CSB

WORKING PAPER

centreforsocialpolicy.eu | December 2010 | No 10 / 09

The standard error of estimates based on EU-SILC. An exploration through the Europe 2020 poverty indicators

Tim Goedemé

Universiteit Antwerpen

CSB

University of Antwerp
Herman Deleeck Centre for Social Policy
Sint-Jacobstraat 2
B – 2000 Antwerp
fax +32 (0)3 265 57 98

The standard error of estimates based on EU-SILC. An exploration through the Europe 2020 poverty indicators

Tim Goedemé

Working Paper No. 10 / 09

December 2010

ABSTRACT

Currently, the European Union Statistics on Income and Living Conditions (EU-SILC) is the single most important data source for cross-national comparative research on income and living conditions in the European Union. As EU-SILC consists of a sample of European households, point estimates should be accompanied by appropriate standard errors and confidence intervals. This is especially so if indicators are constructed for measuring progress towards pre-defined targets such as those of the Europe 2020 strategy. All too often this has been neglected in European poverty research and official publications. In contrast, this paper pays explicit attention to the calculation of standard errors and confidence intervals. Standard errors are strongly dependent on the sample design. Therefore, accurate information on the sample design is crucial, especially for a database like EU-SILC which contains data on about 30 European countries which employ different complex sample designs. However, information on the sample design is incomplete in the EU-SILC User Database for data confidentiality reasons and there are several options for handling this lack of information. In this paper, we document the sample designs used in EU-SILC and compare the information available through different sources, namely the Quality Reports, the User Database and a specific dataset containing additional information about the sample design prepared by Eurostat. Furthermore, on the basis of the specific dataset prepared by Eurostat, we explore which variables are best used when analysing EU-SILC for adequately computing standard errors. We illustrate the importance of various assumptions with regard to the sample design by presenting results for the official Europe 2020 poverty indicators. It is shown that neglecting the sample design can lead to a serious underestimation of the standard errors. In addition, it is discussed how researchers using EU-SILC could best take account of the sample design for appropriately estimating standard errors.

Corresponding author:

Tim Goedemé

Ph.D. Research Fellow of the Research Foundation – Flanders (FWO)
Herman Deleeck Centre for Social Policy (University of Antwerp)
Sint-Jacobstraat 2 (M479)
2000 Antwerp (Belgium)
Tel.: +32 3 265 55 55 - E-mail: tim.goedeme@ua.ac.be

1. Introduction

On 17 June 2010 the European Council agreed on reducing the number of Europeans at-risk-of- poverty or social exclusion by at least 20 million (European Council, 2010). Member States can choose to focus on one of three indicators to achieve the target: an indicator of financial poverty (the so-called at-risk-of-poverty rate), an indicator of material deprivation and an indicator of the number of jobless households[1] (details below). The data underlying these indicators come from the European Union Statistics on Income and Living Conditions (EU-SILC), the principal data source for cross-national comparative research on income and living conditions in the European Union (EU). As EU-SILC is composed of samples in all EU Member States, sampling and non-sampling errors can seriously affect the accuracy of all estimates based on EU-SILC – including the Europe 2020 poverty indicators. However, until now, Eurostat has refrained from consistently publishing standard errors and confidence intervals alongside the official poverty indicators (e.g. Eurostat, 2010b; Wolff, 2010). Unfortunately, this is not a feature unique to Eurostat publications. It seems to be rather common practice to ignore the publication of confidence intervals in the case of descriptive (poverty) statistics (e.g. de Vos and Zaidi, 1998; Kangas and Ritakallio, 2007; Whelan and Maître, 2007; OECD, 2008). Furthermore, in the case of analytical studies using a variety of (regression) methods, it is not always clear whether standard errors are calculated accurately.

Confidence intervals do not address all kinds of survey errors. Nevertheless, the estimation of confidence intervals can save money, time and effort in that they indicate which differences between point estimates are not worth further investigating by showing whether they have a high probability of being due to random error. However, they can only serve this purpose if standard errors have been estimated accurately. In order to do so, among others it is necessary to take account of the sample design and weighting schemes. Previous studies which focused on design effects in the case of poverty measures generally found strong effects of the sample design on the standard error (e.g. Rodgers and Rodgers, 1993: 43; Howes and Lanjouw, 1998: 107; Jolliffe et al., 2004: 563).

Many countries covered by EU-SILC employ complex sample designs involving multiple stages of selection, stratification and clustering. In general, two different types of information are necessary to take account of the sample design: an accurate description of the implemented sample design and adequate variables in the dataset to take account of clustering and stratification. In the case of EU-SILC for many participating countries one or both types of information are lacking. Recently, the documentation of the EU-SILC sample design has become more widely available by the publication of the Intermediate EU-SILC comparative quality report (Eurostat, 2010a) as well as the dissemination of most national quality

1 Expressed as share of people living in households with very low work intensity.

reports[2]. However, as far as the sample design variables in the dataset are concerned, data in the EU-SILC User Database (UDB) are incomplete for many countries. Given the partial nature of the available information, it is a matter of discussion how one can best take account of the sample design and how much bias there is on the estimated standard errors.

This paper aims at offering some guidance on the computation of standard errors for researchers interested in analysing EU-SILC. First, we shortly elaborate on the general principles of the computation of standard errors. Second, drawing on the reports published by Eurostat as well as personal correspondence with the national statistical institutes across Europe, we concisely discuss the sample designs behind EU-SILC and the available sample design variables in EU-SILC. Third, we illustrate the effect on the standard error of various assumptions with regard to the sample design for the Europe 2020 poverty indicators. Moreover, the resulting standard errors are compared to those obtained by using a dataset prepared by Eurostat which contains additional information on the sample design. We conclude with a recommendation on which variables to use when computing standard errors of estimates based on EU-SILC.

2. The estimation of standard errors: some principles

There are several approaches to the estimation of standard errors and the computation of confidence intervals. In the case of linearization, formulae are derived analytically which can be used to estimate the standard error of (complex) indicators based on complex samples. In a second step confidence intervals are computed assuming a certain sampling distribution, usually Student's t-distribution or the normal distribution. A completely different approach is based on re-sampling from the original sample a high number of samples in order to empirically derive a sampling distribution (e.g. Jackknife repeated replication or the bootstrap). Subsequently, on the basis of this 'empirical sampling distribution' standard errors and confidence intervals are computed (cf. Mooney and Duval, 1993 for an introduction; and Biewen, 2002; Trede, 2002; Van Kerm, 2002; Davidson and Flachaire, 2007; and del Mar Rueda and Muñoz, 2009 for an application to poverty and inequality measures). There are various methods in between (see Efron and Tibshirani, 1998: 53-56) and each of these methods has its advantages and shortcomings. This section will not go into the details of the various approaches to variance estimation. Rather, it aims at showing the general principles which always should be taken into account when computing standard errors. Whichever approach is used, in order to get the standard errors right one should replicate as closely as possible the entire procedure of drawing the sample and calculating the desired statistic. There are four main ingredients to this: sample design, weighting, imputation and the

[2] These reports can be downloaded from Circa: http://circa.europa.eu/Public/irc/dsis/eusilc/library?l=/quality_assessment&vm=detailed&sb=Title.

computation of the statistic one is interested in (cf. Eurostat, 2002). In this section we elaborate shortly on each of these issues.

2.1. Sample design

The sample design can seriously affect the standard error. Stakes are high that when assuming a simple random sample when the actual sample design involves clustering and stratification, standard errors will be wrong. Multi-stage designs involve several stages of sampling and sub-sampling and start from the random selection of clusters of elements (e.g. municipalities, census sections, dwelling blocks), i.e. primary sampling units (PSUs). If the design consists of several stages, the next step consists of drawing a subsample within each cluster. The advantage of (geographical) clustering is that interviewers can collect the interviews in a limited number of geographical areas, reducing the costs of the survey (e.g. Sturgis, 2004: 1). However, a major disadvantage is that clustering can seriously increase the standard error if the variance within clusters is small compared to the between-cluster variance with respect to the variable of interest. Intuitively, one could say that if clusters are very homogenous and differ a lot between each other, drawing a sample of clusters at the first stage (instead of elements) increases the risk that not all different kinds of elements are represented in the sample. In other words, estimates risk to vary a lot from sample to sample. However, if clusters are heterogeneous, the within-cluster variance large and the between-cluster variance small, drawing a sample of clusters does not risk missing important types in the population and the effect of clustering on the standard error will be small. Stratification has the opposite effect. Stratification serves the purpose of increasing the representativeness of the sample and decreasing the risk that some parts in the population remain unrepresented. In order to do so, the population is divided into exclusive groups (strata). Subsequently an independent sample is drawn within each of these strata. Especially if the variance between strata is large with respect to the relevant variable, stratification contributes to decreasing the standard error. Usually, the effect of stratification is larger in the case of a clustered sample (cf. Kish, 1965; Kalton, 1983; Howes and Lanjouw, 1998; Lee and Forthofer, 2006: 9). Of course, the effect of the sample design can differ from one variable to another: clusters or strata may differ strongly in the case of one variable and be rather heterogeneous in the case of another. A crucial point is that if the ratio of selected clusters at the first stage to the total number of clusters in the population is small, other stages than the first add little to the standard error (for a mathematical elaboration, see Kish, 1965; Cochran, 1977). Therefore, the common practice is to approximately assess the sampling variance by estimating the variability among the PSUs, since this is the dominating component of the total variance (cf. Eurostat, 2002: 12-13). As a result, for a good approximation only accurate information on the first stage of the sample design is needed, considerably simplifying documentation and computation needs.

2.2. Imputation and weighting

Apart from the sample size and the sample design, standard errors are also influenced by other sources of random error. Among others, these include imputation and weighting. When there is item non-response, sometimes values are imputed (cf. Kalton, 1983: 67-68). If imputation is based on a random procedure, it adds another source of random error[3]. As a result, the computation of standard errors should take this into account (Shao, 1996; Shao and Chen, 1998). However, the inclusion of this source of error is not easy. First, depending on the estimation technique and available software, information is needed on (1) the response / non-response status; (2) the imputation method used and information on the auxiliary variables; (3) information on the 'donor'; (4) and information on the imputation classes (Eurostat, 2002: 19). Second, standard estimation procedures in many software packages do not include routines for taking account of imputation for estimating the standard error in the usual way. Nonetheless, the impact of imputation can be large if there is considerable item non response. A non-response rate of 30% may lead to an under-estimation of the standard error by 10-50% (Kovar and Whitridge, 1995 as cited in Eurostat, 2002: 18). There are few studies on the impact of imputation procedures on standard errors in the case of poverty indicators. One study by Alfons et al. (2009) on the Austrian EU-SILC 2004 data finds that the additional uncertainty introduced by imputation is limited in the case of the EU at-risk-of-poverty indicator, but larger for the average equivalent disposable household income.

