Assessment of socio-demographic sample composition in ESS Round 6[1]

Achim Koch

GESIS – Leibniz Institute for the Social Sciences, Mannheim/Germany, June 2016

Contents

---

[1] The CST of the ESS requests that the following citation for this document should be used: Koch, A. (2016). Assessment of socio-demographic sample composition in ESS Round 6. Mannheim: European Social Survey, GESIS.

1. Introduction

The European Social Survey (ESS) is an academically driven cross-national survey that has been conducted every two years across Europe since 2002. The ESS aims to produce high-quality data on social structure, attitudes, values and behaviour patterns in Europe. Much emphasis is placed on the standardisation of survey methods and procedures across countries and over time. Each country implementing the ESS has to follow detailed requirements that are laid down in the "Specifications for participating countries". These standards cover the whole survey life cycle. They refer to sampling, questionnaire translation, data collection and data preparation and delivery. As regards sampling, for instance, the ESS requires that only strict probability samples should be used; quota sampling and substitution are not allowed. Each country is required to achieve an effective sample size of 1,500 completed interviews, taking into account potential design effects due to the clustering of the sample and/or the variation in inclusion probabilities. Regarding data collection, the ESS specifies – among other things – that face-to-face interviewing is the only mode allowed. Targets are set for the response rate (70%) and the noncontact rate (3% maximum). The fieldwork period is specified (September until December of the survey year), the personal briefing of interviewers is required, and a detailed call schedule for the interviewers is laid down.

The purpose of setting these standards is to achieve accurate and comparable survey data. An important aspect of survey quality refers to the quality of the realised samples in terms of representation of the target population. The sample in each ESS country should reflect the target population of the ESS adequately, which means that bias due to nonresponse should be minimised.[2] Up till now, quality control activities in the ESS were mainly directed at compliance with the prescribed data collection procedures. In each survey round, for instance, it is checked whether or not a country achieved the target response rate, whether the interviewers were adequately briefed, whether the call schedule was adhered to, etc. The (implicit) assumption is that a country that follows the ESS survey procedures and achieves a high response rate will also achieve a sample of good quality. In the present paper we take a step to assessing empirically how "good" the samples actually are. We analyse the socio-demographic sample composition in ESS countries by comparing ESS variable distributions with suitable external benchmark data, for which we chose the European Union Labour Force Survey (LFS). The analyses refer to ESS 6 which was fielded in the years 2012 and 2013. With our analyses we pursue two aims. First, we want to provide an indication of the degree of over-/underrepresentation of certain demographic subgroups in ESS samples. Second, we analyse the correlates of over-/underrepresentation, focusing on two basic parameters, namely the response rate achieved and the type of sample used.

Two years ago, a similar analysis had been conducted for ESS 5 (Koch, Halbherr, Stoop & Kappelhof, 2014). The present analysis goes beyond this previous exercise in two respects. First, the LFS data were taken from two different survey years (2012 vs. 2013), depending on the year in which the sixth round of ESS was actually fielded in a specific country. Second, in order to allow for a better assessment of whether an observed difference between ESS and LFS data is within the limits of sampling error or not, confidence intervals were estimated for the ESS results (taking the geographical clustering of the samples into account, if relevant). Originally, it was also planned to analyse whether the use of post-stratification weights in the ESS might help to reduce deviations from the LFS. However, due to an error in the first

---

[2] As a matter of course, the ESS also requests that sampling error should not exceed a certain level (a minimum effective sample size of 1,500 completed interviews is to be achieved), and over-/undercoverage of certain groups should be avoided in all countries. The focus of the present paper is on the potential negative effect of nonresponse on sample quality.

release of the post-stratification weights for ESS Round 6, this analysis could not be conducted in a timely manner to be included in the present report.

2. Assessing socio-demographic sample composition with external benchmark data

The comparison of survey results with independent and more accurate information about the population parameters is a well-known method to analyse sample quality and the degree of nonresponse bias (Groves, 2006). For this approach no information at the individual level is required. There needs to be another survey or administrative record system containing estimates of variables similar to those being produced from the survey. Then, the survey estimates can be benchmarked with information from the other data source, the so-called gold standard. The difference between estimates from the survey and the other data source can be used as an indicator of bias.

The advantage of this method is that it is in theory relatively simple to implement. Usually, the method is not so expensive since it does not require collecting additional data. The drawback is that normally only a limited set of variables can be compared. In order to draw valid conclusions about nonresponse bias, the benchmark data have to be quite accurate, i.e. they should not be severely affected by, for instance, measurement or nonresponse errors. In addition, the measurement of the relevant variables should match closely between the two data sources (equivalent measurements). Both data sources have to refer to the same target population and also the reference period should be as close as possible. Even if these conditions hold, one has still to be aware that differences between the survey data and the benchmark data might arise from both nonresponse error and sampling error.[3]

It goes without saying that no benchmark information is available for the ESS key survey variables – this is the reason, why the ESS exists! Comparisons have to be restricted to several socio-demographic variables. The results, however, are important beyond these variables. Socio-demographic characteristics are intrinsically important since they are – potentially – related to many attitudes and behaviours. This is the reason, why some of these variables are often used to construct post-stratification weights. From 2014 onwards, post-stratification weights are also provided for the ESS (European Social Survey, 2014).

For a cross-national survey like the ESS the most promising candidate to act as a valid standard for such a comparison is the European Union Labour Force Survey (LFS). Most of the countries that participate in the ESS also conduct the yearly Labour Force Survey for Eurostat.

3. The European Union Labour Force Survey

The European Union Labour Force Survey (LFS) is a large sample survey among residents in private households in Europe.[4] It is an important source for European statistics about the situation and trends in the EU labour market. The LFS is currently fielded in 33 European countries. These include the 28 Member States of the European Union, three EFTA countries (Iceland, Norway and Switzerland), and two EU candidate countries (the Former Yugoslav Republic of Macedonia and Turkey). The sampling units are dwellings, households or

---

[3] In some cases, differential coverage errors may also lead to differences between the two data sources.
[4] http://ec.europa.eu/eurostat/web/microdata/european-union-labour-force-survey

individuals depending on the country-specific sampling frames. Each quarter some 1.8 million interviews are conducted throughout the participating countries to obtain statistical information for some 100 variables. The sampling rates in the various countries vary between 0.2% and 3.3%.

The EU LFS is conducted by the National Statistical Institutes across Europe and is centrally processed by Eurostat (for details of national implementation see Eurostat, 2013, 2014b). The National Statistical Institutes of the Member States are responsible for designing national questionnaires, drawing the sample, conducting interviews and forwarding results to the Commission (Eurostat) in accordance with a common coding scheme. As a rule the data are collected by interviewing the sampled individuals directly, but proxy interviews (through a responsible person in the household) are also possible. Moreover, part of the data can also be supplied by equivalent information from alternative sources, such as e.g. administrative registers (mainly social insurance records and population registers).

Table 1: Timing of fieldwork in ESS 6

| Country | % of interviews completed in year | |
	2012	2013
BE	100.0	
BG		100.0
CH	86.3	13.7
CY	97.6	2.4
CZ		100.0
DE	93.1	6.9
DK		100.0
EE	95.8	4.2
ES		100.0
FI	85.8	14.2
FR		100.0
HU	94.1	5.9
IE	37.9	62.1
IS	79.9	20.1
IT		100.0
LT		100.0
NL	80.2	19.8
NO	89.0	11.0
PL	99.3	.7
PT	18.4	81.6
SE	56.7	43.3
SI	100.0	
SK	76.6	23.4
UK	88.5	11.5
Total	56.6	43.4

Source: ESS 6, ed. 2.1, variable 'inwyys' (start of interview, year)
Highlighted: Countries with all interviews or the majority of interviews completed in the year 2013

The present comparison with the LFS is conducted for the sixth survey round of ESS. In total, 29 countries participated in ESS 6. Among these countries, 24 countries also participated in the LFS. Albania, Israel, Kosovo, Russia and Ukraine were not part of the LFS and had to be

excluded from our analyses. As a rule, fieldwork in each ESS country should take place between September and December of the survey year. Unfortunately, not all countries managed to adhere to this schedule in ESS 6. In nine out of the 24 countries included in our analyses all interviews or the majority of interviews were completed only in 2013 (see Table 1). These countries were Bulgaria, the Czech Republic, Denmark, France, Ireland, Italy, Lithuania, Portugal and Spain. For these countries, we used LFS 2013 data for the comparison. For all the other countries, data from LFS 2012 were used. Table 2 documents a few basic parameters (participation compulsory, response rate[5] and rate of proxy interviews) for LFS 2012 and 2013. For the sake of comparison, ESS 6 response rates are also included.

Table 2: Basic characteristics of LFS 2012, LFS 2013 and ESS 6*

Country	LFS 2012			LFS 2013			ESS 6
	Partici-pation compul-sory	Response rate (%)	Proxy rate among 15-74 years old respondents (%)	Partici-pation compul-sory	Response rate (%)	Proxy rate among 15-74 years old respondents (%)	Res-ponse rate (%)
BE	yes	68.0	20.7	yes	71.2	19.8	58.7
BG	no	78.5	35.3	no	76.4	36.0	70.6
CH**	yes/no	82.9	1.7	no	79.7	1.8	51.7
CY	yes	96.7	32.4	yes	97.3	31.2	76.0
CZ	no	80.7	45.9	no	79.7	45.8	68.2
DE	yes	98.2	25.9	yes	97.9	25.7	33.7
DK	no	52.7	5.4	no	53.0	5.0	56.7
EE	no	68.2	32.0	no	67.6	32.6	67.8
ES	yes	84.5	52.7	yes	84.8	52.5	70.2
FI	no	73.9	3.9	no	72.9	4.0	67.7
FR	yes	84.7	31.4	yes	80.2	28.2	44.9
HU	no	83.8	44.4	no	81.5	45.4	65.1
IE	no	79.7	48.3	no	77.1	49.8	65.0
IS	no	83.2	2.4	no	80.9	0.2	54.7
IT	yes	89.3	16.0	yes	88.3	17.7	36.8
LT	no	82.7	31.6	no	81.0	32.4	74.7
NL	no	79.1	48.8	no	79.4	46.4	55.1
NO	yes	80.5	14.3	yes	78.9	15.6	55.5
PL	no	75.8	40.9	no	71.9	40.7	74.6
PT	yes	84.6	48.7	yes	85.7	48.3	77.1
SE	no	72.8	2.7	no	69.2	2.6	50.7
SI	no	76.3	54.4	no	78.3	55.7	57.8
SK	yes	92.2	44.0	yes	91.8	46.4	73.9
UK	no	56.9	35.0	no	60.6	35.3	53.1
mean		79.4	30.0		78.6	30.0	60.8

* 24 countries which took part both in ESS 6 and in LFS
** CH: shift from compulsory to voluntary participation in quarter 3 in 2012
Source: Eurostat (2013, 2014a, 2014b, 2014c); Beullens, Matsuo, Loosveldt and Vandenplas (2014)

---

[5] In the LFS most countries calculate response rates on the household level, only in a minority of countries response rates are calculated on the person level (which is the standard in ESS).

