and regional level. It is concluded that the power of the Belgian SILC data is not sufficient for closely monitoring the situation of relatively small vulnerable groups, especially at the regional level. Apart from investing more resources in the Belgian SILC data, it is recommended that consistent sample design variables are included in all available SILC datasets and that further research is carried out about the most appropriate estimator of the sampling variance for the Belgian SILC data. In the meantime, the use of the ultimate cluster method for estimating the sampling variance in European research and by Eurostat, appears to be a pragmatic but valid solution, provided the sample design variables are coded adequately.

Table of contents

1 Introduction[1]

In June 2010, the European Council agreed to lift 20 million people out of poverty and social exclusion by 2020 (European Council, 2010). More precisely, the Europe 2020 poverty reduction target indicator "at-risk-of-poverty or social exclusion" (AROPE) is defined as the population covered by the union of three indicators: the percentage of the population confronted with at-risk-of-poverty, several material deprivation and/or very low work intensity. Progress is monitored on the basis of the EU Statistics on Income and Living Conditions (EU-SILC). As EU-SILC is based on a sample survey, an evaluation of progress towards reaching the target should (among other things) take account of the sampling variance. Of course, the sampling variance is not the only, and may even be not the most important source of uncertainty about a sample survey estimate. Nonetheless, in contrast to other (random and non-random) errors, the sampling variance can usually be estimated with a relatively high degree of precision, while other errors are much harder to estimate. So, while descriptions of the limitations of the data may be useful, they may not be a reason for not estimating the sampling variance. Furthermore, I am strongly convinced that the publication of confidence intervals make readers better aware of uncertainties about sample survey estimates in general, random and non-random alike, while a lack of information on the confidence interval, is usually not compensated by warnings in the text of sampling errors when interpreting point estimates[2].

Recently, Eurostat has published a report on standard errors of the Europe 2020 poverty reduction indicators (Osier et al., 2013). In the latter report, the authors present estimates of the standard error of cross-sectional poverty figures, but also of changes over time. The estimation of the standard error of changes over time is confronted with several complications: (1) EU-SILC is a rotational panel, which means that every year, one quarter of the sample is replaced by a new sample; accurate standard error estimation for rotational panels is not included in standard estimation commands; (2) EU-SILC has a complex sample design involving among others unequal probability weighting, several stages of selection and stratification. This needs to be taken into account when estimating the sampling variance; (3) EU-SILC cross-sectional databases are released with randomised identifiers of individuals, households and primary sampling units (PSUs), which means that the databases cannot be linked to one another. Nonetheless, consistent sample design variables are required such that at least PSUs and strata can be identified consistently in all cross-sectional databases that are compared. The fact that PSUs are probably not coded consistently in the various databases may result in strongly biased standard error estimates of changes over time. At the same time, however, given the political salience of the Europe 2020 poverty reduction target, a sufficiently accurate estimation of the sampling variance of changes over time is key to correctly interpreting the observed changes in the sample point estimates.

For some countries (and Belgium in particular), the authors of the report seriously doubt the consistency of the coding of the sample design variables. The standard errors of recent changes in

[1] I am grateful to the EU-SILC team at Statistics Belgium for preparing the data and providing more information on the Belgian SILC data. In addition, I am grateful to Karel Van den Bosch for suggestions and to Guillaume Osier and Yves Berger for our fruitful collaboration on previous papers that have formed the starting point of this report. Of course, I take full responsibility for the views expressed in this report, as well as for any remaining errors and shortcomings.

[2] For a complete overview of different types of errors in surveys, see for instance Groves et al. (2009) and for an application to EU-SILC, see Verma et al. (2010).

AROPE in Belgium that are included in the report are presented in Table 1. In principle, due to the panel character of EU-SILC, one would expect that the standard error of the change in the percentage at-risk-of-poverty or social exclusion, would increase the more distant the years of comparison are from one another. However, we observe exactly the opposite. In addition, the estimated standard error of the change in AROPE for the difference between EU-SILC 2007 and EU-SILC 2011 is very small, implying a very strong covariance between both years, even though at the household level, the sample has been completely renewed within the same time span.

Table 1: The standard error of the difference between EU-SILC 2011 and previous waves for the AROPE indicator, Belgium, percentage points – Estimations Net-SILC2 / Eurostat

	AROPE(x)	AROPE(2011)	difference	standard error difference
2011-2007	21.6	21.0	-0.59	0.12
2011-2008	20.8	21.0	0.18	0.63
2011-2009	20.2	21.0	0.80	0.65
2011-2010	20.8	21.0	0.14	0.70

Source: Eurostat -SILC 2007-2011, calculations by Net-SILC2 & Eurostat (Osier et al., 2013).

Therefore, the question was raised which value should be attached to the standard error estimates of net changes published in Osier et al. (2013), what the order of magnitude is of the correct standard errors and how the estimation of the sampling variance for the Belgian SILC could be improved in the future, so as to avoid the publication of standard errors that are very likely to strongly deviate from the 'real' standard errors.

In order to shed some light on these issues, in collaboration with the EU-SILC team of Statistics Belgium a dataset has been analysed with SILC data of the 2008, 2009, 2010 and 2011 waves. For this dataset (which contains much less variables than the complete EU-SILC User database (UDB)), we can be sure that the sample design variables are coded consistently. In this paper, we report on the results of this analysis.

The report is structured as follows. Section two describes the principal determinants of the sampling variance that should be taken into account when estimating the standard error of changes over time. Particular attention is paid to the Belgian SILC sample design and the quality of the sample design variables. Section three confronts estimates of the standard error of net changes in the at-risk-of-poverty indicator on the basis of the EU-SILC UDB with the standard error of net changes on the basis of database made available by Statistics Belgium for which we can be sure that it contains consistently coded sample design variables. In addition, this section includes a large number of estimates of the change in the at-risk-of-poverty indicator in Belgium and its three regions, while paying special attention to the at-risk-of-poverty rate of some vulnerable groups. The aim is not to publish estimates of the standard error which are as precise as possible, but rather to demonstrate the importance of consistent sample design variables as well as to give a first idea of the magnitude of the standard error of changes over time. Section 4 summarizes the main findings and suggestions and concludes.