Weighting is another potential source of random error. Weights assign more relative importance to some observations than to others in order to restore imbalances in the sample and avoid biased estimates. In general, weights are used to counteract three types of imbalance: unequal selection probabilities, unit non-response and (remaining) differences between the sample and known population data (Kalton, 1983: 69-75). In some cases imbalances in the sample occur on purpose, for instance some small strata may be over-represented to enable reliable estimates of these strata. For obtaining population estimates, respondents are given weights which are inversely proportional to the probability of being selected. Weights are also used to counterbalance unit non-response. This may be done by relying exclusively on data in the sample (e.g. adjusting the achieved sample size within clusters to the total sample size of that cluster); or by relying on external data. In the latter case, the construction of the weights is similar to post-stratification. With post-stratification (or calibration) weights are adjusted such that the estimated distribution corresponds to known population totals, usually regarding demographic variables such as age and sex. In doing so, post-stratification may also increase precision by compensating coverage errors in the sample frame. Even though weights are aimed at increasing the

3 The neglect of imputation generally leads to an under-estimation of the variance: imputation can increase random error, but also the denominator of the variance estimate (n) is overestimated.

precision of survey estimates, they may substantially increase standard errors. This is especially the case if the variance of the weights is large. Therefore, it is important to take weighting into account when computing standard errors[4]. In EU-SILC for all countries weights have been developed to counteract variations in selection probabilities and unit non-response as well as to bring some demographic estimates in line with external population data (cf. Eurostat, 2010c).

2.3. Complex poverty measures

The formulae for calculating standard errors do not only depend on the sample design, imputation and weighting, but also on the computed statistic. Standard errors are more straightforward to compute for some indicators than for others. For instance, the standard error of a proportion or the mean is well known and can easily be adapted to more complex sample designs (Kish, 1965; Cochran, 1977). Many poverty indicators consist of headcounts with a fixed (i.e. not random) poverty threshold. For instance, the Europe 2020 deprivation indicator estimates the proportion of the population which scores badly on at least 4 out of 9 deprivation items. In other words, in this case, the standard error is equal to that of a proportion. The same holds for the proportion of households with very low work intensity. As long as the poverty threshold is not estimated from the survey data itself (i.e. is not random) all poverty measures of the family of the well-known FGT-class of poverty measures have standard error formulae similar to those of a proportion or mean (cf. Foster et al., 1984: 763; Kakwani, 1993; Jolliffe and Semykina, 1999).

However, this is not always the case. In the case of the at-risk-of-poverty indicator, the poverty threshold is estimated on the basis of the survey data: it is equal to 60 percent of the median equivalent household income in the (weighted) sample. Over the past 15 years, using linearization, formulae and software for computing standard errors of many commonly used poverty indicators have been developed, including those which rely on poverty thresholds estimated from the sample. Several authors have derived formulae for standard errors in the case that the poverty line is estimated as a share of average or median income (e.g. Preston, 1995). Some authors have combined this issue with additional considerations such as stochastic dominance over a range of poverty lines (Davidson and Duclos, 2000), the (complex) sample design (Zheng, 2001), the complex sample design and the influence of raking (the use of weights to balance the sample) (Berger and Skinner, 2003) or the fact that household size should be considered a random variable as well (Thuysbaert, 2008).

[4] The effect is also dependent on the kind of weights. In fact, if the variable of interest correlates strongly with the variables taken into account for calibration, post-stratification may reduce the standard error instead of increasing it. However, in order to take account of this effect, one needs a separate probability and calibration weighting variable, whereas most often only one weighting variable which contains the final weights is available (which is also the case in the EU-SILC UDB).

Recently, macros for the linearization of all Laeken poverty indicators have been published for the statistical software package SAS (Osier, 2009). Importantly, linearization relies on asymptotic assumptions, i.e. assumptions regarding a sufficiently large sample size. For population totals based on samples such as those in EU-SILC, with thousands of households included, there is no problem. However, one should be more careful when the method is applied to relatively small subsamples (cf. Osier, 2009: 170).

Many statistical software packages (e.g. SAS, SPSS, Stata) enable in a user-friendly way the computation of standard errors for proportions and means while taking account of the sample design and weighting. Usually, first the survey design variables must be indicated using a specific command (e.g. the *svyset* command in Stata or *CSPLAN* in SPSS). Thereafter specific commands must be used for estimating means and proportions while taking the survey settings into account (e.g. *svy: mean* in Stata, *CSDESCRIPTIVES* in SPSS or *PROC SURVEYFREQ* in SAS). However, ready-made procedures to compute standard errors of more complex poverty indicators and inequality measures are not included in a standard way. Estimation procedures which take account of the sample character of the poverty line, the complex sample design and weighting have been implemented in the freely available software package DAD (Duclos and Araar, 2006; Araar and Duclos, 2009)[5]. Among others, DAD accommodates inference for the FGT class of poverty measures in the case of a relative (estimated) poverty line, while taking the sample design into account. More recently, most of the modules of DAD have been implemented in the software package Stata under the name DASP (Araar and Duclos, 2007)[6]. However, not all of the Laeken poverty indicators can be estimated using DASP (e.g. the relative median at-risk-of-poverty gap). In that case one could turn to SAS and make use of the macro's published by Osier (2009). Another possibility is to apply the bootstrap method, in which case no formulae for computing the standard error and confidence intervals have to be derived[7].

3. The sample design of EU-SILC: what do we know and what is available?

The European Union Statistics on Income and Living Conditions (EU-SILC) is the EU reference source for information on income and living conditions. The dataset includes internationally and cross-temporary comparable

5 http://132.203.59.36/DAD

6 http://132.203.59.36/DASP/index.html

7 For non-smooth indicators such as many of the Laeken poverty indicators the jackknife is not recommended (e.g. Shao and Chen, 1998: 1071; del Mar Rueda and Muñoz, 2009). In the case of the at-risk-of-poverty indicator (FGT0 and FGT1) the resulting standard errors using the bootstrap are very close to those obtained on the basis of linearisation using the DASP module for Stata (figures available from the author).

variables for all EU Member States and some other countries both at the household and the individual level. Many Laeken indicators – designed to monitor poverty and social inclusion in the EU – are based on EU-SILC (e.g. European Commission, 2006; Marlier et al., 2007).

Member states have some freedom in the selection of the sample design, as long as they are in accordance with common guidelines, concepts and definitions as well as rules with regard to probability selection and minimum effective sample sizes (e.g. European Parliament and Council of the European Communities, 2003). The reference population of EU-SILC consists of "all private households and their current members residing in the territory of the member states at the time of data collection. Persons living in collective households and in institutions are generally excluded from the target population."[8] Currently 31 countries are involved in the EU-SILC process, namely all EU Member States plus the four non-EU members Iceland, Norway, Switzerland and Turkey. Nevertheless, the 2008 cross-sectional UDB offers information on only 27 countries, Eurostat being not allowed by the national authorities of France and Malta to disseminate their micro-data and having not received clean data from Switzerland and Turkey.

Considerable differences between participating countries exist in terms of sample design, sample frame and data source (survey vs. register data) (e.g. Eurostat, 2010a). In some countries, single stage designs are in use, whereas in other countries two- or three-stage designs are used. In some countries (notably Hungary and France) two and three-stage designs are combined, depending on the region (stratum) and panel. Most countries apply stratification on at least one stage. Both sampling with equal probabilities and probabilities proportional to size are in use and in some cases systematic sampling is applied (for a detailed account by country see Annex 1). Additionally, it must be noted that EU-SILC has an important panel component, with a 4-year moving rotational panel design in the great majority of countries. In some countries (e.g. Austria, Hungary, Norway) changes in the method of selection have taken place between waves and/or panels, such that exact variance estimation for the cross-sectional data is very complex or even (close to) impossible. Also with regard to the sample frame important differences exist. Sample frames range from censuses to different kinds of population registers while procedures for updating the sample frame as well as the date of the last update vary. A special (and problematic) case is the German sample frame which consists of households included in the Microzensus and who have indicated that they are willing to participate in other surveys as well. Similarly, part of the Dutch sample frame consists of households who have successfully participated in several waves of the Labour Force Survey (LFS). Last but not least in a number of countries many (income)

[8] Source: http://epp.eurostat.ec.europa.eu/cache/ITY_SDDS/EN/sc010_sm1.htm.

variables in EU-SILC are based on (probably more reliable) register data rather than survey data[9].

Table 1: Number of persons, households and PSUs in the EU-SILC UDB, the Eurostat EU-SILC dataset and the sample design as reported by the national statistical offices

country code	country	persons	households	PSUs in UDB	PSUs in Eurostat data	Reported sample design
AT	Austria	13,631	5,711	5,711	5,711	5,711
BE07	Belgium	15,493	6,348	243	243	275
BE08	Belgium	15,108	6,300	6,300	6,300	275
BG	Bulgaria	12,191	4,344	506	1,415	1,415
CY	Cyprus	10,025	3,355	3,355	3,355	3,355
CZ	Czech Republic	26,933	11,294	2,362	2,364	2,362
DE	Germany	28,904	13,312	13,312	13,312	No information
DK	Denmark	14,836	5,778	5,778	5,778	5,778
EE	Estonia	13,032	4,744	4,744	4,744	4,744
ES	Spain	35,970	13,014	1,994	1,994	2,000
FI	Finland	26,481	10,472	10,472	10,472	10,472
FR07	France	25,907	10,498	9,017	9,017	349
FR08	France	25,510	10,418	FR08 not available	349	349
GR	Greece	16,869	6,504	1,064	1,064	1,056
HU	Hungary	22,363	8,818	4,875	5,245	4,184
IE	Ireland	12,551	5,247	1,723	1,723	1,747
IS	Iceland	8,644	2,887	2,887	2,887	2,887
IT	Italy	52,433	20,928	749	749	912
LT	Lithuania	12,150	4,823	4,823	4,823	4,823
LU	Luxembourg	10,147	3,779	3,779	3,779	3,779
LV	Latvia	13,120	5,166	912	912	930
NL	Netherlands	25,448	10,337	462	462	463
NO	Norway	14,216	5,553	5,553	5,553	5,553
PL	Poland	41,200	13,984	468	5,093	5,912
PT	Portugal	11,786	4,454	541	541	542
RO	Romania	19,131	7,805	779	779	780
SE	Sweden	18,825	7,452	7,452	7,452	7,452
SI	Slovenia	28,958	9,028	774	1,672	2,799
SK	Slovakia	16,546	5,450	5,450	5,450	5,450
UK	United Kingdom	21,043	8,936	1,014	1,014	1,065

Notes: PSUs identified by variable DB060. When DB060 is missing PSUs are identified by household ID or DB062 (HU, see text), except for 81 cases in Latvia where both DB060 and DB050 were missing (81 cases deleted from dataset). The number of PSUs has been counted taking into account stratification by DB040 (UDB), respectively DB050 (Eurostat Dataset) in countries where DB060 is not unique across strata; however in some countries PSUs have been regrouped to avoid splitting of PSUs due to households moving from one region to another (see text).