Among the 24 countries, participation in the LFS was mandatory in 9 countries (not counting Switzerland which moved to voluntary participation in the third quarter in 2012). The LFS response rates vary between 52.7% (Denmark 2012) and 98.2% (Germany 2012). Accordingly, the LFS, too, has a nonresponse problem in some countries. The consequences for the nonresponse error of the LFS cannot be assessed here. However, two points can be made in favour of still using LFS as a benchmark for the ESS. First, in each country except Denmark, Estonia and Poland, the LFS response rate is (often considerably) higher than the ESS response rate. On average, the response rate in the LFS is almost 20 percentage points higher than in the ESS (79.4% in LFS 2012 and 78.6% in LFS 2013 vs. 60.8% in ESS 6). Second, it has to be taken into account that the LFS data itself are weighted to adhere to the population distribution. (Nearly) all countries used population information on gender, age and region in their weighting procedure (Eurostat, 2013, 2014b). Several LFS countries included additional variables (like employment status or nationality). Accordingly, at least the distributions of these variables should validly reflect the countries' population.

Apart from the question of nonresponse error, the measurement error properties of the LFS might also be queried. In some LFS countries a large number of proxy interviews are conducted. The proportion of proxy interviews varies between less than 1% (Iceland in 2013) and 56% (Slovenia in 2013). On average across all 24 countries, almost one third (30%) of the interviews were proxy interviews (both in LFS 2012 and LFS 2013). We cannot empirically assess what this means for the quality of the LFS data. However, it seems justifiable to assume that the basic demographic information which we use for our analyses will not noticeably be impaired by this problem (Köhne-Finster & Lingnau, 2009; Thomsen & Villund, 2011; Zühlke, 2008).

4. Data and variables

For our analyses we use ESS round 6 data (edition 2.1)[6] and anonymised EU LFS 2012 and 2013 data (edition 2014)[7]. Comparisons between ESS and LFS were possible for variables which were either measured in an identical way or, if this was not the case, where the measurements could be recoded to a common standard.[8] This was true for six variables: gender, age, marital status, work status, nationality and household size. We deliberately did not include a variable like education in our comparison, which is difficult to measure in a comparable way in a cross-national context.[9]

---

[6] European Social Survey Round 6 Data (2012). Data file edition 2.1. Norwegian Social Science Data Services, Norway - Data Archive and distributor of ESS data. The Core Scientific Team (CST) and the producers bear no responsibility for the uses of the ESS data, or for interpretations or inferences based on these uses.

[7] All results and conclusions are those of the author and not those of Eurostat, the European Commission or any of the national authorities whose data have been used.

[8] The focus here is on comparability between the general standards set in the LFS and the ESS. However, one has to note that the comparability of measurements between countries within the LFS also might be an issue. The LFS sets various standards to ensure that the national surveys provide data that are compatible with the EU definitions. However, the leeway for differences in national questions is larger than in the ESS. Accordingly, the quality report for LFS 2013 states: "As a general conclusion it emerges that, in spite of the progress regarding the adherence to the EU regulations, principles and guidelines (i.e. the explanatory notes), the national questionnaires still largely differ even in the collection of key variables such as WSTATOR (Labour status in the reference week)." (Eurostat, 2014c, p. 23).

[9] Originally, we intended to include also the information on the highest level of education successfully completed. Both ESS and LFS use the ISCED classification of educational attainment. However, whereas the ESS documents in detail how the national degrees were mapped into the international standard (see ESS 6 Data Documentation Report Appendix A1 Education on the ESS website), the respective information is not available for the LFS.

Table 3 shows the variables and the respective categories which we distinguish, plus their source variables in ESS and LFS.

Table 3: Variables of the ESS – LFS comparison

Variable	Categories	ESS source variable	LFS source variable
Gender	• Male • Female	gndr	sex
Age	• 15-24 years • 25-34 years • 35-44 years • 45-54 years • 55-64 years • 65-74 years • 75 years and older	agea (recoded)	age (recoded)
Marital status	• Not married • Married (incl. registered partnership)	maritalb (3-6 = 0) (1-2 = 1)	marstat (0-1 = 0) (2 = 1)
Work status	• Not in paid work in the last 7 days • In paid work (for at least one hour) in the last 7 days	pdwrk + crpdwk	wstator (3-5 = 0) (1-2 = 1)
Nationality	• National of country • No national of country	ctzcntr	national (non-nationals recoded in one category)
Household size	Respondent lives in household comprising • 1 person • 2 persons • 3 persons • 4 persons • 5 or more persons	hhmmb (recoded)	hhnbpers (recoded)

The ESS interviews persons aged 15 years and over resident within private households, regardless of their nationality, citizenship or language. In order to achieve comparable target populations, we excluded persons under 15 years in the LFS. In addition, persons living in an institutional household (which were surveyed in a few LFS countries) were excluded. In Iceland, Norway and Sweden, LFS data are only available for persons aged 74 years or younger. For these three countries, we also restricted the ESS analyses to persons aged 74 years or younger.[10]

ESS data were weighted with the design weight (DWEIGHT). This weight corrects for differences in selection probabilities between sampling units in a country. The design weights are computed as normed inverse of the inclusion probabilities. LFS data were weighted with the standard weight variable COEFF, as recommended by Eurostat. This weight too corrects for differences in selection probabilities. In addition, it includes a post-stratification adjustment to adapt the LFS data to known population characteristics. In (nearly) all LFS countries data on gender, age and region were used for the adjustment. A number of countries

---

[10] At a late stage of preparing this report, we were made aware that the LFS sample in Estonia does not include persons 75 years and older living alone in a household. The exclusion of this group has not been documented by Eurostat in the LFS descriptions (see, e.g., Eurostat, 2014a, p. 5), so that we could not take account of it in the present analyses.

included additional data in weighting, like information on unemployment or nationality (see Eurostat, 2013, 2014b). Using weighted data for the LFS thus should reduce both sampling errors and errors due to nonresponse or non-coverage – at least for the variables included in the weighting procedure.

5. Description of ESS-LFS differences

In order to allow for an overview of which groups are over- or underrepresented in the ESS we provide line charts for each variable. Each chart displays at a time the proportions for one category of a variable both for ESS and LFS. Countries are in ascending order according to their value in the LFS. In order to facilitate comparisons between variables, each chart is scaled to show a range of 40 percentage points (however, often on a different 'level'). The figures show at a glance the absolute differences between ESS and LFS distributions. It can easily be checked whether the structure of over-/underrepresentation is similar across countries, and whether the size of differences differs between variables. For dichotomous variables (gender, marital status, work status, nationality), the proportions for only one category are shown. For age and household size one chart is provided for each category of the variables.

To provide an indication, whether the difference between ESS and LFS is within the limits of sampling error, we calculated 95% confidence intervals for the ESS estimates, taking the geographical clustering of the sample into account where relevant.[11] The confidence intervals were estimated using the complex sample procedure of SPSS.[12] When the confidence intervals do not overlap with the percentage from the LFS, we interpret this as a hint of a significant over- or underrepresentation with respect to that specific estimate.[13]

---

[11] The majority of countries in ESS 6 (17 out of 24) used a geographically clustered sample. Only seven countries used an unclustered design. These were Cyprus, Denmark, Estonia, Finland, the Netherlands, Sweden and Switzerland.

[12] I am very grateful to Stefan Zins, the head of the ESS sampling panel, who provided a data set including identifiers for the PSUs (primary sampling units) for all ESS 6 countries. Michael Blohm provided support in running the 'complex sample' procedure of SPSS.

[13] We could not estimate the sampling errors of the LFS estimates. Due to the rather large sample size (on average, approximately 140.000 persons 15 years and older per country), they tend to be small (see the examples in Eurostat, 2014c, p. 13). In addition, the post-stratification weighting applied in the LFS will reduce sampling errors (see section 4).

Gender

Figure 1 shows the percentage of females in the ESS and LFS samples. In 15 out of the 24 countries, the 95% confidence interval for the ESS estimate includes the estimate from the LFS. In eight countries, however, the proportion of females in the ESS is significantly larger than the proportion observed in the LFS. The largest discrepancy is 7.0 percentage points (Portugal: 60.1% females in ESS vs. 53.1% females in LFS). Hence, in every third country females tend to be overrepresented in the ESS. Only in one country (Norway), the percentage of females in the ESS is significantly below the percentage in the LFS (46.6% vs. 49.1%).

Figure 1: Females in ESS 6 and LFS 2012/2013 (in %)
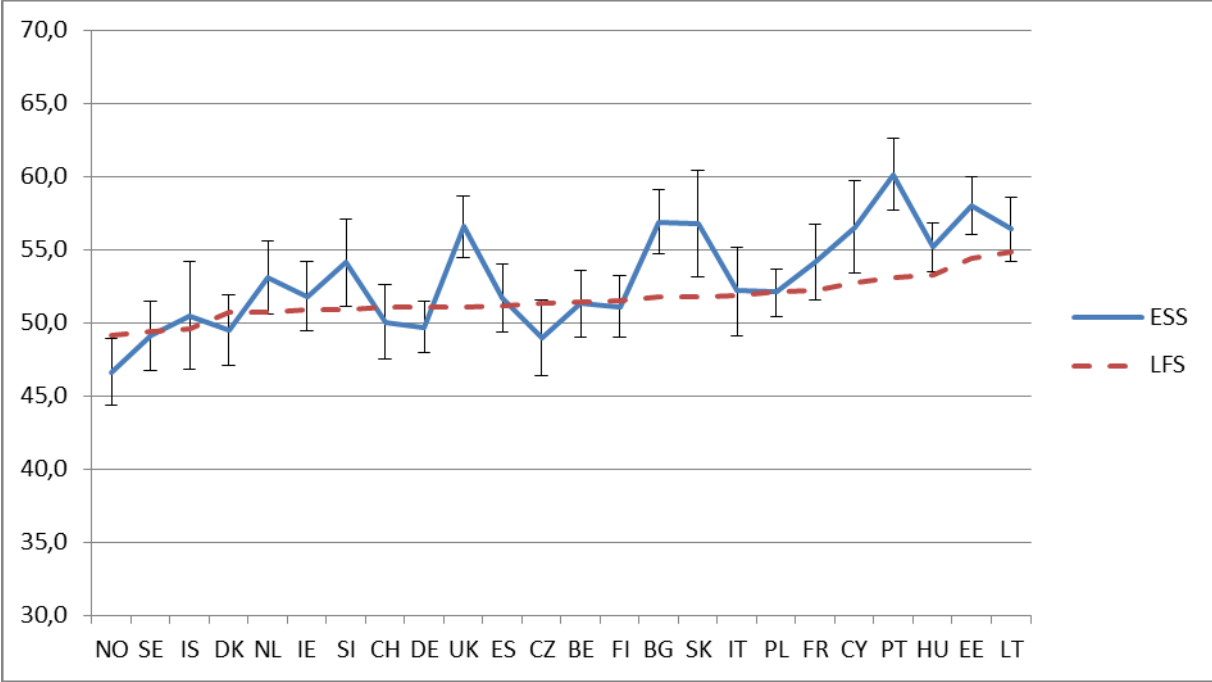
Age

As regards the variable age, we distinguish seven 10-year age groups in the comparison between ESS and LFS (see Figure 2). For the youngest age group, ESS and LFS results do not differ much in 14 out of the 24 countries. In six countries the proportion of 15-24 years old persons is significantly lower in the ESS than in the LFS (the maximum difference is -7.1 percentage points in Slovakia), in four countries it is significantly higher.

