Survey Research Methods (2020)
Vol. 14, No. 3, pp. 301-323
doi:10.18148/srm/2020.v14i3.7621

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

The more similar, the better? How (mis)match between interviewers and respondents in survey situations affects item nonresponse and data quality

Felix Bittmann
University of Bamberg
Germany

Previous research shows that sociodemographic (mis)matches between respondents and interviewers can affect unit and item nonresponse in survey situations. The current paper attempts to deepen our understanding of these findings and investigates the effect of matching with regard to gender and age on item nonresponse, reluctance to answer items and the probability that a third person is interfering with the interview. Using multilevel European Social Survey data from 23 countries, we demonstrate that some types of matching significantly improve data quality. The results corroborate and extend previous findings and underline that sociodemographic matching has the potential to enhance data quality in face-to-face interviewing situations.

Keywords: Item nonresponse; interviewer characteristics; sociodemographic matching

1 Introduction

Previous studies indicate that survey quality and nonresponse are affected by certain qualities of the interviewer, such as gender, age or other sociodemographic characteristics in face-to-face interviews (Durrant, Groves, Staetsky, & Steele, 2010). Beyond that, there is a growing body of evidence showing how respondent-interviewer matching with respect to a number of key variables can improve data quality. The current study follows these results and examines in-depth how matching can contribute to improving data quality and reduce missing data problems in survey situations using European Social Survey (ESS) data from 23 countries. Three main outcomes are used to assess data quality: item nonresponse rates in an interview, measured as the number of questions without a valid response (1), the probability that a respondent is reluctant to answer questions (2) and the probability that a third person is interfering with the interview (3). Thus, the research question motivating the study asks: how does sociodemographic matching with respect to gender, age and their interactions influence data quality in face-to-face interview situations?

This is a crucial, yet underexplored, aspect of survey research design, as demonstrating positive effects of sociodemographic matching on data quality and item nonresponse would provide a relatively simple way of improving survey quality. Finally, although the focus on item nonresponse in comparison to unit nonresponse might seem arbitrary, we choose to investigate the former because the ESS survey provides relevant variables and information regarding item nonresponse. Subsequent studies, however, should also consider unit nonresponse.

Contact information: Felix Bittmann, Universität Bamberg, Feldkirchenstraße 21, 96045 Bamberg, Germany (E-mail: felix.bittmann@uni-bamberg.de)

2 Effects of sociodemographic (mis)matching on data quality

2.1 Theoretical framework and hypotheses

Studies investigating the factors influencing survey data quality indicate that, in addition to the characteristics of the respondent, such as age or level of education (Groves, 2009; Tourangeau, Rips, & Rasinski, 2000), item nonresponse is also affected by the characteristics of the interviewer (Brunton-Smith, Sturgis, & Leckie, 2017; Davis, Couper, Janz, Caldwell, & Resnicow, 2010; Hox, De Leeuw, & Kreft, 1991). In the current study, we explore the interaction of these two aspects, that is, the level of similarity between respondents and interviewers, also called sociodemographic matching, and its effect on data quality, defined as the degree to which survey responses are "true", unbiased and complete.

The main theoretical framework that explains why sociodemographic matching affects data quality is the theory of liking (Groves, Cialdini, & Couper, 1991; Groves, Cialdini, & Couper, 1992), which suggests that a person prefers to interact with others that he or she likes. Important factors that determine whether one likes another person seem to be, among others, similarity of attitudes (Byrne et al.,

1971), background (Stotland & Patchen, 1961) and dress style (Suedfeld, Bochner, & Matas, 1971). This is also known as homophily, which describes the degree to which two individuals are alike with regard to various characteristics, including education, social status, beliefs (Rogers & Bhowmik, 1970).

To underline why this is relevant for the research questions, we will demonstrate two things: firstly, that gender and age refer to the aspects mentioned and secondly that liking should ideally improve the data quality of surveys, which is, due to the ambiguous state of research, not certain until now and the evidence in the literature is mixed (Hornberger, Medley-Proctor, Nettles, Cimporescu, & Howe, 2016; Hurtado, 1994; Poulin, 2010). According to past studies, gender and age influence attitudes (McPherson, Smith-Lovin, & Cook, 2001), due to the fact that similarly aged people from the same birth cohorts share various experiences, such as historic events, observing the same trends, fashions, and slang. Therefore, when the age difference between the respondent and interviewer is small, chances are higher that they share characteristics that make them more similar on average, compared to when the age difference is large. A classic study on the subject indicates that 38% of all friends are within two years of age difference (Fischer, 1977), a finding supported by a more recent analysis which underlines the stability of the effects for race and age with respect to homophily (Smith, McPherson, & Smith-Lovin, 2014). This demonstrates that an interviewer of approximately the same age as a survey respondent might seem friendlier or friend-like compared to interviewers with a larger age difference.

Similar arguments can be drawn for the impact of gender matching on item nonresponse, with the effect being driven primarily by interests, rather than time. Research indicates that men and women differ with regards to the topics in which they show interest, such as their vocational choices or leisure activities (Su, Rounds, & Armstrong, 2009). In other words, on average, men are more similar to men than to women, and vice versa (Brashears, 2015). Consequently, gender matching should, on average, increase the similarity between the respondent and interviewer, which, in turn, should increase liking. Taken together, we assume that the factors mentioned above, like personal preferences, background and dressing style, are more similar when age and gender are similar. Clearly, this does not mean that a perfect sociodemographic match will always result in a high degree of liking, as even within birth cohorts and genders a multitude of interests and attitudes exist, but, on average, we predict the probability of liking to be higher.

Item nonresponse or not answering questions at all is of central interest to understanding how liking and data quality are related. Nonresponse might occur more often when a respondent does not trust the interviewer, or the question is so sensitive that the information will not be shared, especially when guilt or shame are high and the respondent does not expect understanding, or even fears moral judgement (Lahaut, Jansen, Van de Mheen, & Garretsen, 2002; Poulin, 2010). A large degree of similarity between respondent and interviewer could improve these aspects, as similar people might share similar interests and moral attitudes (McPherson et al., 2001). Therefore, when similarity to the interviewer is higher, the fear of moral judgement or lack of understanding might be lower, consequently increasing the probability that the respondent will answer questions posed by the interviewer. Based on these theoretical arguments, we formulate the first hypothesis: The higher the degree of sociodemographic similarity between interviewer and respondent, the lower the probability of item nonresponse.

The second outcome of interest considers how smoothly respondents interact with the interviewer, or whether they are reluctant to answer a question. Following the logic of the theory described above, respondents should be more open and more willing to answer a question truthfully when they trust the interviewer and find him or her sympathetic. When respondents trust the interviewer and interact with him or her more naturally, less probing should be necessary and the question-answer-flow should be more continuous. As this relates to the degree of similarity with respect to gender and age, the second hypothesis is as follows: The larger the degree of similarity between respondent and interviewer, the lower the probability that a respondent is reluctant to answer a question.

The last scenario is slightly different from the first two and considers the context in which the interview takes place. Specifically, respondents might be self-conscious when a third person is present (Reuband, 1992) and, accordingly, situations in which the interviewee is alone with the interviewer might yield better results. Here, the question arises whether the degree of similarity between the respondent and interviewer affects the probability that any other person is present. Of course, there are situations when having a third person present during the interview cannot be avoided, for example, when small children are accompanied by their parent (Aquilino, Wright, & Supple, 2000; Boeije, 2004). In other cases, particularly when the respondent can choose whether to be alone, or accompanied by someone, it could indicate that the interviewer is perceived as dissimilar, or even a threat. When this is the case, it is plausible that a respondent might not want to face the interviewer alone, but have another person by their side, for example their spouse, or a roommate. Consequently, measuring interference by third parties is relevant, since it might negatively influence data quality, making the respondent more self-conscious about their answers and attempt to hide the "true", possibly embarrassing, information from the other person, particularly if they are a close friend or relative (Z. M. Mneimneh, Tourangeau, Pennell, Heeringa, & Elliott, 2015). Thus, investigating the

presence of third parties during interviews is of relevance to measuring survey quality.

Taken together, evidence suggests that increased similarity between the respondent and interviewer increases the probability of liking the interviewer. Therefore, it should not be necessary to call a third person to the interview as a companion, as long as the respondent likes the interviewer and perceives him or her as sympathetic. Furthermore, even if a third person is present, the chance that this person actively interferes is smaller when the interview proceeds smoothly. For example, a third person might agree to attend the interview in silence and interfere only in case of persistent or repeated questioning by the interviewer. As this might be more likely when the similarity between the interviewer and respondent is small, this would explain how similarity and third person interference is related. Consequently, the third hypothesis stipulates that: The larger the degree of sociodemographic similarity between the respondent and interviewer, the lower the probability that a third person is present and interfering with the interviewing process.

2.2 Review of the literature

A large body of literature investigates the effects of respondent-interviewer matching, dating back at least to the 1920s (Rice, 1929). Matching on ethnic characteristics or, especially in the US-context, origin was one of the first aspects to be explored by the early studies (Williams Jr, 1964). While some of them do not find any positive effects of origin/ethnic matching (Axinn, 1989; Dotinga, Van den Eijnden, Bosveld, & Garretsen, 2005), others report positive results (Johnson, Fendrich, Shaligram, Garcy, & Gillespie, 2000; Oyinlade & Losen, 2014; Webster, 1996). A second major factor under investigation is matching with regard to educational levels. Similar levels of education between respondents and interviewers can facilitate communication between the two and enable a more empathetic interview experience, thus improving the quality of collected data. Yet, study results on educational matching are, again, mixed. While Durrant et al. (2010) report that matching on gender and educational level increases cooperation rates, others find null or negative effects (Riphahn & Serfling, 2005; Vercruyssen, Wuyts, & Loosveldt, 2017; Weiss, 1968).

Given the research question of the present study, empirical results regarding matching on gender and age are of central interest. In this context, research shows that women are more open to female interviewers (Fletcher & Spencer, 1984) and both men and women are more responsive to sensitive psychological items when the interviewer is female (Chun, Tavarez, Dann, & Anastario, 2011). Yet, this effect is not stable when the items are about sexual behaviour, or substance abuse (Catania, Binson, Canchola, Pollack, & Hauck, 1996; Johnson & Parsons, 1994). Therefore, it is not possible to conclude that women always perform better as interviewers when sensitive items are concerned, as the specific type of item appears to be relevant.

