

Improving Data Quality with Sociodemographic Matching?: About the Effects and Implications of Age and Gender Matching in Face-to-Face Interviews

Journal of Applied Social Science
2022, Vol. 16(1) 242–260
© The Author(s) 2021



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1936724421998257
journals.sagepub.com/home/jax



Felix Bittmann¹ 

Abstract

According to the theory of liking, data quality might be improved in face-to-face survey settings when there is a high degree of similarity between respondents and interviewers, for example, with regard to gender or age. Using two rounds of European Social Survey data from 25 countries including more than 70,000 respondents, this concept is tested for the dependent variables amount of item nonresponse, reluctance to answer, and the probability that a third adult person is interfering with the interview. The match between respondents and interviewers is operationalized using the variables age and gender and their statistical interactions to analyze how this relates to the outcomes. While previous studies can be corroborated, overall effect sizes are small. In general, item nonresponse is lower when a male interviewer is conducting the interview. For reluctance, there are no matching effects at all. Regarding the presence of other adults, only female respondents profit from a gender match, while age is without any effect. The results indicate that future surveys should weigh the costs and benefits of sociodemographic matching as advantages are probably small.

Keywords

survey quality, gender matching, theory of liking, European Social Survey, statistical matching

Introduction

It is well accepted that survey quality and nonresponse are affected by certain characteristics of the interviewer, like gender, age, or other sociodemographic aspects in face-to-face interviews (Durrant et al. 2010). Beyond that, there is a growing body of evidence showing how respondent-interviewer matching with respect to some key variables can improve data quality (Vercruyssen, Wuyts, and Loosveldt 2017). Especially given the well-researched theory of liking, which predicts similarity of respondents and interviewers regarding some obvious features in face-to-face interview situations like gender, age, education level, style of clothing, or general habitus, these findings are not surprising (Groves, Cialdini, and Couper 1991). The underlying idea is that people like and prefer other people that are similar to themselves. This effect, which probably

¹Leibniz Institute for Educational Trajectories, Bamberg, Germany

Corresponding Author:

Felix Bittmann, Leibniz Institute for Educational Trajectories (LifBi), Wilhelmsplatz 3, 96047 Bamberg, Germany.
Email: felix.bittmann@lifbi.de

operates subconsciously, might benefit survey settings as respondents are more open and might answer questions more freely and less reserved (McPherson, Smith-Lovin, and Cook 2001). The expectation is that this benefits survey quality since it is easier to retrieve true and unfiltered opinions of the respondent and less probing is necessary. The current study follows up these results and examines in depth how matching can contribute to improve data quality and reduce missing data problems in survey situations using European Social Survey (ESS) data from 25 countries. Three main outcomes will be used to assess data quality: the amount of item nonresponse in an interview, the probability that a respondent is reluctant to answer questions, and the probability that a third adult person is interfering with the interview. Summarized: How does sociodemographic matching with respect to gender, age, and their statistical interactions influence data quality in face-to-face interview situations?

The present study attempts to close some research gaps, which concern the influence of gender matching on item nonresponse and item nonresponse bias (West and Blom 2017) and especially the external validity of matching with respect to a large number of countries and cultural settings. First, given the high-quality data available which include more than 70,000 respondents, it should be possible to estimate precise effects and decrease the amount of uncertainty. Second, since multinational data are available from 25 European countries, the findings should have a high external validity and generalize to a wide range of different social and cultural settings. Third, the used multilevel design with three random intercepts (respondents nested within interviewers nested within countries), it is also possible to disentangle interviewer effects from respondent effects. Clearly, interviewers differ with respect to their overall experience and talent, which can influence results and should be investigated with care. In accordance with similar studies (Vercauteren et al. 2017), matching between respondents and interviewers is operationalized using the variables age and gender and tested through statistical interactions in a series of nested models to generate a high level of insight and also account for the influence of control variables.

These research tasks seem highly relevant as matching, if positive effects could be demonstrated, would be a relatively easy to implement way of improving data quality and decreasing the amount of item nonresponse. Finally, it should be admitted that the focus on item nonresponse in comparison to unit nonresponse seems arbitrary, which is true to a certain extent. However, as there are given word limits, it seemed best to focus on item nonresponse as the ESS provides richer data to explore this context. Subsequent studies will also take the unit nonresponse into account.

Review of the Literature

It is generally accepted that not only the characteristics of the respondent such as age or level of education affect aspects of data quality like item nonresponse (Groves 2009; Tourangeau, Rips, and Rasinski 2000) but also the characteristics of the interviewer have an important influence (Brunton-Smith, Sturgis, and Leckie 2017; Davis et al. 2010). The current research deals with the statistical interaction of these two aspects, that is, how the (dis)similarity between respondents and interviewers, also called sociodemographic match, affects data quality. In this context, data quality is given when responses are “true,” unbiased, and complete.

Theory of Liking

The main theoretical framework that explains why sociodemographic matching affects data quality is the theory of liking (Groves et al. 1991; Groves, Cialdini, and 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 attitude

(Byrne et al. 1971), background (Stotland and Patchen 1961), and dress style (Suedfeld, Bochner, and Matas 1971). This is also known as homophily, which states that we like people that are similar to ourselves. To demonstrate why this is relevant for the research questions, two things have to be shown: first, that gender and age refer to the aspects mentioned and second, that liking actually increases data quality. To start with age and gender, it seems logical that these two factors influence certain aspects, for example, attitudes (McPherson et al. 2001). People with approximately the same age share birth cohorts, which means they also experienced certain events and periods together, for example, historic events but also the same trends, fashions, slang, and more. Therefore, when the age difference between respondent and interviewer is small, chances are higher that they share certain social aspects that make them more similar on average, compared to when age difference is large. One study shows that 38 percent of all friends are within two years of age difference (Fischer 1977). The argumentation for gender match is analogous, although interests are here more important than time. Research confirms what is well known: Men and women differ in the topics they show interest in, for example, considering vocational choices or leisure activities (Su, Rounds, and Armstrong 2009). Therefore, it can be assumed that gender matching should, on average, increase the similarity between respondent and interviewer, which should in turn increase liking. In other words, on average, men are more similar to men than to women and vice versa. Summarized, it can be assumed 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 exists, but, on average, the probability of liking is predicted to be higher.

