Survey Research Methods (2018)
Vol. 12, No. 3, pp. 211-230
doi:10.18148/srm/2018.v12i3.7230

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

Continuity Trumps? The Impact of Interviewer Change on Item Nonresponse

Kristin Hajek
Department of Sociology
LMU München, Germany

Nina Schumann
Department of Sociology
LMU München, Germany

Allocating the same interviewer to each respondent over multiple waves is typically recommended for panel surveys. While some studies have investigated the effect of this strategy on wave nonresponse, there is scarce empirical evidence on how interviewer (dis-)continuity affects item nonresponse. This is surprising, given that the amount and pattern of item nonresponse is a crucial aspect of data quality. Using the first seven waves of the German Family Panel pairfam, we investigate whether interviewer continuity indeed influences item nonresponse in a non-experimental setting. Our analysis differentiates between "I don't know" responses and the complete refusal to answer, both with respect to the specific question of household income as well as the entire question program. By applying cross-classified multilevel models and estimating within-respondent effects, we can base our results on an intra-individual comparison, controlling for time-constant unobserved characteristics and taking into account the complex structure of the data. Our analysis shows no detrimental effect of an interviewer change, per se, over the course of the panel; a new interviewer only collects more "I don't know" answers if the respondent belongs to the oldest age group (born 1971–73). Younger respondents, in contrast, have a lower likelihood to answer with "I don't know" to the income question if they experience a change in interviewer. Changes in social distance with respect to age and gender do not prove to be relevant mechanisms for this effect. Only female respondents of the youngest cohort exhibit a lower likelihood of "I don't know" responses on the income question when reassigned to a female interviewer. Older interviewers tend to get more "I don't know" answers, whereas reassignment to a more experienced interviewer (regardless of age) appears to encourage less "I don't know" answers.

Keywords: item nonresponse; interviewer change; interviewer characteristics; panel surveys; cross-classified multilevel models

1 Introduction

A common strategy of survey agencies conducting panel studies is to assign the same interviewer to the same respondent over the course of the panel. While some studies have examined the effect of this approach on wave nonresponse (e.g., Campanelli & O'Muircheartaigh, 1999; Lynn, Kaminska, & Goldstein, 2011; Pickery, Loosveldt, & Carton, 2001), next to no empirical evidence exists as to whether interviewer change affects patterns of item nonresponse in panel surveys. The lack of research on this topic is astonishing, as the extent of item nonresponse determines the amount of useable data and data quality. Prevention of item nonresponse is of particular importance, as it potentially poses serious threats such as (non-random) missing cases, as well as complicated ex-post procedures of imputation and adjustment. Data incompleteness is an issue if respondents who do not answer specific questions significantly differ from respondents who do (Groves et al., 2009; Pickery & Loosveldt, 1998; Pickery et al., 2001); in this case, data is not "missing (completely) at random" (de Leeuw, 2001). One common example for non-random missing data is the higher prevalence of item nonresponse to income questions in the tails of the distribution (Riphahn & Serfling, 2005). In light of this, it is surprising that this issue has generally gained less attention than that of unit and wave nonresponse or deviant, i.e., non-true answers.

In general, item nonresponse can result from data collection methods, the questionnaire, the respondent, and/or the interviewer (de Leeuw, 2001; Groves, 1989). In this article, we focus on the interviewer as the source of item nonresponse. Specifically, we concentrate on the effect interviewer (dis-)continuity has on item nonresponse. The allocation of interviewers to respondents is one factor researchers and agencies can determine to a high degree. Usually, survey

Contact information: Kristin Hajek, Department of Sociology, LMU München, Konradstraße 6, 80801 Munich, Germany (e-mail: Kristin.Hajek@soziologie.uni-muenchen.de)

research agencies prefer to allocate the same interviewer to the same respondent over time. But is this strategy helpful in reducing the amount of missing responses – or under certain conditions even harmful? Do survey researchers need to worry about this aspect in cases for which interviewer change is unavoidable due to interviewer attrition or respondents relocating?

In this article we address the following key questions: First, does an interviewer change between panel waves affect item nonresponse? Second, does the matching of interviewer and respondent characteristics mitigate the effect of interviewer change? Third, is interviewer experience an explanatory mechanism? Fourth, are patterns different for previous wave non-respondents or respondents who moved residence before an interviewer change occurred? In order to answer these questions, this study analyzes nonresponse patterns to the question on household income (which usually produces relatively high amounts of nonresponse and is essential for many analyses) and the overall share of item nonresponse across all survey questions.

Most research on item nonresponse thus far has used cross-sectional data. Taking full advantage of a large, randomly sampled panel study, we employ a combination of multilevel cross-classified models (due to the three non-nested levels: wave, respondent, and interviewer) and longitudinal hybrid models. The latter enables us to base our results on a within-respondent comparison of item nonresponse. One major advantage of this approach is that time-constant respondent characteristics such as education which might influence, for example, the cognitive ability to understand questions and give substantive answers, do not need to be controlled for. Due to the multilevel approach, we can also evaluate the variance share of each of the three levels. To our knowledge, this study is the first to analyze the effect of interviewer change on item nonresponse using a panel design.

2 Item Nonresponse and the Role of the Interviewer

Item nonresponse, i.e., giving a non-substantive answer, can result from two processes: either the respondent lacks the relevant knowledge to provide a meaningful answer, or is unwilling to reveal the true answer. The former can be regarded as a limitation of cognitive resources (e.g., issues understanding the question or being unable to recall relevant information), while the latter is attributable to the respondent's assessment of the interview situation, resulting in a lack of cooperation or motivation (Groves et al., 2009; Schräpler, 2006). Item nonresponse can be regarded as a form of satisficing, where either the respondent is not motivated enough or not able to generate a substantive answer (Krosnick, 1991). In addition, respondents might use the "I don't know" response as a less costly and more polite way to refuse, for example when being confronted with sensitive issues. Following evidence that "I don't know" and refusals should, nonetheless, be considered as different types of item nonresponse (Riphahn & Serfling, 2005; Schräpler, 2006; Tu & Liao, 2007), this paper differentiates between the two.

Interviewer characteristics play a crucial role in obtaining respondent cooperation as participation in face-to-face surveys is a communicative process (de Leeuw, 2001; Groves & Couper, 1998). There is evidence that interviewers' sociodemographic characteristics such as age, gender, or education not only influence unit nonresponse (e.g., Durrant, Groves, Staetsky, & Steele, 2010; Lipps & Pollien, 2011; Müller & Castiglioni, 2015; Pickery & Loosveldt, 2002), wave nonresponse (Lipps & Pollien, 2011; Müller & Castiglioni, 2015), or the consent for linking survey data with administrative records (Korbmacher & Schroeder, 2013), but also item nonresponse (e.g Pickery & Loosveldt, 1998; Riphahn & Serfling, 2005; Schräpler, 2006). The findings on the latter are not unequivocal, however. For example, Riphahn and Serfling (2005) show that female interviewers generate a higher item nonresponse rate than male interviewers, especially if the respondent is female. In contrast, Schräpler (2006) finds that male interviewers have more questions refused than do female interviewers. These two papers differ in both the data used (BHPS vs. GSOEP) and the methods applied (multinomial logit model vs. multilevel probit model), but both investigate income nonresponse. Based on pairfam data, Müller and Castiglioni (2015) report a small positive effect of gender matching on the likelihood to participate in the panel study. Pickery and Loosveldt (1998) find significant interviewer effects on the amount of item nonresponse, but no effects of interviewer age, sex, education, or experience. Therefore, further unmeasured interviewer characteristics seem to influence item nonresponse[1]. Tu and Liao (2007) report effects of social distance between interviewers and respondents concerning age, gender, and ethnicity on item nonresponse to questions on sexual attitudes and behaviors. According to their analyses, interviewer effects are somewhat different for "I don't know" answers and refusals: While age and education distance affect the prevalence of both types of nonresponse (lower item nonresponse with higher age distance, and higher item nonresponse with greater distance in education level), the effects of gender and ethnicity are limited to complete refusals. More specifically, cross-gender interviews, male-male interviews, and larger ethnicity differences produce more such nonresponse.

In addition to sociodemographic characteristics, inter-

[1] Attrition among interviewers might be non-random, for example when interviewer turnover is higher in certain areas (e.g., Campanelli & O'Muircheartaigh, 2002; Watson & Wooden, 2009). This suggests that interviewer change can influence wave nonresponse if the interviewers differ on certain characteristics. Therefore, controlling for interviewer characteristics is important when analyzing interviewer change.

viewer experience seems to play a role not only in survey refusal (e.g., Jäckle, Lynn, Sinibaldi, & Tipping, 2013; Lipps & Pollien, 2011), but also for item nonresponse (e.g., Bilgen, 2011; Essig & Winter, 2009). However, empirical evidence regarding this association is also not uniform. Bilgen (2011), for example, finds slightly higher item nonresponse rates when interviewers are more experienced, both with respect to lifetime and study-specific experience. Similarly, Tu and Liao (2007) report interviewers that have more experience with the same survey organization obtain more refusals and "I don't know" responses to questions on sexual attitudes and behaviors. In contrast, Essig and Winter (2009) report that more experienced interviewers obtain lower rates of item nonresponse for questions on financial assets, and Pickery and Loosveldt (1998) do not find any association between within-study interviewer experience and "no opinion" responses to attitudinal questions. In this paper, we differentiate two groups of interviewers based on the number of interviews they have conducted over all waves of the pairfam panel study.