As regards 25-34 years old persons, in 15 countries the LFS estimate falls outside of the 95% ESS confidence interval. In each of these countries, the ESS estimate is smaller than the LFS estimate, which means that this age group is significantly underrepresented in the respective countries. The maximum difference is -5.2 percentage points in the Czech Republic.

Regarding 35-44 years old persons, the results of ESS are close to the results of the LFS in most countries. A significant difference can only be observed in three out of the 24 countries.

For the next three age groups (45-54 years, 55-64 years, 65-74 years) relevant differences occur in 6, 9, and 10 countries, respectively. In each of these cases, the respective age group is overrepresented in the ESS compared to the LFS data. The largest difference pertains to 65-74 years old persons in Bulgaria (+5.5 percentage points; their proportion is 18.2% in the ESS and 12.7% in the LFS).

The picture is different for the oldest age group. In six countries, persons 75 years or older are significantly underrepresented in the ESS. The largest difference (-5.9 percentage points) pertains to Lithuania: the ESS estimate is 4.6%; the LFS estimate is 10.5%. Only in one country (Estonia), persons in the oldest age group are significantly overrepresented in the ESS.

Taken together, we see considerable differences in the age distribution of some ESS countries. By and large, these differences seem to follow a common pattern: There is a tendency to underrepresent younger age groups (15-24 years old persons, and, in particular, 25-34 years old persons) *and* the oldest age group (75 years and older). In return, persons around and over 50 years (45-54, 55-64 and 65-74 years) tend to be overrepresented. The largest differences in the age distribution between ESS and LFS can be observed in Bulgaria, Czech Republic and Slovakia.

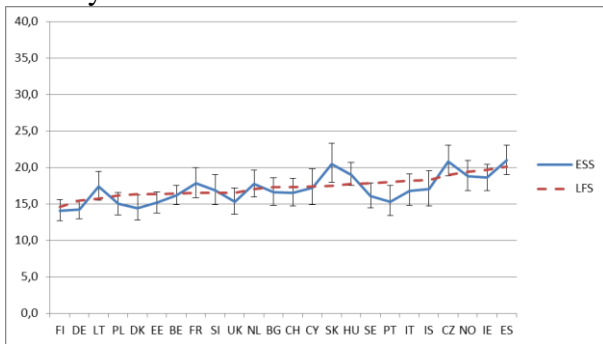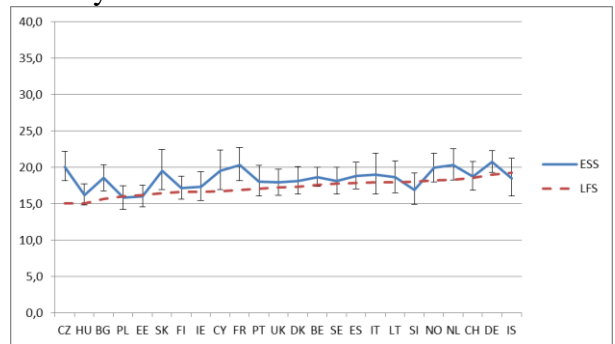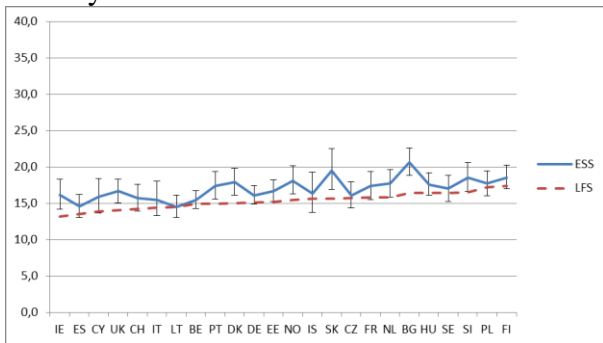**Figure 2: 10-year age groups in ESS 6 and LFS 2012/2013 (in %)**

15-24 years
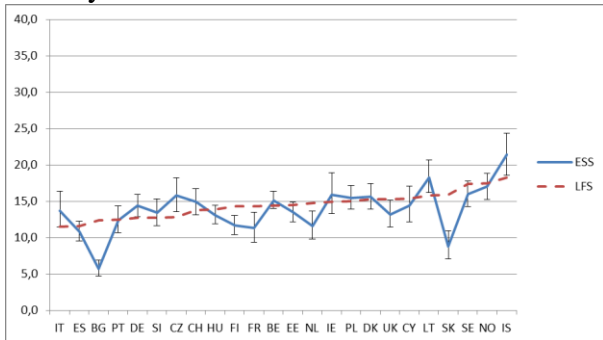
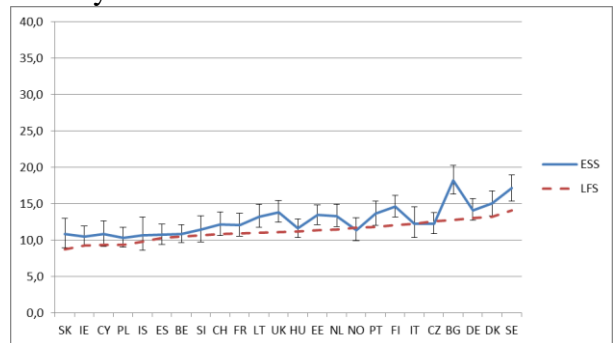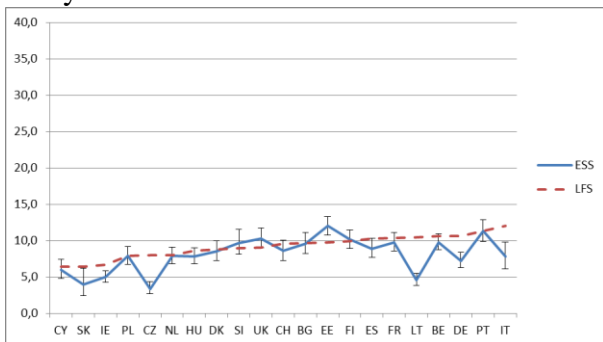## 25-34 years
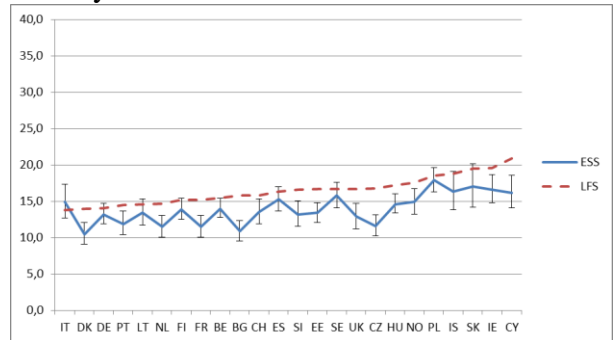
35-44 years


45-54 years

55-64 years

65-74 years

75+ years

Marital status

As regards the marital status, we can only distinguish between married and non-married persons. Both in the ESS and the LFS, the category 'married persons' includes persons living in a registered partnership. In more than half of the countries, the confidence intervals of the ESS estimates do not overlap with the LFS results (see Figure 3). In 12 out of the 24 countries, the proportion of married persons is significantly larger in the ESS than in the LFS. Only in two countries, the ESS proportion is significantly smaller. The biggest discrepancy is observed for Slovakia (+10.1 percentage points). Here, the proportion of married persons is 62.8% in the ESS and 52.7% in the LFS.

**Figure 3: Married persons in ESS 6 and LFS 2012/2013 (in %)***
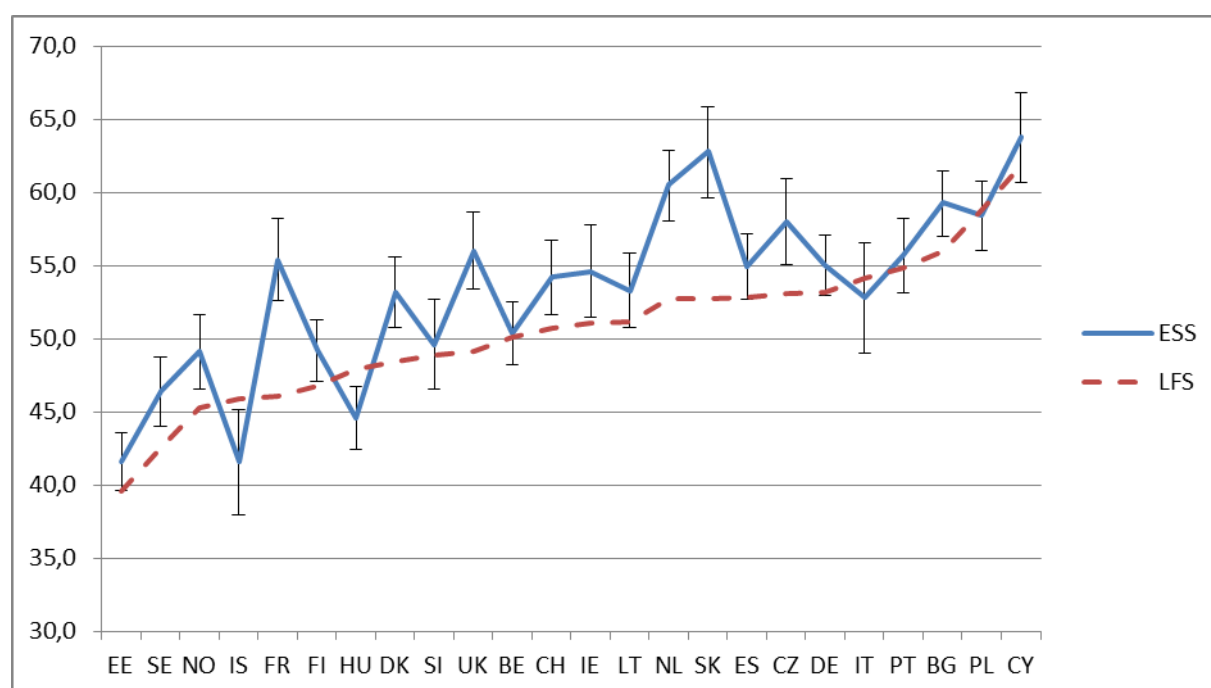
*incl. persons living in a legally registered partnership

Work status

As regards the work status, again a dichotomous distinction can be defined as a common standard for ESS and LFS. We can distinguish persons who have been working for pay for at least one hour in the past seven days from persons who did not. In around half of the countries, the distribution of this variable in the ESS resembles the LFS distribution (see Figure 4).