Item nonresponse or not answering questions at all is of central interest in 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 is high and the respondent does not expect understanding or even fears moral judgment. 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. 2001:429). Therefore, when similarity with the interviewer is higher, the fear of moral judgment or lack of understanding might be lower, which increases the probability that the respondent will answer. Based on this, we can expect that the larger the degree of similarity between respondent and interviewer, the lower the amount of item nonresponse in a face-to-face interview (Hypothesis 1). The second outcome of interest is about how smoothly respondents interact with the interviewer or whether they are reluctant to answer a question. Following the logic of the theory described above, one would conclude that respondents are 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 more naturally, it can be expected that less probing is 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 that the larger the degree of similarity between respondent and interviewer, the lower the probability that a respondent is reluctant to answer a question in a face-to-face interview (Hypothesis 2). The last scenario is slightly different from the first two and does not put the respondent in the focus but rather the situation of the interview itself. It is well known that respondents might be self-conscious when a third person is present (Reuband 1992) and accordingly interview situations are preferred when the respondent is alone with the interviewer. Here the question arises whether the degree of similarity between respondent and interviewer affects the probability that any other person is present. Of course, there are situations when having a third person at the interview cannot be avoided, for example, when small children are present and the parent is the only person at home. In other cases, it might be different, for example, when the respondent can

choose whether to be alone and the interviewer is perceived as a stranger or even a threat. When this is the case, it seems plausible that a respondent opts for not facing the interviewer alone, but having another person participate, for example, the spouse or a roommate. As was shown above, the more the similarity between respondent and interviewer, the larger the probability for liking the interviewer. Therefore, one would expect that there is no need 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, it can be argued that the chance that this person actively interferes is smaller when the interview is proceeding smoothly. For example, a third person might agree to attend the interview in silence, but interferes only when the interviewer is probing the respondent repeatedly. As this might be more likely when the similarity between interviewer and respondent is small, this would explain how similarity and third person interference is related. Based on this reasoning, it is expected that the larger the degree of similarity between respondent and interviewer, the lower the probability that a third person is present and interfering with the interviewing process (Hypothesis 3).

Interviewer Effects

There exists a large body of literature about the effects of respondent-interviewer matching, dating back at least to the 1920s (Rice 1929). Matching on ethnic characteristics or, especially in the U.S. context, origin, was one of the first aspects to be explored (Williams 1964). While some of these studies did not find any positive effect of origin/ethnic matching (Axinn 1989; Dotinga et al. 2005), others reported positive results (Johnson et al. 2000; Oyinlade and Losen 2014; Webster 1996). A second major component is matching with regard to educational levels. One could assume that it should improve interviewing results when respondent and interviewer have approximately the same level of education, as this might make communication easier and enable a more emphatic interview experience. Yet the results are again mixed. While Durrant et al. (2010) report that matching on the gender and educational level increases cooperation rates, others find null or negative effects (Riphahn and Serfling 2005; Vercruyssen et al. 2017; Weiss 1968).

Gender and Sensitive Items

In the current context, matching on gender and age seems most important. Apparently, women are more open to female interviewers (Fletcher and Spencer 1984) and both men and women are more responsive for sensitive psychological items when the interviewer is female (Chun et al. 2011). Yet, this effect is questionable when the items are about sexual behavior or substance abuse (Catania et al. 1996; Johnson and Parsons 1994). Therefore, it is not possible to conclude that women are the better interviewers when sensitive items are concerned, as the specific type of item appears to be relevant. When age is considered, former results are again divided. While some studies report that age matching seems beneficial for older respondents (Herzog and Rodgers 1992), there is no stable effect for sensitive items (Wilson et al. 2002). An ambitious research synthesis, including hundreds of studies, comes to the conclusion that sociodemographic matching seems useful as 4 of the 10 studies considered reporting positive findings (West and Blom 2017). 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 come to the conclusion that age matching reduces item nonresponse, which cannot be attributed to age effects alone, as they control for the age of the respondents. Finally, regarding the effect of sociodemographic matching on the probability that third persons are present, two studies report that females interviewed by a man are more often accompanied by a male in comparison with a female interviewer

(Hartmann 1994; Preetz 2017), which is interpreted as a form of social control. As discussed above, this interpretation is not guaranteed to hold as women may choose freely to have someone with them when they do not trust the interviewer.

A review of the literature reveals a heterogeneous picture; 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 sometimes minor details, for example, the topic of a sensitive item, are also relevant and influence the result. Possibly, a quite complex theoretical framework would be necessary to account for every detail and furthermore all interactions between the relevant aspects, like gender, age, or education. Although the approach of this study will be based on the relatively simple theory of liking as justified above, we hope that including some interactions between age and gender can reveal new insights that might be beneficial for the general development of survey research. Finally, it should be admitted that some of the studies cited are quite dated but are still the best indicators available. New studies would be desirable to check whether the findings still hold nowadays.

Methods

Data

To test the effects of respondent-interviewer matching, data from ESS rounds 8 and 9 will be used. The ESS is a biannual European survey that includes a multitude of questions about all aspects of life with a focus on social aspects and politics, which are highly standardized between countries, and only some questions differ, for example, regarding educational degrees. The topics of the surveys focus on education, income, employment, general standard of living, and political views. Since there are almost no items that are specifically gender related, this might influence the results, which is further addressed in the discussion of the results. There are in total 25 countries included in both rounds of the survey. The ESS releases information about the main questionnaire as well as additional data, which contain 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; however, as presented above, educational matching does not seem to have much influence on outcomes in the majority of studies. All interviews were conducted using computer-assisted personal interview (CAPI) except for the following countries which used paper and pencil interview (PAPI): Spain, Russia, Poland, Lithuania, and Israel (“ESS9 Documentation Report 2018” 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 it is known that nonresponse bias can still affect the results (West and Blom 2017:177–82). As only item nonresponse (and not unit nonresponse) is regarded in this analysis, we therefore decide to control for a set of variables, which is described below. Yet as the main variables of interest, gender and age, are still highly variable even within a region or neighborhood, the bias might be rather small. Finally, the multilevel approach, which includes random intercepts for interviewers, takes into account that respondents nested within interviewers might be more similar to each other than in a complete random design, which is described using the intraclass correlation.

Respondents who were younger than 18 years old at the time of the survey were excluded from the analysis. The ESS pays great attention to maximizing response rates. The average response rate is between 30.6 (Germany) and 74.4 (Israel) percent in Round 8 (Stoop et al. 2016).

Table 1. Descriptive Matching Statistics by Gender of Respondent.

Gender of interviewer	Interviewer younger	Same age	Interviewer older
Male respondents (N = 35,753; 100%)			
Male	4,437 (12.41%)	2,619 (7.33%)	6,490 (18.15%)
Female	8,174 (22.86%)	4,308 (12.05%)	9,725 (27.20%)
Female respondents (N = 39,813; 100%)			
Male	4,730 (11.88%)	2,506 (6.29%)	6,427 (16.14%)
Female	10,259 (25.77%)	5,100 (12.81%)	10,791 (27.10%)

Source. ESS8/9, own calculations.

Note. ESS = European Social Survey.

Variables

Degree of sociodemographic match. There are two central constructs of interest: gender and age match between respondent and interviewer. To operationalize this, three variables will be used: the gender of the respondent (*Female*), indicating whether a respondent is female (1) or male (0). When the gender of respondent and interviewer is identical, the variable *Gender-Match* takes the value 1, 0 otherwise. Finally, the age difference between respondent and interviewer is recoded into a new variable (*AgeMatch*) with three categories: interviewer younger than respondent (0), age difference within five years (1), and interviewer older than respondent (2). Clearly, the cutoff point of five years for a “match” is arbitrary. This form of operationalization is adopted as it is already applied in another similar study (Vercruyssen et al. 2017:4), which maximizes comparability. Also, a continuous variable would be harder to interpret with less concrete statements about their effects.¹ By further including all relevant interactions between these three variables, it is possible to model a quite complex degree of matching. Table 1 (page 6) shows descriptive statistics for the overall match between respondents and interviewers by gender of respondent.

