Survey Research Methods (2016)
Vol. 10, No. 2, pp. 103-118
doi:10.18148/srm/2016.v10i2.6261

ISSN 1864-3361
http://www.surveymethods.org

Interviewer Effects in the European Social Survey

Koen Beullens
Centre for Sociological Research
Leuven, Belgium

Geert Loosveldt
Centre for Sociological Research
Leuven, Belgium

In this paper, we focus on interviewer effects in the European Social Survey, and seek to demonstrate that academic publications seldom take these effects into consideration. An analysis is provided of interviewer effects for 48 continuous items, covering 36 countries in six rounds. The analysis does not only deal with the means of variables. Using multilevel covariance structure analysis, interviewer effects on the relationships between these variables is also assessed. Results indicate that first, countries showing considerable interviewer effects regarding means are also more at risk regarding regression coefficients and second, ignoring interviewer effects leads to an overestimation of the effect size of the relationships between variables and an underestimation of standard errors.

Keywords: Interviewer effects, European Social Survey, multilevel covariance structure analysis

1 Introduction

The ESS is a biennial, multi-country survey (European Social Survey, 2002–2012). The data is readily accessible, and many researchers in the field of social sciences use it for their publications. For 2013, some 381 ESS-based publications (including books, book chapters, journal articles, and working papers) can be found using a Google Scholar search.

The data-collecting countries as well as the Core Scientific Team of the ESS make great efforts to achieve high quality standards with regard to sampling, data collection, quality assessment, archiving, and dissemination. Nevertheless, as in most surveys, many errors may emerge. This paper focuses in particular on interviewer effects.

In literature concerning surveys, there seems to be some agreement about the fact that interviewers affect the answers of respondents. In 1929, Rice observed that interviewers used their own framework in order to interpret the situation of homeless people. Further, Boyd and Westfall (1955) found that interviewers were an important source of error in marketing surveys, as interviewers can fail to recruit respondents, incorrectly recruit them, incorrectly stimulate respondents to give answers, or incorrectly interpret and record answers.

Usually, the intraclass correlation (Kish, 1962) (or intra-interviewer correlation) is used to express the proportion of variance that is attributable to the interviewer-identification variable. Most intraclass correlations (ICCs) that are found throughout relevant literature range between approximately 0.00 and 0.05, sometimes increasing to 0.10 and with some outliers exceeding this figure (see, among others: Freeman & Butler, 1976; Groves & Magilavy, 1980, 1986; Kish, 1962; Mangione, Fowler, & Louis, 1992; Tucker, 1983). However, internal ESS reports have consistently pointed out that for some countries the intra-interviewer correlations can even be much higher (Beullens & Loosveldt, 2013; Loosveldt & Beullens, 2010; Philippens & Loosveldt, 2004). For this reason, we exclusively focus on the ESS in this paper.

Interviewer effects can reflect area effects whenever interviewers are systematically assigned to exclusive areas (Collins & Butcher, 1982; O'Muircheartaigh & Campanelli, 1998; Schnell & Kreuter, 2005). This may be particularly problematic to disentangle, as what are termed interpenetrated designs (Mahalanobis, 1946) are hard to realize in face-to-face surveys because of cost considerations. If a face-to-face survey only needs to cover a relatively small area (for example a city or province), it is feasible to randomly assign interviewers to sample cases. However, in a nationwide survey such as the ESS, interpenetration can only be realized locally. This is why many studies trying to separate interviewer and area effects in nationwide surveys use designs where addresses in local sampling points or areas are randomly assigned to interviewers, and where interviewers are assigned to only one or a few sampling points (Biemer, 2010; Biemer & Stokes, 1985). As an alternative to interpenetrated designs, the same respondents can be interviewed by different interviewers (Biemer & Stokes, 1985; Schaeffer, Dykema, & Maynard, 2010). It is obvious that such reinterviews are only possible in the context of panel surveys or other design that are not easily applied to the ESS.

Contact information: Koen Beullens, Centre for Sociological Research, Parkstraat 45 – box 3601, BE 3000 Leuven, Belgium (e-mail: koen.beullens@soc.kuleuven.be).

Not only might interviewers and areas be confounding factors, but it is also difficult to distinguish between measurement and selection error (e. g. nonresponse). Interviewer measurement effects refer to interviewer characteristics (e. g. ethnicity, gender, and age) or interviewer behavior (e. g. the question reading pace, closely following the interviewer instructions, etc.) that affect respondents' answers. On the other hand, interviewers might achieve differing response rates (Schaeffer et al., 2010; West, Kreuter, & Jaenichen, 2013; West & Olson, 2010) and/or recruit different kinds of respondents. Similarly, observed differences between areas can reflect true differences between local communities, but may also be the consequence of different local response rates.

Whatever the source of correlated responses (interviewer or area, measurement or selection), the fact that respondents' answers are correlated or clustered is troublesome for researchers analyzing ESS data. The first question this paper addresses is whether survey researchers using ESS data take interviewer/area clustering into account in their analyses. We therefore screened 221 journal articles published in 2013 that use data from the ESS. Our expectations are that interviewer effects (or clustering attributable to sampling points) are rarely addressed in substantive research based on ESS data. In a recent publication, Elliott and West (2015) found that among 20 publication, systematically sampled from 1650 publications based on data from the Behavioral Risk Factor Surveillance system (BRFSS), none took into account the clustering of respondent data by interviewers.

Second, this paper seeks to gauge to what extent the ESS is affected by interviewer effects in the different participating countries and over the first six rounds of the survey. Interviewer effects are usually measured over one variable, estimating the amount of variance that is due to the interviewer-identification variable. We also want to expand the research to investigate the effects of interviewers on bivariate analysis. This seems to be most relevant, as the majority of publications based on ESS data do not merely target the mean (or a proportion) of a variable, but aim to measure relationships between survey variables. In fact, interviewer effects on the relationships between variables have only rarely been assessed. Therefore, we apply multilevel covariance structure analysis, in which the relationship between two (continuous) survey variables is modeled at the respondent level as well as at the interviewer-level, assuming that the respondent-level parameter reflects the situation in the absence of interviewer effects. It is expected that interviewers influence the effect sizes of regression parameters, and that interviewer effects also increase the standard errors of these regression parameter estimates.

2 Are interviewer effects being taken into account in journal articles using ESS data?

We used a list of 221 journal articles that were published in 2013, based on a Google Scholar search[1]. Further conditions for inclusion were that the articles should be in English and actively use ESS data. In addition, 24 strictly methodological articles were omitted, so that only substantive research articles were included. Some articles may use the ESS in combination with other sources of data, such as country-level variables (e. g. GDP per capita, crime rates, and immigration data) or other survey sources (e. g. the Survey of Health, Ageing and Retirement in Europe (SHARE) and the European Values Study). Some of the journals that frequently publish empirical articles based on the ESS are Sociology of Health & Illness (3 articles), European Sociological Review (17), Journal of European Social Policy (3), Journal of Cross-Cultural Psychology (3), Party Politics (6), Social Indicators Research (19), European Journal of Ageing (3), International Journal of Public Opinion Research (2), Comparative Political Studies (2), Journal of Happiness Studies (2), Personality and Social Psychology Bulletin (2), PloS one (2) and Electoral Studies (2) (*This fully detailed list is available upon request*).

Using the PDF format of the 221 journal articles, we counted keywords in order to assess the extent to which the articles took interviewer/area effects into consideration. The keywords are "interviewer", "cluster", "municipality", "PSU", "design effect", "intraclass", and "multilevel". These were counted automatically using the *pdfgrep* command in Linux.

In order to make sure that every PDF document was properly legible, checks were performed on whether text strings such as "European Social Survey", "ESS", "table", "figure", "graph", "analysis", or "data" could be found. We found positive results for these keywords in all the PDF files. Moreover, by counting this last set of words, it was possible to evaluate whether ESS data had actually been used for each article. Originally starting with 259 articles, we subsequently omitted 15, because we could not find sufficient indications that these articles used ESS data for original empirical analysis. A few of these papers were strictly theoretical and referred to the ESS without actively using its data. Other articles used ESS data, but only by referring to other publications. As previously stated, a further 23 articles were omitted because they were predominantly methodological, bringing the total number of reviewed articles to 221.