In 13 out of the 24 countries, however, the percentage of persons in paid work differs substantially between the ESS and the LFS. Both an underrepresentation of persons in paid work (6 countries, maximum difference -8.6 percentage points in Slovenia) and an overrepresentation of persons in paid work (7 countries, maximum difference +9.2 percentage points in Italy) can be observed.

**Figure 4: Persons in paid work in ESS 6 and LFS 2012/2013 (in %)**
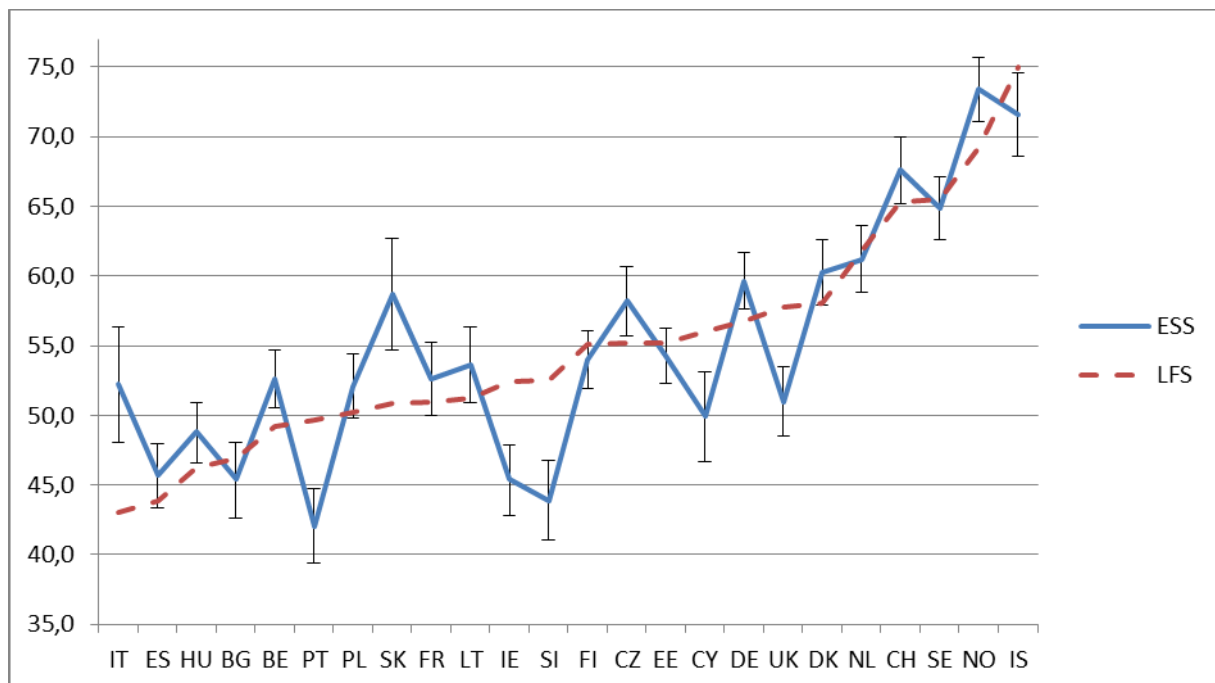
Nationality

As regards the nationality of respondents, we can distinguish between nationals of a country on the one hand (including persons holding a dual citizenship), and non-nationals of that country on the other hand. A special feature of this variable is its rather skewed distribution in some countries. In nine out of the 24 countries the proportion of non-nationals is less than 3 percent (according to the LFS data, see Figure 5). In these countries, as a matter of course, the ESS and LFS proportions of non-nationals do not differ much.

In nearly all other countries, however, the LFS estimate falls outside of the 95% confidence interval of the ESS estimate. In 13 out of the remaining 15 countries, non-nationals are significantly underrepresented in the ESS; only in one country (Norway) they are significantly overrepresented. The largest difference is observed in Cyprus (-11.9 percentage points). According to the LFS, the percentage of non-nationals in Cyprus is 19.3%. The percentage of non-nationals in the ESS, however, is only 7.4%.

Figure 5: Non-nationals in ESS 6 and LFS 2012/2013 (in %)
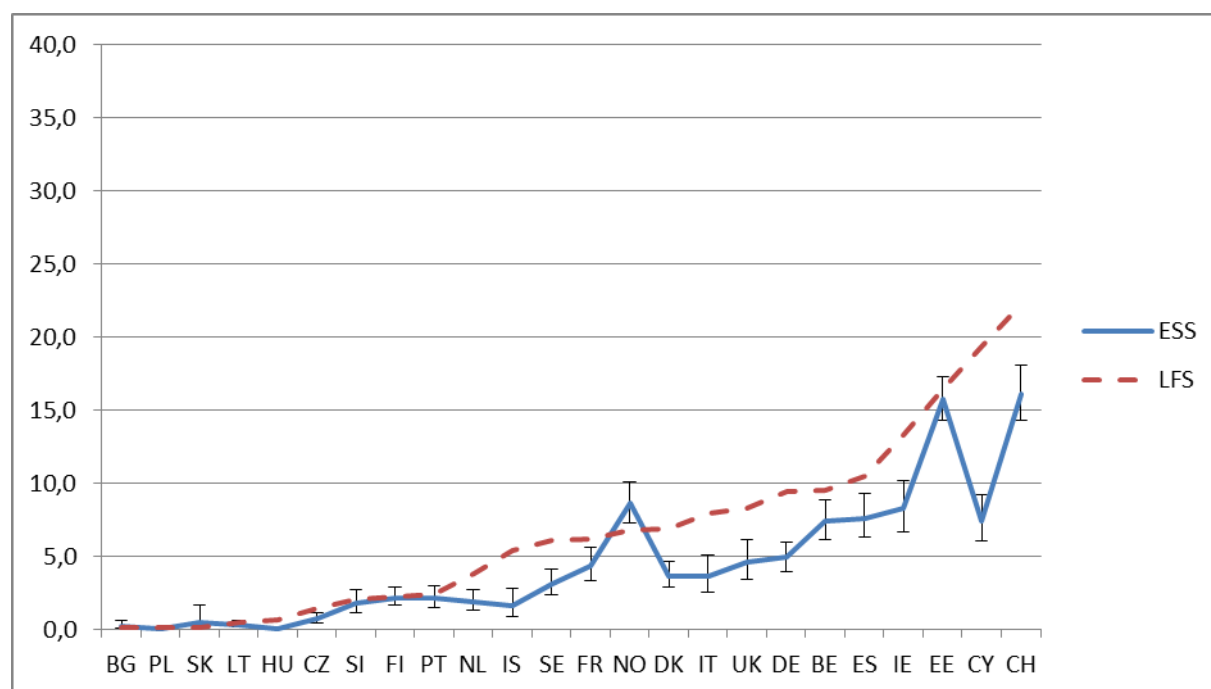
Household size

In order to compare ESS and LFS data on household size, we recoded all persons living in households with five or more persons in one category. LFS data on household size are not available for Denmark, Finland, Iceland, Norway, Sweden and Switzerland. These countries were excluded from the present comparison.

A great number of deviations occur with respect to the proportion of persons living in one-person households (Figure 6). In 10 out of the 18 examined countries, the share of persons living in one-person households is significantly smaller in the ESS than in the LFS. The largest underrepresentation pertains to Lithuania (-6.6 percentage points; ESS: 11.8%, LFS: 18.4%). However, there are also five countries where an overrepresentation of persons living in one-person households can be observed. The gap is largest in Hungary, where the percentage of persons living in one-person households is more than twice as high in the ESS than in the LFS (ESS: 22.3%, LFS: 10.6%).

Between three and six countries exhibit significant differences between ESS and LFS in the proportions of persons living in two-, three- and four-person households. For each of the three household sizes, we see both countries for which the proportions in ESS are below and countries for which the proportions in ESS are above the respective level in the LFS. The largest discrepancy pertains to Czech Republic (+8.2 percentage points): Here, the proportion of persons living in four-person households is 31.2% in the ESS and 23.0% in the LFS.

Persons living in large households (five or more persons) are underrepresented in five ESS countries. Slovakia exhibits the largest discrepancy (-5.9 percentage points): According to the ESS, 14.6% of the respondents live in households with five or more persons. In the LFS, the respective proportion is 20.5%. In another two countries, the proportion of persons living in five or more person households is larger in the ESS than in the LFS.