When age is considered, previous research presents contradictory findings yet again. While some studies report that age matching seems beneficial for older respondents (Herzog & Rodgers, 1992), no reliable effects are found for sensitive items (Wilson, Brown, Mejia, & Lavori, 2002). an ambitious research synthesis including hundreds of studies concludes that sociodemographic matching seems useful, as four out of ten studies considered report positive findings (West & Blom, 2016). This notion is partially supported by Vercruyssen et al. (2017), who report that gender matching for men shows positive effects on item nonresponse (higher item response rates), but not for women. They also ascertain that age matching reduces item nonresponse, which cannot be attributed to age effects alone, as they control for the age of the respondents. A recent study from Germany echoes these findings and reports that, in particular, matching on education between respondent and interviewer significantly increases the rates of success in CATI settings (West et al., 2019).

Finally, regarding the effect of sociodemographic matching on the probability that third persons are present, two studies report that women interviewed by a male, rather than a female interviewer are more often accompanied by a man (Hartmann, 1994; Preetz, 2017), which can be interpreted as a form of social control. However, as discussed earlier, this interpretation may not hold under all circumstances, as women may choose freely whether they wish for someone to accompany them during the interview, if they do not trust the interviewer. A multinational study focussing on cross-country effects concludes that male respondents have a third party present more often than females, regardless of the gender of the interviewer, and that there is a negative correlation between the overall individualism within a country and the chance that a third party is present (Z. Mneimneh, Elliott, Tourangeau, & Heeringa, 2018). With respect to our research question, this means that it might be harder to negotiate and maintain a private interviewing environment when a female interviewer is talking to a male respondent and his wife wants to participate in the interview situation.

A review of the literature reveals a heterogeneous picture, whereby for every relevant aspect there are positive and negative (or null) findings, and no distinct conclusion is possible. A more detailed inspection reveals that these findings are highly complex as minor details, for example, the topic of a sensitive item, are also relevant and could influence the result. Accounting for all of the nuances and potential interactions between the relevant factors, such as gender, age, or education, would require a complex theoretical framework. Although the study is based on the relatively simple theory of liking, we interact age with gender in the analyses in the hopes of generating new insights that might prove beneficial for survey research in general.

3 Methods

3.1 Data

In order to test the effects of respondent-interviewer matching on data quality, we use data from the European Social Survey (ESS) Round 8 (2016, version 2.0). The ESS is a biannual European survey that includes a multitude of questions about all aspects of life, with a particular focus on social aspects and politics. The survey is highly standardized and only a small number of questions differs between countries, for example, those regarding educational degrees. 23 countries are currently involved in the sample (see Table 2 for a complete list of all countries). The effective sample size depends on the dependent variable and can be inferred from the regression tables below. The ESS releases information about the main questionnaire as well information about the interviewers (age and gender), how reluctant respondents were to answer questions and whether any third persons were present and interfering with the interview. Information about the educational level of the interviewer is not available and, therefore, cannot be used for matching. All interviews were conducted using CAPI (computer-assisted personal interview) except for the following countries, which used PAPI (paper and pencil interview): Spain, Russia, Poland, Lithuania and Israel (ESS, 2018). One challenge of the analysis is that the design is not interpenetrated, therefore interviewers are not guaranteed to be assigned randomly to sample units, but rather based on geographic and temporal demand and interviewer availability. As sociodemographic characteristics of interviewers and sample units are not considered for the assignment of sample units to interviewers, there is clearly an element of randomness in the process, although nonresponse bias could still affect the results (West & Blom, 2016). To account for the potential non-random assignment between respondents and interviewers, we include additional control variables (see next section for details). Yet, as the main variables of interest – gender and age – are still highly variable, even within a region or neighbourhood, the bias should be small. Finally, the multilevel approach, which includes random intercepts for interviewers, takes into account the fact that respondents, who are nested within interviewers, might be more similar to each other than in a complete random design, which is described using the intraclass correlation.

The analyses include all observations from the survey, with the exception of respondents who were younger than 18 years old at the time of the survey (1239 cases). The ESS pays great attention to maximizing response rates. The average response rate is between 31% (Germany) and 74% (Israel) (Stoop, Koch, Halbherr, Loosveldt, & Fitzgerald, 2016).

3.2 Variables

Degree of sociodemographic match. The two constructs central to the study include gender- and age-matching between respondent and interviewer, thus, respondents whose age or gender are missing from the data are excluded from the analyses. For interviewers, there are no missing values regarding these variables. To operationalize this, we use three variables: gender of the respondent (Female), indicating whether a respondent is female (1) or male (0); gender of the interviewer (IntFemale), coded analogously; lastly, age difference between the respondent and interviewer, recoded into a new variable (AgeMatch), with three categories: interviewer younger than respondent (0), age difference within five years of range (1) and interviewer older than respondent (2). we used the cut-off point of five years in order to maximize comparability with similar studies in the field, which have previously operationalized age matching in this manner (Vercruyssen et al., 2017). Additionally, we argue that this operationalization facilitates the interpretation of results and enables a clearer assessment of the effects of age matching, compared to using a continuous variable. furthermore, by including all interactions between the three variables, we can model a complex degree of matching. Table 1 shows descriptive statistics for the overall match between respondents and interviewers, separately for female and male respondents.

Item nonresponse rates. Out of all questionnaire items, 315 questions were asked to all respondents across all countries. The aim was to use as many items as possible that were directly posed to respondents. As this was not possible for generated items or paradata, we only included items from the main questionnaire in the sample. Then, we proceeded to count the number of missing items for each person. Answers such as “refusal”, “don’t know” and “no answer” were coded as missing. The variable created is, therefore, a count variable that indicates the absolute number of missing items. The average number of missing items ranged from 1.55 (Belgium) to 11.46 (Hungary) (Table 2).

Reluctance to answer. In order to measure whether respondents had problems answering questions, or were reluctant to answer them, we included a pertinent item from the interviewer questionnaire, which asked the interviewers to rate the degree of reluctance a respondent displayed during the interview, measuring how often respondents were reluctant to answer. The five-point Likert scale included items such as: “very often” (5), “often” (4), “now and then” (3), “almost never” (2), and “never” (1). About 5% of all responses were in the two extreme categories, “often” and “very often”. The averages range from about 1.22 (Norway) to 2.21 (Lithuania) (Table 2).

Third adult person present and interfering with the interview. The interviewer questionnaire also gave the interviewers an option to indicate, whether any third adult persons were interfering with the interview, for example spouses, sib-

Table 1
Descriptive matching statistics by gender of respondent

	Interviewer					
	younger		same age		older	
Gender of interviewer	*n*	%	*n*	%	*n*	%
Male respondents (n=20267)						
Male	2395	12	1519	7	3683	18
Female	4527	22	2536	13	5607	28
Female respondents (n=22520)						
Male	2507	11	1398	6	3632	16
Female	5783	26	3065	14	6135	27

Source: ESS8.

lings, parents-in-law, or other people. As the ESS does not give a clear definition of interference, we assumed the colloquial definition of the term for the purpose of the analyses, which reads, "to interpose in a way that hinders or impedes; come into collision or be in opposition" ("Definition of interfere" 2018). The variable is, therefore, binary, indicating whether interference occurred or not. The percentage of interviews with interference ranged from 3% (Sweden) to 13% (Israel) (Table 2).

We will use three different outcomes in the study that serve as proxies of data quality. The first, rate of item nonresponse, is objective, while the other two rely on the subjective assessment of the interviewer. By analysing all three indicators, we hope to achieve a more complete picture with regard to the interviewing situation and data quality. Table 2 lists descriptive statistics for all indicators by country. Table 3 lists all pairwise correlations for the dependent variables. Statistical analyses reveal that the three variables are correlated, validating them as appropriate measures of the data quality construct. Focussing on the separate countries reveals that the correlation structures are similar over all countries and there is only one negative and statistically significant correlation (Israel). Fewer than 16% of all correlations are either not statistically significant or negative. Thus, we conclude that the overall associations appear to be stable and consistent over most countries in the dataset.

In sum, the indicators chosen are the best available proxies for data quality in the ESS. The number of missing items directly translates to survey quality, as missing items lower the power of any analysis. we deem it important to include a measure of the reluctance to answer questions in order to assess the general atmosphere of the interview situation, as it might not only affect interview quality but also respondents' propensity to participate in follow-up studies. Given the fact that third persons can influence data quality, as discussed earlier in the paper, we also included a variable that measures interference by third parties.

Finally, the aspect of measurement error due to interviewer assessments deserves some attention. The fact that two of the three outcomes of interest we use in the study are based on interviewer ratings could introduce measurement error into the analyses (Eckman, Sinibaldi, & Montmann-Hertz, 2013) For example, interviewers might base their ratings not on the actual interview but on stereotypes they have about their respondents with respect to obvious sociodemographic aspects (age, gender, education, ethnicity). Moreover, even in the absence of a systematic bias, it is possible that interviewers display a large variation in their ratings and perceptions, based on their own background and experiences as an interviewer, which cannot be eliminated through training, resulting in less precise assessments. However, recent empirical studies conclude that interviewer ratings are usually of high quality and validity (Kirchner, Olson, & Smyth, 2017; Plewis, Calderwood, & Mostafa, 2017). Furthermore, as we employ a multilevel (hierarchical) framework of analysis, we can capture and assess variation in responses due to different interviewers, as suggested by existing literature (Hox, 1994; O'Muircheartaigh & Campanelli, 1998).

3.3 Modelling approach

Due to the availability of multilevel data, we use three-level random intercept models in the analyses. Respondents (level 1) are nested within interviewers (level 2), which are finally nested within countries (level 3). This approach, using random intercepts for interviewers and countries, allows us to account for idiosyncratic effects and respective baselines for level 2 and 3 units, since the intercepts are allowed to vary across interviewers and countries.[1] As the propensity to answer a question may be culturally determined, we also take

[1]We additionally introduced a fourth level into the analyses for testing purposes, designating specific regions within countries. However, results indicated that this additional information resulted in worse model fits (AIC / BIC measurements) and required a great amount of additional computational time. Therefore, we retained only the more parsimonious models.