2 The Belgian SILC sample design & the sample design variables

The sample design, weighting scheme, the imputation of missing values and the characteristics of the statistic of interest are crucial determinants of the sampling variance (cf. Eurostat, 2002; Heeringa et al., 2010). Neglecting one or more of these factors, or making wrong assumptions about them, can seriously bias sampling variance estimates (for an illustration on the basis of EU-SILC, see Goedemé, 2013b). Furthermore, it is not always clear whether variance estimates are upwardly or downwardly biased, as both stratification and calibration usually reduce the sampling variance, whereas all the other factors and elements of the sample design have a tendency to increase the sampling variance.

Unfortunately, in the case of EU-SILC in general and changes in the poverty reduction target indicator in particular, users of the EU-SILC user database (UDB) cannot take any of those four determinants properly into account: accurate and consistent sample design variables are for most countries lacking (cf. Goedemé, 2013a); and insufficient information is available in the UDB that would enable researchers to take account of calibration and imputation. Finally, (and probably less importantly), the computation of the standard error of the AROPE indicator is complicated by the fact that the at-risk-of-poverty indicator is estimated with a poverty line as a percentage of median incomes, which in itself is a survey estimate. As is done in the report of Osier et al. (2013), we will neglect this final issue, and assume for variance estimation purposes that the AROPE indicator can be estimated as a simple proportion.

Nonetheless, the quality of the sample design variables has been improved during the past few years, which means that, also for Belgium, to an important extent the sample design can be taken into account when estimating the sampling variance. In the remainder of this section I focus on the sample design of the Belgian SILC data and the quality of the sample design variables, while leaving the issue of calibration and imputation aside.

2.1 The sample design of the Belgian EU-SILC

Most elements (but not all) of the Belgian SILC sample design are discussed in the various quality reports. It is worth noting that the sample design of the 2003 pilot study is somewhat different from the 2004 SILC design onwards. Here, we discuss only the sample design that is currently in use (since SILC 2004).

The Belgian SILC consists of a two-stage stratified, systematic sample. The PSUs are sampled systematically 'with replacement' with a probability proportional to their size (measured by the number of private households) and remain fixed for the entire duration of EU-SILC. In practice this means that the PSUs are ordered on the sampling frame on the basis of their mean income and that the interval used for systematically selecting PSUs is smaller than the size of a number of PSUs. As a result, the probability of selection is for some PSUs larger than 1. The selection of PSUs is stratified by province while Brussels-Capital Region is a separate stratum (resulting in 11 strata in total). At the second stage, within the selected PSUs, households are selected systematically on a sampling frame ordered by the age of the reference person within the household.

Table 2 provides a more detailed overview of the number selected PSUs and the number of multiple hits by stratum (that is, the number of times the same PSU has been selected into the sample). The

table shows that multiple hits are very common in Brussels-Capital Region and nearly non-existent in the other strata. This suggests that the sampled fraction of PSUs could be relatively high, especially in Brussels-Capital Region. In principle, this should put a warning on the use of the so-called 'ultimate cluster approach' for the estimation of the sampling variance. With this approach, the sampling variance is estimated on the basis of the between-PSU variance in the sample (Kalton, 1979; Heeringa et al., 2010; Wolter, 2007; Osier, 2012). It considerably simplifies the computation of the sampling variance, as the within-cluster variance is ignored. For this reason, it is the preferred approach of Osier et al. (2013) for EU-SILC, which includes many countries that employ different sample designs. A crucial condition for evaluating the appropriateness of the ultimate cluster approach is the sampled fraction of PSUs. This must be very low (close to zero), such that the within-PSU variance contributes little to the total sampling variance.

Table 2: Number of times unique PSUs have been selected into the sample and total number of PSUs by stratum code

	Number of times selected						
stratum code	1	2	3	5	6	Total number of unique PSUs	Total number of PSUs
10	9	9	4	1	1	24	50
21	39	2				41	43
22	16					16	16
23	30	1				31	32
24	26					26	26
25	21					21	21
31	8					8	8
32	38					38	38
33	22		1			23	25
34	6					6	6
35	10					10	10
Total	225	12	5	1	1	244	275

Source: BE-SILC Statistics Belgium 2009.

Unfortunately, the sampled fraction of PSUs in the case of the Belgian SILC is not clearly documented. As a proxy, we can estimate a lower bound on the sampled fraction of PSUs assuming that PSUs are so-called 'letters' (geographical statistical sectors, usually parts of communes). Table 3 presents an overview of a lower bound on the sampled fraction of PSUs. The document on which it is based (the minutes of a preparatory meeting on the sample selection procedure for BE-SILC 2004), discusses various scenarios for selecting PSUs. The use of letters without modification as PSUs is ruled out, and a combination of letters is proposed for those letters with a relatively small population size. However, it is not clear which combination of letters has been adopted in the end. Therefore, the use of the number of the unmodified letters as the denominator for calculating the sampled fraction of PSUs is a safe and conservative lower bound of the sampled fraction. From the table it can be observed that the sampled fraction is relatively high (exceeding 0.05) in a large majority of strata and exceedingly so in Brussels-Capital Region and Antwerp. On this basis, the use of the ultimate cluster approach risks to result in a downwardly biased estimate of the sampling variance, especially in the case of domain estimates for the latter two strata. However, given that PSUs are ordered on

the basis of mean income, implicit stratification may strongly reduce the sampling variance of indicators based on variables that strongly correlate with average income at the PSU level.