Several studies have examined the effect of interviewer (dis-)continuity with respect to wave nonresponse, with mixed and even contradictory evidence. Campanelli and O'Muircheartaigh (1999, 2002) as well as Pickery et al. (2001) do not find an effect of interviewer continuity on wave nonresponse. Using pairfam data, Müller and Castiglioni (2015) find a negative effect of interviewer change on continued panel participation (irrespective of the wave in which the interviewer change occurs). Lynn et al. (2011) divide interviewers into low, middle, and high grade interviewers and show that an interviewer change only results in a lower response rate if the interviewer for the first wave was middle grade, and the interviewer of the following wave was a low grade interviewer. They conclude from an experimental design that the relationship between interviewer continuity and wave nonresponse is considerably more complex than typically suggested.

Respondents who do not consistently participate in each panel wave might be affected differently by an interviewer change than continuous participants. Watson and Wooden (2014) specifically analyze the re-engagement of previous wave non-respondents (so-called temporary dropouts) using three panel data sets from different countries. They report that interviewer continuity is detrimental for re-engagement probabilities and attribute this finding to a lack of a positive respondent-interviewer relationship. In addition, a new interviewer might employ different approaches to motivate the respondent to participate in the survey again. Results are different when modelling continued participation: Here, interviewer continuity proves to be beneficial. These findings highlight the importance of regarding temporary dropouts as a unique group of respondents when analyzing interviewer change.

The effect of interviewer discontinuity on item nonresponse has been found to differ depending on the outcome studied. Riphahn and Serfling (2005) include interviewer continuity as a control variable when dealing with item nonresponse on questions regarding income and wealth. Across all items on financial outcomes, they do not find a significant effect of interviewer change. However, they report heterogeneous patterns across outcomes: When differentiating between items, interviewer change increases item nonresponse rates for income, but reduces this for wealth. Their study, however, is based on one wave only[2] (using data from 1988) and is limited to financial outcomes; thus, it does not give insight into which effect of interviewer (dis-)continuity is relevant in a social survey over different types of topics and several panel waves. Analyzing the data of the CATI-administered Swiss Household Panel, Lipps (2007) does not find a significant effect of interviewer continuity on either an index of social desirability, answering extreme categories, or income nonresponse[3]. Moreover, Chadi (2013) found that familiarity with the interviewer in a panel survey can affect respondent answering behavior: With increasing encounters, respondents' overall reported life satisfaction decreases.

As several studies on item nonresponse focus on financial questions, we also analyze the question on household income in order to link our results to prior research. This item typically generates high nonresponse rates and is central to a multitude of research questions. In addition, however, we broaden the scope of our study to include the overall share of nonresponse in order to gain additional insight for survey scientists and practitioners.

3 Theoretical Background and Implications

Two theoretical frameworks regarding mechanisms of survey response behavior and question-answer processes have been discussed in the previous literature. The first is a cognitive model which stresses the different tasks respondents face, such as understanding the question, recalling relevant information from memory, computing a judgment, formatting the judgment according to the given response categories, and editing the response when issues of social desirability, self-representation, or situational adequacy crop up (for overviews see Sudman, Bradburn, & Schwarz, 1996; Tourangeau, Rips, & Rasinski, 2000). The second is based

[2] The majority of studies exploring item nonresponse employ a cross-sectional approach. One exception is Young (2012), who finds that the use of "I don't know" responses to specific questions is fairly stable when asked nine or ten years later. However, no attention is given to the role of interviewer (dis-)continuity in her analysis.

[3] Although this study takes into account the cross-classified structure of the data, the interviewer level is not included in the full longitudinal models due to convergence issues.

on rational choice theory (cf. Groves & Couper, 1998): Respondents first consider the costs and benefits of refusing or deciding to answer specific questions and then choose the option with the highest expected utility. Only if the benefits outweigh the costs will the respondent be willing to participate in a survey and provide a response to any specific question.

Interviewer change can a priori be either detrimental or conducive to the quality of the generated data. We infer this assumption from two different lines of arguments resulting in common survey strategies: If we regard interviews as social exchanges between respondent and interviewer, and if exchange is based on mutual trust (Homans, 1961), we can hypothesize that if respondents trust their assigned interviewer, the costs of responding are lower, and they may be more willing to recall or reveal information (cf. Riphahn & Serfling, 2005). This highlights the (potential) relevance of the personal connection between respondent and interviewer. Presumably, respondents will trust a continuing interviewer more than a new, unfamiliar one (Lynn et al., 2011; Riphahn & Serfling, 2005). In addition, a continuing interviewer might be able to use prior knowledge of the respondent to tailor his/her tactics for persuading respondents to cooperate and for helping give a substantive answer if cognitive demands are high (Lynn et al., 2011).

On the other hand, conventional practice dictates that – prior to the first interview – respondents and interviewers should not be acquainted. Rodriguez, Sana, and Sisk (2015) refer to this as the "stranger-interviewer norm," reflecting the assumption that respondents are less prone to reveal sensitive information to interviewers they already know. While the pairfam data analyzed here stem from a fully standardized face-to-face survey, the topics covered are rather private and the length of the interview (approximately one hour) might foster the establishment of a personal relationship between respondent and interviewer. Thus, at the time of the second interview with the same interviewer, the "stranger-interviewer norm" could be regarded as violated, possibly resulting in a negative influence on data quality.

In general, rapport between respondent and interviewer (i.e., a friendly, harmonious relationship) can be regarded as a "double-edged sword" (Groves et al., 2009, p. 304): Sympathetic feelings toward the interviewer and establishing a personal relationship can motivate respondents, but may also lead to a distortion of answers. So far, little research exists concerning an ideal amount of interviewer rapport. This might be one reason for the unequivocal findings regarding interviewer (dis-)continuity presented above, which would support our presumption that both a positive and a negative effect of interviewer change on the likelihood and share of item nonresponse is conceivable. Temporary dropouts, for example, might benefit from an interviewer change. The fact that they have withdrawn their participation in the previous wave might be due to poor rapport with their past interviewer (Watson & Wooden, 2014). Thus, the effect of interviewer continuity could be negative for this specific group of respondents (also see Lynn et al., 2011 who point to the possibility of non-uniform effects of interviewer change), and we include interaction effects to account for this possibility. In addition, respondents who have moved between waves could expect their interviewer to change, and might therefore be less negatively affected by an interviewer change. Therefore, we include an interaction effect of interviewer change and respondents' geographic mobility.

If the allocated interviewer changes between waves, the combination of sociodemographic characteristics between respondent and interviewer will most likely change as well. Thus, the (mis-)match is a potential mechanism explaining the presumed association between interviewer change and item nonresponse. A match or higher similarity in relevant respondent and interviewer characteristics (e.g., age and gender) can be expected to increase response rates. Respondents might trust their interviewer more and be therefore more willing to respond freely when there is a perceived closeness and low social distance (Riphahn & Serfling, 2005; Tu & Liao, 2007). Thus, we hypothesize that the likelihood for item nonresponse increases with age difference and a difference in gender.

Additionally, interviewer experience is also likely to influence item nonresponse patterns and can be regarded as an intervening mechanism if the previous and the new interviewer differ in their respective experience. Two competing processes are conceivable: On the one hand, interviewers who conduct more interviews might be more efficient in that they work faster and less thoroughly, thereby generating answers of lower quality. On the other hand, more experienced interviewers might be more effective at obtaining substantive answers due to a better knowledge of the survey questions and successful interview tactics (cf. Bilgen, 2011; Pickery et al., 2001).

Figure 1 graphically displays the causal relationships and the potential mechanisms[4] we are interested in. The only control variables included are wave dummies and whether the respondent has moved between waves. Wave dummies are considered to account for effects of the question program as well as general trends of respondent learning or habitualization effects on item nonresponse. Moreover, panel wave could also affect the likelihood of interviewer change if an interviewer no longer works for the survey agency in a particular wave or is reassigned due to the agency's internal policies. A respondent relocation leads to an interviewer change if the previous interviewer is not assigned to the respondent's new living area. Respondent mobility might be followed by changes in household composition, their professional life, or other life events that might influence response

[4] We do not assume that the effect of interviewer change completely disappears once we include these mechanisms.

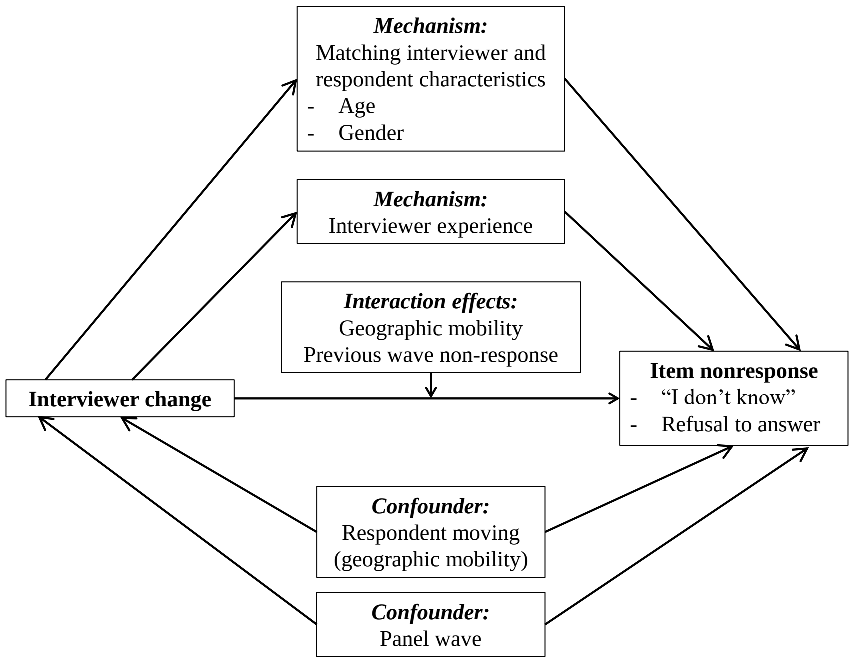

Figure 1. Theoretical model of the effect of interviewer change on item nonresponse

behavior. The question on household income, for example, might yield less "I don't know" answers if respondents move out of the parental home and are more familiar with their own financial situation compared to that of their parents. More refusals might occur if respondents are reluctant to provide information on new household members. Thus, it is necessary to include geographic mobility in our models. Panel wave and geographic mobility are potential confounders, meaning variables that can affect both interviewer change and item nonresponse, and therefore need to be controlled for in order to avoid omitted variable bias. It is a major advantage of our research design that – apart from these two confounders – interviewer change is exogenous, i.e., independent from respondents' characteristics. Controlling for any further variables might lead to overcontrol bias (see Elwert & Winship, 2014). Thus, life events which do not affect the likelihood to experience an interviewer change do not need to be included in the models.