Figure 6: Persons living in households of
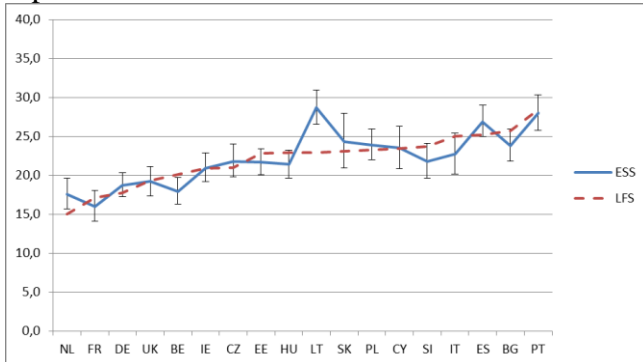   different sizes in ESS 6 and
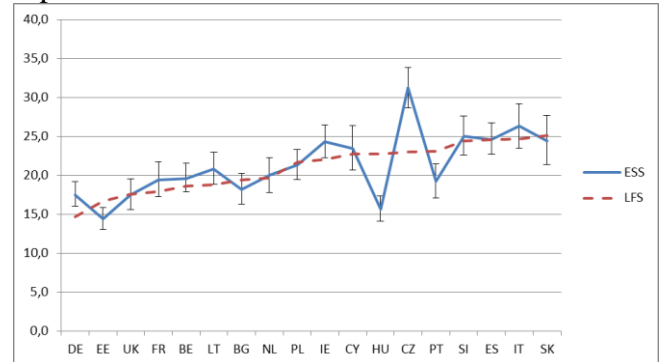   LFS 2012/2013 (in %)

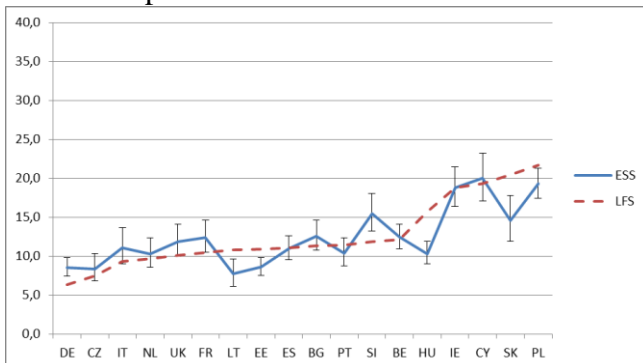## 1-person household
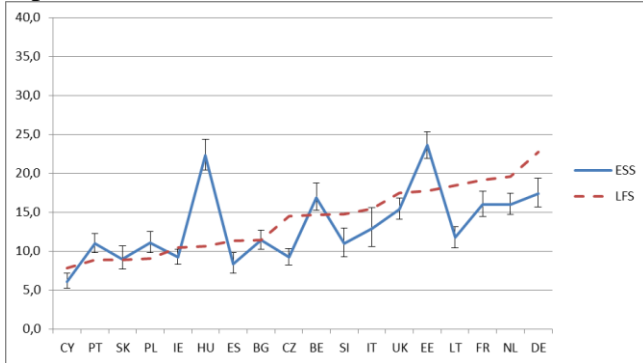
2-person household
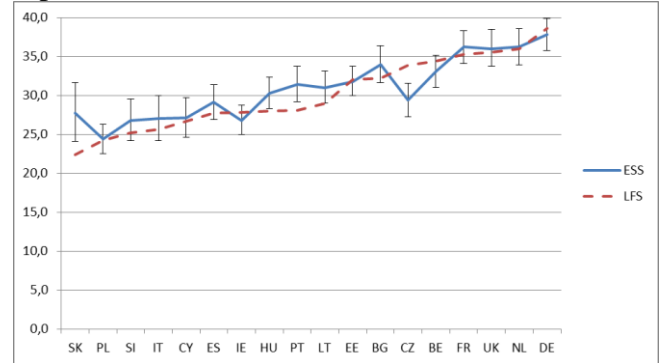
3-person household

4-person household

5-or-more-person household

6. A summary measure of ESS-LFS differences

In order to arrive at a summary measure for the consistency of ESS and LFS variable distributions we calculate the index of dissimilarity (Duncan & Duncan, 1955):

$$D = \frac{1}{2} \sum_{i}^{n} |ESSi - LFSi|,$$

with n = number of categories,
ESSi = percentage in category i of ESS,
LFSi = percentage in category i of LFS.

The index of dissimilarity (D) is a measure widely used in research on segregation. The range of the index is between 0 and 100. In the present context, a value of 0 indicates that there is no dissimilarity between the LFS and the ESS in the relative shares of respondents across the categories of a variable. A value of 100 indicates that the two distributions are completely dissimilar (consider, e.g., a dichotomous variable, where the first category comprises 100% in LFS and 0% in ESS, and the second category comprises 0% in LFS and 100% in ESS). The index of dissimilarity measures the percentage of respondents that would need to move between the categories of a variable to produce exactly the same distribution for the two surveys.

Table 4: Index of dissimilarity (D) between ESS 6 and LFS 2012/2013 distributions

Country	Gender	Age	Marital status	Work status	Nation-ality	House-hold size	mean
BE	0.1	2.6	0.3	3.4	2.1	3.6	2.0
BG	5.1	12.5	3.3	1.5	0.1	3.1	4.3
CH	1.1	4.1	3.5	2.3	6.1		3.4
CY	3.8	6.3	1.9	6.1	11.9	1.8	5.3
CZ	2.3	10.2	4.9	3.0	0.7	9.9	5.2
DE	1.4	5.6	1.8	2.8	4.5	6.1	3.7
DK	1.2	5.8	4.8	2.3	3.3		3.5
EE	3.6	5.7	2.0	0.9	0.7	5.9	3.1
ES	0.5	3.3	2.1	1.8	2.9	3.1	2.3
FI	0.4	4.5	2.4	1.1	0.1		1.7
FR	1.9	7.4	9.3	1.6	1.9	4.3	4.4
HU	1.9	4.2	3.3	2.5	0.6	14.0	4.4
IE	0.9	5.9	3.5	7.0	5.0	2.3	4.1
IS	0.9	4.6	4.3	3.4	3.8		3.4
IT	0.3	5.6	1.3	9.2	4.3	4.9	4.3
LT	1.6	7.1	2.1	2.4	0.2	9.8	3.9
NL	2.4	6.5	7.8	0.6	1.9	3.7	3.8
NO	2.5	4.3	3.8	4.2	1.8		3.3
PL	0.0	2.0	0.4	1.9	0.1	2.8	1.2
PT	7.0	5.4	0.9	7.7	0.3	5.4	4.4
SE	0.3	4.0	3.8	0.6	3.0		2.3
SI	3.2	4.6	0.7	8.6	0.2	5.8	3.8
SK	5.0	12.1	10.1	7.8	0.4	6.6	7.0
UK	5.5	7.2	6.9	6.8	3.7	2.3	5.4
mean	2.2	5.9	3.6	3.7	2.5	5.3	3.8

The mean value of D across all variables and countries is 3.8 (see Table 4).[14] This means that – on average – less than 4% of respondents in ESS would have to change categories in order to achieve the same distribution as in the LFS. D is highest for the variables age (mean 5.9) and household size (mean 5.3). To some extent, this is the consequence of these two variables having a larger number of categories than the remaining variables. The smallest D refers to the variables gender (2.2) and nationality (2.5). The latter is – at least in part – related to the skewed distribution of this variable.

The size of D varies both between countries and between variables. Except for gender, at each of the variables there is at least one country with a D of nearly 10 or even higher. The largest dissimilarity observed pertains to the variable household size in Hungary with a D of 14.0. On the other hand, at each variable there are usually a few countries with a rather low value of D. At four of the six variables (gender, marital status, work status, nationality), for instance, there are between three and ten countries with a D smaller than 1.

The mean value of D across the six variables varies between a low of 1.2 in Poland and a high of 7.0 in Slovakia (see Figure 7). Countries with a rather high average D typically show values well above-average in several variables (see Table 4).

Figure 7: Index of dissimilarity: mean value across six variables

---
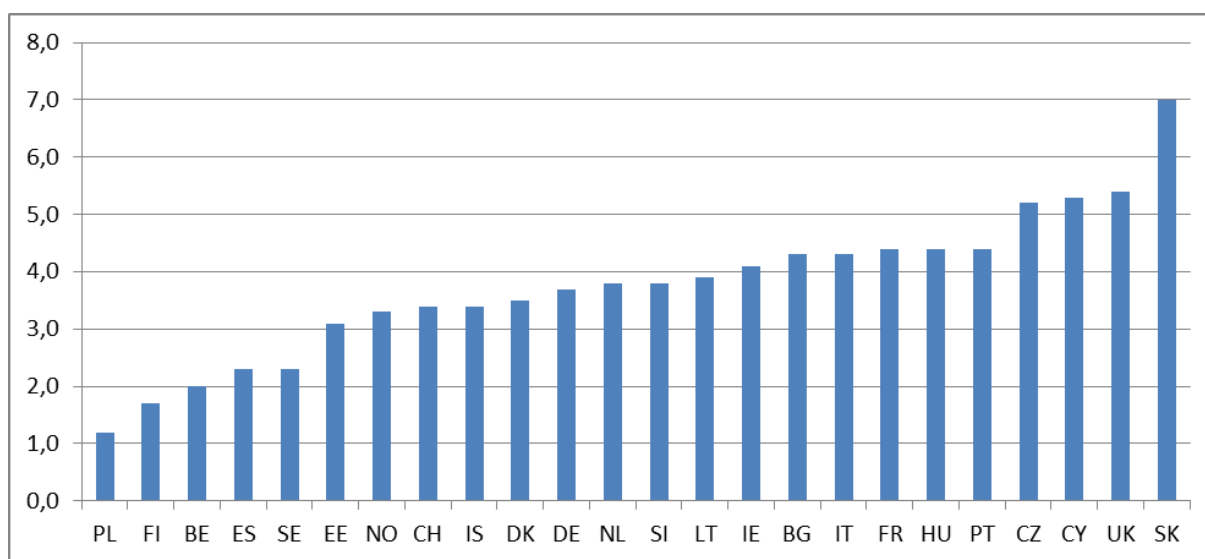
[14] In six countries (Denmark, Finland, Iceland, Norway, Sweden, Switzerland) the index of dissimilarity for household size is not available. In these countries, the average value of D is based on the remaining five variables.

7. Comparison of results for ESS 6 with results for ESS 5

Among the 24 countries included in the present analysis, 22 countries participated also in ESS 5 (only Iceland and Italy did not). For ESS 5, a similar exercise as the present one has been conducted, comparing ESS socio-demographic variable distributions with the respective data from the LFS 2010 (Koch et al., 2014). Table 5 shows, that the correlations (Pearson's r) between the indices of dissimilarity for ESS 6 and ESS 5 are positive for all variables, which means that larger deviations between ESS and LFS in one round come along with larger deviations in the other round. The mean index of dissimilarity, for instance, exhibits a moderate positive relationship of r = .55.

