

Can Interviewer Evaluations Predict Short-Term and Long-Term Participation in Telephone Panels?

Oliver Lipps¹ and Marieke Voorpostel¹

Interviewers often assess after the interview the respondent's ability and reluctance to participate. Prior research has shown that this evaluation is associated with next-wave response behavior in face-to-face surveys. Our study adds to this research by looking at this association in telephone surveys, where an interviewer typically has less information on which to base an assessment. We looked at next-wave participation, non-contact and refusal, as well as longer-term participation patterns. We found that interviewers were better able to anticipate refusal than non-contact relative to participation, especially in the next wave, but also in the longer term. Our findings confirm that interviewer evaluations – in particular of the respondent's reluctance to participate – can help predict response at later waves, also after controlling for commonly used predictors of survey nonresponse. In addition to helping to predict nonresponse in the short term, interviewer evaluations provide useful information for a long-term perspective as well, which may be used to improve nonresponse adjustment and in responsive designs in longitudinal surveys.

Key words: Attrition; postsurvey adjustment; longitudinal weights.

1. Introduction

It is of central importance in longitudinal surveys that respondents participate repeatedly. Only by collecting multiple observations can we assess change over time accurately. Hence, a large body of literature has examined the causes of attrition, as well as strategies to prevent and correct for it (e.g., Couper and Ofstedal 2009; Lipps 2012a; Schonlau et al. 2010). One strategy to improve our understanding of what guides continued participation that has received only little research attention is the use of assessments that interviewers provide after completing an interview. Such assessments consist of questions at the end of the questionnaire that address the interviewer's impression of the respondent's willingness to participate and the quality of the responses. The few studies that use interviewer evaluations to predict later-wave response are mostly based on data from face-to-face interviews (but see Barrett et al. 2006, for a combined telephone and face-to-face approach). They indicate that negative assessments of the interviews and the respondents are associated with subsequent nonresponse in face-to-face surveys (e.g., Plewis et al. 2017).

The extent to which interviewers' evaluations predict subsequent participation in longitudinal household studies conducted by telephone remains unknown. One could

¹ FORS c/o University of Lausanne, 1015 Lausanne, Switzerland. Emails: oliver.lipps@fors.unil.ch and marieke.voorpostel@fors.unil.ch

argue that in the absence of face-to-face contact, telephone interviewers have less information on which to base their evaluations. On the other hand, face-to-face contact with the respondent may also increase the use of stereotypes, decreasing the validity of such evaluations. Our study is the first to assess the relationship between interviewer evaluations and subsequent participation in a longitudinal household panel conducted by telephone.

Interviewer evaluations are useful not only to improve our understanding of what determines (repeated) survey participation, but also to correct for nonresponse. Variables based on the interviewers' assessments of the responsiveness and carefulness of the respondents are rarely included in longitudinal weights, although they have the potential to improve nonresponse adjustments (Peytchev and Olson 2007). In addition, with increasing possibilities for responsive and adaptive design procedures (Groves and Heeringa 2006; Chun et al. 2017; Schouten et al. 2017; Wagner 2008), being able to determine which respondents are more likely to drop out has great potential benefits in terms of spending more resources on these respondents and better tailoring fieldwork efforts to keep them in the longitudinal study. Consequently, we distinguished between nonresponse due to non-contact and due to refusal when looking at next-wave participation, which is an informative distinction for responsive design procedures.

Another contribution this study makes is that we extended our focus from only looking at short-term (wave-to-wave) participation to also including longer-term participation *patterns*. Attrition studies so far have exclusively examined the predictive quality of interviewer evaluations on response at the next wave (Kalton et al. 1990; Plewis et al. 2017; Watson and Wooden 2004). Yet, as longitudinal surveys increasingly include online questionnaires in their designs, interviewer evaluations collected at every wave become less standard. It then becomes useful to have insight in the extent to which interviewer evaluations are able to predict subsequent response patterns, on top of other commonly used predictors.

2. Background

There is a long history of collecting interviewer evaluations and observations in survey research (Feldman et al. 1951). These assessments generally aim to capture two aspects: the interviewer's assessment of the current or future *reluctance* of the respondent to participate (responsiveness), and the interviewer's assessment of the *ability* of the respondent to complete the survey task, which includes the quality of the responses (carefulness).