Amount of item nonresponse. To count the number of items refused by the respondent, all items with the following codings were selected: “refusal,” “don’t know,” and “no answer.” To count the exposure, all items were selected that had either one of these missing codes or a valid response. Doing so allows discounting all questions that were filtered out and are thus not eligible for calculating this measurement. In the following models, the exposure is included in the models to allow a correct estimation of effects. Furthermore, all items that were answered by the interviewers were removed since these cannot give information about the response rate of the respondent. Respondents with more than 50 missing or refused items (which corresponds to the 99th percentile) are excluded from these models since no regular survey situation can be assumed.

Reluctance to answer. To measure whether respondents had problems answering questions or were unwilling and reluctant to answer, an item from the interviewer questionnaire can be used. Interviewers had to rate on a five-point scale the degree of reluctance a respondent displayed during the interview, measuring how often respondents were reluctant to answer. The variable is scaled as following: “very often” (5), “often (4), “now and then” (3), “almost never” (2), and “never” (1). It should be admitted that this variable has some amount of inherent measurement error, which is quite unavoidable in the social sciences, but the multilevel design described below should be able to account for some amount of the variation caused by the inherent (psychological) differences between interviewers.

Table 2. Descriptive Statistics by Country.

Country	N	Item nonresponse (M)	Item nonresponse (SD)	Reluctance to answer (M)	Reluctance to answer (SD)	Third person interfering
AT	4,429	4.537	6.593	1.602	0.832	0.043
BE	3,402	1.981	3.998	1.362	0.719	0.071
CH	2,928	4.552	6.789	1.517	0.868	0.055
CY	762	6.648	6.563	1.909	1.046	0.091
CZ	4,563	6.233	8.083	2.028	1.046	0.037
DE	4,996	3.372	6.045	1.293	0.681	0.056
EE	3,811	2.619	4.845	1.547	0.829	0.046
ES	1,860	6.777	8.493	1.436	0.724	0.115
FI	3,545	1.988	4.973	1.406	0.868	0.036
FR	3,982	3.981	6.193	1.392	0.827	0.058
GB	4,051	5.167	6.473	1.515	0.811	0.049
HU	3,092	9.969	9.771	2.297	1.115	0.053
IE	4,841	4.978	6.432	1.609	0.937	0.061
IL	2,457	9.225	10.047	1.983	1.031	0.133
IS	857	4.098	6.564	2.006	0.919	0.05
IT	5,175	8.052	9.112	2.125	1.034	0.057
LT	2,023	9.101	9.892	2.207	1.029	0.092
NL	3,234	4.232	5.679	1.371	0.7	0.042
NO	2,800	2.208	4.731	1.218	0.675	0.049
PL	3,074	7.541	9.124	1.738	1.071	0.083
PT	1,249	4.877	7.1	1.524	0.874	0.071
RS	2,005	8.051	8.751	2.125	1.083	0.093
RU	2,396	10.543	9.82	2.091	1.094	0.09

Source. ESS8/9, own calculations.

Note. *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 share between 0 and 1. ESS = European Social Survey.

Third adult person present and interfering with the interview. Interviewers indicated in the interviewer questionnaire whether any third adult persons were interfering with the interview, for example, spouses, siblings, in-laws, or other people. As the ESS does not give a clear definition of interference, the colloquial meaning is assumed (“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 this occurred or not.

Three different outcomes will be used. The first, amount of item nonresponse, is objective, while the other two are more subjective and rely on the assessment of the interviewer. By utilizing all three variables, we attempt to achieve a more complete picture with regard to the interviewing situation and data quality. Table 2 (p. 8) lists descriptive statistics for all outcomes by countries. Table 3 lists all pairwise correlations for the dependent variables. The results underline that there is correlation present, which indicates that all three variables can be regarded as appropriate measurements to assess the construct data quality.

To investigate which items are the most sensitive and have the highest amount of invalid responses (counting “Refusal,” “Don’t know,” and “No answer” as item nonresponse), the percentage of missing values is computed. The top three in the ESS are “Total household net income” with 18-percent missing values, “Placement on left right scale” with 12-percent missing values, and “Father: highest level of education” with 9-percent missing values.” This demonstrates that

Table 3. Correlation Matrix for All Three Dependent Variables.

Correlation Matrix	Missing items	Reluctance	Third person present
Missing items	1		
Reluctance	.3475*	1	
Third person present	.0582*	.0836*	1

Source. ESS8/9, own calculations.

Note. Reported are Spearman's Rho correlation coefficients. ESS = European Social Survey.

* $p < .05$.

in the ESS, some question about finances, political orientation, or social background can be regarded as especially sensitive.

In summary, listed outcomes are chosen because they might be the best indicators in the data to measure survey quality. The number of missing items corresponds to survey quality as missing items lower the power of any analysis. Reluctance seems appropriate to measure the atmosphere of the interview situation, which might not only affect interview quality but also the propensity to participate in follow-up studies. Including a variable that measures inference by third adults seems relevant, given the fact that presence of third persons can influence data quality negatively (Mneimneh et al. 2015).

Modeling Approach

As multilevel data are available, three-level random intercept models will be used. Respondents (Level 1) are nested within interviewers (Level 2), which are nested within countries (Level 3). This approach using random intercepts for interviewers and countries allows accounting for idiosyncratic differences between Level 2 and 3 variables.² As the propensity to answer a question may be culturally determined, country differences are taken into account by this design. Furthermore, one would expect differences between interviewers, for example, with respect to work experience, charisma, or openness (West and Blom 2017). By allowing the intercepts to vary, standard errors of variables of interest are estimated correctly. Determined by the scaling of the dependent variables, three different models will be estimated. The first dependent variable, amount of missing items (item nonresponse), is a count variable; therefore, a nonlinear model will be used. As the descriptive statistics suggested conditional overdispersion, we decided to use a multilevel negative binomial regression instead of a multilevel Poisson regression. The statistics shown below underline that this is actually the case and a negative binomial model should be preferred. For the second dependent variable, reluctance to answer questions, a multilevel linear model is estimated. Albeit this variable is being ordinally scaled, this leaves many obstacles for estimation and especially interpretation. Consequently, we decided to use the simpler linear models and compute robustness checks using the ordinal model. However, the differences are minor and the linear model gives good approximation of results. Finally, as the last dependent variable, the probability that a third person is interfering with the interview, is binary, a multilevel logistic regression will be estimated.