Source: EU-SILC 2008 (BE, FR: 2007) UDB, the specific dataset prepared by Eurostat, National Intermediate EU-SILC 2008 Quality Reports and personal communication with Eurostat and national statistical offices.

[9] More information on sampling and non-sampling errors in EU-SILC can be found in Verma et al. (2010).

Response rates vary substantially across countries ranging from 95 percent in Romania to 55 percent in Denmark (Eurostat, 2010a: 14). The resulting sample size is smallest in the case of Iceland, Luxembourg, Cyprus, Latvia, Bulgaria, Portugal and Lithuania (below 5,000 households), and largest in the case of the Netherlands, France, Finland, the Czech Republic, Spain, Germany, Poland and Italy (above 10,000 households) (see Table 1). On average, the number of households is 2.6 times lower than the number of persons, although the average household size differs somewhat from country to country. In countries where the sample design consists of several stages, the number of PSUs (first stage clusters) is substantially lower, ranging from 275 in Belgium to nearly 6,000 in Poland. The number of explicit strata at the first stage varies from 1 (no stratification) to over 500 strata in Hungary (see Table 2). Additionally, in many countries with systematic sampling implicit stratification has been applied (meaning that data have been ordered according to several criteria). This is for instance the case for the UK and Norway.

As mentioned earlier, the data in the UDB referring to the sample design are incomplete. In general, the original stratification variable (DB050) is lacking. As a substitute, the variable identifying NUTS1 / NUTS2 regions (DB040) can be used[10]. However, in many countries this variable is way too rough as it strongly underestimates the number of strata (see Table 2). Additionally, some care in its use is advisable. In the case of Spain, the two regions Ceuta and Melilla must be grouped together as is the case in the real sample design. Furthermore, DB040 may not be used in the case of Finland and Sweden, as the Finish sample is stratified by other than geographical criteria and in Sweden no (explicit) stratification is applied at all. The degree to which the rough stratification using DB040 leads to an over-estimation of the standard errors, largely depends on the extent to which it misses important differences between strata within the regions identified by DB040.

[10] EU-SILC contains a variable on the degree of urbanisation (DB100), however the categories of this variable do not correspond to those used for stratification by rural/urban divisions or size of the settlement (personal communication with national statistical offices).

Table 2: Number of strata in the EU-SILC UDB, the Eurostat EU-SILC dataset and the sample design as reported by the national statistical offices

country	Strata in UDB (using DB040)	Strata in Eurostat Dataset (using DB050)	Reported sample design
AT	3	247	247
BE07	3	11	11
BE08	3	11	11
BG	2	56	56
CY	1	9	9
CZ	8	53	53
DE	1	1	?
DK	1	1	1
EE	1	3	3
ES	18	93	93
FI	1	26	26
FR07	22	22	86
FR08	22	87	86
GR	4	90	90
HU	3	526	529
IE	1	138	138
IS	1	1	1
IT	5	288	288
LT	1	7	7
LU	1	160	160
LV	1	4	4
NL	1	40	40
NO	1	1	1
PL	6	211	211
PT	1	7	7
RO	8	88	88
SE	1	1	1
SI	1	6	6
SK	1	48	48
UK	1	31	30

Source: EU-SILC 2008 (BE, FR: 2007) UDB, the specific dataset prepared by Eurostat, National Intermediate EU-SILC 2008 Quality Reports and personal communication with Eurostat as well as national statistical offices.

In the case of the identification of PSUs (variable DB060), the EU-SILC UDB provides information for most countries. However, this is not the case for Austria and Finland (sampling of dwellings although only households can be identified)[11], Belgium (2008 data does not contain variable DB060), Germany (where the first stage corresponds to the first stage of the Mikrozensus) [12], 'old panels' in French (2007) data and part of the Hungarian EU-SILC. In Latvia, DB060 is missing only for a limited

[11] This is only a problem to the extent that dwellings are occupied by several households at the same time. However, there is no information in the national quality reports on this issue (National Intermediate EU-SILC 2008 Quality reports; cf. Verma et al., 2010: 49).

[12] At the moment of writing it is not entirely clear whether or not this also applies to the Netherlands.

number of cases[13]. As is argued in the next section, the best way to proceed is to assume a sampling of households where the PSU identifier is lacking, except for Hungary, where in many cases secondary sampling unit identification numbers (DB062) can be used. An additional problem is that in a number of countries (Bulgaria, Hungary, Poland and Slovenia) DB060 is not unique across strata. In other words, in the absence of DB050 which identifies (many) more strata, the number of PSUs is underestimated and PSUs may be grouped together while they belong to different strata. Especially in these cases it is impossible to predict which variables are best used to estimate standard errors (just household ID or DB060 and DB040 or even another combination?). Furthermore, from a theoretical point of view it is impossible to say in which direction the bias would go. In order to gain more insight into the bias on the standard errors, results in this paper are compared to those obtained using EU-SILC data prepared by Eurostat (from here 'Eurostat EU-SILC'). In contrast to the EU-SILC UDB, 'Eurostat EU-SILC' contains the variable which identifies primary strata (DB050), and hence also allows for the correct identification of the PSUs in countries where PSU numbers are not unique across strata.

The use of DB040 as a stratification variable poses also another challenge: DB040 refers to the region of residence at the date of the interview, while for the computation of the standard errors the region of residence at the time of the sample selection is needed. If people move from one region to another (which happens in a non-trivial number of cases given the panel component of EU-SILC), original PSUs are split across regions, inflating the number of PSUs (this is especially the case for Belgium, France, Italy, Romania (only 1 case) and Spain. However, it could also be the case in countries where PSU numbers are not unique across strata (Bulgaria, Hungary, Poland and Slovenia). Therefore, it is crucial to discern between the two reasons for split PSUs and to re-group PSUs into the right stratum (region) only where necessary.

In comparison with the UDB, the 'Eurostat EU-SILC' offers much more detailed information. Unfortunately, also these data are not free of errors and remain incomplete for some countries. Although the correct number of strata is available, the region at the moment of interview (DB040) must be used to incorporate stratification in Belgium (2007) as well as in France and Spain (in the latter two countries in combination with the stratification variable DB050). Furthermore, it appears that in some countries DB060 has not been correctly coded, as even with full information on stratification, the number of PSUs is different from that reported by the national statistical institutes (this especially the case for Belgium, Hungary, Italy, Poland and Slovenia, and to some degree also for several other countries). Stratification poses less problems, except for the

[13] In order to prevent variance estimation problems, these cases have been deleted.

(partial) absence of the stratification variable in Belgium (2008) and France (2007); and some possible errors in France, Hungary and the UK[14].

When working with the 'Eurostat EU-SILC' dataset, an additional problem must be tackled. In order to be able to estimate sampling variance it is necessary that each stratum contains at least two PSUs, otherwise it is impossible to obtain an idea of the within-stratum variance. However, when working with the Eurostat EU-SILC dataset, in several countries there are strata containing only one PSU (Austria, Spain, France (2008), Hungary, Ireland, Italy, Luxembourg and the UK). In some countries, this could be because some PSUs are autorepresentative, i.e. they have a probability of 1 of being selected. In other words, these PSUs are rather strata than PSUs. In these cases the next stage of the sample design contains the real PSUs. Unfortunately it is not well documented for which countries and strata this is the case. In other countries, some strata really contain only one PSU because only one PSU has been selected or only one PSU contains respondents. Several methods for dealing with this situation are available. Generally it is recommended to join the stratum to another stratum which is as much as possible similar with regard to the variables of interest (cf. Eurostat, 2002: 51-52). In this paper, all strata which contain one PSU have been re-grouped on the basis of (geographical) proximity and average equivalent net disposable household income. Table 3 shows the number of strata containing only one PSU and the number of strata after joining these strata to a similar stratum.

Table 3: Number of strata containing one PSU in the EU-SILC dataset available to Eurostat and number of PSUs and strata after joining similar strata

country	Number of strata with one PSU	Before re-grouping		After re-grouping	
		PSUs	Strata	PSUs	Strata
AT	3	5,711	247	5,711	244
ES	1	1,994	93	1,994	92
FR08	15	349	87	349	74
HU	66	5,639	526	5,639	470
IE	21	1,723	138	1,723	118
IT	110	749	288	749	210
LU	1	3,779	160	3,779	159
UK	1	1,014	31	1,014	30

Source: 'Eurostat EU-SILC' dataset for the 2007 and 2008 (France) operations.

[14] For an in-depth discussion on the quality of the sample design variables in the 'EU-SILC Eurostat' dataset, see Goedemé (2010).

4. An illustration: the Europe 2020 poverty reduction targets

In section two it has been argued that taking account of the sample design is of crucial importance for estimating standard errors accurately. However, as has been shown in section three, in many cases adequate information with regard to the sample design is lacking in the EU-SILC UDB. In this section we illustrate the importance of various assumptions with regard to the sample design, always taking account of weighting, but ignoring the effect of imputation. Additionally, the standard errors based on the EU-SILC UDB will be compared to those obtained using more complete information available in the 'Eurostat EU-SILC' dataset. In section five, this information will be used to discuss a workable solution to variance estimation for EU-SILC, which is practical in implementation for a wide array of researchers and which leads to estimates as accurate as possible. Results will be compared on the basis of the three indicators that form the Headline Target on social inclusion of the Europe 2020 Strategy. In this section we will first discuss the three indicators. Thereafter we explain the setup of the statistical tests. In the last part the results are presented.

4.1. The indicators

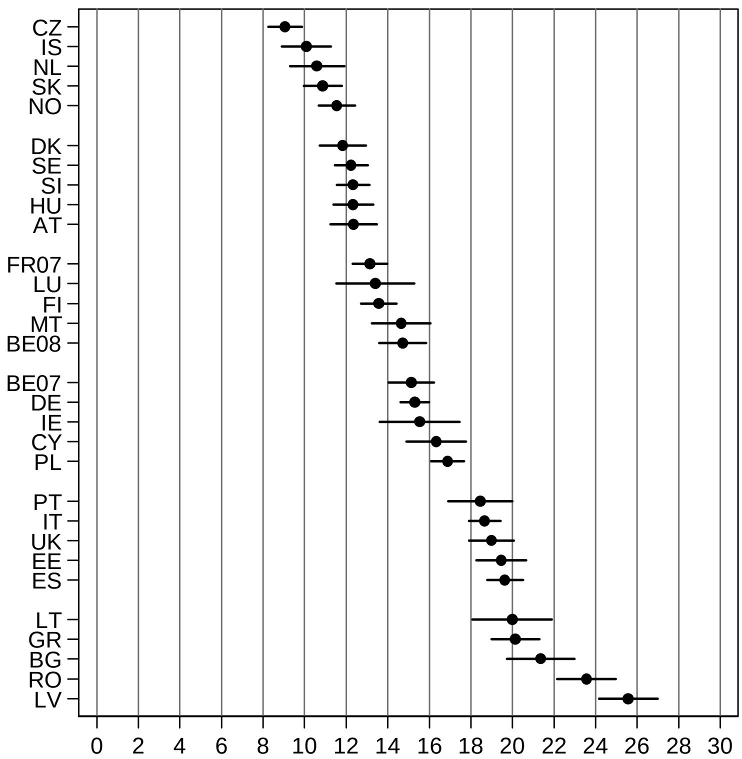
In June 2010 the European Council agreed on lifting 20 million European citizens out of poverty and social exclusion (European Council, 2010). The targeted population consists of persons facing at least one of three following situations: being at-risk-of-poverty , being severely materially deprived, living in a household with very low work intensity.