The frequencies with which the keywords could be found are provided in Table 1. Only 13 articles mention the word "interviewer', of which 10 mention it only once. All the ref-

[1]Source: ESS Annual Bibliographic Report, November 2014, prepared by Brina Malnar, University of Ljubljana. A list of these references can be provided by the authors on request.

Table 1
Frequencies of text strings appearing in 221 journal articles based on ESS data

Text string	Appearing in n articles
interviewer	13
municipality/municipalities	12
PSU	0
cluster	81
intraclass	22
multilevel	106
design effect	2

erences to interviewers in these articles are only for descriptive purposes, e. g. the fact that interviewers needed to use showcards or were instructed to explain some concepts used in the questionnaire, whether or not the interviewer perceived respondents" language problems, etc. In fact, there is no single indication of the use of the word "interviewer" that relates to the concept of interviewer effects.

We found 12 articles that made use of the terms "municipality" or "municipalities", 11 of which only use it once incidentally. One article (Fasel, Green, & Sarrasin, 2013) extensively uses the municipality level as a contextual factor to explain attitudes toward immigrants. Therefore, this level of observation was considered from a substantive point of view, not as a methodological obstacle.

The term "cluster" appears in 81 articles. However, the term never refers to small groups of respondents sampled from the same neighborhood or visited by the same interviewer. Conversely, the concept of clustering usually pertains to cluster analyses (mostly ESS countries) or clusters of theoretical concepts (e. g. political ideologies). In other instances, clustering refers to the grouping of respondents at the country level or sub-country level (e. g. regions, counties, or provinces). In such cases, multilevel analyses were applied, meaning terms such as "intraclass correlation", "intra-class correlation", or "ICC" were also used.

The text string "design effect" only appears in papers that take design weights into consideration due to unequal sampling probabilities. This particularly applies to countries that use address-based or household-based sample frames.

In sum, it seems that in the ESS interviewer/area effects have manifestly escaped the attention of researchers.

3 Taking interviewer effects into account when assessing the relationships between variables

Most of the articles discussed in the previous section estimate relationships between variables, rather than assessing the variables in a univariate way. Therefore, we seek to address the effect interviewers have on bivariate estimates. However, the way univariate statistics are affected by interviewer effects may be seen as an interesting starting point for assessing the effects of interviewers on bivariate analysis.

3.1 Interviewer effects for separate variables

Ideally, intraclass correlations should be (close to) zero, indicating that respondents' answers are not correlated within interviewers. The underlying multilevel null model is denoted as $y_{ij} = \gamma_{00} + \mu_{0j} + \varepsilon_{ij}$ (Hox, 1994). In this model, the continuous survey variable y for respondent i assigned to interviewer j has an overall mean γ_{00}. Interviewer deviations from this overall mean are expressed by μ_{0j}. The more disperse the μ_{0j} values are, then the more the interviewer variance. The random interviewer terms and the residual errors are assumed to be mutually independent and normally distributed with mean zero and variance σ^2_{int} and σ^2_{ε}. The intra-interviewer correlation is the variance σ^2_{int}, proportional to the sum of the interviewer variance σ^2_{int} and the residual variance σ^2_{ε} (Kish, 1962). A similar model can also be formulated to accommodate differently distributed outcomes such as binary variables. However, since the analyses in this paper will be applied to 48 continuous variables in 150 different country-round combinations of ESS, extending the model for continuous variables to other outcomes would make the presentation of the analyses needlessly complicated.

A prevalent problem when investigating interviewer effects is the fact that interviewers are often assigned to locally-demarcated areas. This might result in area effects erroneously being taken as interviewer effects. However, it is important to underline that in this paper we do not seek to prove the existence of interviewer measurement effects in the ESS, but instead aim to outline what the risks are for substantive research if the correlated responses of respondents found in the data are attributable to interviewer measurement effects. Nevertheless, it seems useful to try to disentangle interviewer effects and area effects, in order to assess whether it is adequate or plausible to further develop a model that explores the consequences of interviewer effects.

Therefore, we measured intra-interviewer variance by also taking the area clustering into account. Unfortunately, area IDs are not available for the first four ESS rounds. Further, in some countries, the lack of interpenetration (e. g., Slovakia and Russia) is so manifest that it hinders the convergence of the estimation process of a cross-classified multilevel model where interviewers and areas are random effects. As an alternative, the multilevel model as detailed above is extended, adding two proxy area-defining variables as fixed effects (Hox, 1994; Hox, de Leeuw, & Kreft, 1991). In this regard, Hox (1994) stated that:

> "The hierarchical regression model offers an elegant way of analyzing the simultaneous effect of specific interviewer and respondents characteristics. It is especially attractive if the research design does not pro-

> vide for a random assignment of respondent to interviewers, because it allows the researcher to use statistical rather than experimental control by modeling the interviewer effects conditional on the respondent effects" (p. 300).

In addition, Groves and Couper (1998) indicated that when there is no sufficient interpenetration in a nationwide survey, statistical controls can be used to approximate the randomization. Therefore, we include region (a categorical variable, usually referring to a sub-national entity such as a county or province) in combination with a variable that measures the population density in which the respondents live (self-reported by the respondents). Both fixed-effect variables are also combined into an interaction effect in order to allow for a more refined approximation of the area effect. In line with Hox (1994), we assess the proportional reduction of interviewer variance by comparing this variance before and after the inclusion of the fixed effects.

Two potentially opposing consequences should be taken into account when using this method to separate interviewer and area effects. First, controlling for region and population density (including the interaction) may not take sufficient area effects away from the interviewer effects. In this case, the interviewer effects may still be overestimated. On the other hand, because interviewers are usually assigned to areas close to where they live, in order to save on transportation costs, interviewers may be locally clustered themselves (Groves & Couper, 1998). This might lead to a situation where interviewers and respondents are sometimes unintentionally matched for characteristics such as race (Groves & Couper, 1998). Further, locally-clustered interviewers may have participated in the same interviewer training sessions and briefings, they may have been monitored locally, and they may even share the same local beliefs, attitudes, or accent. Therefore, local interpenetration - or alternatively adding area variables in the multilevel model in order to separate area effects from interviewer effects - may also wrongfully attribute real interviewer effects to area effects.

3.2 Interviewer effects on the relationships between variables

We are particularly interested in the way interviewer effects can affect the relationship between two variables. This is particularly relevant, as most journal articles based on ESS data go beyond univariate analysis. Unfortunately, there is little research in survey literature assessing interviewer effects for bivariate or even multivariate statistics. Davis and Scott (1995) discussed the effects of interviewers on domain comparisons. They found that the effect of interviewer variability on the response variance is smaller when interviewers recruit respondents from two domains. Such domain comparisons can be seen as a form of bivariate association. Using multilevel covariance structure analysis, Beullens and Loosveldt (2014) reported that factor loadings in measurement models can be biased when interviewer effects are ignored, and they therefore speculated on the possibility that interviewer-specific deviations for one variable might be correlated with interviewer-specific deviations for another variable. Hox (1994), Wiggins, Longford, and O'Muircheartaigh (1992) have presented multilevel models in which interviewer effects are measured on a dependent variable (e. g. interview speed in the case of Hox (1994) or annoyance or bother with aircraft noise in the case of Wiggins et al. (1992)), including respondent characteristics as independent variables and using the interviewer as the macro-level. In such a setting, the interviewer impact on the relationship between two variables can be assessed. The particular advantage of such a model is that it allows for random slopes: e. g. is the effect of an individuals' psychiatric status on aircraft annoyance different between interviewers? Unfortunately, this model only allows the dependent variable to be affected by interviewer effects, whereas the independent are assumed to be "taken for granted". In this paper, we wish to explore the situation were both variables are prone to interviewer effects.