Table 5: Correlation between indices of dissimilarity for ESS 6 and ESS 5, 22 countries

Variable	Pearson's r	Number of countries (Missing countries)
Gender	.47	22
Age	.72	22
Marital status	.75	21 (FI)
Work status	.61	22
Nationality	.85	22
Household size	.24	17 (CH, DK, FI, NO, SE)
Mean D across 6 variables	.55	22

Figure 8 depicts the indices of dissimilarity for each of the six variables for the 22 countries that participated in ESS 5 and ESS 6. Despite the positive correlation between ESS 5 and 6 indices of dissimilarity, a number of countries exhibit considerable changes at some variables. Several of the largest discrepancies between ESS and LFS which could be observed for Round 5 tend to be smaller in Round 6. This refers, for instance, to Lithuania (gender discrepancy), Portugal (discrepancy for age and work status), and Cyprus (discrepancy for household size). A strong shift in the opposite direction – that is, a larger discrepancy in ESS 6 – occurred in Slovakia in the case of work status and in Hungary in the case of household size. All in all, the *average size* of the indices of dissimilarity across the 22 countries tends to be similar in ESS 5 and ESS 6 (see Table 6).

Table 6: Average size of indices of dissimilarity, ESS 5 and ESS 6, 22 countries

Variable	Index of dissimilarity ESS 5: average across 22 countries	Index of dissimilarity ESS 6: average across 22 countries
Gender	2.9	2.4
Age	6.3	6.0
Marital status	2.9	3.6
Work status	3.6	3.5
Nationality	2.4	2.3
Household size	5.6	5.3
Mean D across 6 variables	3.8	3.8

Figure 8: Indices of dissimilarity ESS vs. LFS,
        ESS 5 (blue) and ESS 6 (red),
        22 countries

Gender

Age
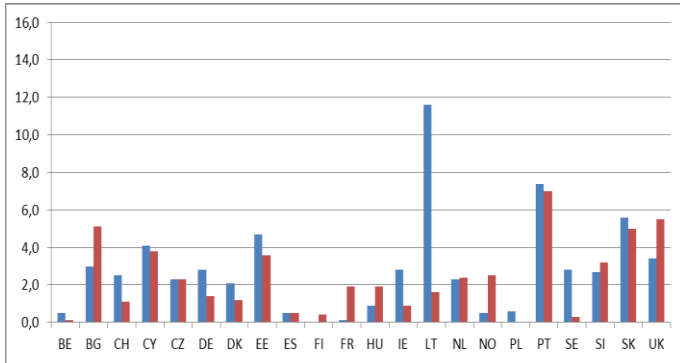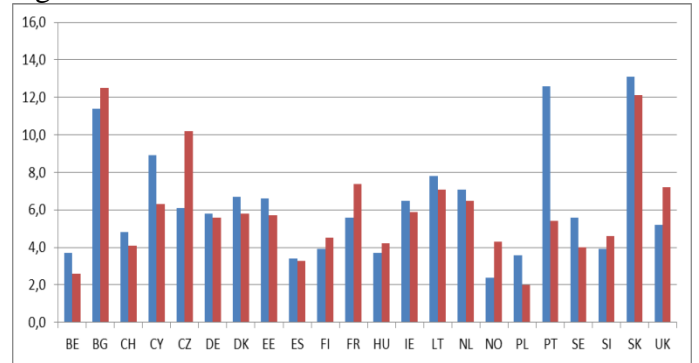
Marital status
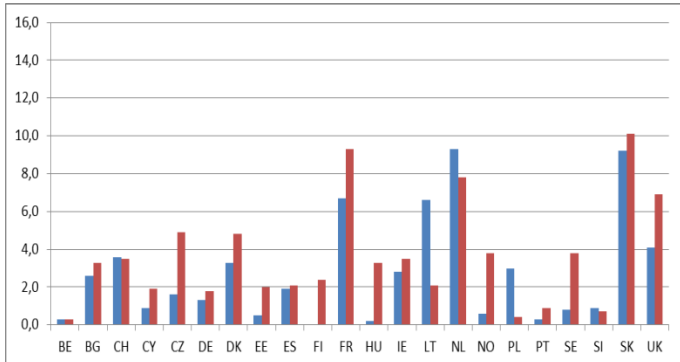
Work status
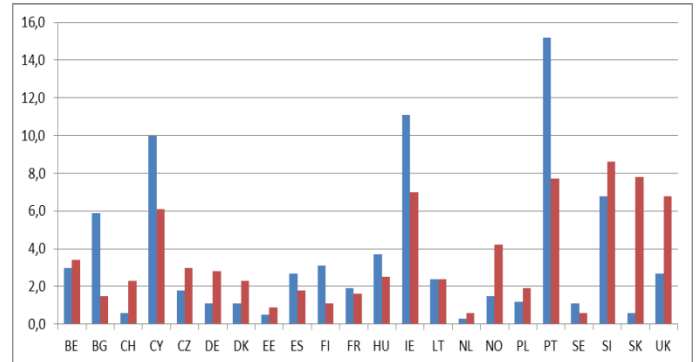
Nationality
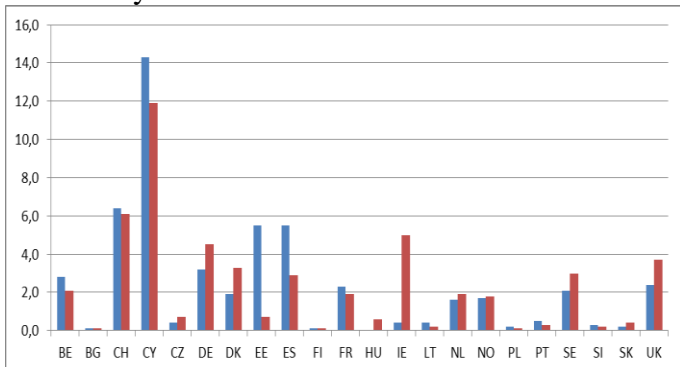
Household size
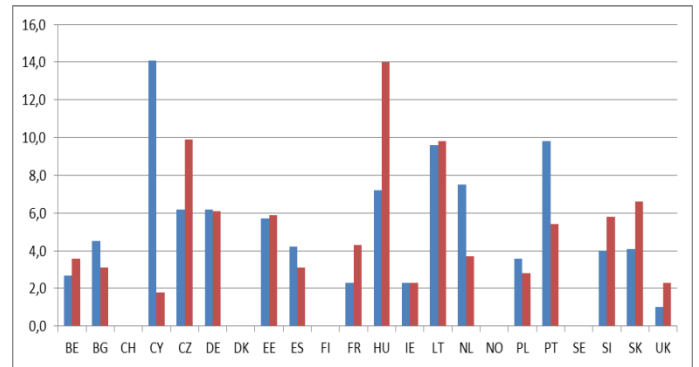
The index of dissimilarity is a non-directional measure. It does not provide an indication of whether certain demographic subgroups are over- or underrepresented. Figure 9 on the next page provides a graphical presentation of the direction and size of differences between ESS and LFS for both ESS round 5 and round 6. In this view, too, a positive relationship between the results of ESS 5 and ESS 6 shows up. On the level of individual countries, the direction of the differences observed for each variable/category usually is the same in both rounds of ESS. In no country, a strong *over*representation (larger than 3 percentage points) of a certain demographic group in Round 5 is followed by a strong *under*representation (larger than 3 percentage points) of the same group in Round 6 and vice versa.

For several variables/categories, the direction of the differences is the same for the large majority of countries in both rounds. This pertains, for instance, to the proportion of 25-34 years old persons, which are underrepresented in all countries in both rounds. In reverse, 55-64 years and 65-74 years old persons tend to be overrepresented in the majority of countries in both rounds. Also women and married persons tend to be overrepresented, whereas non-nationals are underrepresented in many countries in both rounds. The picture is more varied with respect to the proportion of persons in paid work and especially with respect to the proportion of persons living in households of different sizes.

The most obvious explanation for the patterns of under- and overrepresentation we observe will have to recur to differential response propensities of demographic subgroups. If, for instance, a certain group is particularly difficult to contact in a country, and/or the efforts to contact this group are below average in that country, then an underrepresentation of this subgroup will occur. Investigating these processes in more detail will require micro-level analyses of the response behaviour on the level of individual sample units or on the level of individual (interviewer) contact attempts for each country. This is not part of the present task. In the following we will, however, provide a basic analysis of two correlates of the size of dissimilarity between ESS and LFS data on the macro-level of countries.

Figure 9: Size and direction of differences between ESS and LFS data (ESS 5 vs. LFS 2010; ESS 6 vs. LFS 2012/2013; 22 countries)

	female		15-24y		25-34y		35-44y		45-54y		55-64y		65-74y		75+y		married		in paid work		non-national		1pers-hh		2pers-hh		3pers-hh		4pers-hh		5+pers-hh	
ESS round	5	6	5	6	5	6	5	6	5	6	5	6	5	6	5	6	5	6	5	6	5	6	5	6	5	6	5	6	5	6	5	6
BE																																
BG																																
CH																							xx	xx	xx	xx	xx	xx	xx	xx	xx	xx
CY																																
CZ																																
DE																																
DK																							xx	xx	xx	xx	xx	xx	xx	xx	xx	xx
EE																																
ES																																
FI																	xx						xx	xx	xx	xx	xx	xx	xx	xx	xx	xx
FR																																
HU																																
IE																																
LT																																
NL																																
NO															xx	xx							xx	xx	xx	xx	xx	xx	xx	xx	xx	xx
PL																																
PT																																
SE													xx	xx									xx	xx	xx	xx	xx	xx	xx	xx	xx	xx
SI																																
SK																																
UK																																

Light green: overrepresentation 0.1-3.0 percentage points; dark green: overrepresentation 3.1 percentage points and more
Light red: underrepresentation 0.1-3.0 percentage points; dark red: underrepresentation 3.1 percentage points and more
White: no difference at all; xx: information missing

8. Correlates of ESS-LFS differences

The previous sections revealed that the size of the differences between the ESS and the LFS varies between countries. How can we explain this variation in sample quality across countries? In the remainder of the paper we will briefly analyse whether the average index of dissimilarity across six variables (D) is related to two basic survey parameters: the response rate achieved and the sample design used. Since separate analyses for ESS 6 and ESS 5 basically led to the same results, we pool the data from ESS 6 and ESS 5. That way, the number of cases for our analyses can be nearly doubled (24 countries from ESS 6 and 23 countries from ESS 5; 47 cases in total).