Being at-risk-of-poverty means living in a household with an equivalised net disposable household income below 60 percent of the national median. Household income is equivalised using the modified OECD equivalence scale which attaches a weight of 1 to the first adult, a weight of 0.5 to all other household members aged 14 and over and a weight of 0.3 to household members aged less than 14. The equivalised household income is obtained by dividing total household income by the sum of the individual equivalence weights. It is assumed that all household members have the same living standard as they all receive the same equivalised household income (e.g. Atkinson et al., 2002; Marlier et al., 2007). In line with Eurostat practice, no top-bottom coding of income has been applied[15]. It must be stressed that in all countries, except Ireland and the United Kingdom, the income reference period is equal to the year preceding the survey year. Severe material deprivation is measured by an index of nine items relating to financial stress and the enforced lack of some durables. All persons living in a household which at the moment of

[15] In most countries many ways of top-bottom coding would not make a big difference: neither for the estimated number of poor (cf. Van Kerm, 2007), nor for the estimated standard errors (figures available from the author).

the interview lacks at least 4 out of 9 items are considered severely materially deprived. The list of items as well as the threshold is the same across all EU Member States (cf. Guio, 2009; Wolff, 2010)[16]. The third situation relates to the work intensity of the household. It is calculated by adding up the total number of months all household members at working age have worked during the income reference period, expressed in full-time equivalents. This is divided by the total number of months they could have worked. If the ratio is below 0.20, the household is considered having a low work intensity. The share of people facing each of these three situations is respectively recorded in the following indicators: the at-risk-of-poverty rate, the severe material deprivation rate and the share of people living in households with very low work intensity.

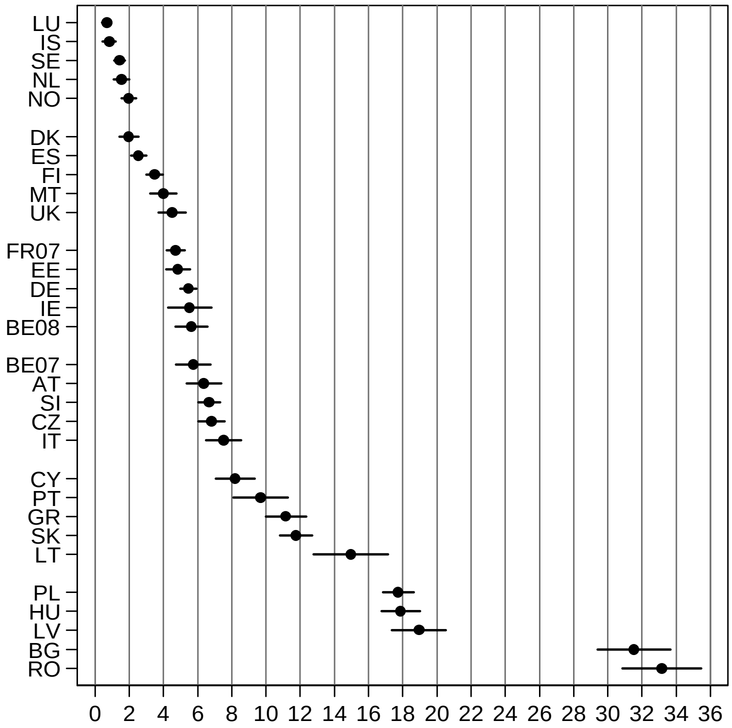
Figure 1: At-risk-of-poverty rate with 95% confidence interval



Source: 'Eurostat EU-SILC' 2008 (BE, FR: 2007), own calculations.

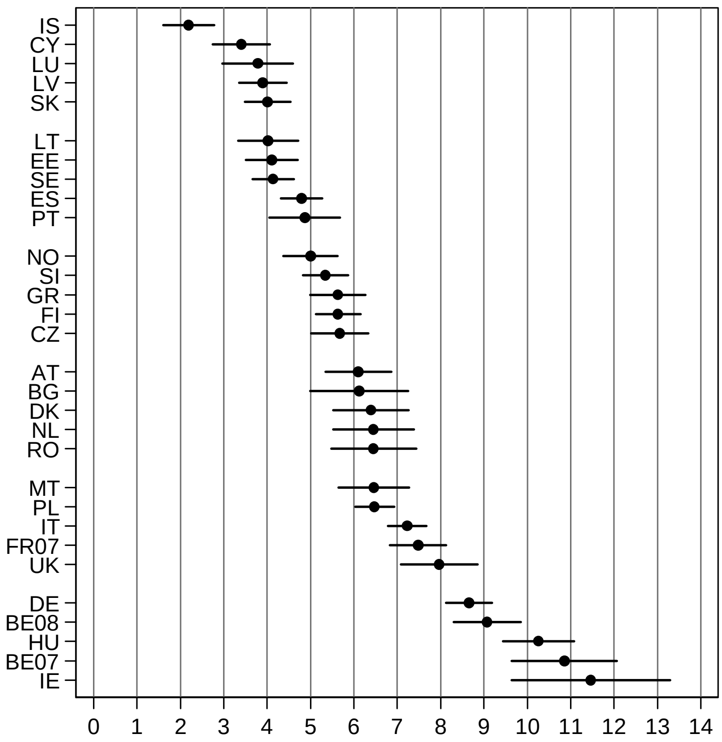
[16] For a theoretical discussion about the validity of the latter two indicators, see Goedemé and Rottiers (2010).

Figure 2: Severe material deprivation rate with 95% confidence interval



Source: 'Eurostat EU-SILC' 2008 (BE, FR: 2007), own calculations.

Figure 3: Share of population living in household with very low work intensity with 95% confidence interval



Source: 'Eurostat EU-SILC' 2008 (BE, FR: 2007), own calculations.

Figures 1-3 depict the point estimates for all three indicators. The at-risk-of-poverty rate ranges from 9 per cent in the Czech Republic to 26 per cent in Latvia. Compared to the at-risk-of-poverty indicator, the differences between European countries are much larger in the case of the severe material deprivation indicator. In Luxembourg and Iceland less than 1 per cent of the population is estimated to be severely materially deprived compared to 33 per cent in Romania. The variation in estimates for the population living in a household with a very low work intensity is somewhat in between, ranging from less than 4 per cent in Iceland, Luxemburg, Cyprus and Latvia to over 10 per cent in Hungary, Belgium (2007) and Ireland. Overall, the country rankings differ much between the indicators. Rankings are most equal in the case of the at-risk-of-poverty and the severe material deprivation rates (Spearman's rank correlation coefficient of about 0.5). The precision of the estimates (here making use of 'Eurostat EU-SILC') strongly depends on the indicator. The width of the 95% confidence intervals ranges between 1.4 and 3.8 percentage points in the case of the at-risk-of-poverty rate, between 0.5 and 4.6 percentage points in the case of the deprivation rate and between 0.9 and 3.6 percentage points in the case of the low work intensity indicator.

The broader usefulness of the comparison of standard errors that will follow depends greatly on the extent to which other variables are correlated in the same way within PSUs and strata as those presented here. Therefore, the weaker the correlation between the three indicators at the micro level, the stronger will be the conclusions if they are the same for all three indicators. Table 4 shows the correlation at the micro level of the underlying variables of all three indicators: equivalised household income, the number of deprived items and the work intensity of the household[17]. Overall correlations are rather weak (but highly significant): for all three possible combinations they range in absolute values between 0.10 and 0.52. As has been observed by Whelan and Maître (2007), the correlation between equivalent household income and deprivation is strongest in the Eastern and Southern European EU member states and weakest in the Nordic countries and the UK. A broadly similar picture arises when we look at the correlation between equivalent household income and the work intensity of the household, although the distinction between South-East and North-West Europe is less clear-cut. However, the country ranking is different and more mixed in the case of the correlation between deprivation and the work intensity of the household. Nevertheless, we can conclude that in most countries the

[17] The results for the binary Europe 2020 poverty indicators are somewhat different in terms of country-groupings. As far as the strength of the correlations is concerned, correlations between the risk of poverty and deprivation as well as between deprivation and low work intensity tend to be weaker and between the risk of poverty and low work intensity somewhat stronger than the correlations obtained for the underlying variables (figures can be obtained from the author). We show the results for the underlying variables, as they could tell also something about the correlation when the thresholds are set differently (e.g. at 50 or 60 per cent of equivalent household income, 3 instead of 4 deprivation items or 0.25 instead of 0.20 of work intensity of the household.

correlation at the individual level between the variables is relatively weak (with a possible exception of Bulgaria). Given the different nature of all three indicators in relation to living conditions and welfare, there is a reasonable chance that other variables in this domain will be correlated to at least one of these three indicators. Therefore, if similar conclusions can be drawn for all three indicators with respect to the effect of the assumed sample design on the standard error, there is a reasonable chance, but no guarantee, that the same conclusions would also apply to other variables.

Table 4: Correlation at the individual level of the underlying variables.