The models are built as following: The baseline model ("Empty") does not include any explanatory variables, only the dependent variable and the multilevel structure. The second model ("Main") includes gender of respondent, the age match variable, the gender match variable, and all pairwise interactions. The final model ("Full") adds all control variables. Based on theoretical considerations, we decided to add the following controls to account for the fact that respondent-interviewer matching might not be absolutely at random: the educational level of the respondent (International Standard Classification of Education [ISCED]), average regional

household income, age of respondent, age squared, number of household members, whether any children live in the household, the marital status of the respondent, and a fixed-effect for the survey round (ESS 8 or 9). Another advantage of adding these variables is that predicted outcomes are more realistic and also account for country differences with respect to age structure or average educational attainment (Colsher and Wallace 1989; De Leeuw, Hox, and Huisman 2003). All calculations are done in Stata 16.1 (commands *menbreg*, *mixed*, and *melogit*) using 10 integration points.³ The covariance structure for all three models is set to “unstructured,” meaning that all variances and covariances for all levels will be estimated distinctly, which seems most appropriate for the data available as no artificial constraints are introduced. Intraclass correlation coefficients (ICC) were computed for all models where this is possible after the estimation to explain the partition of variance. The regression equation for the second model (“Main”) for the binary logistic model is as follows:

$$\text{Logit}(Y_{ijk}) = \beta_0 + \beta_1 \text{Gendermatch}_{ijk} + \beta_2 \text{Agematch}_{jk} + \beta_3 \text{Female}_{ijk} + \beta_4 \text{Gendermatch}_{ijk} \\ \times \text{Agematch}_{jk} + \beta_5 \text{Gendermatch}_{ijk} \times \text{Female}_{ijk} + \beta_6 \text{Agematch} \times \text{Female}_{ijk}.$$

Results

Item Nonresponse

The first analyses consider item nonresponse. The stepwise model building clearly shows that adding the main variables of interest, which are the matching variables and the 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 most cases. This underlines that the matching between interviewers and respondent might not be perfectly random, yet the remaining explanatory power is still significant. Additionally, it should be noted that 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. To ease interpretation in the presence of many interactions, average marginal effects are calculated for the full model, which has the highest predictive power with respect to the Akaike information criterion (AIC), which is lowest. Regression results are displayed in Table 4, average marginal effects of gender and age matching, based on the full model including all control variables, in Figure 1. As an example of how to interpret these figures, let us consider the left panel first which presents the findings for all male respondents. The effect is close to -2 in situations when the interviewer is younger than the respondent. Since the value is negative, we learn that male respondents who were interviewed by a male interviewer have on average two fewer item nonresponse (that means, they answer two items more) than male respondents interviewed by female interviewers. Based on this prediction, we can conclude that gender matching is beneficial for male respondents, regardless of the age matching. However, for female respondents, exactly the opposite is true. For them, gender matching does *increase* the amount of item nonresponse significantly, independent of age matching.

Reluctance to Answer

For the second outcome of interest, the reluctance to answer a question, linear multilevel models are specified. The models are built exactly as for the first outcome. The results are depicted in Table 5 and Figure 2. Clearly, neither for women nor for men is there any statistically significant effect. While some coefficients are statistically significant in the main model, these effects disappear after control variables are introduced. Consequently, we conclude that there is no effect of matching on the reluctance to answer questions as perceived by the interviewers. Regarding the

Table 4. Negative Binomial Regression Results: Amount of Item Nonresponse.

Variables	Empty	Main	Full
Gendermatch		-0.112*** (0.026)	-0.109*** (0.025)
Female		0.174*** (0.026)	0.106*** (0.025)
Gendermatch × Female		0.188*** (0.041)	0.177*** (0.040)
Agematch			
Int. younger		Ref.	Ref.
Same age		-0.177*** (0.021)	-0.004 (0.022)
Int. older		-0.163*** (0.017)	0.051* (0.023)
Gendermatch × Int. younger		0.000 (.)	0.000 (.)
Gendermatch × Same age		0.015 (0.026)	0.013 (0.026)
Gendermatch × Int. older		0.025 (0.021)	0.024 (0.020)
Female × Int. younger		Ref.	Ref.
Female × Same age		-0.095*** (0.026)	-0.018 (0.026)
Female × Int. older		-0.167*** (0.021)	-0.064** (0.020)
Respondent education level (ISCED)			
Less than lower sec. (I)			Ref.
Lower secondary (II)			-0.252*** (0.020)
Lower tier upper sec. (IIIb)			-0.426*** (0.020)
Upper tier upper sec. (IIIa)			-0.515*** (0.020)
Advanced vocational (IV)			-0.578*** (0.021)
Lower tertiary (V1)			-0.727*** (0.022)
Higher tertiary (V2)			-0.712*** (0.021)
Age of respondent			-0.021*** (0.002)
Age of respondent squared			0.000*** (0.000)
Number of people living in household			0.003 (0.004)
Child in household			-0.030* (0.013)
Marital status			
Married			Ref.
Separated/Divorced			0.073*** (0.015)
Widowed			0.107*** (0.018)
Never married			0.054*** (0.013)
Average regional income			-0.074*** (0.015)
ESS Round 9			0.419*** (0.021)
Constant	-4.597*** (0.013)	-5.184*** (0.028)	-2.009*** (0.086)
Ln(α)	-0.034*** (0.008)	-0.057*** (0.008)	-0.125*** (0.008)
Country-level variance	0.111*** (0.004)	0.191*** (0.008)	0.208*** (0.008)
Interviewer-level variance	0.559*** (0.011)	0.571*** (0.012)	0.556*** (0.011)
Observations	73,555	73,555	73,555
AIC	370,714	369,838	366,682

Source. ESS8/9, own calculations.

Note. $N = 73,555$ respondents nested within $J = 5,487$ interviewers nested within $K = 25$ countries. Standard errors are in parentheses. The dependent variable is a count variable. ESS = European Social Survey; AIC = Akaike information criterion; ISCED = International Standard Classification of Education.

* $p < .05$. ** $p < .01$. *** $p < .001$.

partition of variance, it becomes clear that indeed a larger correlation is present at the interviewer level (.35). This is due to mostly two factors: First, interviewers differ in their abilities and experience, so some interviewers might be more talented to acquire responses, which causes this

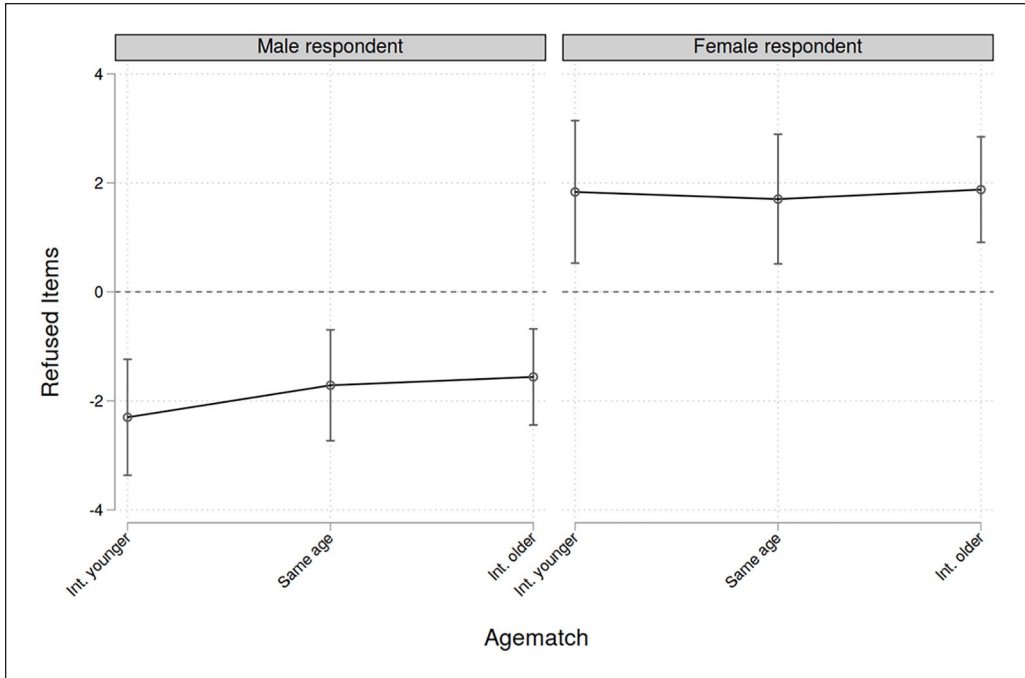


Figure 1. Average marginal effects of gender and age matching regarding amount of item nonresponse. Source. ESS8/9.