Table 3: estimated lower bound of sampled fraction of PSUs

stratum code	nPSUs	N letters	Sampled fraction (lower bound)
10	50	27	1.85
21	43	160	0.27
22	16	202	0.08
23	32	284	0.11
24	26	227	0.11
25	21	242	0.09
31	8	115	0.07
32	38	431	0.09
33	25	347	0.07
34	6	232	0.03
35	10	341	0.03
Total	275	2608	0.11

Source: BE-SILC Statistics Belgium & (N., 2004).

It is crucial to understand that the PSUs remain fixed for the entire duration of SILC. This means that at the first stage of the sample design, an important covariance between various waves of EU-SILC will exist, even when all households have been replaced by other households. If an ultimate cluster approach may be applied (that is, the variance between PSUs is a good proxy of the total sampling variance), for variance estimation purposes it would suffice that the coding of PSUs across time is consistent, without the need to making household and person identifiers consistent between various cross-sections. At the same time, the fact that SILC is a rotational panel, would in that case not complicate the computation of the sampling variance, as rotation is implemented within PSUs, while the PSUs themselves remain the same.

At the second stage, within each PSU households are selected for a duration of four years. With each new wave of SILC, within each PSU one fourth of the households is replaced. During the first wave, 40 households were selected per PSU. Table 4 lists the number of PSUs and observations in the BE-SILC database made available by Statistics Belgium. As can be observed from the table, only in one year of observation (2009), all 275 selected PSUs are included in the sample. The PSU that is missing for the other years varies over time (twice one of Brussels and one in the Walloon region). In 2008 it concerns a PSU that has been selected multiple times, which means that the omission of the PSU does probably not introduce much bias. In the other two cases it concerns a PSU that has been selected only once, so some bias is more likely. The reason for the absence of the PSUs from the sample is that no interviewer could be found for carrying out the interviews in that specific area (information from EU-SILC team at Statistics Belgium).

Table 4: Number of PSUs and number of observations (persons) in EU-SILC 2008 - 2011

Stratum	number of PSUs				number of observations			
code	2008	2009	2010	2011	2008	2009	2010	2011
10	49	50	50	49	1840	1916	1915	2022
21	43	43	43	43	2084	1908	1827	1746
22	16	16	16	16	1218	1163	1182	1151
23	32	32	32	32	1769	1724	1817	1674
24	26	26	26	26	1443	1419	1339	1257
25	21	21	21	21	1761	1605	1686	1649
31	8	8	8	8	441	374	428	436
32	38	38	38	38	2205	2185	2140	2007
33	25	25	24	25	1367	1398	1366	1339
34	6	6	6	6	462	486	474	477
35	10	10	10	10	518	543	580	542
Total	274	275	274	274	15108	14721	14754	14300

Source: BE-SILC Statistics Belgium.

For variance estimation purposes, it is important to know whether the ultimate cluster approach may be applied as it would strongly limit data requirements and significantly simplify variance estimation procedures. It is well known that there are no unbiased estimators of the sampling variance in the case of systematic samples as the one designed for the Belgian SILC survey (cf. Wolter, 2007). Therefore, simulation studies on the basis of (simulated) population data are needed to find out which estimator is most appropriate. Dawagne and Milano (2011) carried out one such study on the basis of the Belgian tax declarations of 2007. They found that the ultimate cluster approach would be somewhat conservative. Apparently, implicit stratification counterbalances sufficiently the neglect of the within-PSU variance. However, the authors report also some counterintuitive results. For instance, assuming a sample design consisting of a simple random sample of households resulted in a higher estimate of the sampling variance and wider confidence intervals than in the case of the ultimate cluster method, the opposite of what is expected in survey sampling theory (e.g. Kish, 1965) and what is found in the study by Goedemé (2013b). Currently, Statistics Belgium uses the jack-knife repeated replication method for estimating the sampling variance, using the ultimate cluster approach with stratification on the basis of the order of selection of the PSUs. Unfortunately, this approach was not included in the simulation study by Dawagne and Milano (2011). It would be useful to carry out some more simulations in the near future that would replicate the previous study to see whether findings can be corroborated. In addition, it would be useful to include the current variance estimation approach used by Statistics Belgium in the simulation study[3] as well as to evaluate estimators that take the effects of non-response, calibration and imputation into account. Finally, a new simulation study should also evaluate the accuracy of the estimators in the case of net changes.

[3] Wolter (2007) concludes that for systematic samples with unequal probabilities of selection further research is needed regarding the appropriateness of various estimators of the variance.

2.2 The available sample design variables

Good sample design variables are needed in order to take account of the sample design (alternatives are discussed in Goedemé, 2013a). With the ultimate cluster approach, only sample design variables referring to the first stage of the sample design are of relevance. More in particular, two variables are required: one that identifies the primary sampling units (that is, a variable with a unique number for every group of households that have been selected as part of the same PSU[4]) and one that identifies the (explicit) strata (that is, a variable with a unique number for all households that have been selected in the same stratum). In addition, if the approach of Statistics Belgium is to be followed, a variable is needed with the order of selection of the PSUs. In EU-SILC, these variables are named DB060, DB050 and DB070 respectively. As the purpose of this report is to evaluate the estimates by Osier et al. (2013), I will follow their approach and only consider variables DB050 and DB060.

At this point it is necessary to make a distinction between the different versions of SILC databases that are developed. There is a difference between the EU-SILC data that are generally available for research (EU-SILC UDB), the data that are available to Eurostat (Eurostat-SILC) and the data that are available at Statistics Belgium (BE-SILC). Only BE-SILC contains consistently coded correct sample design variables. Eurostat-SILC contains correctly coded sample design variables, but the coding is not consistent for various releases of the data: PSU numbers in one wave of the data may refer to other PSUs in another wave. Finally, the EU-SILC UDB contains a correct, but non-consistent, PSU variable (DB060), but does not contain the stratum variable (DB050).