	Income - deprivation		**Income - work intensity**		**Deprivation - work intensity**	
AT	-0.29	***	0.29	**	-0.34	****
BE07	-0.36	****	0.41	****	-0.46	****
BE08	-0.27	***	0.27	**	-0.44	****
BG	-0.52	****	0.39	****	-0.48	****
CY	-0.35	****	0.19	*	-0.23	****
CZ	-0.35	****	0.31	****	-0.36	****
DE	-0.32	****	0.30	****	-0.36	****
DK	-0.19	***	0.17	**	-0.40	****
EE	-0.41	****	0.35	****	-0.33	****
ES	-0.34	****	0.40	****	-0.26	****
FI	-0.29	****	0.30	****	-0.37	****
FR07	-0.34	****	0.28	****	-0.34	****
GR	-0.41	****	0.29	****	-0.26	****
HU	-0.41	****	0.39	****	-0.35	****
IE	-0.33	****	0.35	***	-0.48	****
IS	-0.21	****	0.11	***	-0.21	****
IT	-0.36	****	0.36	****	-0.29	****
LT	-0.38	****	0.34	****	-0.36	****
LU	-0.31	****	0.27	****	-0.13	****
LV	-0.40	****	0.31	****	-0.33	****
NL	-0.23	****	0.21	****	-0.32	****
NO	-0.17	**	0.21	**	-0.32	****
PL	-0.38	****	0.29	****	-0.34	****
PT	-0.42	****	0.25	****	-0.26	****
RO	-0.46	****	0.33	****	-0.26	****
SE	-0.24	****	0.30	****	-0.33	****
SI	-0.40	****	0.35	****	-0.29	****
SK	-0.38	****	0.32	****	-0.32	****
UK	-0.20	**	0.22	*	-0.42	****
Total	-0.35	****	0.22	****	-0.32	****

* p<0.05, **p<0.01, ***p<0.001, ****p<0.0001. P-values computed taking account as much as possible of the sample design using the EU-SILC dataset available to Eurostat. Standard errors have been computed by regressing the variables on each other and taking the highest p-value of the slope as a conservative estimate of the significance level of the correlation[18].

Source: 'Eurostat EU-SILC' 2008 (BE, FR: 2007), own calculations.

[18] See Sribney (2005) http://www.stata.com/support/faqs/stat/survey.html (last accessed on 18/10/2010).

4.2. Setup of the statistical tests

The aim of this section is twofold. The first aim is to show the importance of taking account of the sample design when estimating standard errors. The second is to find out which information in the EU-SILC UDB is best used to take the sample design as much as possible into account. In order to do so, estimated standard errors will be presented for four different scenarios, each containing a different set of assumptions with regard to the sample design: assuming (1) a simple random sample of individuals; (2) a simple random sample of households, (3) a complex sample involving stratification and clustering as can be identified in the EU-SILC UDB (using DB060 and DB040, as explained in section three); (4) a complex sample involving stratification and clustering as can be identified in the 'Eurostat EU-SILC' dataset (using DB060 and DB050 as explained in section three). In all four scenarios it is assumed that PSUs are sampled with replacement and the effect of imputation is ignored, but account is taken of weighting. In the case of the at-risk-of-poverty rate, the fact that the poverty line is estimated from the data is taken into account. The estimation procedure is based on linearization, using the DASP module and the generic commands for the robust estimation of proportions in Stata[19].

All three indicators focus on the proportion of persons living in a poor, deprived and/or workless household. In other words, although all three indicators are analysed at the individual level, they are measured at the household level. Therefore, at least clustering at the household level should be taken into account when computing standard errors: household members do not form independent units of observation, but are clustered within households (cf. Biewen and Jenkins, 2006; Verma et al., 2010: 49). In other words, the least one should do when estimating standard errors of the Europe 2020 poverty indicators at the individual level, is taking clustering within households into account. In order to show the importance of this issue, standard errors assuming a sample of households (scenario 2) will be compared to assuming a simple random sample of individuals (scenario 1).

In some countries, the simple random sample of households is relatively close to the real sample design. This is most notably the case for Denmark, Iceland, Norway and Sweden. In countries for which the variable containing the PSU identification numbers are lacking, assuming a random sample of households is probably the best way forward, even if we know that clustering on a higher level has taken place. This is the case for the Belgian (2008) and German data as well as a part of the French (2007) and Hungarian data (to some extent this also applies to Austria and Finland).

[19] Figures obtained by applying a bootstrap for the at-risk-of-poverty indicator reveal similar results (figures available from the author).

In a third scenario, all possible information regarding clustering and stratification which is available in the EU-SILC UDB is used. This means that DB060 (containing the PSUs) is used where available and household ID where DB060 is not available or not applicable (DB062 in the case of Hungary). Furthermore, the region variable (DB040) functions as stratification variable, as described in section three. For some countries the assumed sample design is rather close to the real one, although the number of strata is underestimated. This is the case in countries with a sample of households such as Cyprus, Estonia, Lithuania, Luxembourg and Slovakia and countries with a multi-stage sample design such as the Czech Republic, Spain, Greece, Latvia, the Netherlands, Portugal and Romania. In other countries both the number of PSUs and the number of strata are underestimated (Belgium, Bulgaria, Ireland, Italy, Poland, Slovenia, the UK). In Hungary and France the number of PSUs is overestimated and the number of strata underestimated given the fact that DB060 is missing in many cases.

In a fourth scenario, standard errors are estimated using the 'Eurostat EU-SILC' dataset. Especially in countries where both the identification of PSUs and of strata is problematic, the fourth scenario is much closer to the real sample design. In this scenario, standard errors are computed using DB060 with the PSUs and the household number where relevant. However, in this scenario full information on stratification (DB050) has been taken into account. For most countries this results in a number of strata and PSUs equal to those reported in the national quality reports. However, as explained in section three, in a number of countries some important differences remain. In the case of strata containing one PSU, similar strata have been grouped in order to have at least two PSUs in each stratum (see section three).

It should be noted that – even in the fourth scenario – in countries with systematic sampling, standard errors probably are somewhat overestimated due to the neglect of implicit stratification. In these cases, special procedures for estimating standard errors could be applied taking account of the order of selection (various methods to deal with systematic sampling are explained by Kish, 1965: 117-120; see also Wolter, 2007). However, the aim of this paper is to come to some 'workable' method for variance estimation. Only for three (Latvia, Slovenia and the UK) out of eight countries (additionally including Estonia, Italy, the Netherlands, Norway, Sweden) in which systematic sampling is applied at the first stage, data on the order of selection are available. Therefore, the most straightforward procedure is assuming simple random sampling (even in Latvia, Slovenia and the UK) as is common practice (cf. Eurostat, 2002: 14). Furthermore, the random error induced by imputation has been ignored, which probably leads to an under-estimation of standard errors (see section two).

4.3. Results

In the previous section it has been argued that when household variables are analysed at the individual level, researchers should take account of clustering at the household level. Results presented in tables 5 to 7 indeed show that if clustering at the household level is ignored, in the case of all three indicators standard errors are seriously underestimated. On average, the standard error which takes account of clustering at the household level is about 70 per cent larger than the standard error which assumes a simple random sample of individuals. Usually, the effect is strongest in the case of the at-risk-of-poverty indicator and weakest in the case of the work intensity indicator. Clustering at the household level has the strongest effect in Romania (standard errors double) and the weakest effect in Denmark (standard errors are around 30 per cent larger).

Table 5: A comparison of estimated standard errors of the at-risk-of-poverty rate, EU-SILC 2008 (BE, FR: 2007)

scenario	point estimate	Standard error				Ratio of standard errors			
		persons	households	UDB	Eurostat data	households / persons	UDB / households	Households / Eurostat data	UDB / Eurostat data
		(1)	(2)	(3)	(4)	(2) / (1)	(3) / (2)	(2) / (4)	(3) / (4)
AT	12.36	0.33	0.57	0.57	0.57	1.76	1.00	1.00	1.00
BE07	15.13	0.30	0.53	0.56	0.55	1.76	1.06	0.96	1.01
BE08	14.72	0.32	0.58	0.57	n/a	1.80	1.00	n/a	n/a
BG	21.36	0.40	0.78	0.93	0.83	1.94	1.19	0.94	1.13
CY	16.34	0.40	0.73	0.73	0.73	1.83	1.00	1.00	1.00
CZ	9.06	0.22	0.39	0.42	0.41	1.81	1.06	0.95	1.00
DE	15.29	0.22	0.35	0.35	n/a	1.58	1.00	n/a	n/a
DK	11.84	0.42	0.57	0.57	n/a	1.36	1.00	n/a	n/a
EE	19.46	0.40	0.61	0.61	0.61	1.55	1.00	1.00	1.00
ES	19.65	0.24	0.43	0.44	0.44	1.83	1.02	0.98	1.00
FI	13.56	0.30	0.45	0.45	0.44	1.50	1.00	1.02	1.02
FR07	13.15	0.25	0.43	0.42	n/a	1.71	0.99	n/a	n/a
GR	20.14	0.32	0.59	0.59	0.59	1.81	1.01	1.00	1.01
HU	12.34	0.25	0.47	0.49	0.49	1.90	1.06	0.95	1.00
IE	15.53	0.48	0.96	0.98	0.98	1.98	1.02	0.98	1.01
IS	10.09	0.38	0.60	0.60	n/a	1.57	1.00	n/a	n/a
IT	18.67	0.20	0.36	0.39	0.39	1.79	1.09	0.94	1.03
LT	19.99	0.52	0.97	0.97	0.97	1.87	1.00	1.00	1.00
LU	13.40	0.53	0.95	0.95	0.95	1.80	1.00	1.00	1.00
LV	25.57	0.37	0.71	0.72	0.72	1.90	1.02	0.98	1.00
NL	10.59	0.35	0.54	0.66	0.66	1.54	1.21	0.82	0.99
NO	11.55	0.32	0.45	0.45	n/a	1.39	1.00	n/a	n/a
PL	16.88	0.20	0.39	0.40	0.40	1.98	1.01	0.98	0.99
PT	18.45	0.40	0.75	0.78	0.78	1.86	1.04	0.96	1.00
RO	23.57	0.32	0.69	0.74	0.72	2.15	1.07	0.96	1.03
SE	12.25	0.27	0.41	0.41	n/a	1.48	1.00	n/a	n/a
SI	12.33	0.23	0.38	0.37	0.40	1.63	0.98	0.96	0.94
SK	10.87	0.24	0.47	0.47	0.46	1.92	1.00	1.00	1.00
UK	19.00	0.29	0.54	0.55	0.55	1.86	1.02	0.98	1.00

n/a: 'Eurostat EU-SILC' dataset does not offer additional information. Standard errors based on linearization using the DASP module for Stata. The fact that the poverty line is estimated from the data has been taken into account.

Source: EU-SILC 2008 (BE, FR: 2007); own calculations.

In 18 countries, the third scenario is different from the second. For these countries it is possible in the EU-SILC UDB to take to some extent account of stratification and there is information on the PSUs. The effect of accounting for this information in comparison to scenario two depends on the indicator. On average, it is strongest in the case of the severe material deprivation rate and weakest in the case of the at-risk-of-poverty rate. Compared to the standard errors which take account of clustering at the household level, standard errors in these countries increase on average with another 5 percent in the case of the at-risk-of-poverty rate, 10 per cent in the case of the work intensity indicator and 18 per cent in the case of the deprivation indicator. In other words, differences between households seem to absorb most of the variance. However, depending on the indicator, for some countries the additional effect is relatively strong. This is most notably the case for Belgium (2007), Italy, Portugal and Romania in the case of the deprivation indicator and Belgium (2007) and the Netherlands in the case of the work intensity indicator (increases of at least 30 per cent). In other words, even though on average clustering at the household level accounts for most of the variance, taking as much as possible account of the sample design cannot be ignored for individual countries.