Table 2
Descriptive statistics by country

	Observations		Item nonresponse			Reluctance to answer		3rd person present
Country	Resp.	Interv.	Mean	Std. Dev.	Median	Mean	Std. Dev.	prop.
AT	1985	107	4.09	6.64	2	1.50	0.76	0.050
BE	1702	137	1.55	3.81	0	1.34	0.70	0.067
CH	1459	61	3.70	6.91	1	1.46	0.86	0.046
CZ	2186	280	4.34	6.86	2	1.84	0.92	0.028
DE	2723	267	3.31	8.03	1	1.26	0.60	0.047
EE	1963	61	3.05	7.67	0	1.58	0.85	0.050
ES	1860	132	7.62	11.15	4	1.44	0.72	0.115
FI	1868	129	2.17	7.77	0	1.40	0.85	0.039
FR	2015	171	3.49	6.96	1	1.32	0.73	0.060
GB	1892	214	4.62	7.47	2	1.45	0.79	0.044
HU	1438	113	11.46	12.11	6	2.16	1.12	0.057
IE	2689	103	4.83	6.60	2	1.61	0.91	0.049
IL	2457	138	10.94	14.02	6	1.98	1.03	0.133
IS	857	41	4.63	8.89	2	2.01	0.92	0.050
IT	2511	191	10.51	15.92	5	2.04	1.04	0.057
LT	2023	143	9.86	11.83	6	2.21	1.03	0.092
NL	1644	119	3.56	6.80	1	1.30	0.65	0.035
NO	1479	81	1.89	5.83	0	1.22	0.74	0.032
PL	1630	129	7.42	12.50	4	1.58	0.98	0.083
PT	1249	54	5.53	9.42	2	1.52	0.87	0.071
RU	2396	245	11.44	12.35	8	2.09	1.09	0.090
SE	1511	75	4.13	9.91	1	1.38	0.92	0.025
SI	1256	53	4.63	8.99	2	1.46	0.86	0.060

Source: ESS8. Item nonresponse is a count variable. Reluctance to answer is measured on a five-point scale from 1 to 5 with higher values standing for larger reluctance to answer questions. Third person interfering is a proportion between 0 and 1.

into account country-differences in this design. Furthermore, one would expect differences between interviewers, for example with respect to work experience, charisma or openness (West & Blom, 2016). By allowing the intercepts to vary, we can correctly calculate the standard errors of estimated parameters of interest. Determined by the scaling of the dependent variables, we will estimate three different models. The first dependent variable, the number of missing items (item nonresponse), is a count variable, for which we run a nonlinear model. As the descriptive statistics suggest conditional overdispersion, we use multilevel negative binomial regression instead of a multilevel Poisson regression. The statistics shown below support this preference for a negative binomial model. For the second dependent variable, reluctance to answer questions, we estimate a multilevel ordered logistic model, which is appropriate for ordinally-scaled variables. Finally, as the last dependent variable, the probability that a third person interfered with the interview, is binary, we compute a multilevel logistic regression.

We build the models as follows: the baseline model ("Empty") does not include any explanatory variables, only the dependent variable and random effects. The second model ("Main") includes the gender of the respondent, gender of the interviewer, the age-match variable between respondent and interviewer, as well as all possible two-way interactions and a three-way interaction. The final model ("Full") adds fixed effects of all control variables to the model. Based on theoretical considerations, the following controls are added to account for the fact that respondent-interviewer matching might not be absolutely random: the educational level of the respondent (ISCED classification), logged age of the respondent, average regional household incomes, number of household members, whether any children live in the household and, finally, the marital status of the respondent. Adding these variables makes the predicted outcomes more realistic, and accounts for country differences with respect to the age structure, or average educational attainment (Colsher & Wallace, 1989; De Leeuw, Hox, &

Table 3
Correlation matrices of dependent variables by country

	Reluctance	Missing items	Reluctance	Missing items	Reluctance	Missing items
	Total Sample		Finland		Lithuania	
Missing items	0.32*	-	0.31*	-	0.19*	-
3rd Person present	0.09*	0.08*	0.11*	0.10*	0.07*	0.04
	Austria		France		Netherland	
Missing items	0.17*	-	0.33*	-	0.23*	-
3rd Person present	0.05*	0.10*	0.10*	0.08*	0.09*	0.03
	Belgium		United Kingdom		Norway	
Missing items	0.22*	-	0.28*	-	0.22*	-
3rd Person present	0.12*	0.11*	0.06*	0.05*	0.02	0.02
	Switzerland		Hungary		Poland	
Missing items	0.24*	-	0.24*	-	0.23*	-
3rd Person present	0.03	0.08*	0.11*	0.04	0.18*	0.04
	Czech Republic		Ireland		Portugal	
Missing items	0.26*	-	0.10*	-	0.23*	-
3rd Person present	0.09*	0.07*	0.07*	−0.02	0.20*	0.03
	Germany		Israel		Russia	
Missing items	0.23*	-	0.05*	-	0.25*	-
3rd Person present	0.04*	0.09*	0.05*	−0.13*	0.03	0.05*
	Estonia		Iceland		Russia	
Missing items	0.17*	-	0.37*	-	0.27*	-
3rd Person present	0.08*	0.07*	0.10*	0.05	0.05*	0.06*
	Spain		Italy		Slovenia	
Missing items	0.28*	-	0.45*	-	0.28*	-
3rd Person present	0.10*	0.11*	0.07*	0.09*	0.15*	0.07*

Source: ESS8. Reported are Spearman's Rho correlation coefficients.
* $p < 0.05$

Huisman, 2003). We perform all calculations in Stata 15 (commands menbreg, meologit and melogit), using 15 integration points. For example, the regression equation for the second model ("Main") for the binary logistic model reads:

$$\begin{aligned}\text{Logit}(Y_{ijk}) = {} & \beta_0 + \beta_1 \text{Female}_{ijk} + \beta_2 \text{IntFemale}_{jk} \\ & + \beta_3 \text{AgeMatch}_{ijk} + \beta_4 \text{Female}_{ijk} \cdot \text{IntFemale}_{jk} \\ & + \beta_5 \text{Female}_{ijk} \cdot \text{AgeMatch}_{ijk} \\ & + \beta_6 \text{IntFemale}_{jk} \cdot \text{AgeMatch}_{ijk} \\ & + \beta_7 \text{Female}_{ijk} \cdot \text{IntFemale}_{jk} \cdot \text{AgeMatch}_{ijk} \quad (1)\end{aligned}$$

Here, the subscript i refers to individuals, j to interviewers and k to countries.

4 Results

4.1 Item Nonresponse

The first model examines how interviewer-matching influences item nonresponse in face-to-face interviews. The stepwise model indicates that adding fixed effects for the main variables of interest, including the matching variables and their interactions, significantly improves the model. After adding the controls, it is evident that not only do these additional variables have a large explanatory power, but also the effects of the main variables become slightly smaller, yet stay significant in almost all cases. This suggests that the matching between interviewers and respondents might not be perfectly random, yet the remaining explanatory power is

Table 4
Negative binomial regression results: Item Nonresponse Rates

	Empty		Main		Full	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Female	-	-	0.437***	0.037	0.319***	0.036
IntFemale	-	-	0.260***	0.053	0.254***	0.044
Female × IntFemale	-	-	−0.138**	0.045	−0.102*	0.044
AgeMatch (Ref.: Interviewer younger)						
Same Age	-	-	−0.096*	0.044	0.013	0.044
Interviewer older	-	-	−0.060	0.037	0.103*	0.041
Female × AgeMatch (Ref.: Female × Interviewer younger)						
Female × Same Age	-	-	−0.250***	0.062	−0.103	0.060
Female × Interviewer older	-	-	−0.282***	0.049	−0.119*	0.047
IntFemale × AgeMatch (Ref.: IntFemale × Interviewer younger)						
IntFemale × Same Age	-	-	−0.178**	0.055	−0.137*	0.054
IntFemale × Interviewer older	-	-	−0.231***	0.046	−0.181***	0.045
Female × IntFemale × AgeMatch (Ref.: Female × IntFemale × Interviewer younger)						
Female × IntFemale × Same Age	-	-	0.192*	0.075	0.133	0.073
Female × IntFemale × Interviewer older	-	-	0.159**	0.060	0.107	0.058
Educational level (ISCED) (Ref.: Less than lower secondary (I))						
Lower secondary (II)	-	-	-	-	−0.384***	0.027
Lower tier upper secondary (IIIb)	-	-	-	-	−0.626***	0.028
Upper tier upper secondary (IIIa)	-	-	-	-	−0.718***	0.027
Advanced vocational (IV)	-	-	-	-	−0.835***	0.028
Lower tertiary education (V1)	-	-	-	-	−1.000***	0.031
Higher tertiary education (V2)	-	-	-	-	−0.996***	0.029
Logged Age of Respondent	-	-	-	-	0.072*	0.029
Number of people living regularly as member of household	-	-	-	-	0.028***	0.007
Children living in the household	-	-	-	-	−0.075***	0.017
Marital Status (Ref.: Married)						
Separated / Divorced	-	-	-	-	0.043*	0.022
Widowed	-	-	-	-	0.245***	0.024
Never Married	-	-	-	-	0.076***	0.019
Mean Income per Region	-	-	-	-	−0.137***	0.019
Constant	1.179***	0.018	0.755***	0.044	3.137***	0.166
Variance Components						
Country variance	0.450***	0.023	0.269***	0.016	0.320***	0.016
Interviewer variance	0.692***	0.021	0.661***	0.021	0.708***	0.021
Alpha	1.14		1.10		1.01	
95%-C.I. Alpha	[1.11; 1.16]		[1.08; 1.12]		[0.99; 1.03]	
N	41,737		41,737		41,737	
AIC	207,872		206,984		205,062	

Source: ESS8. N = 41,737 respondents nested within J = 3,042 interviewers nested within K = 23 countries. The dependent variable is a count variable.
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

still significant. Additionally, every single model presented has a better statistical fit than the corresponding model without multilevel factors, which was to be expected, given the design of the survey. We test this statistically using the AIC, which becomes smaller and signals a better fit. Additionally, we perform likelihood-ratio tests, which lead to the same conclusion (not depicted).