Note. Included are 95-percent confidence intervals. The coefficients display the predicted effect of a successful gender match (i.e., interviewer and respondent have the same gender), separately for gender of respondent and age-matching groups. If a confidence band intersects the zero line, there is no statistically significant effect present on the 5-percent level since zero might be included in the confidence band of the effect. ESS = European Social Survey.

correlation. The second aspect is that interviewers are often assigned due to their geographical availability and the households they visit are often similar to each other on average. This can also explain the correlation since similar households might behave similarly. The ICC regarding the countries is smaller (.12); here, we would expect differences due to cultural and economic differences between countries.

Probability of a Third Person Interfering

The last outcome refers to the research question of whether a sociodemographic matching between respondents and interviewers influences the probability that a third person is interfering with the interview, which could affect the answers given by the respondent. A multilevel logistic regression is estimated. Results can be found in Table 6 (p. 15) and Figure 3. The predictions reveal that gender matching is beneficial for both men and women but only women have statistically significant effects. If they are matched with a female interviewer, the probability that another adult is being present at the interview reduces by about one percentage point. The effects are slightly weaker when the interviewer is younger than the respondent. Summarized, gender matching is beneficial for this aspect of the survey setting. Regarding the ICC, the outcomes are similar to the reluctance models. The only major difference is that the correlation is much lower at the country level, indicating that there are smaller differences between countries when it comes to the interference of a third person at the interview.

Table 5. Linear Regression Results: Reluctance to Answer.

Variables	Empty		Main		Full	
Gendermatch			-0.018	(0.019)	-0.019	(0.019)
Female			0.046*	(0.019)	0.031	(0.019)
Gendermatch × Female			0.019	(0.031)	0.020	(0.031)
Agematch						
Int. younger			Ref.		Ref.	
Same age			-0.067**	(0.014)	-0.004	(0.015)
Int. older			-0.129**	(0.011)	0.005	(0.016)
Gendermatch × Int. younger			Ref.		Ref.	
Gendermatch × Same age			-0.000	(0.017)	0.002	(0.017)
Gendermatch × Int. older			0.005	(0.014)	0.004	(0.014)
Female × Int. younger			Ref.		Ref.	
Female × Same age			-0.037*	(0.017)	-0.022	(0.017)
Female × Int. older			-0.051**	(0.014)	-0.029*	(0.014)
Respondent education level (ISCED)						
Less than lower sec. (I)					Ref.	
Lower secondary (II)					-0.091**	(0.014)
Lower tier upper sec. (IIIb)					-0.169**	(0.014)
Upper tier upper sec. (IIIa)					-0.206**	(0.014)
Advanced vocational (IV)					-0.206**	(0.014)
Lower tertiary (V1)					-0.249**	(0.015)
Higher tertiary (V2)					-0.264**	(0.015)
Age of respondent					0.006**	(0.001)
Age of respondent squared					-0.000**	(0.000)
Number of people living in household					0.001	(0.003)
Child in household					-0.012	(0.009)
Marital status						
Married					Ref.	
Separated/Divorced					-0.007	(0.010)
Widowed					0.018	(0.012)
Never married					0.005	(0.009)
Average regional income					0.015	(0.011)
ESS Round 9					0.147**	(0.016)
Constant	1.695**	(0.065)	1.750**	(0.063)	1.506**	(0.095)
Country-level variance	-0.785**	(0.013)	-0.788**	(0.013)	-0.794**	(0.013)
Interviewer-level variance	-0.255**	(0.003)	-0.258**	(0.003)	-0.264**	(0.003)
ICC country	0.11	0.028	0.10	0.028	0.12	0.030
ICC interviewer in country	0.34	0.021	0.34	0.021	0.35	0.022
Observations	73,896		73,896		73,896	
AIC	180,687		180,232		179,510	

Source. ESS8/9, own calculations.

Note. $N = 73,896$ respondents nested within $J = 5,487$ interviewers nested within $K = 25$ countries. Standard errors are in parentheses. The dependent variable is a count variable. ESS = European Social Survey; ICC = intraclass correlation coefficient; AIC = Akaike information criterion; ISCED = International Standard Classification of Education.

* $p < .05$. ** $p < .001$.

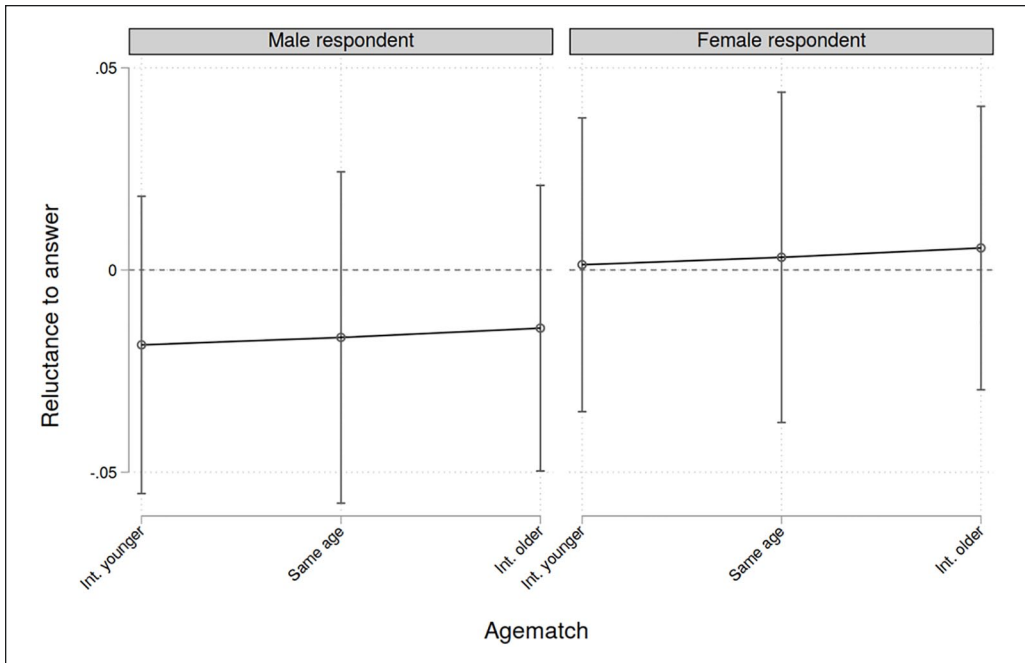


Figure 2. Average marginal effects of gender and age matching regarding the reluctance to answer.