Table 6: A comparison of estimated standard errors of the severe material deprivation rate, EU-SILC 2008 (BE, FR: 2007)

scenario		Standard error				ratio of standard error			
	point estimate	persons	households	UDB	Eurostat data	households / persons	UDB / households	households / Eurostat data	UDB / Eurostat data
		(1)	(2)	(3)	(4)	(2) / (1)	(3) / (2)	(2) / (4)	(3) / (4)
AT	6.36	0.27	0.52	0.52	0.52	1.90	1.00	1.01	1.01
BE07	5.74	0.21	0.34	0.52	0.52	1.68	1.51	0.67	1.01
BE08	5.64	0.23	0.48	0.47	n/a	2.05	0.99	n/a	n/a
BG	31.53	0.49	0.94	1.08	1.08	1.92	1.15	0.87	1.00
CY	8.20	0.32	0.58	0.58	0.58	1.80	1.00	1.00	1.00
CZ	6.81	0.19	0.34	0.39	0.39	1.77	1.15	0.87	1.00
DE	5.46	0.16	0.24	0.24	n/a	1.52	1.00	n/a	n/a
DK	1.97	0.21	0.28	0.28	n/a	1.36	1.00	n/a	n/a
EE	4.85	0.23	0.35	0.35	0.35	1.50	1.00	1.00	1.00
ES	2.55	0.11	0.20	0.23	0.22	1.80	1.13	0.89	1.01
FI	3.47	0.18	0.25	0.25	0.25	1.36	1.00	1.02	1.02
FR07	4.71	0.17	0.27	0.27	n/a	1.59	1.00	n/a	n/a
GR	11.17	0.30	0.54	0.61	0.60	1.79	1.13	0.89	1.01
HU	17.89	0.29	0.56	0.62	0.57	1.89	1.12	0.97	1.09
IE	5.53	0.32	0.59	0.65	0.64	1.83	1.10	0.92	1.01
IS	0.82	0.13	0.20	0.20	n/a	1.48	1.00	n/a	n/a
IT	7.53	0.16	0.30	0.55	0.52	1.87	1.85	0.58	1.06
LT	14.97	0.59	1.11	1.11	1.11	1.90	1.00	1.00	1.00
LU	0.68	0.08	0.13	0.13	0.13	1.58	1.00	1.00	1.00
LV	18.95	0.38	0.69	0.81	0.80	1.84	1.16	0.86	1.00
NL	1.55	0.14	0.23	0.23	0.23	1.65	1.03	0.99	1.02
NO	1.96	0.15	0.22	0.22	n/a	1.43	1.00	n/a	n/a
PL	17.75	0.22	0.42	0.44	0.46	1.95	1.04	0.92	0.96
PT	9.69	0.32	0.62	0.81	0.81	1.90	1.32	0.76	1.00
RO	33.16	0.42	0.90	1.22	1.17	2.12	1.35	0.77	1.04
SE	1.44	0.10	0.16	0.16	n/a	1.49	1.00	n/a	n/a
SI	6.67	0.19	0.32	0.33	0.33	1.66	1.03	0.96	1.00
SK	11.76	0.26	0.48	0.48	0.48	1.89	1.00	1.01	1.01
UK	4.50	0.19	0.39	0.41	0.41	2.03	1.05	0.95	1.01

n/a: Eurostat database does not offer additional information. Standard errors based on linearization using Stata.

Source: EU-SILC 2008 (BE, FR: 2007); own calculations.

The information contained in the UDB and which is used in the third scenario is incomplete. Therefore, it is important to gain insight into the degree of bias on the estimated standard errors. In order to find this out, results can be compared to those obtained by using more complete information on stratification and clustering as available in 'Eurostat EU-SILC' dataset (the fourth scenario). In the case of four countries, the UDB contains full information on the sample design, as the sample design consists of a simple random sample of households (Denmark Iceland, Norway and Sweden), so standard errors can be computed accurately directly from the UDB. Also in the case of Belgium (2008), France (2007) and Germany no information can be added using 'Eurostat EU-SILC', as both the UDB and the 'Eurostat EU-SILC' dataset do not contain the necessary information on the sample design.

In most countries, the difference between third and fourth scenario consists of applying a more detailed stratification. In seven countries this takes place in a context of a sample of households (Austria, Cyprus, Estonia, Finland, Lithuania, Luxembourg and Slovakia) and in 11 countries in a context of a clustered sample (Belgium, the Czech Republic, Spain, Greece, Ireland, Italy, Latvia, the Netherlands, Portugal, Romania and the United Kingdom). The impact of adding stratification is relatively limited. As a result, for all three indicators, independent of the level of clustering the difference between the UDB estimates (scenario three) and estimates based on the Eurostat data (scenario four) is trivial (less than 2 per cent difference). The difference is somewhat larger in Belgium (2007) and Finland (job intensity) as well as Romania (deprivation), but still below 10 per cent. In almost all countries, for all three indicators, the standard errors in scenario three are closer to those obtained using 'Eurostat EU-SILC' (scenario four) than the standard errors in scenario two (assuming a random sample of households). Within countries, the exceptions usually relate to only one indicator. Furthermore, the loss in precision in the case of the exception is always smaller than the gain in precision in the case of the other indicators. As a result, making as much as possible use of the information on the sample design in the EU-SILC UDB as is done in scenario three is highly recommended.

In four countries (Bulgaria, Hungary, Poland and Slovenia) both the number of strata and the number of PSUs increase when going from scenario three to scenario four. In addition, the French 2008 has been used to simulate the situation for the French 2007 data[20]. For all five countries, the difference between UDB standard errors and Eurostat data standard errors tends to be larger. In Bulgaria and Hungary standard errors in scenario three overestimate the standard error compared to those obtained using the 'Eurostat EU-SILC' dataset. Nevertheless, scenario three outperforms scenario two for the Bulgarian data. However, this is not the case for Hungary. Scenario three can be recommended for the at-risk-of-poverty indicator, but not necessarily for the other two indicators. For the latter two indicators one could also consider assuming

[20] More precisely, in the third scenario DB060 has been used only for cases belonging to the newest rotational panel since in the EU-SILC 2007 UDB, DB060 is only available for the newest rotational panel. However, for scenario four all information has been used, including DB060 for all cases.

a random sample of households given the small difference between the standard errors in this scenario and the scenario using the 'Eurostat EU-SILC' dataset. However, it should be borne in mind that in the case of Hungary also the 'Eurostat EU-SILC' dataset is incomplete, necessitating further research. Furthermore, it is probably more prudent to opt for a more conservative estimate of the standard error. In the case of Poland, Slovenia and France (2008) standard errors using the UDB (third scenario) tend to underestimate the standard errors in comparison with the standard errors based on the Eurostat dataset – especially in the case of the at-risk-of-poverty rate and the deprivation rate in France. Nevertheless, in all three countries scenario three outperforms scenario two, except for the at-risk-of-poverty indicator in Slovenia (where the difference between both scenarios is small).

Table 7: A comparison of estimated standard errors of the share of people living in household with very low work intensity EU-SILC 2008 (BE, FR: 2007)

scenario		Standard error				ratio of standard error			
	point estimate	persons	households	UDB	Eurostat data	households / persons	UDB / households	households / Eurostat data	UDB / Eurostat data
		(1)	(2)	(3)	(4)	(2) / (1)	(3) / (2)	(2) / (4)	(3) / (4)
AT	6.10	0.26	0.39	0.39	0.39	1.50	1.00	1.01	1.01
BE07	10.85	0.28	0.49	0.64	0.62	1.76	1.32	0.79	1.04
BE08	9.08	0.26	0.40	0.40	n/a	1.55	0.99	n/a	n/a
BG	6.12	0.25	0.49	0.61	0.58	1.96	1.25	0.85	1.06
CY	3.41	0.21	0.34	0.34	0.33	1.62	1.00	1.00	1.00
CZ	5.68	0.19	0.32	0.34	0.34	1.70	1.05	0.95	1.00
DE	8.66	0.19	0.27	0.27	n/a	1.42	1.00	n/a	n/a
DK	6.39	0.37	0.44	0.44	n/a	1.22	1.00	n/a	n/a
EE	4.11	0.22	0.31	0.31	0.31	1.36	1.00	1.00	1.00
ES	4.80	0.14	0.23	0.24	0.24	1.57	1.08	0.93	1.01
FI	5.64	0.21	0.28	0.28	0.26	1.35	1.00	1.06	1.06
FR07	7.48	0.21	0.33	0.33	n/a	1.59	1.00	n/a	n/a
GR	5.63	0.22	0.31	0.32	0.32	1.41	1.03	0.96	0.99
HU	10.26	0.23	0.41	0.46	0.42	1.77	1.12	0.98	1.10
IE	11.47	0.44	0.86	0.93	0.93	1.96	1.08	0.92	1.00
IS	2.19	0.21	0.30	0.30	n/a	1.41	1.00	n/a	n/a
IT	7.23	0.14	0.22	0.26	0.23	1.57	1.16	0.99	1.14
LT	4.02	0.23	0.35	0.35	0.35	1.54	1.00	1.00	1.00
LU	3.78	0.29	0.42	0.42	0.41	1.46	1.00	1.01	1.01
LV	3.90	0.18	0.27	0.28	0.28	1.48	1.03	0.98	1.01
NL	6.45	0.27	0.36	0.48	0.47	1.34	1.32	0.76	1.01
NO	5.00	0.23	0.32	0.32	n/a	1.37	1.00	n/a	n/a
PL	6.48	0.14	0.23	0.23	0.23	1.68	1.00	0.99	0.99
PT	4.87	0.23	0.39	0.41	0.41	1.68	1.06	0.94	1.00
RO	6.46	0.21	0.41	0.51	0.50	1.95	1.23	0.83	1.01
SE	4.14	0.17	0.24	0.24	n/a	1.41	1.00	n/a	n/a
SI	5.35	0.18	0.26	0.26	0.26	1.44	1.01	0.99	0.99
SK	4.01	0.16	0.27	0.27	0.27	1.71	1.00	1.00	1.00
UK	7.97	0.23	0.44	0.46	0.45	1.91	1.03	0.98	1.01

n/a: Eurostat database does not offer additional information. Standard errors based on linearization using Stata.

Source: EU-SILC 2008 (BE, FR: 2007); own calculations.