In order to explore a possible way in which interviewers might alter the relationship between two variables, consider the two graphs in Figure 1, which are based on fictitious data concerning the relationship between calorie intake and body mass index (BMI)[2].

Suppose that a survey collects data for n = 100 individuals about their calorie intake per day (x-axis) and their BMI (y-axis). Each of the individuals is exclusively and randomly assigned to two interviewers. Furthermore, we assume that there is no nonresponse and both interviewer A and interviewer B have an equal workload of 50 respondents. In the case where these two variables are objectively measured (using the appropriate measurement tools) and automatically recorded, it can be assumed that the interviewers do not affect the measurements. The left-hand panel of Figure 1 reflects this situation, where an existing but rather weak relationship between the two variables is observed. The letters A and B indicate the two interviewers. Because we assume there cannot be any effect of the interviewers, the letters A and B are randomly scattered in the cloud of observations, indicating that the interviewer variances with regard to both variables are negligible and can be ignored.

In the right-hand panel of Figure 1, the same individuals have been interviewed by the two interviewers, but in this situation, the respondents reported their calorie intake and BMI directly to the interviewers without using the appropriate measurement tools. As a result, interviewer effects may

[2] This example is inspired by Eisinga, te Grotenhuis, Larsen, and Pelzer (2011), where evidence was found that interviewer BMI is positively related to the respondents reporting restrained eating.

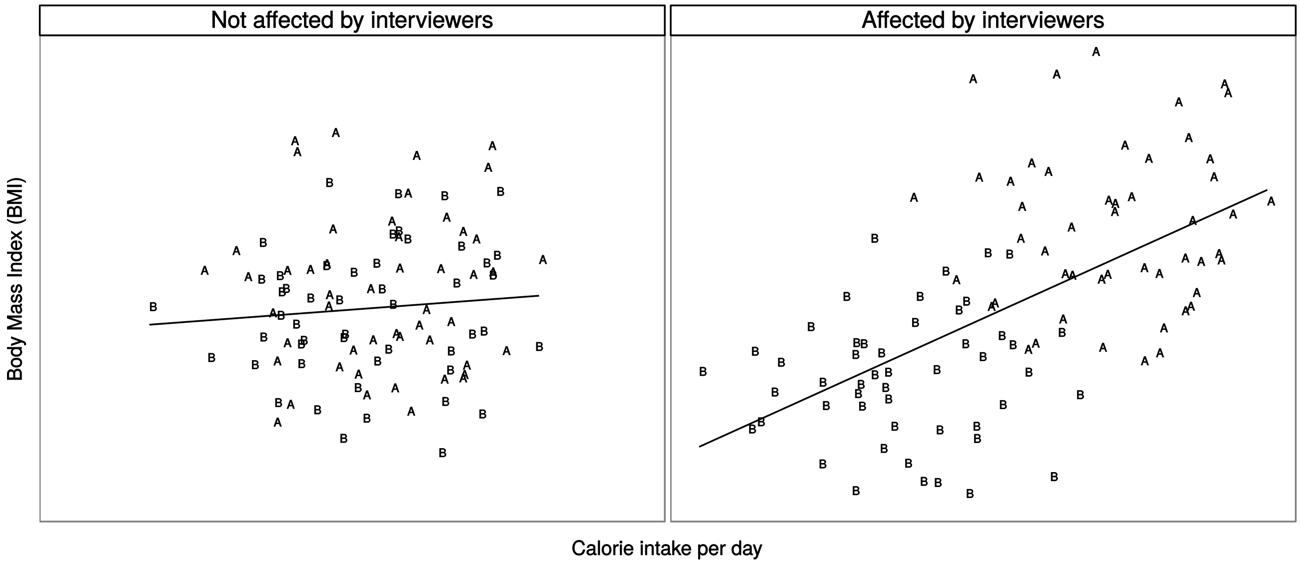

Figure 1. Effect of interviewers on variances of and correlations between the variables calorie intake and Body Mass Index (fictitious data)

emerge. It appears that respondents reporting to interviewer A provide answers that tend toward the positive ends of both variables, whereas interviewer B collects answers tending toward the negative ends of both variables. Not only do these interviewer effects increase the variance of the two variables and thus add "noise" to the variables separately, but their relationship also seems to be affected. In this particular case, the relationship between calorie intake and BMI becomes considerably stronger.

In this example, the relationship between both variables is positive on both the respondent as on the interviewer level. However, it is not inconceivable that the relationship between the two variables on the interviewer level is in the opposite direction as the same relationship on the respondent level. In such a case, when ignoring the interviewer effect the observed correlation between the two variables is weaker or might even become negative as compared to the situation where interviewers did not affect the data.

What could also happen, but is not observable in Figure 1, is the possibility of random slopes. For example, interviewer A might influence interviewees so that the relationship between the variables is stronger compared with those for interviewer B. This might also affect the overall relationship (when interviewer effects are included) compared with the ideal situation where there are no interviewer effects. This might be particularly relevant in a case where interviewers deal with different-sized groups of respondents. In such an unbalanced case, the relationship between two variables might be different before and after taking interviewer effects into account.

Because of the possible risk of interviewer effects, researchers analyzing surveys may want to divide the covariance matrix Σ_T (T referring to the "total") for two survey variables into two separate matrices Σ_B and Σ_W, the first referring to the between-interviewer structure and the second to the within-interviewer structure. When ignoring interviewer effects, researchers (unknowingly) apply their analyses to the total covariance structure, whereas the within-interviewer covariance structure is the appropriate structure to take into account interviewer effects.

We use multilevel covariance structure analysis (B. Muthén, 1994) for the purpose of separating these matrices. It has been used infrequently with regard to interviewer effects, but has gained some popularity throughout recent decades in other fields of research in which hierarchical data structures can be found. For example, Elovainio, Kivimäki, Steen, and Vahtera (2004) used multilevel covariance structure analysis to take the nested responses of employees within work units into account when examining the relationship between job decision latitude, organizational justice, and employee health in Finland. Duncan, Duncan, Hops, and Alpert (1997) used the technique for research on intra-familial substance use.

The total covariance matrix can be separated into the within-level (S_{PW}) and between-level (S_B) counterparts with respect to a multivariate vector y:

$$S_{PW} = (N - J)^{-1} \sum_{j=1}^{J} \sum_{i=1}^{N_j} \left(y_{ij} - \bar{y}_j\right)\left(y_{ij} - \bar{y}_j\right)' \qquad (1)$$

$$S_B = (J - 1)^{-1} \sum_{j=1}^{J} N_J \left(\bar{y}_j - \bar{y}\right)\left(\bar{y}_j - \bar{y}\right)' \qquad (2)$$

where the pooled within matrix S_{PW} is a consistent and unbiased estimator of Σ_W (B. Muthén, 1989). Unfortunately, S_B based on the sample is not simply the estimator for the population counterpart Σ_B. Instead, S_B is a consistent and unbiased estimator of $\Sigma_W + c\Sigma_B$ (B. Muthén, 1994), provided that

$$c = \frac{N^2 - \sum_{j=1}^{J} N_j^2}{N(J-1)} \quad . \tag{3}$$

The index $j = 1, 2, \ldots, J$ identifies the interviewers, and c is a constant that approximates the average number of respondents per interviewer in an unbalanced case. The index $i = 1, 2, \ldots, N$ identifies the respondents. In this paper, however, we are primarily interested in the total and within covariance structures. Since the total covariance does not take any grouping (of interviewers) into account, we can simply use basic statistical models such as OLS regression in the analysis further on.