Source. ESS8/9.

Note. Included are 95-percent confidence intervals. The coefficients display the predicted effect of a successful gender match (i.e., interviewer and respondent have the same gender), separately for gender of respondent and age-matching groups. If a confidence band intersects the zero line, there is no statistically significant effect present on the 5-percent level since zero might be included in the confidence band of the effect.

Discussion

The present study adds to the literature on the effects of sociodemographic matching between survey respondents and interviewers. The first outcome to be considered is the amount of item nonresponse in a face-to-face interview. Multilevel negative binomial regressions demonstrate that some pairings actually increase the amount of missing items and females have, on average, more missing items than men. This indicates that women in general tend to refuse more questions than men, independently of the characteristics of the interviewer. Second, a gender match increases the number of refused items for women, while the opposite is the case for men. This corroborates the findings of Vercruyssen et al. (2017:9), who use the same data set but restrict it to the Belgian subsample. By using information about all countries, it can be demonstrated that the effect is stable and generalizable to a much wider population. The empirical results do not support the theory of liking, as a gender match does not always improve results. Considering age matching, effects are very small. To sum up, Hypothesis 1 is rejected for female respondents, while it seems correct for male respondents.

The second dependent variable of interest is the reluctance to answer questions in an interview. The results indicate that there is no effect at all regarding age or gender matching. Consequently, Hypothesis 2 is rejected as there is no empirical support. The stability of the results is tested using an ordinal logistic regression instead of a linear one, which does not change the findings. The results seem to be rather stable.

The last outcome variable is the probability that a third person is interfering with the interview. As especially women have significantly higher probabilities when gender is mismatched,

Table 6. Logistic Regression Results: Adult Third Persons Interfering.

Variables	Empty		Main		Full	
Gendermatch			-0.059	(0.076)	-0.063	(0.079)
Female			-0.422***	(0.078)	-0.164*	(0.081)
Gendermatch × Female			-0.162	(0.115)	-0.136	(0.119)
Agematch						
Int. younger			0.000	(.)	0.000	(.)
Same age			-0.554***	(0.074)	-0.263**	(0.082)
Int. older			-0.612***	(0.060)	-0.279***	(0.084)
Gendermatch × Int. younger			0.000	(.)	0.000	(.)
Gendermatch × Same age			-0.115	(0.098)	-0.126	(0.101)
Gendermatch × Int. older			-0.085	(0.076)	-0.111	(0.078)
Female × Int. younger			0.000	(.)	0.000	(.)
Female × Same age			0.348***	(0.098)	0.164	(0.101)
Female × Int. older			0.362***	(0.076)	0.137	(0.079)
Respondent education level (ISCED)						
Less than lower sec. (I)					0.000	(.)
Lower secondary (II)					-0.154*	(0.068)
Lower tier upper sec. (IIIb)					-0.458***	(0.074)
Upper tier upper sec. (IIIa)					-0.705***	(0.072)
Advanced vocational (IV)					-0.892***	(0.078)
Lower tertiary (V1)					-1.083***	(0.087)
Higher tertiary (V2)					-1.109***	(0.082)
Age of respondent					-0.079***	(0.006)
Age of respondent squared					0.001***	(0.000)
Number of people living in household					0.147***	(0.017)
Child in household					-0.002	(0.053)
Marital status						
Married					0.000	(.)
Separated/Divorced					-1.147***	(0.080)
Widowed					-2.031***	(0.097)
Never married					-0.644***	(0.058)
Average regional income					-0.109*	(0.046)
ESS Round 9					0.196**	(0.061)
Constant	-3.251***	(0.092)	-2.706***	(0.097)	-0.377	(0.329)
Country-level variance	0.180**	(0.058)	0.167**	(0.054)	0.153**	(0.051)
Interviewer-level variance	1.392***	(0.074)	1.396***	(0.074)	1.497***	(0.079)
ICC country	0.04	0.011	0.03	0.010	0.03	0.009
ICC interviewer in country	0.32	0.013	0.32	0.013	0.33	0.013
Observations	74,039		74,039		74,039	
AIC	31,461		31,146		29,528	

Source. ESS8, own calculations.

Note. $N = 74,039$ respondents nested within $J = 5,487$ interviewers nested within $K = 25$ countries. Standard errors are in parentheses. Reported are logits. ESS = European Social Survey; ICC = intraclass correlation coefficient; AIC = Akaike information criterion; ISCED = International Standard Classification of Education.

* $p < .05$. ** $p < .01$. *** $p < .001$.

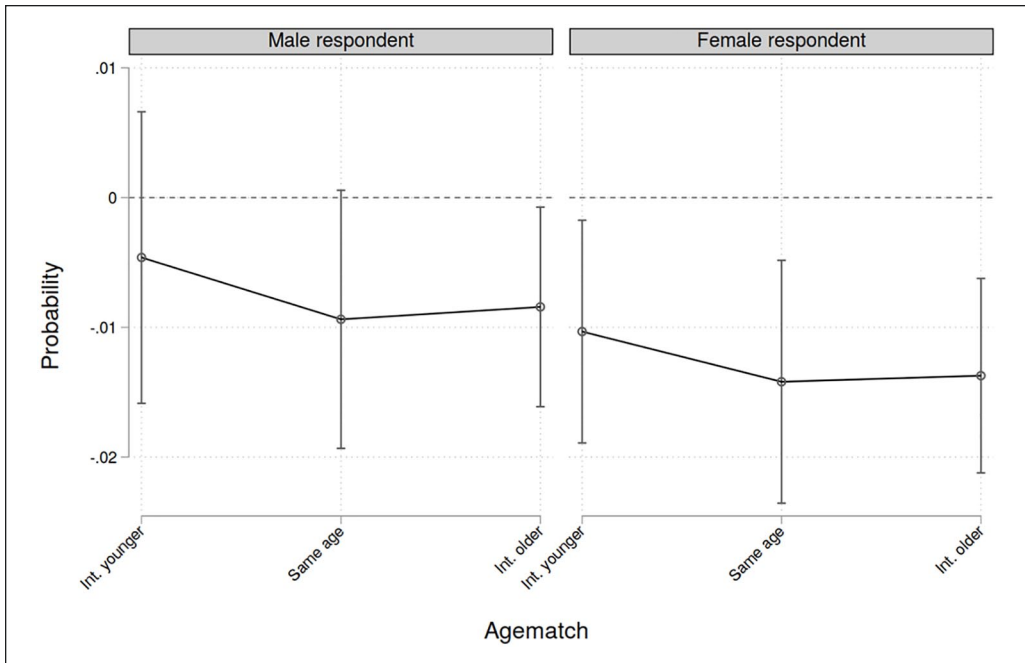


Figure 3. Average marginal effects of gender and age matching regarding the probability that another adult is present.

Source. ESS8/9.