To ease interpretation in the presence of many interactions, we calculate predicted values for the full model, which has the best predictive power. For these predictions, we incorporate random effects into the results by averaging the predicted probabilities that are marginal with respect to the random effects (conditional computation using empirical Bayes means).[2] Table 4 shows the regression results, while

[2]For a more detailed discussion on how Stata computes these marginal predictions in hierarchical models, please

Table 5
Number of missing items (item nonresponse): predictions

Gender of		Interviewer					
		younger		same age		older	
Respondent	Interviewer	Pred.	Std. Err.	Pred.	Std. Err.	Pred.	Std. Err.
Male	Male	14.87	0.97	15.07	1.03	16.48	1.08
	Female	19.17***	1.03	16.94*	0.95	17.74	0.97
Female	Male	20.46	1.33	18.70	1.28	20.13	1.34
	Female	23.82***	1.26	21.68**	1.18	21.77	1.18

Source: ESS8. Predicted values calculated for the Full Model. Asterisks indicate significance levels for each pairwise comparison (Male VS Female interviewer within one gender of respondent).

Table 5 includes the predicted values. The appendix also lists the results calculated for all countries separately.[3] Based on the data, women have, on average, more missing items than men do. Men's item nonresponse rate is lowest when interviewed by younger males (14.87) and the highest when interviewed by a younger female (19.17). For women, data quality is highest when interviewed by a male their age (18.70) and lowest when interviewed by a younger female (23.82). Overall, gender matching with the interviewer significantly reduces item nonresponse in men, while the opposite is true for women. The alpha-statistics indicate that negative binomial models are a better fit than Poisson models, as they are larger than zero (Rabe-Hesketh & Skrondal, 2012).

4.2 Reluctance to answer

Predicted probabilities are displayed for the most extreme category, that is, the probability of answering questions reluctantly "very often". The results can be found in Table 6 and Table 7. Interestingly, a number of main variables lose their statistical significance in the model with the fixed effects for the controls added, yet, even in this case a few gender differences remain. The results indicate that a gender match reduces the probability that men will be reluctant to answer a question "very often" (albeit the differences are not always statistically significant). We find no significant differences for women.

4.3 Probability of a third person interfering

The last outcome investigates whether sociodemographic matching between respondents and interviewers influences the probability that a third person interferes with the interview, potentially affecting the answers given by the respondent. Table 8 and Table 9 present the results of pertinent analyses. As the regression results are difficult to interpret, we use predicted values for a more vivid estimation of the effects. Men are least likely to have a third party interfere when interviewed by a man their age (5%) and most likely when interviewed by a younger woman (8%). For women, the probabilities of interference are lowest when paired with an older woman (5%) and highest with a younger man (7%). Thus, both men and women seem to benefit from gender matching, but age matching is also relevant, as not all combinations show significant differences. Furthermore, the evidence that indicates differences is weak for men.

As a general robustness check, we test potential multicollinearity to account for the many interaction terms present. The mean VIF is at an acceptable level of 5.34. Although a few interaction terms are slightly above the critical value of 10, including the interactions is crucial and unavoidable, given the nature of the analyses. However, the overall bias should be small, since the deviations are not large.

5 Discussion

The present study provides a novel contribution to the literature on the effects of sociodemographic matching between survey respondents and interviewers. The first important finding of the research relates to item nonresponse rates in face-to-face interviews. Multilevel negative binomial regressions demonstrate that some pairings actually increase the number of missing items, and that women have, on average, more missing items than men. This suggests that, in general, women tend to refuse to answer more questions than men, independently of the characteristics of the interviewer. Secondly, gender matching increases the number of missing items for women, while the opposite is the case for men. This corroborates the findings of Vercruyssen et al. (2017), who use the same dataset, but restrict it to the Belgian subsample.

refer to https://www.statalist.org/forums/forum/general-stata-discussion/general/1309934-margins-after-melogit (2019-11-20).

[3]These results should be interpreted with caution, since the power is much lower and it is not possible to estimate some of the effects due to perfect prediction.

Table 6
Ordered logistic regression results: reluctance to answer

	Empty		Main		Full	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Female	-	-	0.266***	0.064	0.238***	0.064
IntFemale	-	-	0.195*	0.077	0.163*	0.080
Female × Interviewer Female	-	-	−0.136	0.077	−0.153*	0.078
AgeMatch (Ref.: Interviewer younger)						
Same Age	-	-	−0.174*	0.078	0.019	0.081
Interviewer older	-	-	−0.459***	0.065	−0.096	0.077
Female × AgeMatch (Ref.: Female × Interviewer younger)						
Female × Same Age	-	-	−0.190	0.108	−0.157	0.109
Female × Interviewer older	-	-	−0.160	0.087	−0.094	0.088
IntFemale × AgeMatch (Ref.: IntFemale × Interviewer younger)						
IntFemale × Same Age	-	-	−0.042	0.096	−0.056	0.097
IntFemale × Interviewer older	-	-	−0.042	0.081	−0.031	0.082
Female × IntFemale × AgeMatch (Ref.: Female × IntFemale × Interviewer younger)						
Female × IntFemale × Same Age	-	-	0.010	0.132	0.048	0.133
Female × IntFemale × Interviewerolder	-	-	0.045	0.108	0.042	0.108
Educational level (ISCED) (Ref.: Less than lower secondary (I))						
Lower secondary (II)	-	-	-	-	−0.229***	0.050
Lower tier upper secondary (IIIb)	-	-	-	-	−0.466***	0.053
Upper tier upper secondary (IIIa)	-	-	-	-	−0.597***	0.051
Advanced vocational (IV)	-	-	-	-	−0.622***	0.054
Lower tertiary education (V1)	-	-	-	-	−0.769***	0.059
Higher tertiary education (V2)	-	-	-	-	−0.794***	0.055
Logged Age of Respondent	-	-	-	-	0.343***	0.059
Number of people living regularly as member of household	-	-	-	-	−0.023	0.013
Children living in the household	-	-	-	-	0.042	0.033
Marital Status (Ref.: Married)						
Separated / Divorced	-	-	-	-	0.025	0.039
Widowed	-	-	-	-	0.092*	0.044
Never Married	-	-	-	-	0.053	0.036
Mean Income per Region	-	-	-	-	0.128*	0.058
Cutpoint 1	−0.236***	0.036	0.598***	0.061	−0.508	0.342
Cutpoint 2	1.415***	0.037	2.264***	0.062	1.178***	0.342
Cutpoint 3	3.259***	0.043	4.123***	0.066	3.056***	0.342
Cutpoint 4	4.403***	0.053	5.270***	0.074	4.207***	0.343
Variance Components						
Country variance	0.572***	0.034	1.243***	0.081	0.792***	0.107
Interviewer variance	1.754***	0.058	1.781***	0.060	1.836***	0.070
N	41,651		41,651		41651	
AIC	74,985		74,597		74,385	

Source: ESS8. N = 41,651 respondents nested within J = 3,042 interviewers nested within K = 23 countries. The dependent variable is ordinally scaled with values from 1 to 5.
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

By using information about all countries, we demonstrate that the effect is stable and generalizable to a much wider population. Furthermore, we do not find support for the theory of liking as the driver of sociodemographing matching effects, as gender matching does not always improve the quality of survey data. Considering age matching, women in particular refuse to answer questions more frequently when age is not matched, while this is not the case for men. To sum up, we reject hypothesis one for female respondents, but find support for it for male respondents.

Going further, the second dependent variable of interest in the study is the reluctance to answer questions in an interview. In the analyses, we presented predictions for the most extreme negative category (reluctance to answer questions "very often". The outcomes reveal that sociodempgraphic matching results in higher quality data for men, but not for

Table 7
Probability of being reluctant to answer "very often": predictions

Gender of		Interviewer					
		younger		same age		older	
Respondent	Interviewer	Pred.	Std. Err.	Pred.	Std. Err.	Pred.	Std. Err.
Male	Male	0.1243	0.0173	0.1259	0.0182	0.1165	0.0173
	Female	0.1384*	0.0189	0.1351	0.0193	0.1273	0.0191
Female	Male	0.1452	0.0190	0.1328	0.0188	0.1283	0.0186
	Female	0.1462	0.0196	0.1329	0.0191	0.1300	0.0193

Source: ESS8. Predicted values calculated for the Full Model, Outcome 5 (respondents "very often" reluctant to answer). Asterisks indicate significance levels for each pairwise comparison (Male VS Female interviewer within one gender of respondent). * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

women. Our conclusion is based on the predicted values for category 5 of the outcome. However, it is also supported by the raw regression coefficients. In the full model, only the gender variable and the gender-matching variable are significant. Therefore, we retain hypothesis two only for the male subsample. Nevertheless, it is important to note that the observed effects are probably weak.

The last dependent variable of interest for this study is the probability that a third person interferes with the interview. Given that women in particular have significantly higher probabilities of allowing such interference when gender is mismatched provides support the theory of liking. The observed effects for men in the sample are small. These results are more or less in line with the theory of liking, which predicts that sociodemographic matching between respondents and interviewers produces better results and less interference. Therefore, we retain hypothesis three. However, it remains unclear why the patterns arise. Future studies should focus on this finding and attempt to identify potential causal mechanisms behind it. For now, we can only speculate that traditional gender roles are the main driver of the effect, putting husbands or male family members at the head of the household and requiring them to protect their spouses or female relatives from strangers during interviews. Indeed, a cursory comparison of country effects indicates that countries in which traditional gender roles dominate show larger effects for interference for female respondents.

All three regressions show significant unexplained variance between countries and interviewers (variances of random effects). This indicates that there are differences in outcomes, and that these are usually larger for interviewers than for countries. As mentioned above, one explanatory factor could be socio-cultural differences between countries and a different way to handle and structure face-to-face interactions with strangers. Individual differences between interviewers can influence their ability to collect quality data. Those include experience as an interviewer, social intelligence, or charisma. Because we do not measure them directly in the model, it is possible that they contribute to the unexplained variance.

Finally, as already discussed before in section two, we include two subjective measurements of data quality in the study. As noted in the literature review, both are widely accepted as useful measures of interview quality. Our study supports previous findings in this domain by highlighting the importance of subjective measurements for survey quality assessments and demonstrating their value for quantitative analyses. Although we do not find unambiguous support for all of our hypotheses, we do find convergence between the three measurements, which appear to provide a coherent picture of the overall survey quality and interview situation. Lastly, concerning cross-country differences, it is worth noting that some countries in the sample exhibit strong effects, while others show no significant effects of matching. This is not surprising, given the potential differences in social and cultural backgrounds. Future studies should focus on this aspect and try to explain this cross-country variation, and explore whether it is possible to sort countries into larger clusters that would allow an explanation of these patterns.