Interviewer evaluations of the respondent's carefulness help to assess the quality of the responses provided. Barrett and colleagues (Barrett et al. 2006) found that interviewers' perceptions of respondents' performance were valid indicators of item nonresponse and frequency of "don't know" answers in a survey of persons with mental and physical disabilities. However, Kirchner et al. (2017) pointed out that this association is not surprising, as interviewers base their evaluations among other things on item nonresponse during the interview. Kaminska et al. (2010) used interviewer evaluations to measure respondents' reluctance and cognitive ability in the European Social Survey. They showed that reluctance was associated with satisficing, but that this relationship was explained by lower cognitive ability. Peytchev and Olson (2007) demonstrated for the US National

Election Study that inclusion of interviewer assessments improved nonresponse adjustments.

A second strength of interviewer evaluations of both responsiveness and carefulness is that they help to predict (continued) survey participation. In cross-sectional studies, interviewer evaluations have been used to predict final participation after each contact (Eckman et al. 2013). The aim of such contact-based evaluations is an ad hoc tailoring to accommodate sample members' possible concerns, which may change across contacts. A number of studies using longitudinal surveys have shown that interviewers' evaluations of respondents are predictive of cooperation at the next wave of data collection, in addition to commonly used predictors. A recent study by Plewis et al. (2017) used interviewer evaluations at the fourth wave of the UK Millenium Cohort Study to predict response behavior in the subsequent wave, which took place four years later. They found that the interviewer's assessment of the likelihood that the respondent would participate in the future, the difficulty the respondent had answering the questions, the enjoyment of and cooperation during the interview all predicted both non-contact and refusal in the next wave. Moreover, they found that dropout of such "difficult" respondents caused bias, suggesting that in order to reduce nonresponse bias, there is benefit to directing additional resources toward keeping such respondents in the panel (Plewis et al. 2017). Kalton et al. (1990) found for the American Changing Lives study that interviewer ratings of the respondent's understanding of the questions, cooperation and the enjoyment of the interview were positively associated with participation in the second wave. Lepkowski and Couper (2002) confirmed this for the National Election Studies. In their study examining sample attrition between the first two waves of the Household, Income and Labour Dynamics in Australia (HILDA), Watson and Wooden (2004) found that the interviewer's evaluation of the respondent's cooperativeness, suspicion of the study, and required assistance to complete the interview were all associated with attrition in the second wave. In a study using 14 waves of the HILDA panel, Perez and Baffour (2018) showed that the interviewer's evaluation of the respondent's suspicion of the study, question understanding, and the respondent's cooperativeness were all associated with personal characteristics that were precursors of panel attrition.

To our knowledge, only two studies, both conducted using the British Household Panel Survey (BHPS), assessed the extent to which interviewers' evaluations are predictive of participation in the longer term. Uhrig (2008) showed that the interviewer's judgement of poor cooperation of the respondent affected both later wave refusal and non-contact, although his models only controlled for interview characteristics and fieldwork operations, and not for any other predictors of nonresponse. Laurie et al. (1999) showed that poor cooperation by the respondent as reported by the interviewer in the first wave was associated with nonresponse at the fourth wave. These findings indicate that interviewer evaluations have predictive power that goes beyond participation in the next wave.

Research findings so far have shown that interviewers' evaluations in face-to-face surveys help predict later response behavior in longitudinal studies, even if commonly used predictors are taken into account. This means that if interviewers' evaluations correlate with the survey variable, they have the potential to improve nonresponse adjustments and need to be included in appropriate models. We examined whether we find this association also in telephone panel surveys. Compared with face-to-face surveys, interviewers in

telephone surveys have less information on which to base their evaluations. There is a higher level of anonymity in telephone surveys (Block and Erskine 2012). Telephone interviewers are not in the home of the respondent, cannot engage in nonverbal communication and their interviews tend to be of a shorter duration, limiting the total time of interaction with the respondent. For our study we might expect, on the one hand, that as a result it would be harder for interviewers to evaluate respondents, limiting predictive power of such evaluations. On the other hand, there may also be less noise in the evaluations, as there is less information (e.g., on the home or physical appearance of the respondent) that may lead to the development of stereotypes. Kirchner et al. (2017) assessed the validity of interviewer evaluations in telephone surveys. They based their expectations on the continuum model of impression formation, which suggests that impressions initially are formed based on observed characteristics (stereotyping), but that one can move beyond this way of processing by using actual behavior to update these pre-existing notions. They found no evidence of stereotyping. Rather, interviewers based their assessments on the quality of the data provided and on other behavior of the respondent. In general confirming the validity of assessments that interviewers are able to make in telephone interviews.