In spite of the lack of stratification variable DB050, users of the EU-SILC UDB can to some extent take account of stratification by making use of variable DB040, which identifies the three regions of Belgium. Given that the provinces and Brussels-Capital Region are perfectly clustered within the three regions, this should not introduce too much bias in the variance estimate. If some bias is present, it would most likely result in an over-estimation of the sampling variance. The use of DB040 as stratification variable is not entirely unproblematic. DB040 refers to the situation at the date of interview rather than the living address at the moment of selection. As a result, when using DB040 as a stratification variable the number of PSUs is inflated because some households have moved from one region to the other between the moment of selection and the moment of interview. In other words, some PSUs are 'split' over various strata and are treated as two or three different PSUs. The effect may be limited for the estimated sampling variance for population estimates, but may be more significant in the case of subpopulation estimates. Therefore, I have developed a procedure to assign households to the region at the moment of selection on the basis of the region with the largest number of households from that PSU, as documented in Goedemé (2013b).

The re-allocation of PSUs over the three regions is illustrated in Table 5. First of all, the table shows that if DB060 would be used to identify the PSUs and DB040 to identify the strata in the EU-SILC UDB, the number of PSUs in total and within each of the three strata would be substantially inflated (see column 'UDB original'). At the same time, the impact on the number of observations is relatively low. The situation is considerably improved when households are re-allocated to their 'original' stratum (see column 'UDB new'): The total number of PSUs corresponds to what we observe in the SILC data provided by Statistics Belgium (see column 'BE-SILC'). In addition, in the case of the 2009 and 2011

[4] Please note that in the case of 'multiple hits', every 'hit' should be counted as a separate PSU.

data, the number of PSUs and observations in every stratum is also correct. For the other two years there is a small deviation which cannot be corrected without exogenous information, as it concerns in 2008 the wrong allocation of a PSU with only one observed household and in 2010 it would concern a PSU that has been allocated to the wrong stratum in spite of the fact that for all PSUs there is a clear majority of households living in one region rather than the other[5]. It could also be the case that a correction has been made at Statistics Belgium in between the distribution of the latest version of the SILC UDB and the data made available by Statistics Belgium for this report. All in all however, after reallocating the PSUs in the EU-SILC UDB, the quality of the sample design variables is considerably improved and reflects rather closely those made available in BE-SILC. This is very useful as it enables us to evaluate the impact of the randomisation of PSU codes on the standard error of changes over time in a comparable manner (see the next section).

Table 5: Allocation of the PSUs and observations over three strata in three different scenarios

		PSUs			Observations		
		UDB original	UDB new	BE-SILC	UDB original	UDB new	BE-SILC
2008	Brussels-Capital	57	48	49	1816	1839	1840
	Flemish Region	153	139	138	8290	8276	8275
	Walloon Region	97	87	87	5002	4993	4993
	Total	307	274	274	15108	15108	15108
2009	Brussels-Capital	53	50	50	1882	1916	1916
	Flemish Region	156	138	138	7855	7819	7819
	Walloon Region	97	87	87	4984	4986	4986
	Total	306	275	275	14721	14721	14721
2010	Brussels-Capital	55	50	50	1858	1915	1915
	Flemish Region	159	139	138	7923	7896	7851
	Walloon Region	103	85	86	4973	4943	4988
	Total	317	274	274	14754	14754	14754
2011	Brussels-Capital	57	49	49	1979	2022	2022
	Flemish Region	163	138	138	7504	7477	7477
	Walloon Region	122	87	87	4817	4801	4801
	Total	342	274	274	14300	14300	14300

Note: 'UDB original' refers to the number of PSUs and observations in the EU-SILC UDB if the original variable DB060 is used as a PSU variable and DB040 as a stratification variable. 'UDB new' contains the same information in the case that the PSU and stratum variables are re-defined as in Goedemé (2013b). 'BE-SILC' refers to the correct number of PSUs and observations per aggregated stratum as observed in the SILC data provided by Statistics Belgium.
Source: EU-SILC UDB 2008 ver5, 2009 ver4, 2010 ver2, 2011 ver2; BE-SILC Statistics Belgium.

In the EU-SILC UDB and in Eurostat-SILC, the sample design variables are not coded consistently over time. This is illustrated for the EU-SILC UDB in Table 6. This table shows that the correlation between the achieved sample size of PSUs in SILC 2011 and the achieved sample size of PSUs in the previous waves is close to zero in the EU-SILC UDB, whereas it is above 0.70 in BE-SILC. In principle, one could

[5] In the future, the procedure could be improved by giving a larger weight to households in the panel that joined most recently the data. Further analysis learns that this would however not have improved the re-allocation of PSUs for the 2010 data.

conceive some matching procedure to make the variables consistent over time. This is not what is currently done at Eurostat, which uses the sample design variables without modification, also for the estimation of the statistical significance of changes over time, as is done in Osier et al. (2013). In principle, there is no reason to use statistical matching which is computationally intensive and not free of errors, as the correct data could be easily provided to Eurostat by the National Statistical Institutes that produce the SILC data.

Table 6: Correlation between the achieved sample size of PSUs (number of persons in the sample) in SILC 2011 and the three previous waves

	2008	2009	2010
EU-SILC UDB	-0.04	-0.09	0.11
BE-SILC	0.71	0.82	0.89

Source: EU-SILC UDB 2008 ver5, 2009 ver4, 2010 ver2, 2011 ver2; BE-SILC Statistics Belgium.