Limitations

The main shortcoming of the research design is that it is not interpenetrated, i.e. it does not guarantee random matching between respondents and interviewers at the onset of the data collection process. Additionally, it presents the possibility of selective refusal rates, based on the originally assigned interviewer. Unfortunately, there is no simple solution for this problem. However, by applying multilevel models and controlling for respondent characteristics, these obstacles should be attenuated. One should also consider that the ESS, despite these challenges, provides excellent data for a large number of respondents, with a wide set of additional information, which makes this analysis possible in the

Table 8
Logistic regression results: adult third persons interfering

	Empty		Main		Full	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Female	-	-	−0.237*	0.114	0.004	0.119
IntFemale	-	-	0.226	0.117	0.220	0.120
Female ×Interviewer Female	-	-	−0.412**	0.141	−0.400**	0.145
AgeMatch (Ref.: Interviewer younger)						
Same Age	-	-	−0.597***	0.149	−0.503**	0.155
Interviewer older	-	-	−0.603***	0.118	−0.386**	0.139
Female × AgeMatch (Ref.: Female × Interviewer younger)						
Female ×Same Age	-	-	0.568**	0.205	0.438*	0.211
Female ×Interviewer older	-	-	0.356*	0.161	0.248	0.166
IntFemale × AgeMatch (Ref.: IntFemale × Interviewer younger)						
IntFemale ×Same Age	-	-	−0.094	0.185	−0.058	0.189
IntFemale ×Interviewer older	-	-	−0.022	0.146	0.001	0.150
Female × IntFemale × AgeMatch (Ref.: Female × IntFemale × Interviewer younger)						
Female ×IntFemale ×Same Age	-	-	−0.262	0.259	−0.236	0.265
Female ×IntFemale ×Interviewer older	-	-	−0.194	0.204	−0.240	0.209
Educational level (ISCED) (Ref.: Less than lower secondary (I))						
Lower secondary (II)	-	-	-	-	−0.358***	0.088
Lower tier upper secondary (IIIb)	-	-	-	-	−0.646***	0.095
Upper tier upper secondary (IIIa)	-	-	-	-	−0.846***	0.092
Advanced vocational (IV)	-	-	-	-	−1.109***	0.100
Lower tertiary education (V1)	-	-	-	-	−1.281***	0.113
Higher tertiary education (V2)	-	-	-	-	−1.332***	0.105
Logged Age of Respondent	-	-	-	-	0.113	0.112
Number of people living regularly as member of household	-	-	-	-	0.210***	0.023
Children living in the household	-	-	-	-	−0.625***	0.065
Marital Status (Ref.: Married)						
Separated / Divorced	-	-	-	-	−1.231***	0.106
Widowed	-	-	-	-	−1.649***	0.121
Never Married	-	-	-	-	−0.606***	0.077
Mean Income per Region	-	-	-	-	−0.131*	0.056
Constant	−3.358***	0.108	−2.975***	0.137	−2.198***	0.591
Variance Components	-	-	-	-	-	-
Country variance	0.217**	0.073	0.205**	0.069	0.180**	0.062
Interviewer variance	1.413***	0.099	1.417***	0.099	1.504***	0.106
Intraclass correlations	-	-	-	-	-	-
Country	0.044	0.014	0.042	0.014	0.036	0.012
Interviewer nested in country	0.331	0.017	0.330	0.017	0.339	0.017
N	41,737		41,737		41,737	
AIC	17,739		17,553		16,739	

Source: ESS8. N = 41,716 respondents nested within J = 3,042 interviewers nested within K = 23 countries. Standard errors in parentheses. Reported are coefficients on the log-odds scale as the dependent variable is binary.
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

first place. Previous research has successfully demonstrated that the ESS is suitable for the research questions considered here.

A second limitation is the previously discussed operationalization of age matching. The cut-off point of five years is arbitrary and different values could be considered. However, using a continuous measure of age might also be problematic, making its interactions with gender, a categorical variable, less intuitive and more difficult to demonstrate. It is unclear within which range respondents perceive an interviewer as their own age, which is further complicated by the fact that only factual age is considered in the study, but not how old someone appears to be. This question, about the perception of interviewer age, deserves more attention. Future studies should compare different operationalisations to test whether these effects are stable. Furthermore, col-

Table 9
Probability of an adult third person interfering: predictions

Gender of		Interviewer					
		younger		same age		older	
Respondent	Interviewer	Pred.	Std. Err.	Pred.	Std. Err.	Pred.	Std. Err.
Male	Male	0.0711	0.0077	0.0474	0.0076	0.0522	0.0066
	Female	0.0842	0.0061	0.0541	0.0061	0.0624	0.0062
Female	Male	0.0713	0.0079	0.0678	0.0061	0.0639	0.0085
	Female	0.0618	0.0054	0.0461**	0.0071	0.0454**	0.0049

Source: ESS8. Predicted values calculated for the Full Model. Asterisks indicate significance levels for each pairwise comparison (Male VS Female interviewer within one gender of respondent).
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

lecting more information about the educational levels of the interviewers might be important to test the influence of educational matching on data quality in face-to-face settings. Since this information was not included in the ESS, it was not possible to investigate this aspect of matching.

Finally, the question arises how these findings can be used to improve the survey process in future studies. Both men and women have lower item nonresponse when interviewed by men, which is highly relevant for data quality, as missing items are always a severe problem. However, recommending the employment of more male interviewers to increase the number of retrieved items is debatable at this point, as more investigations based on experimental designs are required before such practices can be justified.

6 Conclusion

The aim of the current study was to investigate the effects of sociodemographic matching between respondents and interviewers in face-to-face interviews on proxy measures of data quality. The analysis, based on multilevel data from 23 European countries, leads to the conclusion that some matches actually can significantly improve data quality. Specifically, gender matching decreases the amount of item nonresponse for male respondents, but has a negative effect for female respondents. Future studies should build on these results and attempt to investigate causal mechanisms behind these effects. Testing why the theory of liking fails in this context is a highly relevant starting point. Secondly, modelling the process of self-selection into treatment based on interviewer characteristics is another promising area of research, which can help understand whether interviewer characteristics influence unit nonresponse rates. Furthermore, future research should try to uncover patterns of country differences in these effects and analyse whether the patterns found are the same in all European countries, or whether there are significant differences. It is possible that macro-influences, such as economic prosperity, welfare state regime, degree of multiculturalism or other psychosocial factors, affect interactions between respondents and interviewers. Considering that matching between respondents and interviewers could be a relatively simple, yet promising technique to improve data quality, it is critical to explore these avenues further, as they are central to any survey research endeavour.

Acknowledgements

Conflict of Interest: The author declares that he has no conflict of interest.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements: Special thanks goes to Mark Trappmann, Viktoria Sophie Zorn, Agata Kraj and all anonymous reviewers.

References

Aquilino, W. S., Wright, D. L., & Supple, A. J. (2000). Response effects due to bystander presence in CASI and paper-and-pencil surveys of drug use and alcohol use. *Substance Use & Misuse*, *35*(6-8), 845–867. doi:10.3109/10826080009148424

Axinn, W. G. (1989). Interviewers and data quality in a less developed setting. *Journal of Official Statistics*, *5*(3), 265.

Boeije, H. R. (2004). And then there were three: Self-presentational styles and the presence of the partner as a third person in the interview. *Field Methods*, *16*(1), 3–22. doi:10.1177/1525822X03259228

Brashears, M. E. (2015). A longitudinal analysis of gendered association patterns: Homophily and social distance in the general social survey. *Journal of Social Structure*, *16*, 1.

Brunton-Smith, I., Sturgis, P., & Leckie, G. (2017). Detecting and understanding interviewer effects on survey data by using a cross-classified mixed effects location-scale model. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *180*(2), 551–568. doi:10.1111/rssa.12205

Byrne, D., Gouaux, C., Griffitt, W., Lamberth, J., Murakawa, N., Prasad, M., . . . Ramirez III, M. (1971). The ubiquitous relationship: Attitude similarity and attraction: A cross-cultural study. *Human Relations*, *24*(3), 201–207.

Catania, J. A., Binson, D., Canchola, J., Pollack, L. M., & Hauck, W. (1996). Effects of interviewer gender, interviewer choice, and item wording on responses to questions concerning sexual behavior. *Public Opinion Quarterly*, *60*(3), 345. doi:10.1086/297758

Chun, H., Tavarez, M. I., Dann, G. E., & Anastario, M. P. (2011). Interviewer gender and self-reported sexual behavior and mental health among male military personnel. *International journal of public health*, *56*(2), 225–229.

Colsher, P. L., & Wallace, R. B. (1989). Data quality and age: Health and psychobehavioral correlates of item nonresponse and inconsistent responses. *Journal of Gerontology*, *44*(2), P45–P52. doi:10.1093/geronj/44.2.P45

Davis, R. E., Couper, M. P., Janz, N. K., Caldwell, C. H., & Resnicow, K. (2010). Interviewer effects in public health surveys. *Health Education Research*, *25*(1), 14–26. doi:10.1093/her/cyp046

De Leeuw, E. D., Hox, J. J., & Huisman, M. (2003). Prevention and treatment of item nonresponse. *Journal of Official Statistics*, *19*, 153–176.

Dotinga, A., Van den Eijnden, R. J., Bosveld, W., & Garretsen, H. F. (2005). The effect of data collection mode and ethnicity of interviewer on response rates and self-reported alcohol use among Turks and Moroccans in the Netherlands: An experimental study. *Alcohol and Alcoholism*, *40*(3), 242–248.