We used Mplus software (L. Muthén & B. Muthén, 1998-2012) to carry out the deconstruction operations. However, instead of simply reporting the differences between Σ_T and Σ_W, we provide estimates for the regression coefficient β_1 in the model $Y = \alpha + \beta_1 X$ based on both Σ_T and Σ_W, where X and Y can be any two randomly-selected variables from the (continuous) variables that were measured over the six consecutive rounds of the ESS. Using the regression coefficients instead of the elements of the covariance matrix has the advantage of better supporting the interpretative framework of substantive survey research. Much substantive research based on the ESS has used a variety of models that relate to, or are an extension of, regression analysis. Another advantage of using the regression parameters is that we can also assess the standard errors of their estimates. Hence, we can not only assess whether the estimated effect sizes are different for Σ_T as compared with Σ_W, but we can also ascertain whether variance inflation emerges when taking interviewer effects into consideration. An example of the Mplus code used in this paper is provided in the appendix. Using this Mplus code, the parameter estimates for Σ_W can be obtained, the Σ_T equivalents can be obtained using any statistical software. In this paper, the lm-command in R was used.

Because of the risk that responses of respondents are correlated within interviewers, standard errors of parameter estimates (e. g. the mean of a variable) are usually underestimated, and is usually referred to as the interviewer design effect or deff. The increase in standard errors due to interviewers is a function of the intra-interviewer correlation and the average workload. Similarly, it is expected that ignoring interviewer effects in a bivariate case (and thus using Σ_T) implies that standard errors are also underestimated. This means that it is expected that standard errors of correlations or regression coefficient will be smaller under Σ_T than under Σ_W.

Slope coefficients in regression analysis may also be biased when interviewer effects are ignored. Chai (1971) argues that whenever measurement error is correlated between variables, slope coefficient may be underestimated or overestimated, depending on the sign and the magnitude of the of the correlation of the measurement errors. In this respect, consider two variables x and y that can be measured with and without interviewer effects. In the case of x, the answer of respondent i assigned to interviewer j is $x_{ij} = x_i^* + \mu_{x,0j} + \varepsilon_{ij}$, where x_i^* is the true answer for i (without interviewer effect $\mu_{x,0j}$). If x_i^* and y_i^* are positively correlated and $\mu_{x,0j}$ and $\mu_{y,0j}$ are also positively correlated, the magnitude of the relationship between x_{ij} and y_{ij} will overestimate the true correlation between x_i^* and y_i^*. More generally, the slope coefficients based on Σ_T will be overestimated if the slope coefficients based Σ_W and Σ_B have the same direction. This also means that the effect size (absolute value of the parameter) of the slope will be greater under Σ_T as compared to Σ_W. An underestimation will take place if the correlation on the respondent and the interviewer level have different signs. It is expected that such biases may occur, only is it not clear whether this would more frequently lead to an overestimation or an underestimation of the slope parameters.

Again, it should be noticed that the multilevel covariance structure method is not capable to separate interviewer effects from area effects. For this purpose, the survey design needs to be modified in a way that allows either interpenetration of interviewers and areas or reinterviewing the same respondent by different interviewers.

4 Data

Six rounds of completed data collection in the ESS have been used for the analysis. The ESS is a biennial cross-national and cross-sectional survey that has been conducted in 36 European countries. Since not all countries participated in all six rounds, the number of country-round combinations (150) is less than 6 × 36. The topics it deals with cover a wide range of societal issues such as health, religion, work, education, ethnic background, and attitudes toward migration, politics, and justice. A fixed set of survey questions are repeated in each round, although a rotating module provides round-specific survey items. We only use the 48 continuous survey items that are available for each round. Further, we only use continuous variables because of the simplicity of reporting a large set of results that originate from the same model specification. A list of these variables can be found in appendix A. The items are not always measured on the same scale: some use 4-point scales, whereas others use 5, 6, 7, 8, or 11-point scales. In order to make β_1 parameters comparable, we first standardized (mean = 0; standard deviation = 1) the 48 continuous variables.

For the bivariate analysis, we randomly selected 50 pairs of variables out of the 48 that are available. One of the two selected variables takes the role of the dependent variable, the other the role of the independent variable. The selection of the pairs was carried out without replacement, making sure that the two variables in a pair are different. A variable can be part of multiple pairs.

We assess the relationship in the form of a regression for each of the 150 country-round combinations, so that in total, 150×50 regressions are considered, first based on the within-interviewer covariance structure, and second on the total covariance structure. The parameter estimates and their associated standard errors can then be compared. The ratio of the absolute value of the parameter estimates (its value based on Σ_W divided by its value based on Σ_T) then expresses the change in effect size of the parameter estimates. If the ratio is smaller than 1, Σ_T might overestimate the relationship between the two variables. In the same way, the squares of the standard error can be compared. It is expected that the standard errors based on the within-interviewer covariance structure will be greater than those based on the total covariance. Therefore, ratios are likely to be greater than 1, because of variance inflation due to the correlated responses on the interviewer level.

In total, 36 countries participated in at least one round of the ESS. The total number of country-round combinations is 150. Despite the fact that in the first round, Austria, Italy, and Luxembourg were included and provided data, these three country-round combinations are not included because they did not collect data for all 48 selected survey items.

Table 2 provides an overview of the scale on which the ESS is organized. The table gives the number of participating countries, respondents, and interviewers per round. The first round of the ESS had the smallest number of participating countries. Fewer than 40,000 respondents cooperated and less than 3,000 interviewers fielded the survey. The second round represented a large expansion in the number of participating countries, respondents, and interviewers, while the third round was slightly downsized. From round four onwards, the number of participating countries was slightly under 30; each time, more than 50,000 respondents participated and were interviewed by slightly fewer than 4,000 interviewers. Because the survey is not longitudinal, new respondents were sampled for each round. The interviewers, on the other hand, might have worked in more than one round.

We ran the analyses separately for each country and each round. As the fieldwork was funded, organized, and monitored at the country level, we differentiate between countries. Countries might differ considerably with regard to how they selected, trained, monitored, and paid their interviewers. We also differentiate between rounds as countries might have deployed different fieldwork agencies or have modified their fieldwork strategies from round to round.

Table 2
Total number of countries, respondents and interviewers per round in the ESS

Round (year)	Countries	Respondents	Interviewers
1 (2002)	20	38,814	2,961
2 (2004)	25	47,283	3,315
3 (2006)	23	43,000	3,643
4 (2008)	29	56,752	3,951
5 (2010)	27	52,458	3,884
6 (2012)	29	54,673	3,715

5 Results

As indicated in Figure 1, one possible indication of interviewer effects in the relationships between variables is that the variables show evidence of interviewer effects in univariate analyses. As the between-interviewer parts of two variables correlate, bivariate interviewer effects may emerge.

5.1 Univariate analysis

Table 3 presents the average intra-interviewer correlations over 48 continuous survey items for 36 ESS countries in six different ESS rounds. The first six columns of the table show the average intra-interviewer correlations, based on the null model, and the next six columns show the average proportional reduction of interviewer variances after the inclusion of fixed effects (region × population density). For these analyses, the glmer-command from the R-package lme4 has been used.

The overview of intra-interviewer correlations is interesting in many respects. First, considerable differences can be observed between countries. Many western and northern European countries tend to have lower interviewer variance compared with southern or eastern European countries. There are some outlying countries with considerable interviewer effects, including Bulgaria, the Czech Republic, Greece, Kosovo, Lithuania, the Russian Federation, Slovakia, and Ukraine. These countries show average intra-interviewer correlations of 0.15 to 0.20 or even higher. A great majority (87%) of intra-interviewer correlations are found to be significant[3]. Only in Iceland, probably due to a relatively small sample size, only 7 out of the 2×48 tested variables show significant interviewer effects. Even in countries with small intra-interviewer correlation indicated in Table 3, such as Finland, Norway, and the Netherlands, about half of the intra-interviewer correlations are still significant.

Second, the levels of intra-interviewer correlation remain relatively stable within countries, although there are some notable exceptions (such as the Czech Republic in round five

[3]According to likelihood ratio tests, using the lmerTest package in R.