3 The importance of consistent sample design variables: an illustration

As explained in the introduction, questions have been raised about the accuracy of the standard error estimates of changes over time published in Osier et al. (2013). The formula of the sampling variance (VAR) of the difference in the mean (or proportion) (*D*) of two variables y and x with means *Y* and *X* can be written as VAR(*D*) = VAR(*Y-X*) = VAR(*Y*) + VAR(*X*) – 2*COVAR(*Y,X*) (e.g. Heeringa et al., 2010). As becomes clear from the formula, the sampling variance of a difference does not only depend on the variance of the two estimated averages, but also on their covariance (see also Goedemé et al., 2013). In the case of independent samples, the covariance is equal to zero. In the case of EU-SILC, samples of consecutive years are not independent: the composition of one sample has an influence on the composition of the other due to the panel character of the data. In the case of the Belgian SILC, samples are even dependent if there is no overlap in the households included in the sample, due to the fact that the PSUs remain fixed for the entire duration of EU-SILC. Evidently, the covariance at the PSU level between two waves of EU-SILC can only be estimated if PSUs and primary strata receive the same identification number in both waves of EU-SILC.

Table 7 illustrates this point with standard error estimates of the at-risk-of-poverty indicator (AROP60) and the percentage point change over between EU-SILC 2011 and the preceding three waves. The estimation makes use of the ultimate cluster approach, as described above and standard estimation commands of the statistical software package Stata. It is clear that even though the standard errors of the cross-sectional estimates are very similar between the EU-SILC UDB (using the original PSU variable and DB040 as stratification variable) and BE-SILC (assuming only three strata, one for each region), the standard errors of the change in AROP60 differ considerably. In the case of BE-SILC, standard errors of the percentage point change in AROP60 are generally lower and clearly decrease when two years are compared that are closer to one another. In contrast, in the case of the EU-SILC UDB, the standard error of the change is nearly as large as if we would assume that the various waves of EU-SILC consist of independent samples and do not show the same pattern. In this context, it should be noted that an inconsistent coding of the sample design variables may bias the estimated covariance in an unpredictable way. It may decrease the covariance towards close to zero as is the case for the results presented here, but it may also result in a seriously downwardly biased

covariance estimate as is the case for the results presented in Table 1 regarding the difference in AROPE between 2011 and 2007.

Table 7: Standard errors and point estimates of the at-risk-of-poverty indicator (AROP60) and the percentage point change in AROP60, 2008-2011

	EU-SILC UDB		BE-SILC	
	Point	SE	point	SE
2008	14.73	0.74	14.72	0.70
2009	14.57	0.80	14.57	0.77
2010	14.59	0.78	14.59	0.75
2011	15.30	0.90	15.30	0.86
2011-2010	0.71	1.14	0.71	0.64
2011-2009	0.73	1.20	0.73	0.79
2011-2008	0.58	1.19	0.58	0.87

Note: For estimating the standard errors, AROP60 is considered to be a simple proportion.
Source: EU-SILC UDB 2008 ver5, 2009 ver4, 2010 ver2, 2011 ver2; BE-SILC Statistics Belgium

In what follows, I will present some more estimates of changes in several indicators. The purpose is not to present very precise estimates of the sampling variance, but rather to give a rough idea of the order of magnitude of the standard error of changes over time. The sampling variance is estimated with the ultimate cluster approach on the basis of the PSU and primary stratum variable in the data provided by Statistics Belgium. As discussed previously, obtaining more precise variance estimates would require (1) a better estimator of the sampling variance, to be further validated in simulation studies; (2) an estimator which takes also calibration and imputation into account, which is impossible with the dataset provided by Statistics Belgium. I restrict the illustrations to the at-risk-of-poverty indicator with a poverty threshold equal to 60 per cent of the national median equivalent disposable household income (AROP60), by lack of other variables in the data. As noted previously, the variance estimates presented below assume that the at-risk-of-poverty threshold is not estimated on the basis of the data (for a discussion of this issue, see Preston, 1995; Berger and Skinner, 2003; Goedemé, 2012). The first subsection contains estimates for the total Belgian population. In the subsequent subsection, changes in the at-risk-of poverty indicator for the three regions are presented.

3.1 Changes in the at-risk-of-poverty indicator in Belgium

In the following tables, the following figures are presented:

- real changes in the at-risk-of-poverty threshold (at 60% of the median income);
- changes in the at-risk-of-poverty rate with the poverty line anchored in 2008;
- changes in AROP60 for children aged less than 12 years; persons aged 75 and over; persons living in single parent households; the unemployed; and tenants.

Among others, it can be concluded that the at-risk-of-poverty threshold can be estimated with a high degree of precision. Even though it has increased with only about 40 EUR in real terms (prices of 2010) between SILC 2008 and 2011, the change is significant at the 5% confidence level. In comparison, (changes in) the poverty rate of vulnerable groups such as young children, single parent

households, the unemployed and tenants can only be measured with a relatively high level of uncertainty. For these groups, the drop or increase in poverty between 2008 and 2011 should amount to at least between 4 and 8 percentage points[6] in order to be significantly different at the 95% confidence level. In other words, for some vulnerable groups, even at the level of Belgium, the power of EU-SILC to closely monitor the income situation of vulnerable groups is limited. In this context, even relatively large drops or increases in poverty should be interpreted with caution, even though the standard error of year-to-year changes is somewhat smaller for the cases studied below. At the same time, if the confidence level is lowered to 90%, observed changes over the past few years do not all appear to be statistically insignificant.

Table 8: Real changes in the 60 per cent at-risk-of-poverty threshold, expressed in EUR per month of 2011

	point	SE	LB	UB
2008	959.4	10.0	939.7	979.0
2009	986.2	9.6	967.5	1005.0
2010	994.5	13.5	968.1	1020.9
2011	1000.4	12.9	975.0	1025.8
2011-2010	5.9	9.4	-12.6	24.4
2011-2009	14.2	11.2	-7.8	36.2
2011-2008	41.0	12.6	16.4	65.7

Note: 95% confidence intervals

Source: BE-SILC Statistics Belgium.

Table 9: Changes in the at-risk-of-poverty rate, with the poverty line anchored in 2008

	point	SE	LB	UB
2008	14.72	0.69	13.36	16.08
2009	13.07	0.71	11.67	14.47
2010	12.97	0.70	11.60	14.34
2011	13.33	0.81	11.74	14.92
2011-2008	-1.39	0.87	-0.03	0.32
2010-2008	-1.75	0.71	-3.14	-0.35
2009-2008	-1.65	0.59	-2.82	-0.49

Note: 95% confidence intervals

Source: BE-SILC Statistics Belgium.