For several decades the response rate achieved in a survey has been used as a proxy for the degree of nonresponse bias and the quality of the realised sample (Kreuter, 2013). With respect to the present analysis, we therefore should expect that higher response rates come along with smaller ESS-LFS differences. Empirically this is not the case. Figure 10 provides no evidence for a negative relationship between the (average) D per country and the response rate (Pearson's r = .14).
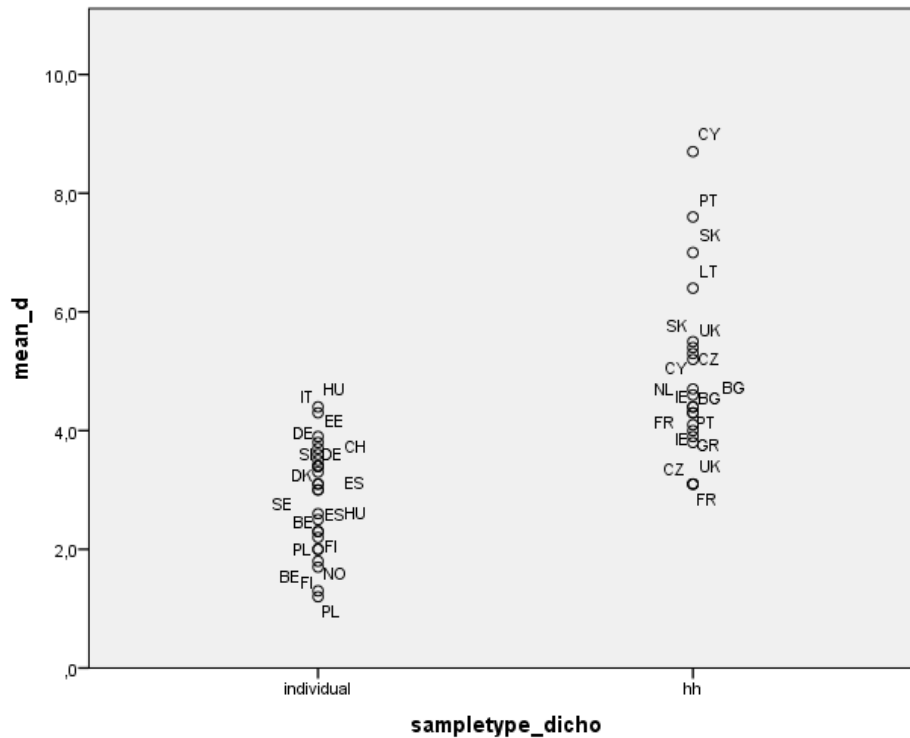
Figure 10: Index of dissimilarity by response rate (in %), ESS 6 and ESS 5
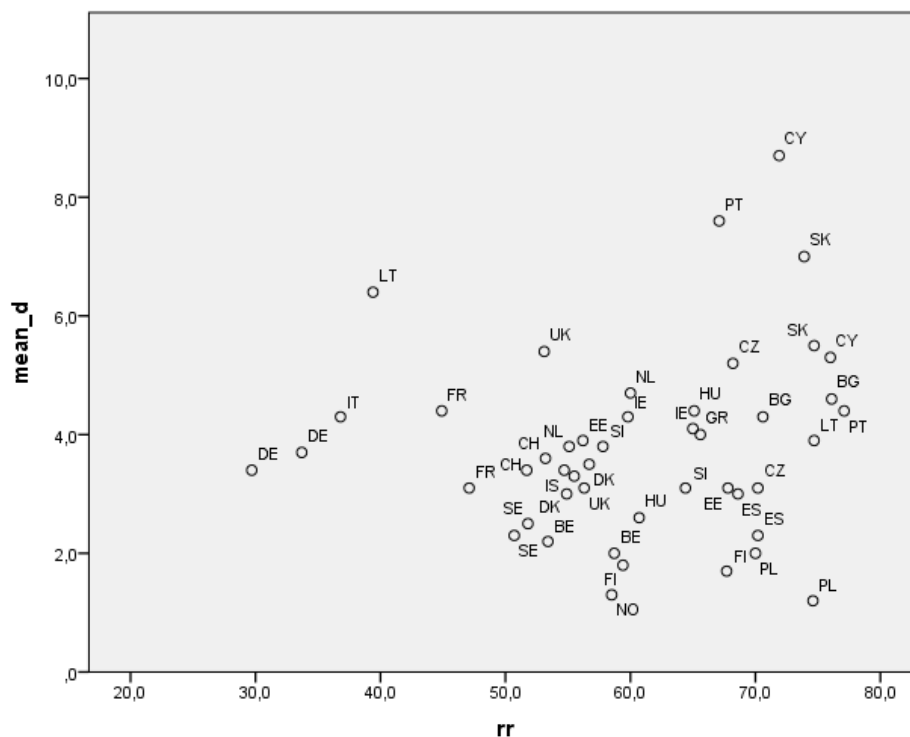
More pronounced differences can be found when we turn to the type of sample used (see Figure 11). ESS prefers countries to use a sample of named individuals from a register (European Social Survey, 2013). If such a sample is used, the sample is drawn without any involvement of the interviewers. Where a sampling frame of individuals is not available, countries may use a sampling frame of households or addresses. This can take various forms, differing among other things, in the degree of interviewer involvement required. One possibility is that the households/addresses come from a list, like a registry of telephone numbers or the customer directory of an electricity provider. If such a list is not available, the survey organisation and their interviewers have to enumerate the households, usually before

fieldwork starts. In all household/address designs, the interviewers have to select a respondent in the household, since the ESS interviews only one person per household. Among the 47 cases/countries included in our analyses, 26 countries used a sample of individuals and 21 countries used a sample of households or addresses in ESS 5 or 6. Figure 11 shows the average index of dissimilarity, separately for countries using a sample of individuals and for countries using a household/address sample. In the group of countries using a sample of households/addresses, the average size of D is considerably higher (mean = 4.9) than in the group of countries using a sample of individuals (mean = 2.9).

Figure 11: Index of dissimilarity by type of sample, ESS 5 and ESS 6 (n = 47)


A different pattern reveals itself when the relationship between the response rate and the index of dissimilarity is analysed separately for countries with a sample of individuals on the one hand, and countries with an address/household sample on the other hand (see Figure 12). For countries with samples of individuals a moderate negative relationship between the response rate and D can be observed (r = -.47). In line with conventional wisdom, the sample composition of countries with a higher response rate corresponds closer with the LFS data than the sample composition of countries with a lower response rate. For countries using a sample of addresses/households, however, this is not true. Here, a weak tendency to larger discrepancies with higher response rates shows up (r = .18).

Figure 12: Index of dissimilarity by response rate (in %), separately for countries with samples of individuals and countries with samples of households/addresses, ESS 5 and ESS 6

Sample of individuals: Pearson's r = -.47 (n = 26)
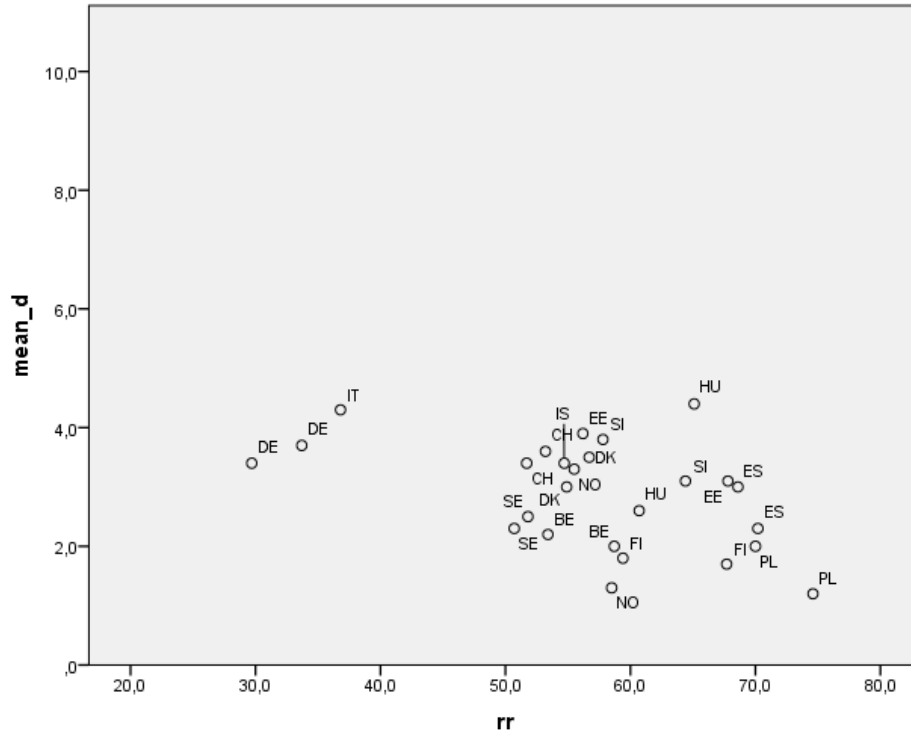
Sample of households/addresses: Pearson's r = .18 (n = 21)
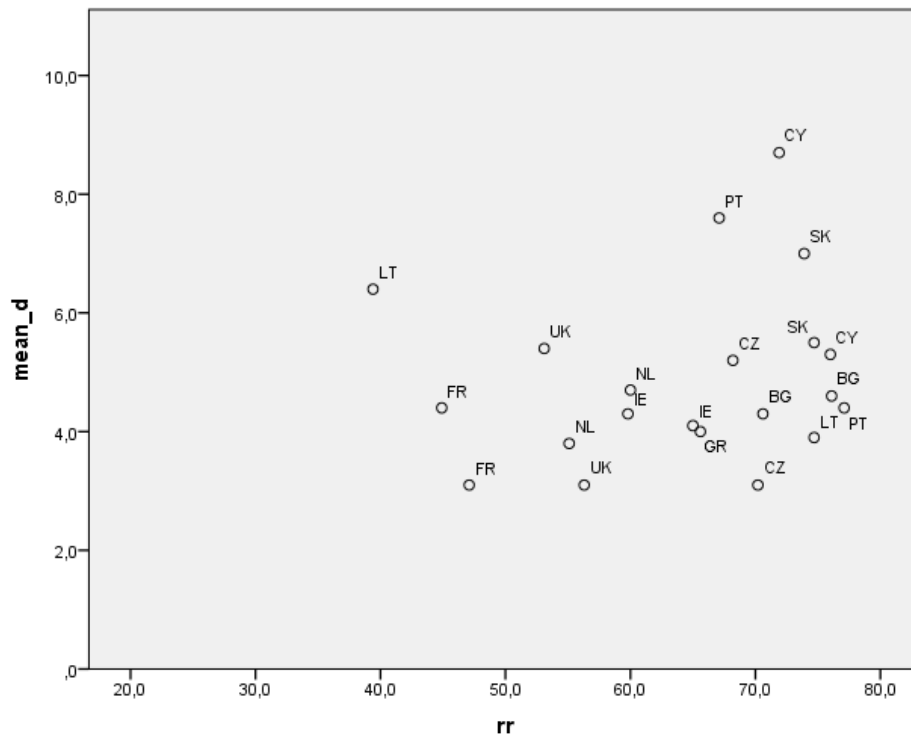
An explanation of these results might be achieved, when we take the role of the interviewers into account. As mentioned earlier, interviewers play an important role in sample selection when a sample of households/addresses is used. If interviewers do not follow the rules of random sampling properly – for instance during the listing of households before fieldwork, or during the selection of target persons within households – this might contribute to the patterns of results we observed. Some interviewers may preferentially select households and persons who are cooperative and at home, in order to keep their response rate high and to reduce the number of visits required. If interviewers, for instance, tend to substitute a reluctant male target person by his cooperative wife when selecting a respondent within a household, this will lead to an overrepresentation of women in the final sample (see Kohler, 2007; Sodeur, 1997). In addition, such a misconduct of interviewers will not only increase sample bias, but will also undermine the reliability of the response rate calculated. Undocumented substitution of reluctant target persons, as this behaviour might be termed, will lead to inflated response rates. Such processes might explain why sometimes (seemingly) high response rates come along with large biases in sample composition.