Table 3

Summary of intra-interviewer correlations over 48 survey items for 36 countries in six ESS rounds and average proportional reduction of interviewer variance before and after controlling for fixed area effects (region × self-reported population density)

	Average intra-interviewer correlation						Average proportional reduction of interviewer variance after inclusion fixed effects					
	1	2	3	4	5	6	1	2	3	4	5	6
Albania	-	-	-	-	-	0.05	-	-	-	-	-	-
Austria	-	0.09	0.09	-	-	-	-	0.09	0.08	-	-	-
Belgium	0.04	0.06	0.05	0.05	0.05	0.05	0.33	0.24	0.39	0.26	0.34	0.38
Bulgaria	-	-	0.22	0.23	0.23	0.24	-	-	0.06	0.20	0.21	0.19
Croatia	-	-	-	0.11	0.14	-	-	-	-	0.11	0.04	-
Cyprus	-	-	0.13	0.14	0.18	0.15	-	-	0.21	0.22	-	-
Czech Republic	0.16	0.18	-	0.21	0.04	0.26	0.04	0.04	-	0.04	0.31	0.02
Denmark	0.03	0.03	0.03	0.03	0.03	0.02	0.15	0.26	0.17	0.23	0.22	0.28
Estonia	-	0.11	0.12	0.01	0.07	0.10	-	0.09	0.07	0.14	0.18	0.06
Finland	0.02	0.02	0.02	0.02	0.02	0.02	0.33	0.14	0.32	0.35	0.44	0.46
France	-	0.04	0.05	0.04	0.04	0.04	-	0.17	0.11	0.18	0.11	0.18
Germany	0.08	0.10	0.13	0.08	0.07	0.05	0.14	0.10	0.11	0.19	0.24	0.41
Greece	0.16	0.19	-	0.23	0.22	-	0.11	0.13	-	0.02	0.06	-
Hungary	0.06	0.08	0.10	0.12	0.10	0.16	0.09	0.09	0.05	0.03	0.07	0.14
Iceland	-	0.01	-	-	-	0.01	-	-	-	-	-	0.09
Ireland	0.10	0.08	0.07	0.05	0.15	0.16	0.07	0.13	0.09	0.10	0.11	0.07
Israel	0.11	-	-	0.18	0.15	0.12	0.06	-	-	0.01	-	-
Italy	-	-	-	-	-	0.07	-	-	-	-	-	0.40
Kosovo	-	-	-	-	-	0.27	-	-	-	-	-	0.10
Latvia	-	-	-	0.15	-	-	-	-	-	0.12	-	-
Lithuania	-	-	-	-	0.16	0.28	-	-	-	-	0.03	0.05
Luxembourg	-	0.09	-	-	-	-	-	-	-	-	-	-
Norway	0.02	0.02	0.03	0.01	0.02	0.02	0.46	0.25	0.21	0.31	0.23	0.39
Poland	0.08	0.09	0.09	0.10	0.11	0.10	0.12	0.05	0.13	0.10	0.15	0.12
Portugal	0.16	0.15	0.19	0.16	0.19	0.13	0.06	0.05	0.08	0.10	0.08	0.18
Romania	-	-	-	0.23	-	-	-	-	-	0.05	-	-
Russian Fed.	-	-	0.18	0.22	0.20	0.22	-	-	0.02	0.04	0.04	0.03
Slovakia	-	0.08	0.12	0.19	0.17	0.22	-	0.07	0.03	0.04	0.03	0.04
Slovenia	0.05	0.03	0.06	0.07	0.09	0.09	0.29	0.05	0.19	0.05	0.13	0.28
Spain	0.15	0.11	0.09	0.13	0.07	0.05	0.16	0.18	0.03	0.17	0.14	0.10
Sweden	-	0.01	0.02	0.02	0.02	0.07	-	0.48	0.25	0.28	0.15	0.08
Switzerland	0.05	0.06	0.06	0.06	0.06	0.05	0.22	0.17	0.08	0.07	0.15	0.28
The Netherlands	0.02	0.02	0.03	0.03	0.02	0.02	0.25	0.17	0.09	0.15	0.22	-
Turkey	-	0.15	-	0.21	-	-	-	-	0.06	0.13	-	-
UK	0.03	0.04	0.04	0.04	0.05	0.06	0.14	0.24	0.17	0.25	0.13	0.16
Ukraine	-	0.21	0.23	0.22	0.24	0.26	-	0.12	0.11	0.11	0.09	0.07

and Sweden in round six). In some countries, indications of increasing interviewer effects can be observed, for example in Hungary, Slovakia, and Slovenia. By contrast, Spain seems to have strongly reduced its interviewer effects across the consecutive ESS rounds. Intra-interviewer correlations in Spain seems to peak in rounds 1 and 4 and this may be related to the fact that in these two rounds, Spain deployed far more interviewer (168 in round 1 and 138 in round 4), as compared to the other 4 rounds where only 82 or less interviewers have been deployed. This does not explain the evolution of intra-interviewer correlation in Spain, but at least suggests that the survey design has a possible impact on the survey outcomes.

Third, and probably most important, the average proportional reduction of interviewer variance after taking the covariates region $\times$ population density into account only appears to be small in countries that have considerable interviewer effects. The countries of Croatia, Cyprus, the Czech Republic, Greece, Hungary, Ireland, Israel, Latvia, Lithuania, Poland, Portugal, Romania, the Russian Federation, Slovakia, Turkey, and Ukraine all seem to have average intra-interviewer correlations of about 0.10 or above, whereas including the area-related covariates only slightly reduces the interviewer variance. The strongest average proportional reductions of interviewer variances can be found in countries that already show low levels of intra-interviewer correlation. Apparently, our attempt to filter area effects out of the presumed interviewer effects shows that interviewer effects are mostly retained in the high-risk countries. In these countries, respondents who were interviewed in the same region (province or county) and population density conditions provided answers that still strongly depended on the interviewer they worked with. Further, if the interviewers are very likely to be locally clustered themselves in terms of training, supervision, or even local accents or attitudes, it is even possible that too much interviewer variance may have been taken away. These results align with the findings of Schnell and Kreuter (2005) who showed in their study of crime victimization that more design effect could be attributed to interviewers than to sampling points, given an interpenetrated design. However, O'Muircheartaigh and Campanelli (1998), also based on interpenetrated design, found roughly similar levels of variance attributable to interviewers and areas.

Our results suggest that a plausible explanation for the correlated responses appears to be interviewer effects, which should therefore be investigated further for their potential effects on the bivariate relationships between the survey variables.

5.2 Bivariate analysis

For each of the 50 $\times$ 150 bivariate analyses, the estimates for β_1 and their standard errors are examined in order to assess how they behave based on the within-interviewer covariance structure compared with the total covariance structure. For a concise presentation of the results, we only provide results per country (see Table 4). It should be noted that all variables were first standardized (mean = 0; standard deviation = 1).

The three columns of Table 4 headed "average effect size for β_1" show what happens to these parameter estimates before and after taking interviewer effects into account. We consider the absolute values of the parameter estimates and consider this to be the effect size of the strength of the relationship between two variables. The first column illustrates the average effect size when interviewer effects are ignored (Σ_T). The second column provides the equivalent averages when taking interviewer effects into account (Σ_W), and the third column shows the ratio of the first two. For each country, the average effect sizes decrease when interviewer effects are taken into account. In other words, ignoring interviewer effects may imply overestimated effect sizes. For some countries, these results are quite striking. For example, in Kosovo, Romania, Turkey, and Ukraine, the parameter effect sizes decrease on average by 20% or more. Another notable aspect is that the average effect sizes bring the countries closer together if interviewer effects are not ignored. The standard deviation over all the average effect sizes in the first column is 0.0131 (based on Σ_T), whereas the same standard deviation over all the countries in the second column reduces to 0.0096. For a cross-national survey, this is an important finding, because observed effect differences between countries could erroneously be attributed to real country differences.

These results only reflect the average effect size over all the countries. Of course, this does not mean that all parameter estimates necessarily become weaker when going from the total covariance to the within-interviewer covariance. Of all the regression parameters that are significant under Σ_T, 70% become smaller under Σ_W.