Durrant, G. B., Groves, R. M., Staetsky, L., & Steele, F. (2010). Effects of interviewer attitudes and behaviors on refusal in household surveys. *Public Opinion Quarterly*, *74*(1), 1–36. doi:10.1093/poq/nfp098

Eckman, S., Sinibaldi, J., & Montmann-Hertz, A. (2013). Can interviewers effectively rate the likelihood of cases to cooperate? *Public Opinion Quarterly*, *77*(2), 561–573. doi:10.1093/poq/nft012

ESS. (2018). ESS documentation report 2018. The ESS Archive. Retrieved August 22, 2018, from http://www.europeansocialsurvey.org/docs/round8/survey/ESS8_data_documentation_report_e02_0.pdf

Fischer, C. S. (Ed.). (1977). *Networks and places: Social relations in the urban setting*. New York: Free Press.

Fletcher, C., & Spencer, A. (1984). Sex of candidate and sex of interviewer as determinants of self-presentation orientation in interviews an experimental study. *Applied Psychology*, *33*(3), 305–313. doi:10.1111/j.1464-0597.1984.tb01437.x

Groves, R. M. (Ed.). (2009). *Survey methodology* (2nd ed). OCLC: ocn302189175. Hoboken, N.J: Wiley.

Groves, R. M., Cialdini, R. B., & Couper, M. (1991). *Toward a useful theory of survey participation*. US Census Bureau [custodian].

Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly*, *56*(4), 475. doi:10.1086/269338

Hartmann, P. (1994). Interviewing when the spouse is present. *International Journal of Public Opinion Research*, *6*(3), 298–306. doi:https://doi.org/10.1093/ijpor/6.3.298

Herzog, A. R., & Rodgers, W. L. (1992). The use of survey methods in research on older americans. *The epidemiologic study of the elderly*, 60–90.

Hornberger, A. P., Medley-Proctor, K., Nettles, C. D., Cimporescu, M. A., & Howe, G. W. (2016). The influence of the racial/ethnic match of interviewer and respondent on the measurement of couples relationship quality and emotional functioning. *Couple and Family Psychology: Research and Practice*, *5*(1), 12–26. doi:10.1037/cfp0000053

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

Hox, J. J., De Leeuw, E. D., & Kreft, G. (1991). The effect of interviewer and respondent characteristics on the quality of survey data: A multilevel model. In *Measurement errors in surveys* (pp. 439–461). John Wiley & Sons.

Hurtado, A. (1994). Does similarity breed respect? Interviewer evaluations of Mexican-descent respondents in a bilingual survey. *Public Opinion Quarterly*, *58*(1), 77–95.

Johnson, T. P., Fendrich, M., Shaligram, C., Garcy, A., & Gillespie, S. (2000). An evaluation of the effects of interviewer characteristics in an RDD telephone survey of drug use. *Journal of Drug Issues*, *30*(1), 77–101.

Johnson, T. P., & Parsons, J. A. (1994). Interviewer effects on self-reported substance use among homeless persons. *Addictive Behaviors*, *19*(1), 83–93.

Kirchner, A., Olson, K., & Smyth, J. D. (2017). Do interviewer postsurvey evaluations of respondents' engage-

ment measure who respondents are or what they do? a behavior coding study. *Public Opinion Quarterly*, *81*(4), 817–846. doi:10.1093/poq/nfx026

Lahaut, V. M., Jansen, H. A., Van de Mheen, D., & Garretsen, H. F. (2002). Non-response bias in a sample survey on alcohol consumption. *Alcohol and Alcoholism*, *37*(3), 256–260.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444. doi:10.1146/annurev.soc.27.1.415

Mneimneh, Z., Elliott, M. R., Tourangeau, R., & Heeringa, S. G. (2018). Cultural and interviewer effects on interview privacy: Individualism and national wealth. *Cross-Cultural Research*, *52*(5), 496–523. doi:10.1177/1069397118759014

Mneimneh, Z. M., Tourangeau, R., Pennell, B.-E., Heeringa, S. G., & Elliott, M. R. (2015). Cultural variations in the effect of interview privacy and the need for social conformity on reporting sensitive information. *Journal of Official Statistics*, *31*(4), 673–697.

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.

Oyinlade, A. O., & Losen, A. (2014). Extraneous effects of race, gender, and race-gender homo-and heterophily conditions on data quality. *SAGE Open*, *4*(1), 1–13. doi:10.1177/2158244014525418

Plewis, I., Calderwood, L., & Mostafa, T. (2017). Can interviewer observations of the interview predict future response? *Methods, Data, Analyses*, *11*(1). doi:10.12758/mda.2016.010

Poulin, M. (2010). Reporting on first sexual experience: The importance of interviewer-respondent interaction. *Demographic Research*, *22*, 237–288. doi:10.4054/DemRes.2010.22.11

Preetz, R. (2017). Why were there three? – determinants of the presence of an intimate partner during face-to-face interviews. *Survey Methods*. doi:10.13094/smif-2017-00007

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling using Stata* (3rd ed.). OCLC: ocn787853271. College Station, Texas: Stata Press Publication.

Reuband, K.-H. (1992). On third persons in the interview situation and their impact on responses. *International Journal of Public Opinion Research*, *4*(3), 269–274.

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

Riphahn, R. T., & Serfling, O. (2005). Item non-response on income and wealth questions. *Empirical Economics*, *30*(2), 521–538. doi:10.1007/s00181-005-0247-7

Rogers, E. M., & Bhowmik, D. K. (1970). Homophily-heterophily: Relational concepts for communication research. *Public opinion quarterly*, *34*(4), 523–538.

Smith, J. A., McPherson, M., & Smith-Lovin, L. (2014). Social distance in the United States: Sex, race, religion, age, and education homophily among confidants, 1985 to 2004. *American Sociological Review*, *79*(3), 432–456. doi:10.1177/0003122414531776

Stoop, I., Koch, A., Halbherr, V., Loosveldt, G., & Fitzgerald, R. (2016). Field procedures in the European Social Survey round 8: Guidelines for enhancing response rates and minimising nonresponse bias. ESS ERIC Headquarters. Retrieved August 23, 2018, from http://www.europeansocialsurvey.org/docs/round8/methods/ESS8_guidelines_enhancing_response_rates_minimising_nonresponse_bias.pdf

Stotland, E., & Patchen, M. (1961). Identification and changes in prejudice and in authoritarianism. *The Journal of Abnormal and Social Psychology*, *62*(2), 265.

Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological bulletin*, *135*(6), 859.

Suedfeld, P., Bochner, S., & Matas, C. (1971). Petitioner's attire and petition signing by peace demonstrators: A field experiment 1. *Journal of Applied Social Psychology*, *1*(3), 278–283.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. doi:10.1017/CBO9780511819322

Vercruyssen, A., Wuyts, C., & Loosveldt, G. (2017). The effect of sociodemographic (mis)match between interviewers and respondents on unit and item nonresponse in Belgium. *Social Science Research*, *67*, 229–238. doi:10.1016/j.ssresearch.2017.02.007

Webster, C. (1996). Hispanic and anglo interviewer and respondent ethnicity and gender: The impact on survey response quality. *Journal of Marketing Research*, *33*(1), 62–72.

Weiss, C. H. (1968). Validity of welfare mothers' interview responses. *Public Opinion Quarterly*, *32*(4), 622–633.

West, B. T., & Blom, A. G. (2016). Explaining interviewer effects: A research synthesis. *Journal of Survey Statistics and Methodology*, *5*(2), 175–211. doi:10.1093/jssam/smw024

West, B. T., Elliott, M. R., Mneimneh, Z., Wagner, J., Peytchev, A., & Trappmann, M. (2019). An examination of an interviewer-respondent matching protocol in

a longitudinal CATI study. *Journal of Survey Statistics and Methodology*. doi:10.1093/jssam/smy028

Williams Jr, J. A. (1964). Interviewer-respondent interaction: A study of bias in the information interview. *Sociometry*, *27*(3), 338–352.

Wilson, S. R., Brown, N. L., Mejia, C., & Lavori, P. W. (2002). Effects of interviewer characteristics on reported sexual behavior of California Latino couples. *Hispanic Journal of Behavioral Sciences*, *24*(1), 38–62.

Appendix
Tables

Table A1
Item Nonresponse Rates (Number of Missing Items) by country

	Gender of		Interviewer					
			younger		same age		older	
	Respondent	Interviewer	Pred.	Std. Err.	Pred.	Std. Err.	Pred.	Std. Err.
AT	Male	Male	4.724	0.687	4.522	0.649	5.130	0.654
	-	Female	3.978	0.475	3.362	0.444	2.255***	0.236
	Female	Male	7.581	1.036	5.731	0.826	4.821	0.608
	-	Female	3.439***	0.404	2.383***	0.292	3.529	0.335
BE	Male	Male	0.967	0.203	1.026	0.276	1.681	0.371
	-	Female	2.539*	0.728	1.391	0.414	1.654	0.359
	Female	Male	1.589	0.343	1.412	0.354	1.957	0.435
	-	Female	1.724	0.460	2.331	0.705	1.653	0.383
CH	Male	Male	2.805	0.615	2.778	0.544	3.762	0.374
	-	Female	3.559	0.717	4.523	1.006	2.552	0.509
	Female	Male	5.371	1.012	2.943	0.593	3.668	0.336
	-	Female	5.332	1.104	6.351*	1.420	3.721	0.657
CZ	Male	Male	5.047	0.831	2.946	0.633	4.169	0.675
	-	Female	4.684	0.570	5.761**	0.758	5.709	0.735
	Female	Male	5.326	0.848	4.820	0.880	5.161	0.915
	-	Female	4.882	0.565	5.654	0.742	5.022	0.647
DE	Male	Male	3.456	0.622	2.326	0.395	3.106	0.382
	-	Female	3.217	0.608	2.718	0.512	3.153	0.422
	Female	Male	3.859	0.710	3.546	0.647	3.088	0.397
	-	Female	5.374	0.961	4.388	0.869	3.960	0.543
EE	Male	Male	1.880	0.716	0.488	0.267	2.394	0.735
	-	Female	1.876	0.231	1.778***	0.289	2.964	0.398
	Female	Male	4.391	1.267	1.674	0.775	4.868	1.427
	-	Female	2.561	0.273	4.094***	0.546	3.964	0.554
ES	Male	Male	6.320	0.935	5.247	0.962	6.236	1.009
	-	Female	6.945	0.672	5.461	0.628	6.097	0.659
	Female	Male	7.989	1.154	8.005	1.276	7.958	1.206
	-	Female	9.044	0.895	7.851	0.943	8.413	0.933
FI	Male	Male	1.154	0.564	1.668	0.772	1.269	0.593
	-	Female	2.009	0.362	1.236	0.275	1.608	0.338
	Female	Male	1.800	0.793	4.882	2.896	1.415	0.760
	-	Female	2.445	0.443	2.682	0.564	2.415	0.532
FR	Male	Male	2.352	0.421	2.857	0.598	2.150	0.387
	-	Female	3.468	0.469	3.349	0.508	3.244*	0.434
	Female	Male	2.388	0.414	3.122	0.616	3.749	0.636
	-	Female	4.643***	0.571	4.034	0.576	3.741	0.493
GB	Male	Male	3.837	0.682	3.841	0.783	4.049	0.601
	-	Female	4.462	0.762	4.984	0.937	5.327	0.821
	Female	Male	3.175	0.556	3.596	0.727	4.874	0.700
	-	Female	5.893**	0.970	4.459	0.760	7.449*	1.148
HU	Male	Male	9.856	1.083	10.126	1.464	7.647	0.767
	-	Female	12.663*	0.998	11.475	1.024	10.789**	0.811
	Female	Male	11.727	1.120	9.412	1.212	10.684	1.177
	-	Female	14.259	0.989	15.003***	1.024	12.352	0.866