4 Data and Sample

The German Family Panel pairfam is a multidisciplinary survey focusing on partnership and family dynamics. Randomly sampled "anchor" persons from three birth cohorts (1991–93, 1981–83, and 1971–73) have been surveyed annually since 2008. Furthermore, anchors' partners, parents, and children are also included in the survey. Our analysis on item nonresponse is based on anchor data from Release 7.0 (Brüderl et al., 2016).[5] A more detailed description of the study can be found in Huinink et al. (2011).

The core pairfam topics include the process of partnership formation, development, and separation, expectations regarding partnerships, parenthood, the timing and spacing of births, parenting, child development, intergenerational relationships, as well as social embeddedness. Although most questions across the first seven waves are stable, some instruments rotate every two years, and others have only been asked once. The majority of questions in the anchor survey are asked face-to-face (CAPI, i.e., computer-assisted personal interview), but a self-administered module (CASI) has been included for more sensitive questions such as sexuality and personality traits. As the dependent variables in our analysis we use the question on household income as well as the percentage of "I don't know" responses and question refusals across all CAPI-administered questions for which these responses were permitted. We exclude all CASI questions as we assume that interviewer effects should account for less bias in this self-administered module.

The main pairfam sample for waves 1 to 7 consists of 53,447 person-year observations, and permits one-wave temporary dropouts; for example, it is possible to participate in wave 1 and then again in wave 3, skipping wave 2. We

[5] To order the scientific use file of the German Family Panel follow the instructions given on the pairfam homepage (http://www.pairfam.de/en/data/data-access/).

only consider the first interviewer change in our sample and censor cases which experience a second change, thus eliminating 1,929 observations. Moreover, we exclude eleven observations due to an invalid gender indication and 1,363 more with missing values on any of the examined variables. For respondents who experience an interviewer change, we delete all observations except for the last with the first interviewer (dropping 4,698 observations). We do this to be able to directly compare the last interview with the first interviewer and the following interviews with the second interviewer. Without doing so we would obtain a comparison of the average of all interviews with the first interviewer to the following ones with the second interviewer, which is not our substantive aim. We further restrict our sample to respondents who participated in at least two waves (dropping 2,620 observations), as only then can within-person effects be estimated. Furthermore, we exclude 1,659 observations with more than 10% "I don't know" responses or a 10% question refusal rate for at least one wave. Our final sample consists of a total number of 41,167 person-year observations for 9,048 anchor respondents.

When possible, the survey agency assigns the same interviewer to the same respondents in each pairfam wave. However, if an interviewer dropped out of the interviewer staff, was temporarily unavailable, or either interviewer or respondent moved to another sampling area, an interviewer change would occur. Furthermore, interviewers with a poor response rate were replaced in some cases (Brix, Wich, & Schneekloth, 2015). To model the effect of an interviewer change over time, four dummy variables were generated indicating the number of encounters with the second interviewer. The variable "first encounter with second interviewer" is coded 1 if an interviewer change occurred in the current wave, and 0 again in the following waves. Similarly, the variable "second encounter with the second interviewer" is 1 in the interview following the change, otherwise 0. The same pattern is applied to the variables "third encounter" and "fourth or more". Due to the limited numbers of cases, the fourth and all following encounters with the second interviewer are combined into one dummy variable. With this strategy, we can explore how an interviewer change effects item nonresponse depending on the number of encounters with the new interviewer and whether potential effects of an interviewer change attenuate or even strengthen over time.

To determine possible effects of interviewer characteristics and interaction effects with respondents' characteristics, we include the interviewer's age, the age difference to the respective respondent, changing from a male to a female interviewer (for male and female respondents, respectively), and interviewer experience. We generate a dummy variable indicating whether the interviewer has already completed more or less than the cumulative median number of pairfam interviews in each wave.

We identified respondent geographic mobility as a possible influence of both the probability of item nonresponse as well as the likelihood for interviewer change and include this variable as a potential confounder. Moving to a different area, to which a different interviewer might be assigned, could lead to an interviewer change, especially if the distance between original and new locations is large. We construct two dummy variables for a change in respondent's main residence since the previous wave: "respondent mobility up to 100 km" and "respondent mobility >100 km".

5 Analytical Strategy

At each interview occasion, panel observations (i.e., "person years") are nested both within respondents and within interviewers. However, not all respondents maintain the same interviewer over time. Thus, respondents and interviewers are crossed at level 2, leading to a non-hierarchical, parallel nesting structure. To account for this specific data structure, we employ cross-classified multilevel modelling using Bayesian estimation procedures as implemented via Markov Chain Monte Carlo (MCMC) methods. The analyses were conducted using MLwiN (version 2.36) using the runmlwin command (Leckie & Charlton, 2013) in Stata (version 14.2). This software (Rasbash, Browne, Healy, & Cameron, 2009) is especially suitable for complex multilevel modelling such as cross-classified models. We chose to use MCMC modelling (Browne, 2012) as it is computationally more efficient than IGLS (Iterative Generalised Least Squares) and extendable to cross-classified models. By using multilevel modelling, systematic differences (i.e., dependency) between clusters of both respondents and interviewers can be adequately accounted for, and the amount of variation due to each source can be quantified.

Substantively, we are interested in within-respondent estimates, meaning observed changes in item nonresponse patterns for the same individual over time. While conventional respondent fixed-effects models used for panel data have the advantage of controlling for (measured and unmeasured) unobserved heterogeneity at the respondent level (Brüderl & Ludwig, 2014; Wooldridge, 2010), such models cannot handle the cross-classified multilevel structure apparent in our data. To simultaneously account for this specific structure and base our conclusions on a within-respondent comparison, we employ hybrid models (Allison, 2009) by decomposing between and within estimations (see Schunck (2013) for an applied discussion of these models) with cross-classified random effects. Thus, our estimation strategy is an extension of conventional hybrid / between-within random effects models as there is more than one source of clustering (for a very recent discussion and application of such models in the context of survival analyses see Cafri and Fan, 2018)[6]. Mixed-

[6] To our knowledge, this is the only statistical discussion of this

effects models as well as cross-classified extension have been used by a variety of studies detecting interviewer effects in survey data (e.g., Brunton-Smith, Sturgis, & Leckie, 2017; Durrant et al., 2010; Lipps, 2007).

In a cross-classified multilevel model, observations i ($i = 1, \ldots, N$) simultaneously belong to two non-hierarchical contexts. In our case, these are respondents j ($j = 1, \ldots, J$) and interviewers k ($k = 1, \ldots K$). The cross-classification for the response measurement $y_{i(jk)}$ is indicated by placing the indices in parentheses:

$$y_{i(jk)} = \beta X_{i(jk)} + u_j + v_k + e_{i(jk)} \quad , \tag{1}$$

with u_j and v_k representing random-intercept effects (i.e., unobserved respondent- and interviewer-specific influences). The remaining observation-specific residual is $e_{i(jk)}$. These parameters are assumed to be independent of each other and normally distributed, with a mean of 0 and estimated variances $\sigma^2_u, \sigma^2_v, \sigma^2_e$:

$$u_j \sim N(0, \sigma^2_u), \quad v_k \sim N(0, \sigma^2_v), \quad e_{i(jk)} \sim N(0, \sigma^2_e) \quad .$$

To estimate our models, each time-varying predictor variable (i.e., fixed effect) is split into three components: $\bar{X}_j$ is the mean of the j^{th} cluster (between-cluster component), $\bar{X}_k$ is the mean of the k^{th} cluster (between-cluster component), and these cluster means are then subtracted from the individual covariate values (within-cluster component):

$$y_{ijk} = \beta_W(X_{ijk} - \bar{X}_j - \bar{X}_k) + \beta^{(j)}_B \bar{X}_j + \beta^{(k)}_B \bar{X}_k + u_j + v_k + e_{i(jk)} \tag{2}$$

In the results tables, we report within effects β_W since this is our substantive interest. In addition, we show variance contributions (i.e., random-effect variances σ^2_u, σ^2_v, σ^2_e) to evaluate the amount of variation due to the three levels. The intraclass correlation coefficient (ICC) for interviewers and respondents is calculated to evaluate the dependence in responses due to these sources. The intra-interviewer correlation coefficient ρ_k displays the correlation in the outcome between two observations of different respondents generated by the same interviewer. In contrast, the intra-respondent correlation coefficient ρ_j gives the correlation in the outcome between two observations of the same respondent generated by different interviewers:

$$\rho_j = \frac{\sigma^2_u}{\sigma^2_u + \sigma^2_v + \sigma^2_e}; \quad \rho_k = \frac{\sigma^2_v}{\sigma^2_u + \sigma^2_v + \sigma^2_e} \quad .$$

We estimate linear models for item nonresponse proportions for all CAPI questions. For the analysis of income nonresponse, we use linear probability models to ease the interpretation of the coefficient estimates, which may thus be interpreted as marginal effects. In addition, we replicate the income analyses as logistic regression models and report the results in the appendix.