In this study, we first explored the bivariate relationship between interviewer evaluations and participation *patterns* to see if there was indeed a longer-term association. Then we controlled for a rich set of predictors of nonresponse to see if these relationships persist. Next, we analyzed interviewer evaluation predictive power on participation in the next wave in the same way. We expected the interviewer to be better able to predict participation at the next wave than in the long term. Nevertheless, we expected the interviewer evaluation to have a nonzero longer-term predictive power. In addition, we expected that interviewers are better at anticipating subsequent refusal than non-contact (Uhrig 2008; Plewis et al. 2017; Lipps 2012a).

We assessed the contribution that interviewer evaluations make above other commonly used predictors. If, after controlling for such predictors, interviewer evaluations do not have any explanatory power, there is no need to collect such data for the purpose of predicting participation. In line with most previous studies, we took into account commonly used socio-demographic characteristics, as well as the likelihood that the household would move. These covariates are commonly included in studies linking interviewer evaluations to longitudinal survey participation. Our study controlled for a number of additional variables not included in previous studies that also capture the respondent's survey language competence, engagement, characteristics of the participation history (i.e., number of waves in the panel) and indicators of response quality (i.e., proportion of "don't know" answers). Finally, we also added measures of the social engagement of the respondent (Groves et al. 2009).

3. Data and Methods

3.1. Data

We used data from the Swiss Household Panel (SHP) (Tillmann et al. 2016). The SHP is an ongoing large-scale, nationwide, annual, centralized computer-assisted telephone interview (CATI) panel survey that started in 1999 with a sample of 5,074 households and

added a refreshment sample in 2004, consisting of 2,538 households. Both samples were randomly drawn from the telephone register and cover the Swiss residential population. A second refreshment sample, drawn at random from the Swiss population register, started in 2013, consisting of 3,989 households. Households were contacted via landline and mobile numbers. Households that split up or moved remain in the study as long as they reside in Switzerland. New household members are included in the study. In addition, households that refused in a certain wave are re-approached for a number of waves in an effort to keep them in the sample. Each year, the household reference person is asked to first complete the household roster using a grid questionnaire. All listed household members of at least 14 years old are then approached to complete individual questionnaires. As household members could only participate if the household reference person participated, we focused only on participation at the household level, measured by whether the household reference person completed the grid questionnaire. Interviewer evaluations are available for all individual interviews in every wave since 2004. Upon completion of each individual questionnaire, the interviewer answers a number of questions on the impression the interviewer had of the respondent and the interview.

We have two different analytical samples (for details, see Subsection 3.5, Methods): a sample to investigate the predictive power of the interviewer evaluations for long-term participation patterns (Sample 1), and a sample to investigate the predictive power of the interviewer evaluations for wave-to-wave response (Sample 2). For the long-term model based on Sample 1, we used data on subsequent participation from all household reference persons who completed an individual questionnaire in 2004 ($N = 4,394$). For the short-term model (Sample 2), we use all households for which we have at least one interviewer evaluation, which implies that the reference person completed an individual questionnaire at least once between 2004 and 2016. We then looked at participation in the next wave, including data from 2005 to 2017. We disregarded observations in which a household became ineligible (left the country, all members institutionalized or deceased) or was no longer approached for other reasons (no valid landline or mobile telephone number available, written refusal, not contactable or refusal for several waves in a row) and were left with 10,336 households and 61,844 observations. After dropping observations with any missing values on one of the covariates, 60,298 observations (from 10,185 households) remained in our short-term analytical Sample (2). The distributions of the participation patterns and the interviewer evaluations differed only slightly between the full sample and the analytical sample (where observations with any missing covariates were dropped). For example, while the next wave grid is completed by 92.9% of the households in the full sample, this is the case for 93.1% in the analytical sample. For the interviewer evaluation variables, the biggest relative difference occurs for 'easy to convince the sample member', which is true of 96.3% in the full sample and 96.6% in the analytical sample. We thus abstain from imputing missing covariates.