Continues on next page

Continued from previous page

	Gender of		Interviewer					
			younger		same age		older	
	Respondent	Interviewer	Pred.	Std. Err.	Pred.	Std. Err.	Pred.	Std. Err.
IE	Male	Male	4.315	0.476	3.541	0.348	3.446	0.308
	-	Female	5.157	0.749	3.899	0.525	4.227	0.419
	Female	Male	5.095	0.603	3.921	0.486	4.007	0.443
	-	Female	7.959*	1.166	4.280	0.573	4.852	0.510
IL	Male	Male	10.647	0.878	13.303	1.151	12.484	0.924
	-	Female	6.990***	0.842	6.013***	0.809	6.040***	0.586
	Female	Male	14.854	1.238	14.266	1.263	15.366	1.156
	-	Female	9.240***	0.989	11.146	1.512	6.766***	0.579
IS	Male	Male	3.340	0.875	2.417	1.007	2.384	1.031
	-	Female	4.075	0.713	2.976	0.646	2.094	0.407
	Female	Male	4.153	1.075	5.330	2.336	3.015	1.816
	-	Female	6.818	1.166	2.921	0.586	3.513	0.694
IT	Male	Male	7.479	0.860	8.345	1.209	6.505	0.769
	-	Female	9.127	0.856	7.486	0.664	7.373	0.527
	Female	Male	11.037	1.338	10.079	1.529	8.466	0.988
	-	Female	9.728	0.829	8.805	0.802	8.747	0.619
LT	Male	Male	7.363	1.578	4.692	1.058	8.866	2.045
	-	Female	14.944***	0.903	16.806***	1.182***	19.657	1.400
	Female	Male	8.931	0.901	7.944	1.431	12.424	2.519
	-	Female	17.839***	0.972	16.632***	1.031	21.789***	1.490
NL	Male	Male	3.707	0.914	2.457	0.485	2.491	0.342
	-	Female	3.170	0.583	2.978	0.570	3.685	0.581
	Female	Male	3.447	0.742	2.993	0.562	3.643	0.498
	-	Female	4.534	0.816	5.044	0.880	4.629	0.666
NO	Male	Male	1.160	0.251	1.514	0.349	1.799	0.294
	-	Female	1.755	0.395	1.931	0.503	2.040	0.395
	Female	Male	1.780	0.376	1.474	0.366	2.719	0.477
	-	Female	2.429	0.550	1.661	0.453	2.043	0.385
PL	Male	Male	4.809	0.632	5.762	0.915	4.066	0.494
	-	Female	7.745**	0.847	6.899	0.891	6.169**	0.578
	Female	Male	8.248	1.032	8.073	1.250	5.509	0.658
	-	Female	10.375	1.084	8.642	1.002	7.821**	0.721
PT	Male	Male	1.171	0.309	1.900	0.951	4.523	3.849
	-	Female	2.717**	0.337	5.269	1.619	4.577	2.209
	Female	Male	2.766	0.820	3.779	1.717	9.025	4.569
	-	Female	4.045	0.461	7.132	1.649	17.108	6.530
RU	Male	Male	12.531	2.037	7.466	2.029	10.174	2.156
	-	Female	11.185	0.663	10.400	0.752	10.134	0.700
	Female	Male	11.416	1.752	8.659	3.045	13.661	3.919
	-	Female	12.312	0.656	12.256	0.771	13.668	0.971
SE	Male	Male	2.089	0.479	3.736	1.086	2.568	0.563
	-	Female	4.228*	0.839	2.570	0.529	3.540	0.593
	Female	Male	3.519	0.822	3.187	0.765	4.070	0.862
	-	Female	5.354	1.006	5.713	1.154	5.089	0.827
SI	Male	Male	1.695	0.415	5.070	1.621	2.174	0.630
	-	Female	3.835***	0.454	4.370	0.703	7.419***	1.327
	Female	Male	2.874	0.633	6.851	2.339	3.067	1.043
	-	Female	5.579***	0.610	4.656	0.740	5.512	0.954

Source: ESS8. Predicted values calculated for the Full Model. Asterisks indicate significance levels for each pairwise comparison (Male VS Female interviewer within one gender of respondent). Standard errors in parentheses.

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table A2
Reluctance to answer questions ("very often") by country

	Gender of		Interviewer					
			younger		same age		older	
	Respondent	Interviewer	Pred.	Std. Err.	Pred.	Std. Err.	Pred.	Std. Err.
AT	Male	Male	0.004	0.002	0.004	0.002	0.004	0.002
	-	Female	0.003	0.001	0.003	0.001	0.004	0.002
	Female	Male	0.005	0.002	0.006	0.002	0.003	0.001
	-	Female	0.004	0.001	0.004	0.001	0.004	0.002
BE	Male	Male	0.007	0.003	0.007	0.003	0.007	0.002
	-	Female	0.010	0.004	0.008	0.004	0.010	0.004
	Female	Male	0.013	0.005	0.007	0.003	0.006	0.002
	-	Female	0.011	0.004	0.008	0.004	0.010	0.004
CH	Male	Male	0.023	0.008	0.015	0.006	0.014	0.004
	-	Female	0.049	0.016	0.024	0.009	0.028	0.008
	Female	Male	0.025	0.009	0.015	0.006	0.015	0.004
	-	Female	0.040	0.013	0.042	0.014	0.023	0.007
CZ	Male	Male	0.020	0.006	0.019	0.006	0.017	0.005
	-	Female	0.017	0.004	0.020	0.005	0.018	0.005
	Female	Male	0.015	0.004	0.017	0.006	0.028	0.009
	-	Female	0.012	0.003	0.016	0.004	0.018	0.004
DE	Male	Male	0.005	0.002	0.005	0.002	0.006	0.002
	-	Female	0.006	0.002	0.005	0.002	0.005	0.002
	Female	Male	0.008	0.003	0.011	0.004	0.006	0.002
	-	Female	0.008	0.003	0.008	0.003	0.007	0.002
EE	Male	Male	0.016	0.006	0.022	0.009	0.023	0.008
	-	Female	0.019	0.005	0.017	0.005	0.021	0.005
	Female	Male	0.028	0.009	0.010	0.005	0.027	0.009
	-	Female	0.016	0.004	0.019	0.005	0.017	0.005
ES	Male	Male	0.004	0.002	0.004	0.002	0.005	0.002
	-	Female	0.004	0.002	0.005	0.002	0.004	0.002
	Female	Male	0.006	0.003	0.003	0.001	0.006	0.003
	-	Female	0.006	0.003	0.005	0.002	0.004	0.002
FI	Male	Male	0.015	0.008	0.032	0.016	0.017	0.009
	-	Female	0.027	0.006	0.016	0.005	0.020	0.005
	Female	Male	0.035	0.016	0.076	0.037	0.024	0.015
	-	Female	0.022	0.005	0.024	0.006	0.023	0.006
FR	Male	Male	0.011	0.004	0.012	0.005	0.011	0.004
	-	Female	0.007	0.002	0.011	0.004	0.013	0.004
	Female	Male	0.013	0.004	0.007	0.003	0.013	0.005
	-	Female	0.014	0.004	0.011	0.003	0.011	0.004
GB	Male	Male	0.013	0.004	0.013	0.005	0.004	0.001
	-	Female	0.010	0.003	0.009	0.003	0.006	0.002
	Female	Male	0.010	0.004	0.007	0.003	0.007	0.002
	-	Female	0.015	0.005	0.014	0.004	0.008	0.002
HU	Male	Male	0.025	0.009	0.052	0.018	0.024	0.008
	-	Female	0.042	0.011	0.052	0.013	0.037	0.010
	Female	Male	0.023	0.007	0.031	0.012	0.034	0.011
	-	Female	0.048*	0.011	0.042	0.009	0.046	0.011
IE	Male	Male	0.012	0.003	0.010	0.003	0.011	0.002
	-	Female	0.006*	0.001	0.008	0.002	0.007*	0.001
	Female	Male	0.009	0.003	0.012	0.003	0.011	0.003
	-	Female	0.006	0.002	0.004**	0.001	0.007*	0.001