6 Descriptive Findings

6.1 Interviewer Change and Interviewer Characteristics

In our sample, 6,560 respondents were surveyed by a total of 417 interviewers in wave 1 (Table 1). The sample decreased to 4,270 respondents and 247 interviewers by wave 7. Note that the number of respondents does not continually decrease in our sample due to our strategy of keeping only the last encounter with the first interviewer for respondents who experience an interviewer change at any point. The number of new interviewers joining the staff after wave 1 is very small (6-35 interviewers per wave), meaning the majority of respondents facing an interviewer change are interviewed by existing staff members with survey-specific experience from previous waves. The number of temporary dropouts decreases from 449 in wave 3 to 141 in wave 7, while the number of respondents' long-distance moves varies between 58 and 114 per wave. Short-distance moves occur considerably more often. The percentage of interviewer changes, identified using interviewer ID numbers, goes down from 15.7% in wave 2 to 5.7% in wave 7. Most respondents are interviewed by the same interviewer, but a considerable share (36%) of respondents experience at least one interviewer change between waves 1 and 7. The probability of experiencing an interviewer change is higher if respondents did not participate in the previous wave (29.9% compared to 7.2% averaged over all waves). Moreover, 58% of interviewer changes go hand in hand with the previous interviewer leaving the interviewer staff. However, it is not known if they leave due to retirement, health issues, better job opportunities, or low performance.

Between 40 and 44% of interviewers are women (depending on the wave), compared to approximately 53% of female respondents, with no systematic allocation of interviewers to respondents based on gender or age. The average interviewer age in wave 1 is 56 years, with a minimum of 22 and a maximum of 82. As the pairfam sample stems from only three birth cohorts (1991–93, 1981–83, and 1971–73), most respondents are considerably younger than their interviewers (Figure A1, see appendix).

6.2 Item Nonresponse in pairfam

Neither "I don't know" answers nor refusals were explicitly offered as response categories for questions conducted by the interviewer, but recorded if stated by the respondent. For household income[7] we distinguish between the refusal to answer and an "I don't know" response. Concerning the entire

between-within extension of cross-classified models.

[7] The wording of the question is the following: "Combining all income types: How much was the total monthly household income for all household members last month? Please indicate the monthly net amount, after deduction of taxes and contributions to retirement,

Table 1
Number of respondents and interviewers, percentage of interviewer change per wave

	Total		New interviewers	Previous wave non-resp.	Resp.' short-distance moves	Resp.' long-distance moves	Resp. with interviewer change
	Respondents	Interviewers					
Wave (year)	N	N	N	N	N	N	%
1 (2008/09)	6,560	417	-	-	-	-	-
2 (2009/10)	7,302	335	35	-	725	58	15.7
3 (2010/11)	6,526	315	19	449	831	114	10.7
4 (2011/12)	6,028	298	21	280	784	106	7.1
5 (2012/13)	5,471	286	19	233	740	107	8.0
6 (2013/14)	5,010	266	6	190	726	100	6.0
7 (2014/15)	4,270	247	11	141	556	67	5.7

Source: pairfam waves 1–7, Release 7.0 (own calculations)

CAPI question program, we also generated two dependent variables: the percentage of “I don’t know” answers over all CAPI variables, and the percentage of refusals over all CAPI variables. By doing so, we can differentiate the effects of an interviewer change on the respective type of item nonresponse. We do not regard questions skipped due to routing or by mistake; only responses to questions read aloud by the interviewer.

In each panel wave, “I don’t know” responses occur on average more often than refusals - overall and for the specific income item (see Table 2). Non-valid answers decrease over the panel, most pronounced for “I don’t know” responses to the income item: About 37% of respondents gave an “I don’t know” answer to the household income question in wave 1 and only about 16% in wave 7. One mechanism behind this pattern could be increased familiarity with the survey and the specific questions asked. Knowing which questions are asked every year, respondents can look up the necessary information before the interview takes place. Also, selective attrition might occur throughout the waves, with reluctant or less knowledgeable respondents dropping out at higher rates. As Table A1 (see appendix) shows, “I don’t know” response and refusal rates vary over cohorts: while 57% of respondents in cohort 1 compared to 7% in cohort 3 give an “I don’t know” response to the income question, 6% in cohort 1 and 12% in cohort 3 refuse to answer. Many younger respondents still attend school and live with their parents or with friends/peers. It is likely that they do not possess the relevant information on diverse household characteristics – even if they are willing to provide a valid answer. Older respondents, on the other hand, might have a higher income and may therefore be more reluctant to disclose their household income. Given these marked differences between cohorts, we will also present separate models for each of the three cohorts.

Item nonresponse propensity cannot be regarded as a stable personality trait. Rather, a significant number of respondents does not answer a question in one wave, but does in another: 49% of respondents answered with “I don’t know” in at least one wave, but only 2% of respondents always gave an “I don’t know” response to the household income question; 22% of respondents ever refused to answer this question, but less than 1% always refused to answer. Less than 1% of respondents never gave an “I don’t know” answer, while about 6% of respondents never refused to answer to any question across all seven waves[8]. Therefore, there must be circumstances which have an influence on people’s response pattern and those might be associated with an interviewer change or interviewer characteristics.

7 Multiple Regression Results

Table 3 shows the results of our multilevel analyses concerning household income, whereas Table 4 lists the estimations over all CAPI questions. To explicitly focus on the contrast between “I don’t know”/refusal and a valid answer (which is the distinction made by the theoretical arguments above), we discard changes from “I don’t know” to refusals and vice versa[9]. Due to our interest in changes within respondents over time, we present only within-cluster estimates (i.e., the effects of the demeaned variables). For single-level data, these within components are equivalent to coeffi-

unemployment, and health insurance. Please include regular payments like pensions, housing allowances, child allowances, student loans/allowances, child support, etc.”

[8] Depending on respondents’ personal situation, interview duration varies over waves and across respondents. On average, respondents answered 267 CAPI questions, the overall minimum at 84 and maximum at 567.

[9] This is why the number of observations differs between the models on “I don’t know” and refusals for the income analyses.

Table 2
Item nonresponse rates per wave

	Income: Don't know		Income: Refusals		CAPI: Don't know		CAPI: Refusals	
Wave (year)	N	%	N	%	mean %	SD[a]	mean %	SD[a]
1 (2008/09)	2,391	36.5	693	10.6	1.6	1.7	0.6	1.3
2 (2009/10)	2,443	33.5	747	10.2	1.0	1.2	0.5	1.1
3 (2010/11)	2,083	31.9	531	8.1	1.4	1.5	0.6	1.2
4 (2011/12)	1,563	25.9	451	7.5	0.7	1.1	0.5	1.0
5 (2012/13)	1,334	24.4	368	6.7	1.1	1.4	0.4	0.9
6 (2013/14)	982	19.6	383	7.6	0.8	1.1	0.4	0.9
7 (2014/15)	664	15.6	355	8.3	0.8	1.1	0.4	0.9

Source: pairfam waves 1–7, Release 7.0 (own calculations)
[a] SD = standard deviation

cients of standard panel fixed-effects models (see Schunck, 2013). The reported random coefficients of the three (non-hierarchical) levels allow for the computation of the interviewer and respondent variance share (i.e., intraclass correlation coefficients[10]). Each estimation includes two models for each type of item nonresponse, respectively: Models (a) include variables on interviewer change, differentiating between consecutive encounters with the second interviewer, and control for panel wave and respondent mobility. The reference point for respondents who experiences an interviewer change is the last encounter with the first interviewer. Models (b) add interviewer characteristics as well as variables indicating interviewer-respondent (dis-)similarities.

The most striking outcome of our analyses is a negative or null effect of interviewer change on the likelihood for all types of item nonresponse. However, the effect is only significant for "I don't know" responses with respect to household income (first encounter: −0.029; third encounter: −0.030). The effects do not change substantively when interviewer characteristics are accounted for (see Models b). For both types of nonresponse, the included interviewer and respondent characteristics do not explain much of the effect of interviewer change on item nonresponse. Contrary to the widespread strategy of striving for interviewer continuity in panel studies, interviewer change in fact does not seem to foster the likelihood for item nonresponse.

With respect to interviewer experience, we find that if the interviewer has completed more than the cumulative median number of interviews up to the respective interview, the likelihood of "I don't know" answers is significantly smaller for the income question (−0.016) as well as over all CAPI questions (−0.082). This might reflect a higher effectiveness of more experienced interviewers in getting valid answers. We do not find any effects of age distance to respondents on item nonresponse behavior. If we observed more changes to interviewers younger than the respondent, we might observe a different trend, but given the socio-demographic structure of the interviewer staff these changes are extremely rare. Over all CAPI questions, older interviewers get more "I don't know" answers. Female respondents tend to provide less "I don't know" answers on the income question when reassigned to a female interviewer (−0.040). In contrast, we find no significant effect of interviewer age or gender on refusal rates to either the income question or over all CAPI questions.

Over the course of the panel, we see a clear pattern of respondent learning (i.e., a decrease in item nonresponse) with respect to household income. The pattern is less linear with respect to all CAPI questions, reflecting the changing questionnaire program over the panel lifetime. The interviewer intraclass correlation coefficients (ICC) reported in Tables 3 and 4 are higher for refusals (0.22–0.24) than for "I don't know" answers (0.09–0.14). Thus, interviewers more strongly affect the probability of refusals. Respondent correlations amount to relatively high values of at least 0.3 for the question on household income. Over all CAPI questions, these respondent correlations are considerably smaller, especially for refusals. This highlights the fact that response behaviors are influenced by situational factors, which include the characteristics and behaviors of the respective interviewer. Still, the largest share of the variance is found on the level of the person-year observation (the sum of respondent and interviewer ICCs amounts to around .57 at the maximum), reflecting the existence of considerable between-wave variation.