5. Discussion: recommended sample design

Formulating a recommendation always runs the risk that following the recommendation does not always result in the best solution. This is not different in the case of taking account as much as possible of the sample design in EU-SILC. Nevertheless, we believe that the results presented in the previous section allow for some recommendations. First of all, it should be stressed that in all circumstances one should take account of clustering within households when household variables are analysed at the individual level. Otherwise, standard errors are severely underestimated, regardless of the sample design. This really makes a difference. As an example the number of non-significant country-by-country comparisons could be compared in the case of a 95% confidence interval around the at-risk-of-poverty indicator. If a sample of individuals is assumed (scenario 1) 69 out of 406 country-by-country comparisons are not significant. In contrast, if a sample of households is assumed (scenario 2), the number of *non*-significant differences amounts to 112 (an increase of over 60%). The number of non-significant country-by-country differences in the case of the deprivation indicator amounts to 31 (scenario 1) instead of 64 (scenario 2) and in the case of the work intensity indicator it amounts to 87 (scenario 1) instead of 133 (scenario 2).

Second, where applicable the use of all available information in the UDB (i.e. using the region variable DB040 for stratification and the variable DB060 for identifying the PSUs) as applied in the third scenario outperforms in most cases the assumption of a simple random sample of households. Additionally, in comparison with scenario two in many cases it offers more conservative estimates of the standard error. The latter is not always the optimal choice, but from the perspective of scientific prudence it surely is. Furthermore, except for Bulgaria (at-risk-of-poverty indicator) the over-estimation is limited and in any case outweighs the under-estimation if a sample of households would be assumed. In many cases the difference with assuming a sample of households is not very large. For instance, it makes no difference when counting the number of non-significant country-by-country differences in the case of the at-risk-of poverty rate and only a small difference in the case of the other two indicators (69 non-significant differences for the deprivation indicator and 140 for the low work intensity indicator compared to respectively 64 and 133 in the case of scenario 2)[21]. Nevertheless, we believe that working with the third scenario is worth the effort (compared to assuming a sample of households as in scenario two). First of all, estimated standard errors are more precise, which means that in other conditions where there are larger differences between both scenarios, scenario three is likely to outperform scenario two. Second, once the PSU and stratification variables are computed, taking account of these is not more difficult than

[21] The same results are obtained when using the 'Eurostat EU-SILC' dataset (scenario 4).

taking account of clustering within households: for practical applications the same commands must be used. Furthermore, in some cases the difference between both scenarios is quite substantial. Therefore, scenario three is highly recommended. This is even true for countries where the information on PSUs and stratification is very partial, at best. It turns out that the available PSUs and strata account for the major part of the variance such that estimated standard errors come quite close to those estimated on the basis of the specific dataset containing additional information about the sample design prepared by Eurostat.

Table 8: Recommended sample design to be implemented for the computation of standard errors using EU-SILC 2008 UDB and comparison with standard errors based on 'Eurostat EU-SILC' dataset

	Primary sampling units			Stratification	Standard error in scenario three (recommended) compared to scenario four ('Eurostat EU-SILC' dataset)	Comment
	HID	DB060	DB062	DB040		
AT	x			x	trivial	DB040 in principle only since 2008; re-group PSUs
BE07		x		x	slight over-estimation	Re-group PSUs
BE08	x			(x)	non-trivial under-estimation	BE07 can give idea of bias
BG		x	(x)	x	conservative	Slight to large over-estimation. Do not re-group PSUs
CY	x				trivial	
CZ		x		x	trivial	Re-group PSUs if necessary
DE	x				non-trivial under-estimation	No information on real standard errors
DK	x				none	
EE	x				trivial	
ES		x		x	trivial	Transformed DB040; re-group PSUs
FI	x			(x)	conservative	Slight to large over-estimation
FR07	x	x	(x)	x	Non-trivial under-estimation	Re-group PSUs. More research needed on strata containing one PSU. over-estimation in case of work intensity estimator
GR		x	(x)	x	trivial	Re-group PSUs if necessary
HU		x	x	x	conservative	Use DB062 when DB060 is missing. More research needed on strata containing one PSU. Problem with 'Eurostat EU-SILC'.
IE		x			trivial	
IS	x				none	
IT	x	x	(x)	x	conservative	Re-group PSUs. More research needed on strata containing one PSU.
LT	x				trivial	
LU	x				trivial	
LV		x	(x)		trivial	drop cases with no information on DB060
NL		x			slight bias	Direction of bias depending on indicator
NO	x				none	
PL		x	(x)	x	slight under-estimation	do not re-group PSUs; problem with 'Eurostat EU-SILC'
PT		x			trivial	
RO		x	(x)	x	slight over-estimation	re-group PSUs
SE	x			(x)	none	
SI		x	(x)		under-estimation	under-estimation in case of at-risk-of-poverty. HID could be preferred for that indicator
SK	x				trivial	
UK		x			slight over-estimation	

(x) means that data on regions are available in EU-SILC 2008 UDB, but they should not be used..

Trivial means an observed bias of less than 1 per cent; slight means less than 5 per cent.

Note that in the case of Spain, the regions Ceuta and Malilla should be grouped together to form one stratum.

To what extent are these recommendations also applicable in other circumstances (e.g. other variables, regression analyses etc.)? The results presented in this paper show that the effect of clustering (and stratification) depends on the variable of interest. However, although all three indicators analysed here are not strongly correlated, in most cases results consistently indicate that applying the third scenario with regard to the sample design is the most accurate way forward when analysing the EU-SILC UDB. Therefore, one can have some confidence in the recommended use of variables presented in table 7. Nonetheless, it offers no guarantee. In fact, in principle the accuracy of estimates of standard errors should be evaluated for every analysis separately. An easier solution would be that the original PSU and stratification variables would directly be provided along with the UDB. Even though concerns with privacy should not be downplayed, the degree to which the provision of the original PSU and stratification variables would breach confidentiality clearly merits a more thorough discussion.

6. Conclusion

If estimates are based on samples, they should be accompanied by appropriate standard errors and confidence intervals. This is true for scientific research in general, but even more important if these estimates are used to inform and evaluate policy measures such as those aimed at attaining the Europe 2020 poverty reduction target. Unfortunately, in many cases standard errors are lacking and no idea is given of the precision of the estimates, which is namely true for estimates based on EU-SILC. In order to compute accurate standard errors, sample design, weighting, imputation and the complexity of the indicator should be taken into account. This requires adequate information on all these factors, proper variables in the dataset and adequate software which takes these issues into account.

In this paper we have argued that adequate and user-friendly software is available to take account of the sample design and weighting schemes as well as the complexity of relative poverty indicators. As the first stage of the sample design is crucial for estimating standard errors, the necessary information for each EU-SILC participating country has been gathered and analysed in this paper using different information sources, i.e. the quality reports, the EU-SILC User Database and a specific dataset containing additional information on the sample design prepared by Eurostat. Although even in the dataset prepared by Eurostat variables identifying primary strata and PSUs are not fully accurate, they enable a much more precise replication of the real sample design than the data available in the UDB. Therefore, on the basis of these more complete data, standard errors have been compared to those estimated on the basis of the information in the UDB. For all three Europe 2020 poverty indicators it has been shown that for many countries reasonable approximations of the real standard errors can be achieved if optimal use is made of the information available in the UDB. Although further research on the wider application of the findings in this paper is a necessary complement, we believe the recommendations presented in this paper offer a good starting point for researchers wishing to appropriately inform their readers about the precision of their EU-SILC estimates.

Acknowledgements

I would like to thank Karel Van den Bosch, Guillaume Osier, Ulrich Kohler, Lina Salanauskaite and Joris Ghysels for stimulating discussions, comments and suggestions. I am most grateful to Fabienne Montaigne and Pascal Wolff for offering me the opportunity to run my Stata programmes on a dataset containing additional information about the sampling design during a short stay at Eurostat. In addition I am also grateful to Abdelkrim Araar, Koen Decancq, Peter Thijssen, Philippe Van Kerm, Vijay Verma and my colleagues at the Herman Deleeck Centre for Social Policy who supported me in some way or another in writing this paper. An early version of the paper has been presented at the Final Equalsoc Conference, June 2010 in Amsterdam. Comments and suggestions from the participants are gratefully acknowledged. All opinions expressed in this paper as well as any remaining errors and shortcomings are my own responsibility. Opinions and suggestions expressed in this paper do not necessarily reflect those of Eurostat.

Appendix 1: EU-SILC Sample Design

Table 9: Sample design of EU-SILC by country

	Type of sampling design	No. of Stages	First-stage			Final stage		
			Type of unit	Selected by	Stratification	Type of unit	Selected by	Stratification
AT	Simple random sampling	1				Dwellings	Simple random sampling	2008: NUTS2, socio-economic, interviewer region
BE	Stratified two-stage sampling	2	Municipalities	Prop. sampling	NUTS2 Region	Households	Systematic sampling	No
BG	Stratified two-stage sampling	2	Census enumeration units	Prop. sampling	Administrative territorial regions (NUTS3)	Households	Systematic sampling	no
CY	Stratified simple random sampling	1				Households	Simple random sampling	Geographical criteria
CZ	Stratified two-stage sampling	2	Census sections	Prop. sampling	NUTS3 and size of municipality	Dwellings	Simple random sampling	no
DE	Stratified two-stage design	3 (?)	Dwelling blocks (Mikrozensus)	Prop. sampling (?)	Region and type of building block	Households	Simple random sampling (?)	Regional and socio-economic criteria
DK	Simple random sampling	1				Persons 14+	Simple random sampling	no
EE	Stratified systematic sampling	1				Persons 14+	Systematic sampling	County level ("big" counties, "small" counties and Hiiu)
ES	Stratified two-stage sampling	2	Census sections	Prop. sampling	Administrative region and size of the municipality	Dwellings	Systematic sampling	no
FI	Post-stratified unequal probability sampling	1				Dwellings	Prop. sampling	Socio-economic criteria
FR	Stratified three-stage sampling	3	Groups of municipalities	Prop. sampling	NUTS2, degree of urbanisation and rural/urban	Dwellings	Systematic sampling	no
GR	Stratified two-stage sampling	2	Dwelling blocks	Prop. sampling	NUTS2 and degree of urbanisation	Households	Systematic sampling	no
HU	Stratified two-stage sampling	2	Localities	Prop. sampling	Election district and number of dwellings	Dwellings	Systematic sampling	no
IE	Stratified two-stage sampling	2	Dwelling blocks	Simple random sampling without replacement	NUTS2 and degree of urbanisation	Households	Simple random sampling	no
IS	Simple random sampling	1				Persons 16+	Simple random sampling	no
IT	Stratified two-stage sampling	2	Municipality	Systematic Prop. sampling	Administrative region and number of residents	Households	Systematic sampling	no
LT	Stratified simple random sampling	1				Persons 16+	Simple random sampling	Degree of urbanisation

Table 10: Sample design of EU-SILC by country (continued)