In terms of interviewer involvement, we can introduce a further distinction among countries using a sample of households/addresses and check whether our considerations also apply when using this more detailed classification. The degree of interviewer involvement in the sampling process is highest, when a household sample is used where the households have to be enumerated by the interviewers or other field staff (either before or during fieldwork). In this situation, the interviewer is involved in selecting households *and* persons. When a sample of households from a list is used, the influence of the interviewer is limited to the selection of a person within the household. Samples of individuals from a register are drawn without any interviewer involvement. Table 7 shows the average D and the average response rate for countries in ESS 5 and 6, separately for these three types of samples. As can be seen, the size of the average D increases the higher the interviewer involvement in the sampling process is. Also the average response rate becomes larger, the stronger the interviewer involvement in sample selection is. Both patterns are in accordance with our hypothesis that interviewer deviations from the rules of random selection might contribute to the differences we observe.

Table 7: Average D and average response rate by type of sample, ESS 5 and ESS 6

Type of sample	Average D	Average response rate (in %)	Number of cases (countries)
Sample of individuals	2.9	57.0	26 (14)
Sample of households/addr., from a list	4.7	60.1	14 (7)
Sample of households/addr., interviewer enumerates	5.3	72.2	7 (4)

However, a note of caution is in place when interpreting these macro-level results. The observed relationships rest on a small number of countries only, and they can be strongly influenced by (including or excluding) individual countries. What is more, is the observational (non-experimental) nature of our data. Sampling methods were not randomly assigned to the countries which participate in the ESS. As a consequence, countries and sampling methods are confounded, and the differences we observe between different types of samples do not necessarily imply a causal effect of the method 'per se'. It might be the case,

for instance, that some countries are more strongly affected by differential nonresponse of certain population subgroups than other countries (independently from the type of sample they use), or that countries differ in the degree of rigour they put on the correct implementation of the particular sampling method they deploy.

9. Summary and conclusions

(1) This paper used external benchmark data to analyse the socio-demographic sample composition in ESS 6. In several ESS 6 countries, the comparison with data from the LFS revealed only small differences for six socio-demographic variables (gender, age, marital status, work status, nationality and household size). At the same time, large differences were observed in a number of countries – sometimes only with respect to one variable, sometimes with respect to several of the variables examined. The basic pattern of results received for ESS 6 was rather similar to the one received from an analogous exercise for ESS 5. Notwithstanding this, on the level of individual countries large changes could be observed between the two ESS rounds a number of times.[15]

A detailed follow up of the reasons for the differences in sample quality between individual countries was not part of the present task. Generally speaking, the most obvious explanation for differences in sample quality between countries is that countries differ with respect to the response propensities of socio-demographic subgroups. If, for instance, a certain group is particularly difficult to contact in a country, and/or the efforts to contact this group are below average in that country, then an underrepresentation of this subgroup will occur. Each ESS country is invited to check its own results and to judge whether they make sense, given the available insights into the societal conditions and the fieldwork efforts exerted in that country. As the case may be, further country-specific analyses should be considered.

(2) A correlational analysis including ESS 6 and ESS 5 data showed that the type of sample seemed to be important in predicting the size of the deviation from the LFS. ESS countries with a sample of individuals exhibited smaller discrepancies from the LFS on average than countries using a household/address sample. When the relationship between response rates and sample quality was analysed separately for countries with a sample of individuals on the one hand, and countries with a sample of households/addresses on the other hand, different patterns were observed. Whereas for countries with a sample of individuals sample quality was positively related to the response rate (as one usually expects), the direction of this correlation was the other way round for countries with a sample of households/addresses: Among this group, countries with high response rates revealed a somewhat lower sample quality than countries with low response rates.

Such a pattern cannot be explained sufficiently by referring exclusively to potential differences in subgroups' response propensities. A satisfactory explanation might be achieved by also taking the interviewer and his/her behaviour into account. Interviewers play an important role in sample selection when a sample of households/addresses is used. If interviewers do not follow the rules of random sampling properly this might contribute to the patterns of results we observed. In particular countries with high response rates and large

---

[15] In the course of preparation for ESS round 8, the results of the sample composition assessment both for Round 6 and for Round 5 were fed back to countries on an individual basis (this was done in March 2016). That way, the information could be taken into account when planning national fieldwork in the countries participating in ESS round 8.

discrepancies from the LFS should therefore check whether their system of quality control back-checks is sufficient to prevent and/or detect interviewer misbehaviour in sample selection.

(3) Finding out about the causes of our results is but one part of the challenge. The other is to discuss the implications these results may have for the planning of future ESS survey rounds. If we assume, that the observed differences can be interpreted as indications of nonresponse bias, an obvious question is whether dedicated efforts to balance response rates for socio-demographic subgroups are desirable at least in some ESS countries. A study of Peytcheva and Groves (2009) might cast doubt as to whether aiming for balanced response rates in demographic subgroups is important. They found that bias in demographic variables is not predictive of the difference between respondents and nonrespondents in substantive variables of the same survey. If we nevertheless came to the conclusion that balancing response rates is a good thing, further questions arise: How should balanced response rates be achieved? Every effort to balance response rates needs additional data on the gross sample to allow for targeted fieldwork efforts. In many surveys – in particular in surveys based on samples of households/addresses – such information is not (routinely) available. And even if the relevant information is available, it still has to be decided how to proceed. For instance: What should be done if balancing is desirable for several variables? Which measures are most appropriate to raise response rates in specific subgroups, etc.? In a nutshell: There are many obstacles to overcome in order to achieve a balanced sample in the end. A discussion within the 'ESS community' about the desirability and feasibility of efforts to achieve balanced response rates seems recommendable.

(4) To complicate things further, one might finally ask whether a balanced sample might not just as well be achieved by simply applying post-stratification (PS) weights. Using PS-weights would be much easier to implement, and relying on them would be a less expensive way to deal with the issue of nonresponse bias (and sampling error) in socio-demographic variable distributions. Originally, it was planned to analyse as part of the present task whether the application of the PS-weights provided in ESS in the recent past would help to reduce the discrepancies with the LFS. Due to an error in the PS-weights provided for ESS 6, this could not be done. Corrected weights have been made available in 2016, and it still seems worthwhile to investigate their effect in the next ERIC work programme.

(5) Another strand of potentially promising work is to move into the direction of a process-oriented approach. The present investigation only had a look at the composition of the final sample in each country. More detailed insights might be gained by analysing how the discrepancies which we observed evolved during fieldwork. This requires some kind of sequencing of the completed interviews for each country, for instance simply by the date at which the interviews were conducted. A few preliminary analyses of this kind have already been made. In order to receive a more complete picture and to evaluate in more detail the potentialities of such an approach, however, further analyses are needed. Such an exercise could be seen in line with the request of the ESS Methods Advisory Board (MAB) for a stronger emphasis on monitoring and control of fieldwork in ESS countries. It might therefore be discussed whether these process-oriented analyses should be included, provided a continuation of the sample quality assessment work is planned for future rounds of ESS.

References:

Beullens, K., Matsuo, H., Loosveldt, G. & Vandenplas, C. (2014). Quality report for the European Social Survey, Round 6. European Social Survey ERIC

Duncan, O. D. & Duncan, B. (1955). A Methodological Analysis of Segregation Indexes. American Sociological Review, 20, 210-217

European Social Survey (2013). Round 7 Specification for Participating Countries. London: Centre for Comparative Social Surveys, City University London

European Social Survey (2014). Documentation of ESS Post-Stratification Weights. www.europeansocialsurvey.org, 25th April 2014

Eurostat (2013). Labour force survey in the EU, candidate and EFTA countries. Main characteristics of national surveys, 2012. 2013 edition. eurostat Statistical working papers

Eurostat (2014a). Quality report of the European Union Labour Force Survey – 2012. 2014 edition. eurostat Statistical working papers

Eurostat (2014b). Labour force survey in the EU, candidate and EFTA countries. Main characteristics of national surveys, 2013. 2014 edition. eurostat Statistical working papers

Eurostat (2014c). Quality report of the European Union Labour Force Survey – 2013. 2014 edition. eurostat Statistical working papers

Groves, R. M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. In: Public Opinion Quarterly, 70, 646-675

Koch, A., Halbherr, V., Stoop, I.A.L. & Kappelhof, J.W.S. (2014). Assessing ESS sample quality by using external and internal criteria. Mannheim: European Social Survey, GESIS

Köhne-Finster, S. & Lingnau, A. (2009). Untersuchung der Datenqualität erwerbsstatistischer Angaben im Mikrozensus. Ergebnisse des Projekts "Nachbefragung im Mikrozensus/LFS". Wirtschaft und Statistik, 12/2008, 1067-1088

Kohler, U. (2007). Surveys from inside: An assessment of unit nonresponse bias with internal criteria. Survey Research Methods, 1, 55-67

Kreuter, F. (2013). Facing the Nonresponse Challenge. The ANNALS of the American Academy of Political and Social Science, 645, 23-35

Peytcheva, E. & Groves, R. M. (2009). Using Variation in Response Rates of Demographic Subgroups as Evidence of Nonresponse Bias in Survey Estimates. Journal of Official Statistics, 25, 193-201

Sodeur, W. (1997). Interne Kriterien zur Beurteilung von Wahrscheinlichkeitsauswahlen. ZA-Information, 41, 58-82

Thomsen, I. & Villund, O. (2011). Using Register Data to Evaluate the Effects of Proxy Interviews in the Norwegian Labour Force Survey. Journal of Official Statistics, 27, 87-98

Zühlke, S. (2008). Auswirkungen von Proxy-Interviews auf die Datenqualität des Mikrozensus. P. 3-10 in: Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen (ed.): Statistische Analysen und Studien NRW, Band 53