The differentiation by respondent age group (i.e., birth cohort) yields some interesting patterns (see Tables 5 and 6; full models (b) including interviewer characteristics): The negative effects of interviewer change and of reassignment to female interviewers on the likelihood of item nonresponse found for the income question seem to mainly stem from the youngest cohort. For cohorts 2 and 3, the effects are considerably smaller or even positive, albeit not significant. Over all CAPI questions, the oldest of the three cohorts shows a

[10] We refrain from including random effects of explanatory variables in order not to overburden our models.

Table 3
Within-cluster regression analyses predicting item nonresponse for household income

	I don't know		Refusal	
Variable	Model 1a	Model 1b	Model 2a	Model 2b
(a) Fixed Effects				
Encounters with second interviewer				
Last encounter with first interviewer (ref. categ.)	-	-	-	-
1st	−0.029*	−0.029*	−0.001	−0.002
2nd	−0.020	−0.020	0.018	0.017
3rd	−0.030	−0.030	0.013	0.013
4th or more	−0.009	−0.008	0.017	0.017
Respondent geographic mobility				
No move of main residence (ref. categ.)	-	-	-	-
Respondent mobility up to 100 km	−0.090***	−0.090***	−0.009	−0.009
Respondent mobility > 100 km	−0.128***	−0.127***	−0.017	−0.017
Panel wave				
Wave 1, 2008/09 (ref. categ.)	-	-	-	-
Wave 2, 2009/10	−0.003	−0.013	−0.002	−0.003
Wave 3, 2010/11	−0.022***	−0.039	−0.019***	−0.019
Wave 4, 2011/12	−0.080***	−0.103***	−0.035***	−0.035
Wave 5, 2012/13	−0.097***	−0.125**	−0.046***	−0.045
Wave 6, 2013/14	−0.138***	−0.173***	−0.047***	−0.045
Wave 7, 2014/15	−0.167***	−0.208***	−0.043***	−0.040
Interviewer characteristics and interaction effects with respondents				
Interviewer age (in years)	-	0.006	-	−0.001
Age difference to respondent	-	−0.003	-	0.005
Male respondent, female interviewer (ref.: male-male)	-	−0.028	-	0.001
Female respondent, female interviewer (ref.: female-male)	-	−0.040*	-	−0.012
Interviewer experience	-	−0.016*	-	−0.009
(b) Random Effects				
Respondent σ^2	0.096***	0.046***	0.038***	0.038***
Interviewer σ^2	0.018***	0.016***	0.027***	0.027***
Panel observation σ^2	0.091***	0.091***	0.050***	0.050***
ICC interviewer	0.088	0.106	0.235	0.236
ICC respondent	0.468	0.299	0.329	0.329
DIC	23,671	22,675	980	979
N (respondent-years)	37,639	37,639	29,707	29,707
N (respondents)	8,847	8,847	7,935	7,935

Fixed and random effects from linear probability cross-classified multilevel models. All coefficients represent within-cluster effects (demeaned variables, i.e., cluster means subtracted). The deviance information criterion (DIC) indicates the model fit with higher values indicating a poorer fitting.
Source: pairfam waves 1-7, Release 7.0 (own calculations)
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

positive effect of interviewer change on the likelihood to answer with "I don't know" for the first two encounters with the new interviewer. If female respondents of cohort 2 are reassigned to female interviewers they even give more "I don't know" answers. However, if male respondents of cohort 3 are reassigned to female interviewers they refuse less often to answer over all CAPI questions.

Since it is possible that interviewer changes which have an obvious, "self-induced" reason have a different impact on respondents, we included interaction terms between the first encounter with a new interviewer and simultaneously occurring short-distance or long-distance moves (see models (c) in Tables A2 and A3). To test whether respondents react differently to an interviewer change if they did not participate in the previous wave, we also introduced an interaction term of first-time contact with the new interviewer and having missed the previous wave (see models (d) in Tables A2 and A3). While previous wave non-respondents have a higher

Table 4
Within-cluster regression analyses predicting the percentage of item nonresponse over all CAPI questions

	I don't know		Refusal	
Variable	Model 3a	Model 3b	Model 4a	Model 4b
(a) Fixed Effects				
Encounters with second interviewer				
Last encounter with first interviewer (ref. categ.)	-	-	-	-
1st	0.002	−0.011	−0.035	−0.039
2nd	−0.042	−0.052	−0.012	−0.017
3rd	−0.035	−0.044	−0.011	−0.018
4th or more	−0.039	−0.047	−0.003	−0.009
Respondent geographic mobility				
No move of main residence (ref. categ.)	-	-	-	-
Respondent mobility up to 100 km	−0.058**	−0.058**	−0.026	−0.026
Respondent mobility > 100 km	−0.101	−0.100	0.120**	0.120**
Panel wave				
Wave 1, 2008/09 (ref. categ.)	-	-	-	-
Wave 2, 2009/10	−0.600***	−0.725***	−0.132***	−0.111**
Wave 3, 2010/11	−0.136***	−0.329***	0.026	0.066
Wave 4, 2011/12	−0.840***	−1.102***	−0.048**	0.010
Wave 5, 2012/13	−0.433***	−0.761***	−0.148***	−0.074
Wave 6, 2013/14	−0.666***	−1.063***	−0.144***	−0.052
Wave 7, 2014/15	−0.665***	−1.128***	−0.139***	−0.029
Interviewer characteristics and interaction effects with respondents				
Interviewer age (in years)	-	0.068*	-	−0.018
Age difference to respondent	-	0.004	-	0.015
Male respondent, female interviewer (ref.: male-male)	-	−0.002	-	−0.061
Female respondent, female interviewer (ref.: female-male)	-	−0.008	-	−0.078
Interviewer experience	-	−0.082***	-	0.024
(b) Random Effects				
Respondent σ^2	0.542***	0.444***	0.235***	0.234***
Interviewer σ^2	0.238***	0.245***	0.296***	0.287***
Panel observation σ^2	1.107***	1.107***	0.775***	0.775***
ICC interviewer	0.126	0.136	0.227	0.222
ICC respondent	0.287	0.247	0.180	0.180
DIC	127,185	126,827	111,587	111,569
N (respondent-years)	41,167	41,167	41,167	41,167
N (respondents)	9,048	9,048	9,048	9,048

Fixed and random effects from linear cross-classified multilevel models. All coefficients represent within-cluster effects (demeaned variables, i.e., cluster means subtracted).The deviance information criterion (DIC) indicates the model fit, with higher values indicating a poorer fitting.
Source: pairfam waves 1-7, Release 7.0 (own calculations)
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

propensity to refuse over the course of all CAPI questions, no significant interaction effects are detectable. Neither interaction effect yielded a significant effect. With respect to our main explanatory variables of interviewer change, the logistic regression models for the income analyses confirm the results reported above (see Table A4). The results for the interviewer characteristics and their matching with the respondent characteristics show different significance patterns. The direction of these effects, however, are the same.

8 Discussion

The purpose of this study was to investigate whether a change of interviewer affects item nonresponse in panel surveys. To our knowledge, this is the first study to specifically investigate this link employing sophisticated multilevel cross-classified models showing within-respondent longitu-

Table 5
Within-cluster regression analyses predicting the percentage of item nonresponse for household income, separately for the three birth cohorts (full models 1b and 2b)

Variable	I don't know Cohort 1 (1991-93)	I don't know Cohort 2 (1981-82)	I don't know Cohort 3 (1971-73)	Refusal Cohort 1 (1991-93)	Refusal Cohort 2 (1981-82)	Refusal Cohort 3 (1971-73)
(a) Fixed Effects						
Encounters with second interviewer						
Last encounter with first interviewer (ref. categ.)	-	-	-	-	-	-
1st	−0.064**	0.010	−0.009	0.001	−0.007	−0.002
2nd	−0.050	0.031	0.004	0.014	0.026	0.014
3rd	−0.042	0.020	−0.023	0.042	0.020	−0.003
4th or more	−0.037	0.047	0.007	0.032	0.027	0.005
Respondent geographic mobility						
No move of main residence (ref. categ.)	-	-	-	-	-	-
Respondent mobility up to 100 km	−0.183***	−0.027**	0.005	−0.021*	−0.001	−0.003
Respondent mobility > 100 km	−0.162***	−0.017	−0.010	−0.037	0.026	−0.012
Panel wave						
Wave 1, 2008/09 (ref. categ.)	-	-	-	-	-	-
Wave 2, 2009/10	−0.070*	−0.041	0.040*	−0.012	0.013	−0.006
Wave 3, 2010/11	−0.128*	−0.089	0.038	−0.058	0.001	−0.009
Wave 4, 2011/12	−0.282***	−0.124	0.035	−0.077	−0.011	−0.025
Wave 5, 2012/13	−0.345***	−0.148	0.039	−0.105	−0.011	−0.029
Wave 6, 2013/14	−0.484***	−0.175	0.049	−0.118	−0.004	−0.022
Wave 7, 2014/15	−0.601***	−0.188	0.053	−0.123	0.017	−0.020
Interviewer characteristics and interaction effects with respondents						
Interviewer age (in years)	0.041	0.007	−0.008	−0.001	−0.008	−0.000
Age difference to respondent	−0.021	0.005	−0.006	0.019	−0.003	0.006
Male resp., female interv. (ref.: male-male)	−0.061	0.051	−0.043	0.022	−0.004	−0.030
Female resp., female interv. (ref.: female-male)	−0.110**	0.034	−0.012	−0.039	0.009	0.001
Interviewer experience	0.017	−0.016	0.004	0.001	−0.014	0.004
(b) Random Effects						
Respondent σ^2	0.034***	0.037***	0.023***	0.023***	0.034***	0.045***
Interviewer σ^2	0.043***	0.008***	0.007***	0.046***	0.017***	0.025***
Panel observation σ^2	0.132***	0.072***	0.049***	0.053***	0.048***	0.049***
ICC interviewer	0.205	0.070	0.089	0.377	0.170	0.210
ICC respondent	0.164	0.316	0.293	0.187	0.344	0.376
DIC	14,105	3,946	−391	788	−118	204
N (respondent-years)	14,933	10,707	11,999	6,771	10,341	12,595
N (respondents)	3,511	2,577	2,759	2,525	2,574	2,836