Note. Included are 95-percent confidence intervals. The coefficients display the predicted effect of a successful gender match (i.e., interviewer and respondent have the same gender), separately for gender of respondent and age-matching groups. If a confidence band intersects the zero line, there is no statistically significant effect present on the 5-percent level since zero might be included in the confidence band of the effect.

this would support the theory of liking. For men, there are only small effects. These results are approximately in line with the theory of liking that predicts more similar interviewers produce better results and less interference. Hypothesis 3 is therefore accepted. However, this only refers to matching gender since there are only very little effects for matching age groups.

Ultimately, the question arises why the theory of liking mostly fails in the current analyses. Since the sample size is very large, many different social and cultural settings are included, and relevant control variables are available, the data quality is probably not the source of the null findings (however, some caveats are discussed further below). Possibly, the definition of “match” is too crude in the given operationalization. While gender is surely an important aspect of similarity, gender alone will only capture a tiny effect of the overall sociodemographic similarity. Including age in this definition will surely increase precision and allows a better matching but is still probably not very precise as other main factors like educational level, intelligence, migration status, race, political views, health status, or marriage are disregarded, to name only a few. To test this in more detail, studies must be conducted that provide more details about the interviewers since these usually do not answer detail questionnaires. However, the practical implications of this finding must also be regarded since it is probably not practical to match respondents and interviewers perfectly with respect to a large number of dimensions. A second major aspect to consider is the *content* of the items. As the literature review reveals, matching effects are especially large when gender-related topics are included, for example, items about the societal roles of males and females or views on gender egalitarianism. As the ESS 8 and 9 include almost no such items, this might affect the outcomes and underestimate the influence of gender matching.

It is conceivable that results might be much more pronounced when sensitive items about gender-related topics itself are asked. Further studies should attempt to look more into this issue.

Limitations

The main shortcoming of the design is that it cannot be guaranteed that respondents and interviewers are originally matched at random. This is because the design is not interpenetrated and, additionally, there might be selective refusal rates based on the originally assigned interviewer. This issue implies that potential interviews with the most extreme outcomes (i.e., largest amount of item nonresponse and most reluctance to answer) were never conducted since the respondents refused any questioning altogether. If we assume that this decision to participate is based on some characteristics of matching, this means that the effects reported in this study are conservative since these extreme outcomes were never realized. Unfortunately, there is no simple solution for this problem; it is clearly not possible to enforce participation, yet by applying multilevel models and controlling for certain respondent characteristics, these obstacles should be attenuated. One should also consider that the ESS, despite these challenges, does provide excellent data for a large number of respondents with a large set of additional datasets, which makes this analysis possible in the first place. Previous research papers have successfully demonstrated that the ESS is suitable for posed research questions. A second limitation is the operationalization of age match as already discussed above. The cutoff point of five years is arbitrary and different values could be considered. 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. Future studies might compare different ways of operationalization to test whether effects are stable. Finally, the question arises how these findings can be used to improve the survey process in future studies. Apparently, both men and women have lower item nonresponse when interviewed by men, which seems highly relevant for the data quality, as missing items are always a severe problem. However, stating that more male interviewers should be employed as to increase the number of retrieved items might be highly controversial at this point and would require more detailed investigations that are, optimally, based on experimental designs. Furthermore, there is contrary evidence from past research where respondents were *more* likely to answer sensitive questions by female interviewers (Heer and Passel 1987). The conclusion is that gender effects of interviewers are either quite volatile or depend on the type of questions asked. Since the number of sensitive questions is rather small in the ESS, the findings are probably not generalizable to other surveys with different topics of interest.

Conclusion and Practical Advice

The aim of the research was to investigate how sociodemographic matching between respondents and interviewers in face-to-face interviews affects data quality of the responses given. The analysis, which is based on multilevel data from 25 European countries, leads to the conclusion that some matches actually can improve data quality significantly (at least statistically). Gender matching does decrease the number of item nonresponse only for male respondents and even has a negative effect for female respondents. While there is no effect at all for the reluctance to answer questions, the effects regarding the presence of another adult are clear and highlight that gender matching does have small effects and benefits the survey settings since being alone with the respondent should be beneficial. Based on these findings, which stem from a very large and multinational European sample, giving practical advice is not straightforward. The first finding is that male interviewers seem to reduce the amount of item nonresponse for both male and female respondents, as men might probe more and longer if necessary when no clear answer is given. However, regarding the presence of other adults, a gender match is actually beneficial for

the setting. Insofar, it appears that the practical advice should be not to overestimate the effects of gender matching. Effects are small and putting much effort into the matching does probably not yield large benefits. Especially regarding age match, which would be even more difficult to implement, there are basically no effects and any efforts might be wasted. However, it might be beneficial to implement a gender matching when the questionnaire contains many sensitive items and the presence of third persons should be minimized.

Future studies should build on these results and attempt to investigate causal mechanisms. Testing why the theory of liking fails in this context seems like a highly relevant starting point. Given that previous findings were also rather mixed, it might be beneficial to modify or even abandon this theory, at least for practical advice since only little positive findings are available. Second, modeling the process of self-selection into treatment, based on interviewer characteristics, would be another highly desirable area of research to understand whether interviewer characteristics influence the chance for unit nonresponse. Another outcome to consider are *patterns* of missing data in future research to better understand which questions are considered especially sensitive and how this relates to matching. As shown in the limitations, current evidence is mixed and no clear answers are possible at the moment. Furthermore, it could be interesting to put a focus on country differences and analyze whether the patterns found are the same in all European countries or whether there are significant differences. It might be that macroinfluences, like economic prosperity, type of social welfare systems, degree of multiculturalism, or other psychosocial factors, impact how respondents and interviewers interact. These aspects seem central as matching between respondents and interviewers could be a relatively simple yet promising technique to improve data quality, which should be essential for any survey research.

Acknowledgments

The author would like to thank Mark Trappmann, Viktoria Sophie Zorn, and the two anonymous reviewers for their helpful comments.


Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Felix Bittmann  <https://orcid.org/0000-0003-0802-5854>

Notes

1. Exploring nonlinear effects seems like a relevant starting point for subsequent analyses.
2. At this point, it should be noted that introducing a fourth level (region within country) was tested. However, the results indicated that this additional information resulted in worse model fits (AIC/BIC measurements) and required a great amount of additional computational time. Therefore, it was decided to keep the more parsimonious models.
3. Complete do-files are available on request.

References

- Axinn, William G. 1989. "Interviewers and Data Quality in a Less Developed Setting." *Journal of Official Statistics* 5(3):265–80.