[6] This is a crude estimate on the basis of the results presented in the table below. A proper analysis of the power could provide more precise estimates.

Table 10: Change in AROP60 of children less than 12 years old, 90% confidence intervals

	point	SE	LB	UB
2008	15.77	1.49	13.32	18.22
2009	17.18	1.69	14.39	19.96
2010	19.57	1.69	16.77	22.36
2011	19.44	1.77	16.52	22.35
2011-2010	-0.13	1.50	-2.60	2.34
2011-2009	2.26	1.88	-0.85	5.37
2011-2008	3.67	2.04	0.31	7.03

Source: BE-SILC Statistics Belgium.

Table 11: Change in AROP60 of persons aged 75 years and over, 90% confidence intervals

	point	SE	LB	UB
2008	23.81	1.90	20.66	26.95
2009	24.70	1.97	21.45	27.94
2010	20.89	1.65	18.16	23.62
2011	21.36	1.85	18.31	24.41
2011-2010	0.47	1.96	-2.77	3.71
2011-2009	-3.34	2.48	-7.43	0.76
2011-2008	-2.45	2.41	-6.43	1.53

Source: BE-SILC Statistics Belgium.

Table 12: Change in AROP60 of persons living in single parent households, 90% confidence intervals

	point	SE	LB	UB
2008	39.47	3.08	34.38	44.55
2009	36.89	3.17	31.66	42.11
2010	35.26	3.34	29.75	40.76
2011	38.51	3.11	33.38	43.65
2011-2010	3.26	3.48	-2.48	9.00
2011-2009	1.63	3.89	-4.79	8.04
2011-2008	-0.95	4.21	-7.90	5.99

Source: BE-SILC Statistics Belgium.

Table 13: Change in AROP60 of the unemployed, 90% confidence intervals

	point	SE	LB	UB
2008	34.77	1.93	31.59	37.95
2009	33.43	2.01	30.11	36.75
2010	30.37	1.81	27.38	33.35
2011	37.84	2.31	34.02	41.65
2011-2010	7.47	2.38	3.54	11.40
2011-2009	4.41	2.64	0.04	8.77
2011-2008	3.07	2.74	-1.46	7.59

Source: BE-SILC Statistics Belgium.

Table 14: Change in AROP60 among tenants, 90% confidence intervals

	point	SE	LB	UB
2008	28.46	1.63	25.78	31.14
2009	28.62	1.68	25.85	31.40
2010	29.52	1.58	26.90	32.13
2011	33.09	1.77	30.16	36.01
2011-2010	3.57	1.55	1.01	6.13
2011-2009	4.46	2.00	1.16	7.77
2011-2008	4.63	2.19	1.01	8.25

Source: BE-SILC Statistics Belgium.

3.2 Changes in AROP60 in the three regions of Belgium

The following tables contain similar estimates as in the previous subsection, but for the three regions of Belgium. The data provided by Statistics Belgium do not contain DB040. Instead, the stratification variable has been used to construct a variable for identifying the regions. In practice this means that households that have moved from one region to another between the moment of selection and the moment of interview are considered to be still living in the region they inhabited at the moment of selection. This may bias somewhat both the point estimates and the estimation of the sampling variance.

Not surprisingly, standard errors are generally larger for regional estimates than for estimates of the total population. This is especially so for Brussels-Capital, and to a less extent for the Walloon region. Remarkably, the covariance for regional estimates seems to be somewhat stronger than in the case of estimates for Belgium. As a result, the standard error of the changes are not much larger than the standard errors of the cross-sectional estimates for the respective region. In addition, especially in Brussels-Capital Region, observed changes have been so large, that even for relatively small vulnerable groups they appear to be statistically significant at the 90% confidence level. In general, however, EU-SILC is not sufficiently precise to closely monitor at the regional level (changes in) the situation of important vulnerable groups.

Table 15: Change in AROP60 by region, 90% confidence intervals

	Brussels-Capital Region				Flemish Region				Walloon Region			
	Point	SE	LB	UB	point	SE	LB	UB	point	SE	LB	UB
2008	25.9	2.9	21.1	30.8	10.1	0.8	8.8	11.4	19.5	1.2	17.5	21.5
2009	27.7	3.5	21.8	33.5	10.1	0.8	8.7	11.5	18.4	1.2	16.3	20.5
2010	27.7	3.3	22.2	33.3	10.3	0.8	9.0	11.7	18.0	1.3	15.9	20.1
2011	33.6	3.2	28.2	39.0	9.8	0.9	8.3	11.3	19.2	1.5	16.6	21.8
2011-2010	5.9	2.4	1.9	9.9	-0.5	0.8	-1.8	0.8	1.2	1.1	-0.6	3.0
2011-2009	5.9	2.7	1.4	10.5	-0.3	1.0	-1.9	1.3	0.8	1.4	-1.6	3.2
2011-2008	7.7	3.1	2.4	13.0	-0.3	1.0	-2.0	1.4	-0.3	1.5	-2.8	2.2

Source: BE-SILC Statistics Belgium.

Table 16: Change in AROP60 among children aged less than 12 by region, 90% confidence intervals

	Brussels-Capital Region				Flemish Region				Walloon Region			
	point	SE	LB	UB	point	SE	LB	UB	point	SE	LB	UB
2008	26.3	4.7	18.5	34.1	8.9	2.0	5.6	12.2	22.7	2.4	18.6	26.8
2009	32.9	5.0	24.6	41.2	10.8	2.3	7.0	14.7	21.4	2.7	16.9	25.9
2010	38.3	5.2	29.5	47.0	12.0	1.8	9.0	14.9	24.8	3.3	19.4	30.3
2011	43.2	4.8	35.1	51.4	11.4	2.2	7.7	15.2	23.1	3.0	18.1	28.0
2011-2010	5.0	5.4	-4.0	14.0	-0.5	1.8	-3.5	2.5	-1.8	2.5	-6.0	2.4
2011-2009	10.3	5.5	1.1	19.6	0.6	2.7	-3.9	5.1	1.6	2.8	-3.1	6.4
2011-2008	16.9	6.1	6.8	27.1	2.5	2.9	-2.3	7.3	0.3	3.1	-4.7	5.4

Source: BE-SILC Statistics Belgium.