Fixed and random effects from linear cross-classified multilevel models. All coefficients represent within-cluster effects (demeaned variables, i.e., cluster means subtracted). The deviance information criterion (DIC) indicates the model fit, with higher values indicating a poorer fitting.
Source: pairfam waves 1-7, Release 7.0 (own calculations)
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

dinal estimates. Using data from the German Family Panel pairfam, we did not observe any detrimental overall effect of an interviewer change on item nonresponse, neither for the question regarding household income nor over all questions administered by an interviewer. Only the oldest birth cohort shows an increase in "I don't know" responses when considering the percentage over all CAPI questions. In contrast, we find a lower likelihood for "I don't know" answers in the youngest cohort with respect to the income question if the respondent is reassigned to a new interviewer. The popular assumption that a continuing interviewer leads to increased trust and less item nonresponse can therefore not be confirmed. Rather, the effect of an interviewer change on item nonresponse seems to differ depending on the respondent's age group and the questions asked. Older respondents might be more suspicious of changes and younger respon-

Table 6
Within-cluster regression analyses predicting the percentage of item nonresponse over all CAPI questions, separately for the three birth cohorts (full models 3b and 4b)

Variable	I don't know Cohort 1 (1991-93)	I don't know Cohort 2 (1981-82)	I don't know Cohort 3 (1971-73)	Refusal Cohort 1 (1991-93)	Refusal Cohort 2 (1981-82)	Refusal Cohort 3 (1971-73)
(a) Fixed Effects						
Encounters with second interviewer						
Last encounter with first interviewer (ref. categ.)	-	-	-	-	-	-
1st	−0.104	−0.036	0.159*	−0.065	−0.119*	0.041
2nd	−0.153	−0.105	0.172*	−0.080	−0.100	0.128
3rd	−0.126	−0.055	0.096	−0.014	−0.111	0.062
4th or more	−0.137	−0.031	0.051	−0.020	−0.049	0.042
Respondent geographic mobility						
No move of main residence (ref. categ.)	-	-	-	-	-	-
Respondent mobility up to 100 km	−0.145***	−0.037	0.089*	−0.051	−0.037	0.033
Respondent mobility > 100 km	−0.087	0.068	−0.187	0.119*	0.095	0.075
Panel wave (ref.: 1 (2008/09))						
Panel wave						
Wave 1, 2008/09 (ref. categ.)	-	-	-	-	-	-
Wave 2, 2009/10	−1.060***	−0.571***	−0.480***	−0.067	−0.105	−0.216**
Wave 3, 2010/11	−0.574***	−0.380∗	−0.027	0.080	0.036	−0.027
Wave 4, 2011/12	−1.693***	−0.921***	−0.610***	−0.056	−0.026	−0.035
Wave 5, 2012/13	−1.353***	−0.733**	−0.136	−0.154	−0.088	−0.166
Wave 6, 2013/14	−1.683***	−1.003**	−0.436	−0.207	−0.084	−0.092
Wave 7, 2014/15	−1.855***	−1.163**	−0.320	−0.219	−0.057	−0.078
Interviewer characteristics and interaction effects with respondents						
Interviewer age (in years)	0.121	0.099	−0.013	0.031	−0.022	−0.021
Age difference to respondent	0.017	0.012	0.007	−0.021	0.012	0.002
Male resp., female interv. (ref.: male-male)	0.055	0.038	−0.103	0.008	−0.115	−0.304**
Female resp., female interv. (ref.: female-male)	−0.147	0.244*	0.038	−0.076	−0.069	−0.186
Interviewer experience	−0.041	−0.031	0.003	−0.034	0.052	0.017
(b) Random Effects						
Respondent σ^2	0.460***	0.452***	0.320***	0.138***	0.290***	0.305***
Interviewer σ^2	0.432***	0.162***	0.106***	0.287***	0.245***	0.276***
Panel observation σ^2	1.327***	0.984***	0.891***	0.802***	0.654***	0.813***
ICC interviewer	0.195	0.102	0.080	0.234	0.206	0.198
ICC respondent	0.207	0.283	0.243	0.113	0.244	0.219
DIC	51,569	35,109	38,835	43,077	30,308	37,671
N (respondent-years)	15,795	11,788	13,584	15,795	11,788	13,584
N (respondents)	3,529	2,650	2,869	3,529	2,650	2,869

Fixed and random effects from linear cross-classified multilevel models. All coefficients represent within-cluster effects (demeaned variables, i.e., cluster means subtracted). The deviance information criterion (DIC) indicates the model fit, with higher values indicating a poorer fitting.
Source: pairfam waves 1-7, Release 7.0 (own calculations)
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

dents more open to meeting a new interviewer. Moreover, after the second encounter with a new interviewer, no significant long-term effects are visible (with one exception for the income question): Once respondents are acquainted to their new interviewer, no effects on item nonresponse seem to remain. A temporary drop-out, short- or long-distance moves do not seem to alter the effects of an interviewer change.

More experienced interviewers tend to elicit less "I don't know" answers. We do not find an effect of age difference between interviewer and respondent, but admit that this might be due to the fact that there is a large age disparity between interviewers and respondents in the pairfam study, and that this difference might not change significantly when an interviewer change occurs. Our analyses do show that the overall percentage of "I don't know" answers rises if respondents are reassigned to older interviewers. The percentage of "I don't know" answers on the income question decreases if female respondents of the youngest cohort are reassigned to a female interviewer. In contrast, over all CAPI questions female respondents of cohort 2 give more "I don't know" answers if interviewed by women. In addition, male respondents of the oldest cohort tend to refuse less often to answer if reassigned to female interviewers. Thus, effects of interviewer gender do not seem to be a matter of interviewer-respondent matching, i.e., perceived closeness. It rather seems to depend on the age group and the questions asked, which might explain the ambiguous results reported in other research concerning interviewer characteristics. The effect might be different for specific questions and questionnaires; therefore, instead of analyzing one specific question (as is often the case when focusing on items regarding personal finances), it is important to also look at the overall percentage of item nonresponse in a survey in order to evaluate the total impact of interviewer (dis-)continuity.

While core questions are repeated annually, the question battery changes somewhat over time in the pairfam panel. We tried to account for this by controlling for panel wave in our regression models. The alternative would have been to completely focus on selected questions only, but our aim was to gain insight into the impact of interviewer discontinuity for surveys on the whole. By doing so, we present results that are new in the field, as previous research has thus far primarily focused on unit/wave nonresponse only, or has been limited to a specific subset of questions, in particular income and wealth.

Previous findings have shown that interviewer continuity seems to be beneficial for reducing wave nonresponse, i.e., panel attrition (see literature review above). Panel respondents who are contacted by a new interviewer and still participate in the survey might be a special, more intrinsically motivated population. Therefore, we might not find a detrimental effect of interviewer change on item nonresponse as respondents who do not want to be interviewed by a new interviewer do not participate in the survey at all. Further research is needed to verify this hypothesis. It is possible that selective respondent and interviewer attrition throughout the panel are the major disadvantages of the data at hand. As shown in the descriptive analyses, item nonresponse is highest in the first wave and decreases with time, which might be due to learning processes or selective respondent attrition.

The only interviewer-based information included in our analyses were age, gender, and the number of interviews they had thus far conducted for the pairfam study. Interviewer behavior as well as his / her motivation and personality traits can either reduce or induce item nonresponse, and explaining a question or asking it again (i.e., probing the respondent) can lead to a higher probability of a valid answer. In contrast, failing to probe, not asking a question at all, or purposely noting a missing answer without reading the question aloud (e.g., to shorten interview time) lead to higher rates of missing data (de Leeuw, 2001). Thus, these interviewer characteristics could account for the effects of interviewer change if respondents are reassigned to interviewers of higher or lower skills or motivation.

Our analyses assume that a new interviewer does not differ in any such characteristics from the previous interviewer. However, if there is selective, non-random attrition among the interviewers resulting in, for example, less qualified interviewers dropping out of the interviewer staff, interviewer change might then foster item response rates. Non-random interviewer attrition could cause a reassignment of respondents to more efficient interviewers, leading to a lower likelihood of item nonresponse. Pickery and Loosveldt (1998) and Pickery et al. (2001) find considerable interviewer variances in their analyses on item nonresponse, but these effects cannot be explained by interviewer age, sex, education, or experience. Our results also suggest that these characteristics do not fully explain interviewer variance. Thus, including easily-measured interviewer characteristics might not be sufficient to fully account for interviewer effects on nonresponse. Certainly, further interviewer characteristics influencing the propensity for item nonresponse are integral to our understanding. Unfortunately, no further interviewer characteristics are available for examination in the pairfam panel survey (e.g., interviewer's experience in total, persuasiveness, charisma, etc.), which leads us to suggest that future studies investigate these further. The interviewer intraclass correlation coefficients show that a considerable proportion of variation in our outcomes is due to differences across interviewers, especially when generating refusals. Thus, unobserved influences operating at this level indeed affect response outcomes and need to be accounted for.