3.2. *Dependent Variables*

We used household-level information on participation to construct the categories of our dependent variables. Participation in the survey implied that the household reference person completed the grid questionnaire in a given wave. Non-participation was either

the result of non-contact or refusal at the household level. The distinction between these two states is often not clear-cut, as noncontact can be a form of hidden refusal (Stoop 2005). We coded the outcome in a given wave as refusal when the final disposition code was refusal. Whether or not to assign to refusal or noncontact is less clear when an interviewer managed to make contact with the household, the household reference person did not give a hard refusal, but the interview had not taken place by the end of the six-month fieldwork period. We decided to code these cases as a final non-contact (about half of all final non-contacts) because the contact with the household could have been with another household member. In addition, the outcome was coded as final non-contact if the household had a valid telephone number (landline or mobile number), but the interviewer could not establish any contact with the household. We cannot completely rule out the possibility that some non-contacts were hidden refusals, but these cases should be few. In addition, there is no information about the mode of contact (landline or mobile number). In the analytical sample for the short-term response (Sample 2), of the 6.9% nonresponse, 5.0% were refusals and 1.9% non-contacts, meaning that non-contacts were relatively rare.

After assigning participation statuses to the observations, we constructed two dependent variables for the two analyses in this study: for the long-term model (Sample 1), we used participation patterns. For each household reference person who completed an individual questionnaire in 2004, we established patterns of participation (grid completion) in subsequent waves, hence one pattern for each reference person. We distinguished between five exhaustive and mutually exclusive participation patterns that distinguish between full participation, immediate dropout and three different irregular patterns. Our choice of patterns acknowledges the fact that the determinants of participation and therefore the process of attrition differs across respondents. Some respondents participate mostly loyally, others drop out definitely sooner or later, and still others participate infrequently (see Lugtig 2014):

1. The household is either highly committed to the survey or developed a habit of taking part (Lugtig 2014) and participated at every wave in which it was eligible (full participation, 53.9%),
2. The household refused at least once, for example due to a temporary “shock” caused by a life-changing event (Lemay 2009), but participated at the most recent wave in which it was eligible (there may have been non-contacts at other waves) (refusal and re-entry, 9.1%),
3. The household never refused, but at least once no contact could be established, probably caused by a temporary absence (see Lepkowski and Couper 2002, who show that refusal and non-contact have different determinants), and the household participated in the most recent wave in which it was eligible (non-contact and re-entry, 2.6%),
4. The household participated at least in one later wave, but dropped out in or before the most recent wave in which it was eligible, which may be the result of panel fatigue (Lemay 2009) (participation and dropout, 30.1%),
5. The household did not participate in any later wave, probably because it experienced participation negatively (Lemay 2009) (immediate dropout, 4.3%).

For the short-term model (Sample 2), we used next-wave participation status as the dependent variable, distinguishing participation, refusal, and non-contact. We accounted for the multilevel structure of the data with observations nested in households.

3.3. Independent Variables

The main independent variables concerned the interviewer evaluation of the respondent's *ability* and *reluctance*, which we will use to predict next-wave response and response patterns. Ability was measured with the question "Was the respondent's understanding of the questions. . .?" with responses including good (2), fair (1), and poor (0). Three questions measured reluctance: "In general, what was the respondent's attitude toward the interview?" (Friendly and cooperative (3), cooperative but not particularly interested (2), impatient and restless (1), and hostile (0)); "How difficult was this case to get?" (Somewhat easy (0), somewhat difficult (1), and very difficult (2)); and "Do you expect this respondent to participate in the next wave?" (Absolutely (3), probably yes (2), maybe (1), and no (0)). We dichotomized the separate indicators of ability and reluctance, each coded 1 if the respondent was fully cooperative or able (the respective first categories), and 0 otherwise.

The distribution of these indicators in the short-term analytical sample (Sample 2) was as follows (in brackets for the long-term analytical Sample 1 for the year 2004): 92.9% (89.9%) fully understood the questions well, 96.4% (94.5%) were friendly and cooperative, 96.6% (93.3%) were somewhat easy to get, and 84.3% (66.0%) were expected to absolutely participate in the next wave. The interviewer evaluation indicators correlated only weakly with each other with an absolute correlation coefficient between .25 ('respondent friendly' and 'respondent understands questions well', in the short-term analytical sample) and .36 ('respondent easy to get' and 'respondent will participate in next wave', in the short-term analytical sample). We therefore included all four indicators in the models.