Continues on next page

Continued from previous page

	Gender of		Interviewer					
			younger		same age		older	
	Respondent	Interviewer	Pred.	Std. Err.	Pred.	Std. Err.	Pred.	Std. Err.
IL	Male	Male	0.046	0.007	0.039	0.006	0.038	0.006
	-	Female	0.017***	0.004	0.015***	0.003	0.013***	0.003
	Female	Male	0.050	0.007	0.058	0.009	0.040	0.006
	-	Female	0.019***	0.004	0.021***	0.005	0.014***	0.003
IS	Male	Male	0.035	0.014	0.033	0.019	0.044	0.025
	-	Female	0.013	0.004	0.016	0.006	0.014	0.006
	Female	Male	0.039	0.016	0.025	0.016	0.047	0.033
	-	Female	0.022	0.007	0.022	0.008	0.030	0.012
IT	Male	Male	0.014	0.004	0.020	0.007	0.018	0.006
	-	Female	0.017	0.004	0.016	0.004	0.016	0.004
	Female	Male	0.017	0.005	0.016	0.006	0.021	0.007
	-	Female	0.017	0.004	0.018	0.004	0.018	0.004
LT	Male	Male	0.015	0.005	0.010	0.005	0.008	0.005
	-	Female	0.019	0.005	0.015	0.004	0.014	0.004
	Female	Male	0.008	0.002	0.007	0.003	0.005	0.003
	-	Female	0.015	0.004	0.010	0.003	0.011	0.003
NL	Male	Male	0.004	0.002	0.009	0.004	0.006	0.002
	-	Female	0.006	0.002	0.006	0.003	0.005	0.002
	Female	Male	0.005	0.003	0.008	0.003	0.010	0.004
	-	Female	0.009	0.003	0.009	0.003	0.008	0.003
NO	Male	Male	0.014	0.006	0.019	0.008	0.025	0.008
	-	Female	0.036	0.013	0.052	0.020	0.033	0.013
	Female	Male	0.037	0.012	0.032	0.012	0.016	0.006
	-	Female	0.066	0.021	0.007	0.006	0.037	0.014
PL	Male	Male	0.029	0.007	0.028	0.009	0.031	0.008
	-	Female	0.032	0.009	0.029	0.010	0.024	0.006
	Female	Male	0.050	0.011	0.037	0.011	0.040	0.009
	-	Female	0.039	0.011	0.034	0.010	0.032	0.008
PT	Male	Male	0.012	0.007	0.026	0.020	0.024	0.031
	-	Female	0.010	0.005	0.012	0.008	0.003	0.004
	Female	Male	0.023	0.014	0.009	0.009	0.027	0.024
	-	Female	0.011	0.005	0.021	0.013	0.027	0.021
RU	Male	Male	0.067	0.021	0.042	0.021	0.021	0.011
	-	Female	0.027	0.005	0.036	0.006	0.033	0.006
	Female	Male	0.044	0.014	0.069	0.040	0.114	0.059
	-	Female	0.027	0.004	0.023	0.004	0.028	0.005
SE	Male	Male	0.033	0.012	0.027	0.012	0.042	0.016
	-	Female	0.025	0.008	0.036	0.011	0.037	0.011
	Female	Male	0.059	0.018	0.037	0.015	0.052	0.019
	-	Female	0.037	0.011	0.045	0.014	0.047	0.013
SI	Male	Male	0.011	0.004	0.021	0.014	0.024	0.013
	-	Female	0.025*	0.007	0.013	0.005	0.038	0.012
	Female	Male	0.015	0.005	0.010	0.008	0.019	0.011
	-	Female	0.022	0.005	0.023	0.008	0.019	0.007

Source: ESS8. Predicted values calculated for the Full Model. Asterisks indicate significance levels for each pairwise comparison (Male VS Female interviewer within one gender of respondent).

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table A3
Probability of a third person being present and interfering with the interview by country

	Gender of		Interviewer					
			younger		same age		older	
	Respondent	Interviewer	Pred.	Std. Err.	Pred.	Std. Err.	Pred.	Std. Err.
Austria								
AT	Male	Male	0.096	0.030	0.061	0.023	0.024	0.015
	-	Female	0.078	0.027	0.046	0.023	0.065	0.028
	Female	Male	0.060	0.023	0.072	0.028	0.032	0.015
	-	Female	0.053	0.020	0.049	0.022	0.022	0.013
BE	Male	Male	0.075	0.025	0.044	0.022	0.054	0.017
	-	Female	0.078	0.031	0.089	0.038	0.053	0.018
	Female	Male	0.108	0.034	0.071	0.027	0.055	0.016
	-	Female	0.070	0.031	0.059	0.033	0.035	0.015
CH	Male	Male	0.068	0.033	0.038	0.027	0.019	0.009
	-	Female	0.066	0.031	0.031	0.022	0.049	0.020
	Female	Male	0.095	0.042	0.042	0.030	0.070	0.020
	-	Female	0.039	0.024	0.052	0.031	0.046	0.019
CZ	Male	Male	0.008	0.008	0.039	0.027	0.088	0.039
	-	Female	0.027	0.008	0.028	0.013	0.042	0.018
	Female	Male	0.011	0.008	0.051	0.030	0.055	0.032
	-	Female	0.022	0.007	0.011	0.008	0.044	0.019
DE	Male	Male	0.052	0.018	0.026	0.012	0.044	0.010
	-	Female	0.056	0.021	0.042	0.020	0.040	0.013
	Female	Male	0.096	0.029	0.072	0.023	0.040	0.010
	-	Female	0.067	0.023	0.027	0.016	0.037	0.013
EE	Male	Male	-	-	0.031	0.034	-	-
	-	Female	0.080	0.025	0.071	0.024	0.065	0.018
	Female	Male	-	-	-	-	0.034	0.027
	-	Female	0.052	0.019	0.057	0.019	0.036	0.013
ES	Male	Male	0.118	0.033	0.054	0.031	0.049	0.023
	-	Female	0.135	0.027	0.097	0.025	0.076	0.019
	Female	Male	0.189	0.043	0.071	0.031	0.103	0.034
	-	Female	0.156	0.031	0.077	0.026	0.113	0.025
FI	Male	Male	0.041	0.029	-	-	0.159	0.083
	-	Female	0.045	0.011	0.016	0.011	0.055	0.024
	Female	Male	0.038	0.027	0.049	0.047	0.165	0.101
	-	Female	0.033	0.010	0.021	0.012	0.027	0.016
FR	Male	Male	0.083	0.029	0.050	0.025	0.021	0.013
	-	Female	0.064	0.019	0.061	0.021	0.062*	0.019
	Female	Male	0.065	0.024	0.116	0.038	0.069	0.025
	-	Female	0.063	0.019	0.022*	0.012	0.039	0.014
GB	Male	Male	0.110	0.040	-	-	0.019	0.009
	-	Female	0.066	0.027	0.025	0.018	0.035	0.012
	Female	Male	0.079	0.036	0.018	0.018	0.036	0.012
	-	Female	0.077	0.029	0.013	0.013	0.055	0.015
HU	Male	Male	0.088	0.041	0.076	0.044	0.041	0.019
	-	Female	0.141	0.038	0.092	0.032	0.039	0.015
	Female	Male	0.062	0.032	0.026	0.025	0.040	0.020
	-	Female	0.084	0.029	0.036	0.016	0.035	0.013
IE	Male	Male	0.061	0.019	0.037	0.016	0.090	0.025
	-	Female	0.029	0.012	0.023	0.011	0.052	0.015
	Female	Male	0.084	0.027	0.081	0.027	0.061	0.022
	-	Female	0.029	0.012	0.023	0.011	0.035	0.012

Continues on next page

Continued from previous page

	Gender of		Interviewer					
			younger		same age		older	
	Respondent	Interviewer	Pred.	Std. Err.	Pred.	Std. Err.	Pred.	Std. Err.
IL	Male	Male	0.125	0.027	0.096	0.025	0.106	0.023
	-	Female	0.195	0.044	0.124	0.041	0.174	0.037
	Female	Male	0.096	0.025	0.151	0.032	0.155	0.030
	-	Female	0.173	0.041	0.206	0.049	0.154	0.031
IS	Male	Male	0.091	0.048	0.168	0.104	0.020	0.025
	-	Female	0.122	0.041	0.072	0.042	0.023	0.020
	Female	Male	0.102	0.056	0.121	0.084	-	-
	-	Female	0.058	0.026	0.039	0.036	0.017	0.018
IT	Male	Male	0.068	0.026	0.041	0.024	0.032	0.018
	-	Female	0.094	0.022	0.055	0.017	0.059	0.015
	Female	Male	0.068	0.028	0.086	0.040	0.068	0.029
	-	Female	0.066	0.017	0.030	0.013	0.027	0.009
LT	Male	Male	0.197	0.077	0.151	0.098	0.347	0.138
	-	Female	0.108	0.021	0.080	0.020	0.109	0.023
	Female	Male	0.074	0.044	0.076	0.068	0.170	0.105
	-	Female	0.087	0.017	0.072	0.016	0.052	0.015
NL	Male	Male	0.093	0.046	0.024	0.018	0.049	0.018
	-	Female	0.061	0.029	-	-	0.035	0.014
	Female	Male	0.105	0.053	0.062	0.032	0.019	0.010
	-	Female	0.041	0.026	0.027	0.019	0.016	0.009
NO	Male	Male	0.062	0.023	0.041	0.021	0.022	0.012
	-	Female	0.029	0.019	0.023	0.023	0.048	0.027
	Female	Male	0.022	0.014	0.040	0.023	0.046	0.021
	-	Female	0.030	0.022	0.022	0.022	0.039	0.025
PL	Male	Male	0.062	0.019	0.093	0.032	0.067	0.021
	-	Female	0.111	0.032	0.063	0.029	0.084	0.024
	Female	Male	0.072	0.023	0.052	0.023	0.084	0.025
	-	Female	0.065	0.025	0.102	0.035	0.056	0.022
PT	Male	Male	0.112	0.056	-	-	-	-
	-	Female	0.147	0.037	0.116	0.098	0.384	0.200
	Female	Male	0.221	0.087	-	-	-	-
	-	Female	0.109	0.031	-	-	-	-
RU	Male	Male	0.065	0.043	-	-	0.059	0.048
	-	Female	0.128	0.019	0.099	0.020	0.104	0.021
	Female	Male	0.111	0.053	0.151	0.136	0.127	0.116
	-	Female	0.066	0.012	0.073	0.016	0.072	0.018
SE	Male	Male	0.036	0.021	0.017	0.017	0.016	0.011
	-	Female	0.075	0.035	0.051	0.026	0.019	0.011
	Female	Male	0.034	0.022	0.018	0.018	0.017	0.011
	-	Female	0.028	0.019	-	-	0.008	0.006
SI	Male	Male	0.104	0.043	-	-	0.119	0.073
	-	Female	0.100	0.025	0.040	0.023	0.081	0.039
	Female	Male	0.021	0.016	-	-	-	-
	-	Female	0.045	0.014	0.107	0.039	0.061	0.034

Source: ESS8. Predicted values calculated for the Full Model. Asterisks indicate significance levels for each pairwise comparison (Male VS Female interviewer within one gender of respondent). Missing entries are due to perfect prediction of the outcome.

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$