What lessons can be learned with regard to design and practical implementation of future panel surveys and interviewer allocation strategies? Our study shows no detrimental effect of an interviewer change per se on item nonresponse. In fact, contrary to conventional wisdom, reassigning interviewers might reduce the propensity of item nonresponse for some age groups. This certainly depends on the characteristics of the interviewer staff, but the finding is reassuring, as no large-scale survey can completely prevent interviewer attrition or reassignment. However, as our analysis are based on non-experimental data and a non-random assignment of interviewers, further replications and experimental approaches are needed to confirm these results.

Acknowledgements

This paper uses data from the German Family Panel pairfam, coordinated by Josef Brüderl, Sonja Drobnič, Karsten Hank, Bernhard Nauck, Franz Neyer, and Sabine Walper.

pairfam is funded as long-term project by the German Research Foundation (DFG).

References

Allison, P. D. (2009). *Fixed effects regression models*. Quantitative Applications in the Social Sciences. Los Angeles: SAGE.

Bilgen, I. (2011). *Is less more & more less . . . ? the effect of two types of interviewer experience on "don't know" responses in calendar and standardized interviews.* Doctor of Philosophy Dissertation, University of Nebraska, Lincoln.

Browne, W. J. (2012). *MCMC estimation in MLwiN, v2.26.* Centre for Multilevel Modelling, University of Bristol.

Brüderl, J., Hank, K., Huinink, J., Nauck, B., Neyer, F. J., Walper, S., . . . Wilhelm, B. (2016). *The German Family Panel (pairfam)*. ZA5678 Data file Version 7.0.0.

Brüderl, J. & Ludwig, V. (2014). Fixed-effects panel regression. In H. Best & C. Wolf (Eds.), *Sage handbook of regression analysis and causal inference* (pp. 327–356). London: Sage.

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.

Cafri, G. & Fan, J. (2018). Between-within effects in survival models with cross-classified clustering: Application to the evaluation of the effectiveness of medical devices. *Statistical Methods in Medical Research*, *27*(1), 312–319.

Campanelli, P. & O'Muircheartaigh, C. (1999). Interviewers, interviewer continuity, and panel survey nonresponse. *Quality and Quantity*, *33*(1), 59–76.

Campanelli, P. & O'Muircheartaigh, C. (2002). The importance of experimental control in testing the impact of interviewer continuity on panel survey nonresponse. *Quality and Quantity*, *36*(2), 129–144.

Chadi, A. (2013). The role of interviewer encounters in panel responses on life satisfaction. *Economics Letters*, *121*(3), 550–554. doi:10.1016/j.econlet.2013.10.024

de Leeuw, E. D. (2001). Reducing missing data in surveys: An overview of methods. *Quality and Quantity*, *35*(2), 147–160.

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.

Elwert, F. & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology*, *40*, 31–53.

Essig, L. & Winter, J. (2009). Item nonresponse to financial questions in household surveys: An experimental study of interviewer and mode effects. *Fiscal Studies*, *30*(3–4), 367–390.

Groves, R. M. (1989). *Survey errors and survey costs.* New York: Wiley.

Groves, R. M. & Couper, M. P. (1998). *Nonresponse in household interview surveys.* New York: Wiley.

Groves, R. M., Fowler, F. J. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology.* Hoboken, New Jersey: John Wiley & Sons Inc.

Homans, G. C. (1961). *Social behaviour: Its elementary forms.* London: Routledge & Kegan Paul.

Huinink, J., Brüderl, J., Nauck, B., Walper, S., Castiglioni, L., & Feldhaus, M. (2011). Panel Analysis of Intimate Relationships and Family Dynamics (pairfam): Conceptual framework and design. *Zeitschrift für Familienforschung – Journal of Family Research*, *23*(1), 77–101.

Jäckle, A., Lynn, P., Sinibaldi, J., & Tipping, S. (2013). The effect of interviewer experience, attitudes, personality and skills on respondent co-operation with face-to-face surveys. *Survey Research Methods*, *7*(1), 1–15.

Korbmacher, J. M. & Schroeder, M. (2013). Consent when linking survey data with administrative records: The role of the interviewer. *Survey Research Methods*, *7*(2), 115–131.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*(3), 213–236.

Leckie, G. & Charlton, C. (2013). Runmlwin – a program to run the MLwiN multilevel modelling software from within stata. *Journal of Statistical Software*, *52*(11), 1–40.

Lipps, O. (2007). Interviewer and respondent survey quality effects in a CATI panel. *Bulletin de méthodologie sociologique (Bulletin of sociological methodology)*, *95*, 5–25.

Lipps, O. & Pollien, A. (2011). Effects of interviewer experience on components of nonresponse in the European Social Survey. *Field Methods*, *23*(2), 156–172.

Lynn, P., Kaminska, O., & Goldstein, H. (2011). *Panel attrition: How important is it to keep the same interviewer?* ISER Working Paper Series 2011-02.

Müller, B. & Castiglioni, L. (2015). Attrition im Beziehungs- und Familienpanel pairfam. In J. Schupp & C. Wolf (Eds.), *Nonresponse Bias: Qualitätssicherung sozialwissenschaftlicher Umfragen* (pp. 283–308). Wiesbaden: Springer VS.

Pickery, J. & Loosveldt, G. (1998). The impact of respondent and interviewer characteristics on the number of "no opinion" answers. *Quality and Quantity*, *32*(1), 31–45.

Pickery, J. & Loosveldt, G. (2002). A multilevel multinomial analysis of interviewer effects on various components of unit nonresponse. *Quality and Quantity*, *36*(4), 427–437.

Pickery, J., Loosveldt, G., & Carton, A. (2001). The effects of interviewer and respondent characteristics on response behavior in panel surveys: A multilevel approach. *Sociological Methods & Research*, *29*(4), 509–523.

Rasbash, J. C., Browne, W., Healy, M., & Cameron, B. (2009). *MLwiN version 2.1.* Centre for Multilevel Modelling, University of Bristol.

Riphahn, R. T. & Serfling, O. (2005). Item non-response on income and wealth questions. *Empirical Economics*, *30*(2), 521–538.

Rodriguez, L. A., Sana, M., & Sisk, B. (2015). Self-administered questions and interviewer-respondent familiarity. *Field Methods*, *27*(2), 163–181.

Schräpler, J.-P. (2006). Explaining income nonresponse – a case study by means of the british household panel study (BHPS). *Quality and Quantity*, *40*(6), 1013–1036.

Schunck, R. (2013). Within and between estimates in random-effects models: Advantages and drawbacks of correlated random effects and hybrid models. *The Stata Journal*, *13*(1), 65–76.

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers.* The Application of Cognitive Processes to Survey Methodology. San Francisco: Jossey-Bass.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response.* Cambridge: Cambridge University Press.

Tu, S.-H. & Liao, P.-S. (2007). Social distance, respondent cooperation and item nonresponse in sex survey. *Quality and Quantity*, *41*(2), 177–199.

Watson, N. & Wooden, M. (2009). Identifying factors affecting longitudinal survey response. In P. Lynn (Ed.), *Methodology of longitudinal surveys* (John Wiley and Sons). Chichester.

Watson, N. & Wooden, M. (2014). Re-engaging with survey non-respondents: Evidence from three household panels. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *177*(2), 499–522.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data.* Cambridge, MA: MIT Press.

Young, R. (2012). *Don't know responses in survey research.* Doctor of Philosophy Dissertation, The Pennsylvania State University.

Appendix
Figures and tables

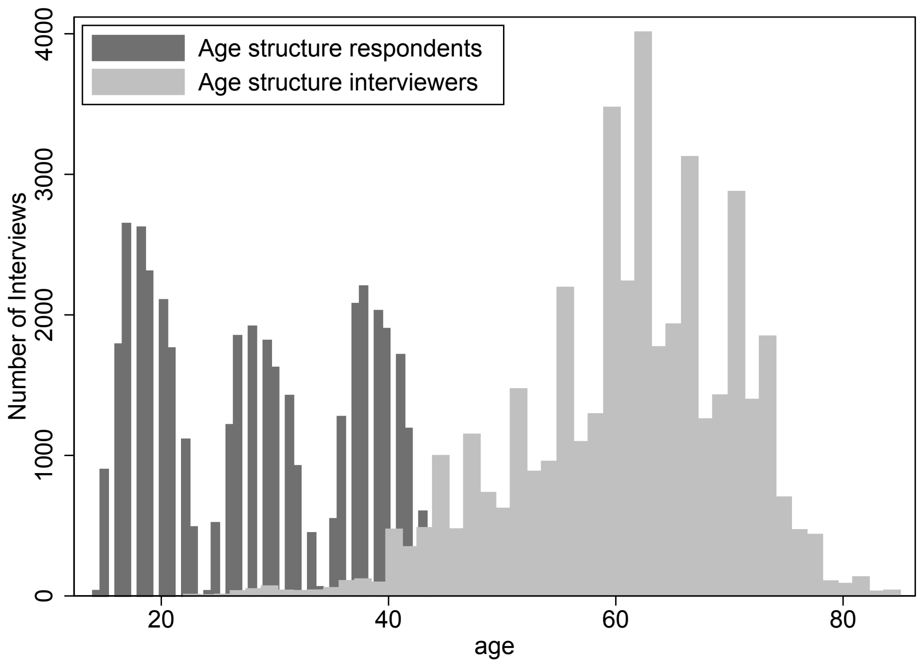

Figure A1. Age structure of respondents and interviewers by number of interviews conducted.
Source: pairfam waves 1-7, Release 7.0 (own calculations)