	Type of sampling design	No. of Stages	First-stage			Final stage		
			Type of unit	Selected by	Stratification	Type of unit	Selected by	Stratification
LU	Stratified simple random sampling	1				"Tax" households	Simple random sampling	Social security status variables
LV	Stratified two-stage sampling	2	Census sections	Systematic samling (Prop. to size)	Degree of urbanisation	Dwellings	Simple random sampling	no
NL	Stratified two-stage sampling	3	Municipalities	Systematic Prop. sampling	COROP and interviewer region	address	Simple random sampling	no
NO	Systematic sampling	1				Persons 16+	Systematic random sampling	One-year age Group (until SILC2006)
PL	Stratified two-stage sampling	2	Census sections	Prop. sampling	NUTS2 and degree of urbanisation	Dwellings	Simple random sampling	no
PT	Stratified two-stage sampling	2	Census sections	Prop. sampling	NUTS3	Dwellings	Simple random sampling	no
RO		2	Census sections	Prop. sampling	Urban / rural and county (NUTS3)	Dwellings	Systematic sampling	no
SE	Systematic sampling	1				Persons 16+	Systematic sampling	no
SI	Stratified two-stage sampling	2	Census sections	Systematic prop. sampling	Size of the settlement and proportion of agricultural households	Persons 16+	Systematic sampling	no
SK	Stratified simple random sampling	1				Households	Simple random sampling	NUTS3 and degree of urbanisation
UK	Stratified two-stage sampling	2	Postcode sectors	Systematic Prop. sampling	Criteria based on 2001 Census data	Dwellings	Systematic sampling	no

Prop. sampling: sampling proportional to size (i.e. number of dwellings or households)

Until EU-SILC 2007, the German SILC included at least one subsample generated by quota sampling. For that part, the calculation of confidence intervals was somewhat pointless.

Source: Adapted from (Eurostat, 2009: 67-69), integrating information of the intermediate national quality reports of EU-SILC 2008[22] and the comparative intermediate quality report (Eurostat, 2010a). Where conflicts between the information of both types of documents arose, preference has been given to the information in the intermediate national quality reports. For Germany also DESTATIS (2006) and DESTATIS (2009) and for Portugal also Statistics Portugal (2009).

22 See http://circa.europa.eu/Public/irc/dsis/eusilc/library.

References

Alfons, A., Temple, M. and Filzmoser, P. (2009), *On the Influence of Imputation Methods on Laeken Indicators: Simulations and Recommendations*, Conference of European Statisticians, Neuchatel, Switzerland, 5-7 October 2009, 9p.

Araar, A. and Duclos, J.-Y. (2007), *DASP: Distributive Analysis Stata Package*: PEP, CIRPÉE and World Bank, Université Laval.

Araar, A. and Duclos, J.-Y. (2009), 'DAD: A Software for Poverty and Distributive Analysis' in *Journal of Economic & Social Measurement*, 34(2/3): 175-189

Atkinson, A. B., Cantillon, B., Marlier, E. and Nolan, B. (2002), *Social Indicators: the EU and Social Inclusion*, Oxford: Oxford University Press, 240p.

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

Biewen, M. (2002), 'Bootstrap Inference for Inequality, Mobility and Poverty Measurement' in *Journal of Econometrics*, 108(2): 317-342

Biewen, M. and Jenkins, S. P. (2006), 'Variance Estimation for Generalized Entropy and Atkinson Inequality Indices: the Complex Survey Data Case' in *Oxford Bulletin of Economics and Statistics*, 68(3): 371-383

Cochran, W. G. (1977), *Sampling Techniques*, New York: John Wiley & Sons.

Davidson, R. and Duclos, J.-Y. (2000), 'Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality' in *Econometrica*, 68(6): 1435-1464

Davidson, R. and Flachaire, E. (2007), 'Asymptotic and Bootstrap Inference for Inequality and Poverty Measures' in *Journal of Econometrics*, 141(1): 141-166

de Vos, K. and Zaidi, A. M. (1998), 'Poverty measurement in the European Union: Country-specific or Union-wide poverty lines?' in *Journal of Income Distribution*, 8(1): 77-92

del Mar Rueda, M. and Muñoz, J. (2009), 'Estimation of poverty measures with auxiliary information in sample surveys' in *Quality and Quantity*, online first: 1-14

DESTATIS (2006), *Mikrozensus*, Wiesbaden: Statistisches Bundesamt.

DESTATIS (2009), *Gemeinschaftsstatistik über Einkommen und Lebensbedingungen*, Wiesbaden: Statistisches Bundesamt, 8p.

Duclos, J.-Y. and Araar, A. (2006), *Poverty and Equity. Measurement, Policy, and Estimation with DAD*, New York: Springer, 393p.

Efron, B. and Tibshirani, R. J. (1998), *An Introduction to the Bootstrap*, Boca Raton: Chapman & Hall/CRC, 436p.

European Commission (2006), *Portfolio of Overarching Indicators and Streamlined Social Inclusion, Pensions, and Health Portfolios*, Brussels, 50p.

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

European Parliament and Council of the European Communities (2003), *Regulation (EC) No 1177/2003 of the European Parliament and of the Council of 16 June 2003 concerning Community statistics on income and living conditions (EU-SILC)*, OJ L 165, 3.7.2003, pp. 1-9

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

Eurostat (2009), *EU-SILC User Database Description*, Version 2007.1 from 01-03-09: European Commission, 69p.

Eurostat (2010a), *2008 Comparative EU Intermediate Quality Report. Version 2 - June 2010*, Luxembourg: Eurostat, 59p.

Eurostat (2010b), *Combating poverty and social exclusion. A statistical portrait of the European Union 2010*, Luxembourg: Publications Office of the European Union, 111p.

Eurostat (2010c), *Description of Target Variables: Cross-sectional and Longitudinal 2008 operation (Version January 2010)*, EU-SILC 065 (2008 operation): European Commission, 391p.

Foster, J., Greer, J. and Thorbecke, E. (1984), 'A Class of Decomposable Poverty Measures' in *Econometrica*, 52(3): 761-766

Goedemé, T. (2010), *The construction and use of sample design variables in EU-SILC. A users's perspective*, Report prepared for Eurostat, November 2010, Antwerp: Herman Deleeck Centre for Social Policy, University of Antwerp, 16p.

Goedemé, T. and Rottiers, S. (2010), *Poverty in the enlarged European Union. A discussion about definitions and reference groups*, CSB Working Paper Series, WP. 10/06, Antwerp: Herman Deleeck Centre for Social Policy, University of Antwerp, 24p.

Guio, A.-C. (2009), *What Can Be Learned From Deprivation Indicators in Europe?*, Paper Presented at the Indicator Subgroup of the Social Protection Committee, 10th February 2009, 33p.

Howes, S. and Lanjouw, J. O. (1998), 'Does Sample Design Matter for Poverty Rate Comparisons?' in *Review of Income & Wealth*, 44(1): 99-109

Jolliffe, D., Datt, G. and Sharma, M. (2004), 'Robust Poverty and Inequality Measurement in Egypt: Correcting for Spatial-price Variation and Sample Design Effects' in *Review of Development Economics*, 8(4): 557-572

Jolliffe, D. and Semykina, A. (1999), 'sg117 - Robust standard errors for the Foster–Greer–Thorbecke class of poverty indices' in *Stata Technical Bulletin*, STB-51: 34-36

Kakwani, N. (1993), 'Statistical Inference in the Measurement of Poverty' in *The Review of Economics and Statistics*, 75(4): 632-639

Kalton, G. (1983), *Introduction to Survey Sampling*, Quantitative Applications in the Social Sciences, Sage University Paper No. 35, Beverly Hills: Sage Publications, 96p.

Kangas, O. E. and Ritakallio, V.-M. (2007), 'Relative to What? Cross National Pictures of European Poverty Measured by Regional, National and European Standards' in *European Societies*, 9(2): 119-145

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

Lee, E. S. and Forthofer, R. N. (2006), *Analyzing Complex Survey Data. Second Edition*, Quantitative Applications in the Social Sciences, 71, Thousand Oaks: Sage Publications, 91p.

Marlier, E., Atkinson, A. B., Cantillon, B. and Nolan, B. (2007), *The EU and Social Inclusion. Facing the challenges*, Bristol: The Policy Press, 303p.

Mooney, C. Z. and Duval, R. D. (1993), *Bootstrapping: A Nonparametric Approach to Statistical Inference*, Quantitative Applications in the Social Sciences, Sage University Paper No. 95, Newbury Park: Sage Publications, 72p.

OECD (2008), *Growing Unequal? Income Distribution and Poverty in OECD Countries*, Paris: OECD, 308p.

Osier, G. (2009), 'Variance Estimation for Complex Indicators of Poverty and Inequality Using Linearization Techniques' in *Survey Research Methods*, 3(3): 167-195

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

Rodgers, J. R. and Rodgers, J. L. (1993), 'Chronic Poverty in the United States' in *The Journal of Human Resources*, 28(1): 25-54

Shao, J. (1996), 'Invited Discussion Paper Resampling Methods in Sample Surveys' in *Statistics: A Journal of Theoretical and Applied Statistics*, 27(3): 203-237

Shao, J. and Chen, Y. (1998), 'Bootstrapping Sample Quantiles Based on Complex Survey Data Under Hot Deck Imputation' in *Statistica Sinica*, 8(4): 1071-1085

Statistics Portugal (2009), *Inquérito às condições de vida e rendimento (Statistics on Income and Living conditions)*: Instituto Nacional de Estatística, 120p.

Sturgis, P. (2004), 'Analysing Complex Survey Data: Clustering, Stratification and Weights' in *Social Research Update*, 43: 1-6

Thuysbaert, B. (2008), 'Inference for the Measurement of Poverty in the Presence of a Stochastic Weighting Variable' in *Journal of Economic Inequality*, 6(1): 33-55

Trede, M. (2002), 'Bootstrapping inequality measures under the null hypothesis: Is it worth the effort?' in *Journal of Economics*, 77(Supplement 1): 261-282

Van Kerm, P. (2002), 'Inference on Inequality Measures: A Monte Carlo Experiment' in *Journal of Economics*, 77(Supplement 1): 283-306

Van Kerm, P. (2007), *Extreme Incomes and the Estimation of Poverty and Inequality Indicators from EU-SILC*, IRISS Working Paper Series, NO. 2007-01: CEPS-Instead, 51p.

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

Whelan, C., T. and Maître, B. (2007), 'Income, Deprivation and Economic Stress in the Enlarged European Union' in *Social Indicators Research*, 83(2): 309-329

Wolff, P. (2010), *17% of EU Citizens were at-risk-of-poverty in 2008*, Statistics in Focus, 9/2010: Eurostat, 8p.

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

Zheng, B. (2001), 'Statistical Inference for Poverty Measures with Relative Poverty Lines' in *Journal of Econometrics*, 101(2): 337-356