- Brunton-Smith, Ian, Patrick Sturgis, and George Leckie. 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* 180(2):551–68.
- Byrne, Dorm, Charles Gouaux, William Griffith, John Lamberth, N. Murakawa, M. Prasad, Atma Prasad, and M. Ramirez, III. 1971. "The Ubiquitous Relationship: Attitude Similarity and Attraction: A Cross-cultural Study." *Human Relations* 24(3):201–207.
- Catania, Joseph A., Diane Binson, Jesse Canchola, Lance M. Pollack, and Walter Hauck. 1996. "Effects of Interviewer Gender, Interviewer Choice, and Item Wording on Responses to Questions Concerning Sexual Behavior." *Public Opinion Quarterly* 60(3):345–75.
- Chun, Helen, Maria I. Tavaréz, Grace E. Dann, and Michael P. Anastario. 2011. "Interviewer Gender and Self-reported Sexual Behavior and Mental Health among Male Military Personnel." *International Journal of Public Health* 56(2):225–29.
- Colsher, P. L. and R. B. Wallace. 1989. "Data Quality and Age: Health and Psychobehavioral Correlates of Item Nonresponse and Inconsistent Responses." *Journal of Gerontology* 44(2):P45–52.
- Davis, R. E., M. P. Couper, N. K. Janz, C. H. Caldwell, and K. Resnicow. 2010. "Interviewer Effects in Public Health Surveys." *Health Education Research* 25(1):14–26.
- "Definition of Interfere." 2018. Merriam-Webster. Retrieved April 8, 2019 (<https://www.merriam-webster.com/dictionary/interfere>).
- De Leeuw, Edith D., Joop J. Hox, and Mark Huisman. 2003. "Prevention and Treatment of Item Nonresponse." *Journal of Official Statistics* 19:153–76.
- Dotinga, Aafje, Regina J. J. M. van den Eijnden, Willem Bosveld, and Henk F. L. Garretsen. 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):242–48.
- Durrant, G. B., R. M. Groves, L. Staetsky, and F. Steele. 2010. "Effects of Interviewer Attitudes and Behaviors on Refusal in Household Surveys." *Public Opinion Quarterly* 74(1):1–36.
- "ESS9 Documentation Report 2018." 2018. Retrieved https://www.europeansocialsurvey.org/docs/round9/survey/ESS9_data_documentation_report_e03_1.pdf.
- Fischer, Claude S., ed. 1977. *Networks and Places: Social Relations in the Urban Setting*. New York: Free Press.
- Fletcher, Clive and Amanda Spencer. 1984. "Sex of Candidate and Sex of Interviewer as Determinants of Self-presentation Orientation in Interviews an Experimental Study." *Applied Psychology* 33(3):305–13.
- Groves, Robert M., ed. 2009. *Survey Methodology*. 2nd ed. Hoboken, NJ: John Wiley.
- Groves, Robert M., Robert B. Cialdini, and Mick P. Couper. 1991. *Toward a Useful Theory of Survey Participation*. Washington, DC: U.S. Census Bureau [Custodian].
- Groves, Robert M., Robert B. Cialdini, and Mick P. Couper. 1992. "Understanding The Decision to Participate in a Survey." *Public Opinion Quarterly* 56(4):475–95.
- Hartmann, Petra. 1994. "Interviewing When the Spouse Is Present." *International Journal of Public Opinion Research* 6(3):298–306.
- Heer, David M. and Jeffrey S. Passel. 1987. "Comparison of Two Methods for Estimating the Number of Undocumented Mexican Adults in Los Angeles County." *International Migration Review* 21(4): 1446–73.
- Herzog, A. Regula and Willard L. Rodgers. 1992. "The Use of Survey Methods in Research on Older Americans." Pp. 60–90 in *The Epidemiologic Study of the Elderly*, edited by R. B. Wallace and R. F. Woolson. Oxford, England: Oxford University Press.
- Johnson, Timothy P., Michael Fendrich, Chitra Shaligram, Anthony Garcy, and Samuel Gillespie. 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, Timothy P. and Jennifer A. Parsons. 1994. "Interviewer Effects on Self-reported Substance Use among Homeless Persons." *Addictive Behaviors* 19(1):83–93.
- McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27(1):415–44.

- Mneimneh, Zeina M., Roger Tourangeau, Beth-Ellen Pennell, Steven G. Heeringa, and Michael R. Elliott. 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):673–97.
- Oyinlade, A. Olu and Alex Losen. 2014. "Extraneous Effects of Race, Gender, and Race-gender Homo- and Heterophily Conditions on Data Quality." *SAGE Open* 4(1):1–13.
- Preetz, Richard, and Malte Langeheine. 2017. "Why Were There Three? Determinants of the Presence of an Intimate Partner during Face-to-face Interviews." *Survey Methods: Insights from the Field*: 1–17. Retrieved, https://www.ssoar.info/ssoar/bitstream/handle/document/55044/ssoar-insights-2017-preetz_et_al-Why_were_there_three_Determinants.pdf?sequence=1.
- Reuband, Karl-Heinz. 1992. "On Third Persons in the Interview Situation and Their Impact on Responses." *International Journal of Public Opinion Research* 4(3):269–74.
- Rice, Stuart A. 1929. "Contagious Bias in the Interview: A Methodological Note." *American Journal of Sociology* 35(3):420–23.
- Riphahn, Regina T. and Oliver Serfling. 2005. "Item Non-response on Income and Wealth Questions." *Empirical Economics* 30(2):521–38.
- Stoop, Ineke, Achim Koch, Verena Halbherr, Geert Loosveldt, and Rory Fitzgerald. 2016. "Field Procedures in the European Social Survey Round 8: Guidelines for Enhancing Response Rates and Minimising Nonresponse Bias." Retrieved February 18, 2021 (https://www.europeansocialsurvey.org/docs/round8/methods/ESS8_guidelines_enhancing_response_rates_minimising_nonresponse_bias.pdf).
- Stotland, Ezra and Martin Patchen. 1961. "Identification and Changes in Prejudice and in Authoritarianism." *The Journal of Abnormal and Social Psychology* 62(2):265–74.
- Su, Rong, James Rounds, and Patrick Ian Armstrong. 2009. "Men and Things, Women and People: A Meta-analysis of Sex Differences in Interests." *Psychological Bulletin* 135(6):859–84.
- Suedfeld, Peter, Stephen Bochner, and Carol Matas. 1971. "Petitioner's Attire and Petition Signing by Peace Demonstrators: A Field Experiment 1." *Journal of Applied Social Psychology* 1(3):278–83.
- Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. Cambridge, England: Cambridge University Press.
- Vercruyssen, Anina, Celine Wuyts, and Geert Loosveldt. 2017. "The Effect of Sociodemographic (Mis) Match between Interviewers and Respondents on Unit and Item Nonresponse in Belgium." *Social Science Research* 67:229–38.
- Webster, Cynthia. 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, Carol H. 1968. "Validity of Welfare Mothers' Interview Responses." *The Public Opinion Quarterly* 32(4):622–633.
- West, Brady T, and Annelies G, Blom. 2017. "Explaining Interviewer Effects: A Research Synthesis." *Journal of Survey Statistics and Methodology* 5(2):175–211.
- Williams Jr, J Allen. 1964. "Interviewer-Respondent Interaction: A Study of Bias in the Information Interview." *Sociometry* 27(3):338–352.
- Wilson, Sandra R, Nancy L, Brown, Carolina, Mejia, and Philip W, Lavori. 2002. "Effects of Interviewer Characteristics on Reported Sexual Behavior of California Latino Couples." *Hispanic Journal of Behavioral Sciences* 24(1):38–62.

Author Biography

Felix Bittmann, sociologist by training, is a research assistant and doctoral candidate at the Leibniz Institute for Educational Trajectories in Bamberg, Germany. In his dissertation, he investigates the role of educational aspirations in the context of educational trajectories with quantitative panel data. His research interests comprise the role of education in the life course, social inequality, quantitative methods, and philosophy of science.