Table 17: Change in AROP60 among elderly persons aged 75 and over by region, 90% confidence intervals

	Brussels-Capital Region				Flemish Region				Walloon Region			
	Point	SE	LB	UB	point	SE	LB	UB	point	SE	LB	UB
2008	23.7	5.0	15.3	32.2	22.4	2.5	18.2	26.6	26.5	3.4	20.8	32.1
2009	25.5	5.5	16.2	34.7	22.6	2.6	18.2	26.9	28.4	3.5	22.6	34.2
2010	24.9	4.9	16.7	33.1	19.8	2.2	16.2	23.4	21.7	3.0	16.8	26.7
2011	21.7	5.6	12.3	31.1	20.2	2.4	16.2	24.1	23.6	3.3	18.1	29.2
2011-2010	-3.2	5.3	-12.1	5.7	0.3	2.5	-3.8	4.4	1.9	3.8	-4.5	8.3
2011-2009	-3.8	6.5	-14.7	7.2	-2.4	3.2	-7.7	2.8	-4.7	4.8	-12.7	3.2
2011-2008	-2.0	6.4	-12.7	8.6	-2.2	3.1	-7.3	2.9	-2.8	4.6	-10.5	4.8

Source: BE-SILC Statistics Belgium.

Table 18: Change in AROP60 among persons living in single parent households by region, 90% confidence intervals

	Brussels-Capital Region				Flemish Region				Walloon Region			
	point	SE	LB	UB	point	SE	LB	UB	point	SE	LB	UB
2008	39.4	7.0	27.6	51.1	28.5	5.1	20.1	36.9	49.2	4.6	41.5	56.9
2009	34.0	7.1	22.1	45.9	22.3	4.3	15.2	29.5	51.3	5.0	43.0	59.6
2010	39.9	7.3	27.6	52.2	21.9	4.6	14.3	29.6	44.8	4.9	36.6	52.9
2011	44.2	7.2	32.1	56.2	22.4	4.4	15.1	29.6	52.3	4.6	44.6	60.0
2011-2010	4.2	9.2	-11.1	19.6	0.5	4.2	-6.5	7.4	7.5	5.5	-1.6	16.6
2011-2009	10.2	8.6	-4.3	24.7	0.1	4.7	-7.7	7.8	1.0	6.3	-9.4	11.5
2011-2008	4.8	9.2	-10.7	20.2	-6.1	6.4	-16.8	4.5	3.1	6.5	-7.6	13.8

Source: BE-SILC Statistics Belgium.

Table 19: Change in AROP60 among the unemployed by region, 90% confidence intervals

	Brussels-Capital Region				Flemish Region				Walloon Region			
	point	SE	LB	UB	point	SE	LB	UB	point	SE	LB	UB
2008	56.5	4.0	49.9	63.1	20.2	2.8	15.5	24.9	41.8	2.7	37.3	46.4
2009	59.3	4.8	51.2	67.3	19.0	2.7	14.6	23.4	40.0	3.0	35.1	45.0
2010	47.5	5.4	38.6	56.5	23.0	2.6	18.6	27.3	32.9	2.6	28.6	37.3
2011	56.7	4.4	49.3	64.2	23.4	3.5	17.7	29.1	44.7	3.7	38.5	50.8
2011-2010	9.2	4.7	1.3	17.1	0.4	3.4	-5.2	6.0	11.7	4.0	5.0	18.5
2011-2009	-2.6	6.6	-13.6	8.5	4.4	3.7	-1.8	10.6	4.6	4.1	-2.1	11.4
2011-2008	0.2	5.2	-8.5	8.9	3.1	4.0	-3.5	9.8	2.8	4.4	-4.4	10.1

Source: BE-SILC Statistics Belgium.

Table 20: Change in AROP60 among tenants, by region, 90% confidence intervals

	Brussels-Capital Region				Flemish Region				Walloon Region			
	point	SE	LB	UB	point	SE	LB	UB	point	SE	LB	UB
2008	34.2	3.6	28.1	40.3	19.8	2.3	16.0	23.6	38.2	2.9	33.4	43.0
2009	38.4	4.0	31.6	45.1	18.9	2.4	14.9	22.9	36.7	2.5	32.5	40.9
2010	38.3	3.6	32.2	44.4	21.6	2.2	18.0	25.2	34.7	2.5	30.6	38.8
2011	47.8	3.3	42.3	53.4	21.0	2.5	16.9	25.2	41.1	3.0	36.2	46.0
2011-2010	9.6	3.4	3.8	15.3	-0.6	2.0	-3.8	2.7	6.4	3.0	1.4	11.5
2011-2009	9.5	3.9	2.9	16.0	2.1	2.9	-2.7	7.0	4.4	3.6	-1.6	10.4
2011-2008	13.7	4.1	6.8	20.6	1.3	3.2	-4.0	6.5	2.9	4.0	-3.7	9.5

Source: BE-SILC Statistics Belgium.

4 Conclusion and suggestions for improvement

The purpose of this report is to shed some light on the statistical precision of the Belgian EU-SILC data for monitoring changes over time. Recently, Eurostat has published a report on the standard error of changes in the AROPE indicator (Osier et al., 2013), but clearly, the estimated standard errors of changes over time are not very reliable, due to a lack of consistency in the coding of the sample design variables. This report explains why correct sample design variables are crucial for getting a proper idea of the size of the sampling variance for changes over time. For doing so, the Belgian SILC sample design is described, and the quality of the Belgian sample design variables reviewed. In addition, on the basis of a dataset with consistent sample design variables prepared by Statistics Belgium, some estimates are presented to get a rough idea of the size of the standard error of changes over time for the at-risk-of-poverty indicator.