The next three columns in Table 4 provide a similar comparison to the first three, but here the comparison is applied to the respective standard errors of the estimated regression parameters. In addition to the fact that the effect sizes of parameters reduce, their respective standard errors apparently increase when examining the ones based on the within-interviewer covariance structures. All countries seem prone to variance inflation because of interviewer effects. The variance inflation factor is determined by taking the square of the standard error based on the within-interviewer covariance structure divided by its total covariance structure counterpart. A minimal increase of variance of about 4% to 5% needs to be accounted for, with some outlying countries notably higher, such as the Czech Republic (+23%), the Russian Federation (+15%), and Ukraine (+17%). Some 86% of all the estimated parameters have increased standard errors when going from the total covariance to the within-interviewer covariance structure. It should be noted that the standard errors are obtained using Maximum Likelihood estimation, assuming that the data are normally distributed. However, as some of the variables used in the analysis sometimes deviate from this assumption, more robust estimation techniques could be used (Yuan & Bentler, 2006), which may increase the variance inflation even more.

Two graphs illustrate the findings of Table 4. In Figure 2, the x-axis represents the estimates under Σ_T, the y-axis represents the estimates under Σ_W. The bisector is added indicating the situation where the absolute value of the estimate is the same under Σ_T and Σ_W. The graph shows the variability of the effect sizes, ranging from the origin to about 0.60. About 60% of the estimates is found significant (alpha <

Table 4
Summary of interviewer effects on regression slopes by ESS country

	Pairs of variables tested	Average effect size for b_1			Average standard error for b_1		
		Σ_T	Σ_W	Σ_W/Σ_T	Σ_T	Σ_W	VIF
Albania	50	0.111	0.105	0.947	0.030	0.030	1.044
Austria	100	0.120	0.109	0.907	0.021	0.023	1.171
Belgium	300	0.099	0.094	0.945	0.024	0.025	1.077
Bulgaria	200	0.129	0.105	0.816	0.024	0.025	1.125
Croatia	100	0.124	0.110	0.891	0.026	0.027	1.066
Cyprus	200	0.108	0.097	0.903	0.031	0.032	1.066
Czech Republic	250	0.119	0.115	0.972	0.023	0.026	1.233
Denmark	300	0.097	0.092	0.953	0.026	0.027	1.058
Estonia	250	0.107	0.100	0.929	0.024	0.025	1.074
Finland	300	0.105	0.103	0.982	0.023	0.024	1.058
France	250	0.125	0.118	0.944	0.023	0.024	1.089
Germany	300	0.110	0.103	0.937	0.019	0.019	1.078
Greece	200	0.119	0.102	0.854	0.020	0.021	1.080
Hungary	300	0.122	0.114	0.938	0.026	0.027	1.117
Iceland	100	0.100	0.099	0.989	0.046	0.049	1.131
Ireland	300	0.117	0.101	0.860	0.023	0.024	1.071
Israel	200	0.104	0.089	0.857	0.021	0.022	1.092
Italy	50	0.124	0.115	0.927	0.033	0.035	1.095
Kosovo	50	0.123	0.077	0.627	0.029	0.030	1.120
Latvia	50	0.116	0.108	0.933	0.023	0.024	1.055
Lithuania	100	0.148	0.120	0.813	0.024	0.025	1.080
Luxembourg	50	0.094	0.088	0.934	0.025	0.026	1.061
Netherlands	300	0.103	0.101	0.982	0.023	0.024	1.046
Norway	300	0.104	0.102	0.980	0.025	0.026	1.055
Poland	300	0.096	0.090	0.937	0.024	0.025	1.108
Portugal	300	0.138	0.118	0.861	0.022	0.023	1.055
Romania	50	0.137	0.108	0.788	0.023	0.024	1.107
Russian Fed.	200	0.119	0.097	0.814	0.021	0.022	1.147
Slovakia	250	0.109	0.103	0.940	0.025	0.026	1.090
Slovenia	300	0.115	0.104	0.905	0.027	0.028	1.086
Spain	300	0.104	0.094	0.901	0.023	0.024	1.065
Sweden	250	0.111	0.107	0.967	0.025	0.025	1.059
Switzerland	300	0.095	0.092	0.966	0.024	0.025	1.050
Turkey	100	0.135	0.101	0.747	0.022	0.024	1.102
UK	300	0.111	0.105	0.953	0.022	0.023	1.089
Ukraine	250	0.118	0.094	0.800	0.024	0.026	1.172

0.05) under both Σ_T and Σ_W (indicated by gray "+"-signs). It can be observed that many cases are situated below the bisector. This means that the absolute parameter estimates under the total covariance structure are usually greater as compared to their within covariance structure counterparts. About 27% of the estimates are never significant (under Σ_T or Σ_W), as indicated by the gray circles. The remaining cases are only significant depending on whether Σ_T or Σ_W are considered. 620 estimates are significant under Σ_T but are not significant under Σ_W (black squares), whereas only 336 estimates are not significant under Σ_T but are significant under Σ_W (black triangles). Of all parameters in the analyses, 68.35% of the estimates are significant under Σ_T, whereas only 64.56% of the estimates are significant under Σ_W.

Additional to this last figure, Figure 3 provides a scatter plot of the standard errors under Σ_T and Σ_W. Clearly, a vast majority cases are above the bisector, indicating that the standard errors under Σ_W are greater than under Σ_T. The dots in

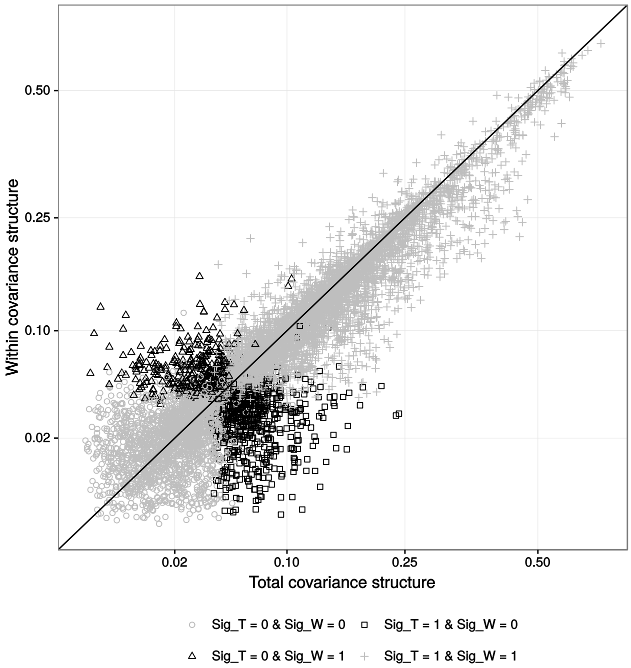

Figure 2. Absolute parameter estimates under the total and within covariance structure, ESS

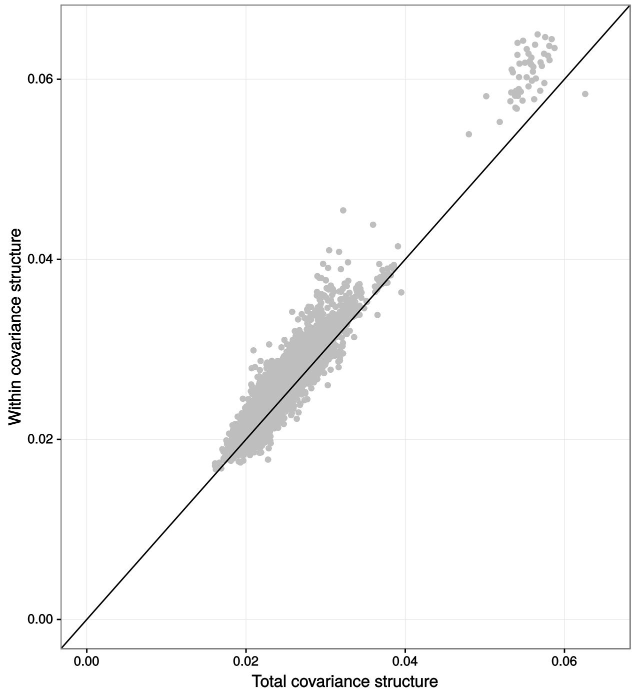

Figure 3. Standard errors under the total and within covariance structure, ESS

the upper right corner belong to Iceland, that had a remarkably smaller sample size than the other countries.