Table A1
Item nonresponse rates per cohort

	Income: Don't know		Income: Refusals		CAPI: Don't know		CAPI: Refusals	
Cohort (Birth year)	N	%	N	% mean	%	SD	mean %	SD
Cohort 1 (1991–1993)	9,024	57.1	862	5.5	1.5	1.5	0.5	1.1
Cohort 2 (1981–1983)	1,447	12.3	1,081	9.2	0.9	1.3	0.4	1.0
Cohort 3 (1971–1973)	989	7.3	1,585	11.7	0.8	1.2	0.5	1.1

SD = standard deviation
Source: pairfam waves 1-7, Release 7.0 (own calculations)

Table A2

Within-cluster regression analyses predicting item nonresponse for household income including interaction effects with geographic mobility and previous wave non-response

	I don't know		Refusal	
Variable	Model 1c	Model 1d	Model 2c	Model 2d
(a) Fixed Effects				
Encounters with second interviewer				
Last encounter with first interviewer (ref. categ.)	-	-	-	-
1st	−0.037**	−0.025	−0.004	−0.002
2nd	−0.021	−0.019	0.017	0.017
3rd	−0.032*	−0.030	0.013	0.013
4th or more	−0.011	−0.008	0.017	0.017
Interaction effects				
First encounter with second interviewer × respondent mobility up to 100 km	0.033	-	−0.001	-
First encounter with second interviewer × respondent mobility > 100 km	0.025	-	0.025	-
First encounter with second interviewer × previous wave non-response	-	−0.039	-	−0.003
Previous wave nonresponse	-	0.019	-	−0.004
Respondent geographic mobility				
No move of main residence (ref. categ.)	-	-	-	-
Respondent mobility up to 100 km	−0.093***	−0.091***	−0.009	−0.009
Respondent mobility > 100 km	−0.135***	−0.126***	−0.029	−0.017
Panel wave				
Wave 1, 2008/09 (ref. categ.)	-	-	-	-
Wave 2, 2009/10	−0.012	−0.013	−0.003	−0.003
Wave 3, 2010/11	−0.038	−0.039	−0.019	−0.019
Wave 4, 2011/12	−0.101**	−0.102**	−0.034	−0.035
Wave 5, 2012/13	−0.123**	−0.124**	−0.044	−0.044
Wave 6, 2013/14	−0.171***	−0.173***	−0.043	−0.044
Wave 7, 2014/15	−0.206***	−0.207***	−0.038	−0.040
Interviewer characteristics and interaction effects with respondents				
Interviewer age (in years)	0.006	0.006	−0.001	−0.001
Age difference to respondent	−0.003	−0.003	0.005	0.005
Male respondent, female interviewer (ref.: male-male)	−0.028	−0.026	−0.001	0.001
Female respondent, female interviewer (ref.: female-male)	−0.041*	−0.038*	−0.014	−0.012
Interviewer experience	−0.016*	−0.016*	−0.009	−0.009
(b) Random Effects				
Respondent σ^2	0.046***	0.046***	0.038***	0.038***
Interviewer σ^2	0.016***	0.016***	0.027***	0.027***
Panel observation σ^2	0.091***	0.091***	0.050***	0.050***
ICC interviewer	0.106	0.106	0.233	0.234
ICC respondent	0.299	0.299	0.330	0.330
DIC	22,678	22,678	982	979
N (respondent-years)	37,639	37,639	29,707	29,707
N (respondents)	8,847	8,847	7,935	7,935

Fixed and random effects from linear probability cross-classified multilevel models. All coefficients represent within-cluster effects (demeaned variables, i.e., cluster means subtracted). The deviance information criterion (DIC) indicates the model fit with higher values indicating a poorer fitting.

Source: pairfam waves 1-7, Release 7.0 (own calculations)

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

Table A3
Within-cluster regression analyses predicting the percentage of item nonresponse over all CAPI questions including interaction effects with geographic mobility and previous wave non-response

	I don't know		Refusal	
Variable	Model 3c	Model 3d	Model 4c	Model 4d
(a) Fixed Effects				
Encounters with second interviewer				
Last encounter with first interviewer (ref. categ.)	-	-	-	-
1st	−0.001	−0.020	−0.032	−0.038
2nd	−0.048	−0.052	−0.016	−0.014
3rd	−0.039	−0.043	−0.016	−0.014
4th or more	−0.042	−0.047	−0.008	−0.005
Interaction effects				
First encounter with second interviewer × respondent mobility up to 100 km	−0.003	-	−0.020	-
First encounter with second interviewer × respondent mobility > 100 km	−0.120	-	−0.040	-
First encounter with second interviewer × previous wave non-response	-	0.023	-	−0.072
Previous wave nonresponse	-	0.052	-	0.109**
Respondent geographic mobility				
No move of main residence (ref. categ.)	-	-	-	-
Respondent mobility up to 100 km	−0.058**	−0.061**	−0.025	−0.031
Respondent mobility > 100 km	−0.052	−0.105	0.132*	0.109*
Panel wave				
Wave 1, 2008/09 (ref. categ.)	-	-	-	-
Wave 2, 2009/10	−0.726***	−0.725***	−0.111**	−0.113**
Wave 3, 2010/11	−0.331***	−0.334***	0.065	0.057
Wave 4, 2011/12	−1.105***	−1.107***	0.008	0.001
Wave 5, 2012/13	−0.764***	−0.766***	−0.076	−0.085
Wave 6, 2013/14	−1.067***	−1.068***	−0.054	−0.065
Wave 7, 2014/15	−1.133***	−1.134***	−0.031	−0.042
Interviewer characteristics and interaction effects with respondents				
Interviewer age (in years)	0.069*	0.069*	−0.018	−0.017
Age difference to respondent	0.003	0.004	0.015	0.014
Male respondent, female interviewer (ref.: male-male)	−0.005	−0.003	−0.063	−0.063
Female respondent, female interviewer (ref.: female-male)	−0.013	−0.010	−0.081	−0.080
Interviewer experience	−0.083***	−0.082***	0.023	0.025
(b) Random Effects				
Respondent σ^2	0.443***	0.442***	0.234***	0.233***
Interviewer σ^2	0.242***	0.233***	0.287***	0.278***
Panel observation σ^2	1.107***	1.108***	0.775***	0.774***
ICC interviewer	0.135	0.131	0.221	0.217
ICC respondent	0.247	0.248	0.181	0.181
DIC	126,835	126,833	111,574	111,547
N (respondent-years)	41,167	41,167	41,167	41,167
N (respondents)	9,048	9,048	9,048	9,048

Fixed and random effects from linear cross-classified multilevel models. All coefficients represent within-cluster effects (demeaned variables, i.e., cluster means subtracted). The deviance information criterion (DIC) indicates the model fit, with higher values indicating a poorer fitting.
Source: pairfam waves 1-7, Release 7.0 (own calculations)
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table A4
Within-cluster regression analyses predicting item nonresponse for household income; logistic model

	I don't know		Refusal	
Variable	Model 1a	Model 1b	Model 2a	Model 2b
(a) Fixed Effects				
Encounters with second interviewer				
Last encounter with first interviewer (ref. categ.)	-	-	-	-
1st	−0.366*	−0.280*	0.010	0.066
2nd	−0.210	−0.124	0.405	0.476
3rd	−0.332	−0.250	0.286	0.362
4th or more	−0.107	−0.033	0.349	0.431
Respondent geographic mobility				
No move of main residence (ref. categ.)	-	-	-	-
Respondent mobility up to 100 km	−0.967***	−0.923***	−0.181	−0.185
Respondent mobility > 100 km	−1.058***	−0.983***	−0.367	−0.312
Panel wave				
Wave 1, 2008/09 (ref. categ.)	-	-	-	-
Wave 2, 2009/10	−0.014	−0.183**	−0.017	−0.069
Wave 3, 2010/11	−0.218**	−0.490***	−0.336**	−0.397***
Wave 4, 2011/12	−0.857***	−1.215***	−0.681***	−0.750***
Wave 5, 2012/13	−1.034***	−1.464***	−0.931***	−1.005***
Wave 6, 2013/14	−1.535***	−2.075***	−0.935***	−1.018***
Wave 7, 2014/15	−1.965***	−2.578***	−0.839***	−0.931***
Interviewer characteristics and interaction effects with respondents				
Interviewer age (in years)	-	0.093***	-	0.005
Age difference to respondent	-	−0.000	-	0.027*
Male respondent, female interviewer (ref: male-male)	-	−0.240	-	0.075
Female respondent, female interviewer (ref.: female-male)	-	−0.398*	-	−0.095
Interviewer experience	-	−0.040	-	−0.202
(b) Random Effects				
Respondent σ^2	9.394***	3.304***	6.535***	6.493***
Interviewer σ^2	2.148***	1.781***	4.502***	4.390***
DIC	26,288	25,338	12,662	12,660
N (respondent-years)	37,639	37,639	29,707	29,707
N (respondents)	8,847	8,847	7,935	7,935

Fixed and random effects from logistic cross-classified multilevel models. All coefficients represent within-cluster effects (demeaned variables, i.e., cluster means subtracted). No ICCs computed since random effects of level 1 (panel observation) are a function of the mean, which depends on the values of the explanatory variables. The deviance information criterion (DIC) indicates the model fit with higher values indicating a poorer fitting.
Source: pairfam waves 1-7, Release 7.0 (own calculations)
$^{*}\ p < 0.05$ $^{**}\ p < 0.01$ $^{***}\ p < 0.001$