It is clear that EU-SILC is not sufficiently precise for closely monitoring small changes in the income situation of relatively small, but non-negligible vulnerable groups, especially not at the regional level. Nonetheless, some trends can be observed. In general, this calls for a bigger investment in EU-SILC, if policy makers take the point of evidence-based policy making to fight poverty and social exclusion seriously. I would like to stress that the problem is particularly present for monitoring trends at the regional level. Obviously, an increase in the sample size is only realistic if also more resources are invested in the EU-SILC team at Statistics Belgium so as to ensure that the collection and processing of the SILC data can be carefully monitored. In any case, increases in the sample size in general and changes in the sample design in particular should be implemented very carefully, so that it would still be possible to estimate the sampling variance and the sample size is increased in the most efficient way.

Apart from increasing the power of the SILC survey, it is highly recommended to ensure that consistent sample design variables are included in the SILC database sent to Eurostat and to the research community. In other words, every PSU and stratum should receive a unique code that does not change from one dataset to another. Only by doing so, optimal use can be made of the panel character of EU-SILC by estimating the covariance between various waves.

Finally, it is recommended to expand the simulation study of Dawagne and Milano (2011) in order to gain more insight into which estimator would be most appropriate for estimating the sampling variance and confidence intervals in the case of the Belgian SILC. Such a study should cover also other estimators, such as the estimator that is currently used at Statistics Belgium and the one used in Osier et al. (2013) and consider also issues such as non-response, calibration and imputation as well as the sampling variance of changes over time. In addition, it would be useful to expand the study to cover also the other AROPE variables and use a wider timeframe, such that the covariance is estimated also for waves of SILC with no overlap in sampled households.

References

Berger, Y. G., and Skinner, C. J. (2003), 'Variance Estimation for a Low Income Proportion' in *Journal of the Royal Statistical Society. Series C (Applied Statistics),* 52(4): 457-468.

Dawagne, J.-M., and Milano, R. (2011), 'Estimation de la variance dans le cas d'un tirage systématique d'unités primaires avec remise', in Tremblay, M.-E., Lavallée, P. and El haj Tirari, M. (eds.), *Pratique et méthodes de sondages*, Paris: Dunod, pp. 342-348.

European Council (2010), *European Council 17 June 2010 Conclusions*, Brussels: European Council, 15p.

Eurostat (2002), *Monographs of official statistics. Variance estimation methods in the European Union*, Luxembourg: Office for Official Publications of the European Communities, 63p.

Goedemé, T. (2012), *Essays on poverty and minimum income protection for Europe's elderly*, PhD dissertation, Antwerp: University of Antwerp, 262p.

Goedemé, T. (2013a), *The EU-SILC sample design variables: critical review and recommendations*, CSB Working Paper Series, CSB WP 13/02, Antwerp: Herman Deleeck Centre for Social Policy, University of Antwerp, 36p.

Goedemé, T. (2013b), 'How much confidence can we have in EU-SILC? Complex sample designs and the standard error of the Europe 2020 poverty indicators' in *Social Indicators Research,* 110(1): 89-110. doi: 10.1007/s11205-011-9918-2.

Goedemé, T., Van den Bosch, K., Salanauskaite, L., and Verbist, G. (2013), *Testing the Statistical Significance of Microsimulation Results: Often Easier than You Think. A Technical Note*, ImPRovE Working Paper No. 13/10, Antwerp: Herman Deleeck Centre for Social Policy - University of Antwerp, 27p.p.

Groves, R. M., Fowler, F. J. J., Couper, M. P., Lepkowski, J. M., Singer, E., and Tourangeau, R. (2009), *Survey Methodology (Second edition)*, New Jersey: John Wiley & Sons, 461p.

Heeringa, S. G., West, B. T., and Berglund, P. A. (2010), *Applied Survey Data Analysis*, Boca Raton: Chapman & Hall/CRC, 467p.

Kalton, G. (1979), 'Ultimate Cluster Sampling' in *Journal of the Royal Statistical Society. Series A (General),* 142(2): 210-222.

Kish, L. (1965), *Survey Sampling*, New York: John Wiley & Sons, 643p.

N. (2004), *Niveau des unités primaires pour SILC (& EBM &EFT?)). Revu et complété après remarques P. Slock & G. Vekeman 29 janvier; remarques J.M. Dawagne 2 février; remarques J.P. Ntwa, P. Slock, C. Vanderhoeft 4 février; fichier superficie par commune C. Sajotte 4 février*.

Osier, G. (2012), *The linearization approach implemented by Eurostat for the first wave of EU-SILC: what could be done from the second wave onwards?*, Paper presented during the Workshop on standard error estimation and other related sampling issues in EU-SILC, organized in the context of the EU-funded “Net-SILC2” project, Eurostat, Luxembourg, 29-30 March 2012.

Osier, G., Berger, Y., and Goedemé, T. (2013), *Standard error estimation for the EU–SILC indicators of poverty and social exclusion*, Eurostat Methodologies and Working papers, Luxembourg: Publications Office of the European Union, 54p.

Preston, I. (1995), 'Sampling Distributions of Relative Poverty Statistics' in *Journal of the Royal Statistical Society. Series C (Applied Statistics),* 44(1): 91-99.

Verma, V., Betti, G., and Gagliardi, F. (2010), *An assessment of survey errors in EU-SILC*, Eurostat Methodologies and Working Papers, Luxembourg: Eurostat, 70p.

Wolter, K. M. (2007), *Introduction to Variance Estimation*, New York: Springer, 447p.