Both figures illustrate that when interviewers are ignored (using Σ_T), the parameter effect sizes tend to be overestimated whereas their standard errors are systematically underestimated.

6 Discussion

This paper provides estimates of interviewer effects, initially in terms of the means of 48 continuous variables in the 150 country-round combinations of the ESS up to round six. The scope of the paper is also extended by taking into account the bivariate relationships (regression coefficients) between these variables. Assuming that the observed differences between the interviewers are actually real interviewer effects, the results suggest that for some countries, the quality of substantive analyses may be affected substantially by interviewer effects. This is particularly the case in eastern and southern European countries, where intra-interviewer correlations reach levels of 0.20 and higher. In these countries, the effect sizes of the relationships between survey variables tend to be overestimated, whereas their standard errors tend to be underestimated when interviewer effects are ignored. Therefore, there is a risk that the relationships between survey variables may be incorrectly inferred.

Most (if not all) survey researchers who use the ESS data for substantive research do not seem to take interviewer effects into account when performing analyses, although it seems appropriate to do so. Of the 221 substantive journal articles published in 2013 and based on ESS data, not one considered the problem of interviewer effects. Hence, there may be a risk that some of these authors might have drawn invalid conclusions from the data.

What can survey researchers do if they care about interviewer effects? Neither the interviewer nor the sampling point identification variables are included in the main ESS data, but should be found in datasets that need to be downloaded separately and linked to the main datasets. This may hinder researchers from devoting the necessary attention to interviewer (or area) effects. Apart from these obstacles, researchers should be aware that working with data for some of the high-risk countries (e. g. Bulgaria, Cyprus, the Czech Republic, Greece, Ireland, Israel, Kosovo, Lithuania, Portugal, Romania, the Russian Federation, Slovakia, Turkey, and Ukraine) may jeopardize their findings. As many articles based on ESS data have used the survey in a cross-national way, and given the finding that the effect sizes of the relationships between variables converge between countries after taking interviewer effects into account, survey researchers should be aware of the relatively high risk caused by interviewer effects. One possible precaution that can be taken is to deliberately increase the standard errors of the analysis, or to anticipate the possibility that the effect sizes of the parameters are lower than observed. Alternatively, the significance level could also be altered (for example, 0.01 instead of 0.05). The magnitude of the intra-interviewer correlations, as shown in Table 3, may be a good guideline in terms of where to expect country-round combinations that are at risk of interviewer effects between survey variables.

One of the issues that are not discussed in this paper is why countries differ so strongly in terms of interviewer effects. If some countries are substantially different from others, it may be worthwhile to investigate what is done in low-risk countries in order to keep interviewer effects at an acceptable level, and what may have been overlooked in high-risk countries. In countries where interviewer effects seem to change over time (e. g. the Czech Republic, Hungary, Ireland, Slovakia, and Spain), strategic variables such as interviewer selection, training and briefing, remuneration, or supervision may be useful ways of explaining interviewer effects.

This last issue also relates to the question of whether interviewer effects are a specific problem only for the ESS. The issue of interviewer effects seems to be country-specific in the first place, despite the fact that the ESS has robustly standardized its questionnaire, showcards, and interviewer instructions. In addition, training material was made available across all the participating countries. The ESS should further elaborate on the issue of interviewer effects by providing more interviewer training material and carrying out a survey among interviewers to identify relevant interviewer differences. This might lead to crucial insights regarding interviewer effects, across and within ESS countries.

The way in which interviewer effects can specifically affect relationships between variables is a research area that requires more empirical and theoretical work. The issue that should be addressed is why variables correlate at the interviewer level. There may be various reasons as to why interviewers systematically create deviations in respondent answers, causing biased correlations between variables. Here, we speculate that interviewers may deviate from the principles of standardized interviewing and take recourse to interviewer-specific principles, such as instructions not being properly followed, questions not being completely read, or possible answers being poorly conveyed to the respondent. Other issues that may create biased results include suggestive interviewing, speeding, or straightlining. Revealing the mechanisms behind the observed interviewer effects may therefore be prioritized more in the ESS. Some interviewer characteristics (e. g. interviewer experience) as well as interviewer behaviour during the interview (e. g. reading pace, suggestiveness) may elicit interviewer effects. Collecting more information about interviewers and recording some interviews per interviewer may be a good suggestion to inspire future research.

Another possibility is that the interviewer effects we find may actually be area effects. Although this is certainly a potential valid argument, it should not be seen as a reason to ignore correlated responses in the data. In this paper, we assume that the clustering originates from interviewer measurement effects. However, even if the true source of the clustering in the data can be attributed to area effects, the clustering remains a threat to the quality of the data, as variance inflation should still be accounted for. Alternatively, differences in nonresponse patterns for areas and/or interviewers might require weighting strategies in order to adjust for possible bias, which can also alter parameters and standard errors. The fact that it is (almost) impossible to correctly diagnose the root cause(s) of the clustering should make researchers even less confident about the data. Therefore, this contribution is essentially an urgent reminder that survey researchers should consider the correlated responses within interviewers/areas to be a serious problem in survey research, and the ESS in particular. Regarding the survey design, it is worthwhile for the ESS to consider ways to assign interviewers to areas that accommodate a more profound interpenetration. Currently, in the ESS in many cases it can be observed that interviewers are exclusively assigned to areas and vice versa. Such (local) interpenetration might at least partially remedy the interviewer-area confounding.

References

Beullens, K. & Loosveldt, G. (2013). *Assessing interviewer variance and its implication for data collection, Deliverable 12.10*. KU Leuven, ESS-DACE.

Beullens, K. & Loosveldt, G. (2014). Interviewer effects on latent constructs in survey research. *Journal of Survey Statistics and Methodology*, *2*, 433–458. doi:10.1093/jssam/smu019

Biemer, P. (2010). Overview of design issues: total survey error. In P. Marsden & J. Wright (Eds.), *Handbook of survey research* (2nd, pp. 27–58). Bingley: Emerald.

Biemer, P. & Stokes, L. (1985). Optimal design of interviewer variance experiments in complex surveys. *Journal of the American Statistical Association*, *80*, 158–166. doi:10.2307/2288066

Boyd, H. & Westfall, R. (1955). Interviewers as a source of errors in survey. *Journal of Marketing*, *19*, 311–324.

Chai, J. (1971). Correlated measurement errors and the least squares estimator of the regression coefficient. *Journal of the American Statistical Association*, *66*(335), 478–483.

Collins, M. & Butcher, B. (1982). Interviewer and clustering effects in an attitude survey. *Journal of the Market Research Society*, *25*(1), 38–58.

Davis, P. & Scott, A. (1995). The effect of interviewer variance for domain comparisons. *Survey Methodology*, *21*, 99–106.

Duncan, T., Duncan, S., Hops, H., & Alpert, A. (1997). Multilevel covariance structure analysis of intra-familial substance use. *Drug and Alcohol Dependence*, *46*(3), 167–180. doi:10.1016/S0376-8716(97)00057-4

Eisinga, R., te Grotenhuis, M., Larsen, G., & Pelzer, B. (2011). Interviewer BMI effects on under- and over-reporting of restrained eating: evidence from a national Dutch face-to-face survey and a postal follow-up. *International Journal of Public Health*, *57*(3), 643–647.

Elliott, M. & West, B. (2015). ‘‘clustering by interviewer”: A source of variance that is unaccounted for in single-stage health surveys. *American Journal of Epidemiology*, *182*(2), 118–126.

Elovainio, M., Kivimäki, M., Steen, N., & Vahtera, J. (2004). Job decision latitude, organizational justice and health: Multilevel Covariance Structure Analysis. *Social Science & Medicine*, *58*(9), 1659–1669. doi:10.1016/S0277-9536(03)00366-6

European Social Survey. (2002–2012). European Social Survey Round 1-6 Data. Norwegian Social Science Data Services, Norway – Data Archive and Distributor of ESS data. Retrieved from http://www.europeansocialsurvey.org/

Fasel, N., Green, E., & Sarrasin, O. (2013). Facing cultural diversity: anti-immigrant attitudes in Europe. *European Psychologist*, *18*(4), 253–262. doi:10.1027/1016-9040/a000157

Freeman, J. & Butler, E. (1976). Some sources of interviewer variance in surveys. *Public Opinion Quarterly*, *40*(1), 79–91.

Groves, R. & Couper, M. (1998). *Nonresponse in household interview surveys*. Wiley Series in Probability and Statistics. New York (N.Y.): Wiley.

Groves, R. & Magilavy, L. (1980). Estimates of interviewer variance in telephone surveys. In *Proceedings of the American Statistical Association, section on surrey research methods* (pp. 622–627).

Groves, R. & Magilavy, L. (1986). Measuring and explaining interviewer effects in centralized telephone surveys. *Public Opinion Quarterly*, *50*(2), 251–266.

Hox, J. (1994). Hierarchical regression models for interviewer and respondent effects. *Sociological Methods and Research*, *22*(3), 300–318.

Hox, J., de Leeuw, E., & Kreft, I. (1991). The effect of interviewer and respondent characteristics on the quality of survey data: a multilevel model. In P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 439–462). New York: Wiley.

Kish, L. (1962). Studies of interviewer variance for attitudinal variables. *Journal of the American Statistical Association*, *57*, 92–115.

Loosveldt, G. & Beullens, K. (2010). *Evaluation of interviewer effects in round four of the european social survey, Deliverable 12.2*. KU Leuven, ESS-DACE.

Mahalanobis, P. (1946). Recent experiments in statistical sampling in the Indian statistical institute. *Journal of the Royal Statistical Society*, *109*, 325–370.

Mangione, T., Fowler, F., & Louis, T. (1992). Question characteristics and interviewer effects. *Journal of Official Statistics*, *8*(3), 293–307.

Muthén, B. (1989). Latent variable modeling in heterogeneous populations. *Psychometrica*, *54*(4), 376–398.

Muthén, B. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research*, *22*(3), 376–398.

Muthén, L. & Muthén, B. (1998-2012). *Mplus user’s guide* (7th). Los Angelos, CA: Muthén.

O’Muircheartaigh, C. & Campanelli, P. (1998). The relative impact of interviewer effects and sample design effects on survey precision. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *161*(1), 63–77. doi:10.111/1467-985X.00090

Philippens, M. & Loosveldt, G. (2004). *Interviewer-related variance in the European Social Survey*. Paper Presented at the Sixth International Conference in Logic and Methodology. Amsterdam, August.

Rice, S. (1929). Contagious bias in the interview: a methodological note. *American Journal of Sociology*, *35*(3), 420–423.

Schaeffer, N., Dykema, J., & Maynard, D. (2010). Interviewers and interviewing. In P. Marsden & J. Wright (Eds.), *Handbook of survey research* (2nd, pp. 437–470). Bingley: Emerald.

Schnell, R. & Kreuter, F. (2005). Separating interviewer and sampling-point effects. *Journal of Official Statistics*, *21*(3), 389–410.

Tucker, C. (1983). Interviewer effects in telephone surveys. *Public Opinion Quarterly*, *47*, 84–95.

West, B., Kreuter, F., & Jaenichen, U. (2013). Interviewer effects in face-to-face surveys: a function of sampling, measurement error or nonresponse? *Journal of Official Statistics*, *29*(2), 277–297. doi:10.2478/jos-2013-0023

West, B. & Olson, K. (2010). How much of interviewer variance is really nonresponse error variance? *Public Opinion Quarterly*, *74*(5), 1004–1026. doi:10.1093/poq/nfq061

Wiggins, R., Longford, N., & O’Muircheartaigh, C. (1992). A variance components approach to interviewer effects. In A. Westlake, R. Banks, C. Payne, & T. Orchard (Eds.), *Survey and statistical computing* (pp. 243–254). Amsterdam: North-Holland.

Yuan, K.-H. & Bentler, P. (2006). Asymptotic robustness of standard errors in multilevel structural equation models. *Journal of Multivariate Analysis*, *97*, 1121–1141.

Appendix A
List of variables used

(Table follows on next page)

Table A1
Overview of 48 substantive core survey variables in the ESS

Variable	Label	points
AESFDRK	Feeling of safety of walking alone in local area after dark	4
IMDFETN	Allow many/few immigrants of different race/ethnic group from majority	4
IMPCNTR	Allow many/few immigrants from poorer countries outside Europe	4
IMSMETN	Allow many/few immigrants of same race/ethnic group as majority	4
POLINTR	How interested in politics	4
FREEHMS	Gays and lesbians free to live life as they wish	5
GINCDIF	Government should reduce differences in income levels	5
HEALTH	Subjective general health	5
IMPDIFF	Important to try new and different things in life	5
IMPENV	Important to care for nature and environment	5
IMPFREE	Important to make own decisions and be free	5
IMPFUN	Important to seek fun and things that give pleasure	5
IMPSAFE	Important to live in secure and safe surroundings	5
IMPTRAD	Important to follow traditions and customs	5
IPBHPRP	Important to behave properly	5
IPCRTIV	Important to think new ideas and being creative	5
IPEQOPT	Important that people are treated equally and have equal opportunities	5
IPFRULE	Important to do what is told and follow rules	5
IPGDTIM	Important to have a good time	5
IPHLPPL	Important to help people and care for others well-being	5
IPLYLFR	Important to be loyal to friends and devote to people close	5
IPMODST	Important to be humble and modest, not draw attention	5
IPRSPOT	Important to get respect from others	5
IPSHABT	Important to show abilities and be admired	5
IPSTRGV	Important that government is strong and ensures safety	5
IPSUCES	Important to be successful and that people recognize achievements	5
IPUDRST	Important to understand different people	5
SCLACT	Take part in social activities compared to others of same age	5
PRAY	How often pray apart from at religious services	7
RLGATND	How often attend religious services apart from special occasions	7
SCLMEET	How often socially meet with friends, relatives or colleagues	7
TVTOT	Tv watching, total time on average weekday	8
HAPPY	How happy are you	11
IMBGECO	Immigration bad or good for country's economy	11
IMUECLT	Country's cultural life undermined or enriched by immigrants	11
IMWBCNT	Immigrants make country worse or better place to live	11
PPLFAIR	Most people try to take advantage of you, or try to be fair	11
PPLHLP	Most of the time people helpful or mostly looking out for themselves	11
PPLTRST	Most people can be trusted or you can't be too careful	11
STFDEM	How satisfied with the way democracy works in country	11
STFECO	How satisfied with present state of economy in country	11
STFEDU	State of education in country nowadays	11
STFHLTH	State of health services in country nowadays	11
STFLIFE	How satisfied with life as a whole	11
TRSTLGL	Trust in the legal system	11
TRSTPLC	Trust in the police	11
TRSTPLT	Trust in politicians	11
TRSTPRL	Trust in country's parliament	11

Appendix B
Code Example

The following shows an example of Mplus code for modeling a simple regression within and between interviewers. The first four lines of the code specify the data and variables. The interviewer identification variable ‘INT’ is specified as a cluster variable, which is the grouping variable in order to separate the within and between covariance structure. Thereafter, it is important to specify the same model (v2 on v1) on the respondent level (%within%) as well as on the interviewer level (%between%).

```
Title: <TITLE>;
Data:File is <FILE>;
Variable:Names are v1 v2 INT;
Usevar are v1 v2;
cluster = INT;
Analysis:type = twolevel;
estimator = ML;
Model:
\%within\%
v2 on v1;
\%between\%
v2 on v1;
```