3.4. Control Variables

We included the following control variables known to be associated with attrition (Voorpostel 2010; Voorpostel and Lipps 2011):

- Geographical mobility: ownership of the house (yes/no), degree of intention to move during the coming year (0–10), nationality (Swiss/from a neighboring country, that is, Germany, Austria, Liechtenstein, Italy, and France / from another country), whether the respondent has lived in Switzerland for at least 14 years (yes/no).
- Demographic characteristics: age in categories (14–39, 40–49, 50–59, 60–69, 70+), survey language competence (first language, second language, other), education (less than high school level, equivalent to high school level, more than high school level), gender, partner status (living with partner, does not live with partner, no partner), presence of children aged up to seven years in the household (yes/no), number of household members eligible for an interview.
- Social engagement/inclusion/participation history/income: member of a club (yes/no), political interest (0–10), trust in people (0–10), number of waves, equivalized household income.

Table 1. Descriptive statistics independent variables analytical samples for long-term patterns (2004) and for wave-to-wave sample (2005–2017). Data: SHP 2004–2017.

Variables	2004		2005–2017	
	Mean	Std dev.	Mean	Std dev.
Independent variables: Interviewer evaluations				
Respondent is friendly [0,1]	0.945	0.227	0.964	0.187
Respondent understands questions well [0,1]	0.899	0.302	0.929	0.258
Respondent is difficult to convince [0,1]	0.067	0.249	0.034	0.181
Respondent will repeat next wave [0,1]	0.660	0.474	0.843	0.364
Control variables: geographical mobility				
Respondent owns residence [%]	0.464	0.499	0.518	0.500
Intention to move in next 12 months [0 = not at all, . . . , 10 certainly]	1.337	2.895	1.185	2.677
Nationality: Swiss [%]	0.896	0.306	0.906	0.292
Nationality: from a neighbouring country [%]	0.060	0.238	0.058	0.234
Nationality: from another country [%]	0.044	0.205	0.036	0.187
In Switzerland for more than 14 years [%]	0.948	0.221	0.974	0.159
Control variables: demographic characteristics				
Age [14–39 years] [%]	0.292	0.455	0.210	0.408
Age [40–49 years] [%]	0.249	0.433	0.225	0.418
Age [50–59 years] [%]	0.196	0.397	0.217	0.412
Age [60–69 years] [%]	0.140	0.347	0.176	0.380
Age [70+ years] [%]	0.123	0.329	0.172	0.378
Survey language is first language [%]	0.954	0.210	0.971	0.167
Survey language is second language [%]	0.039	0.194	0.026	0.159
Survey language is not first or second language [%]	0.007	0.085	0.003	0.051
Education level: less than high school equivalent [%]	0.163	0.369	0.147	0.354
Education level: high school equivalent [%]	0.556	0.497	0.510	0.500
Education level: more than high school [%]	0.282	0.450	0.343	0.475
Gender: Male [%]	0.365	0.481	0.383	0.486
Partner: yes, living together [%]	0.644	0.479	0.651	0.477
Partner: yes, but not living together	0.100	0.301	0.097	0.295
Partner: no [%]	0.256	0.436	0.253	0.435
Children under seven years in household [%]	0.132	0.338	0.102	0.302
Number of interview eligible household members	2.047	0.967	2.063	0.967
Control variables: social engagement				
Member of a club [%]	0.497	0.500	0.409	0.500
Political interest [0 = not at all, . . . , 10 very interested]	5.690	2.836	5.625	2.794
Trust [0 = can't be too careful, 10 most people can be trusted]	5.623	2.507	6.193	2.273
Number of waves [1, . . . , 6], [1, . . . , 18]	3.666	2.447	8.040	4.656
Equalised household income [Sfr.]	117829	95868	131026	112500
Response quality variables				
Proportion of don't knows [%]	0.008	0.013	0.007	0.013
Proportion of refused items [%]	0.002	0.008	0.002	0.007
Survey year [2004], [2004, . . . , 2016]	2004	0	2010.569	3.837
Sample (observations)	4,394	60,298		
Sample (households)	4,394	10,185		

- Response quality in current wave: proportion of “don’t know” answers, proportion of refused answers.
- Survey-related variables: survey year (only in the short-term analysis).

Table 1 presents the descriptive statistics of all variables in the study for both analytical samples.

3.5. Methods

We analyzed the relationship between interviewer evaluations and survey participation separately for participation patterns and wave-to-wave participation. For both, we first assessed the *bivariate* relationship between interviewer evaluations and participation outcomes, followed by logistic multinomial models controlling for a number of covariates. For the bivariate analyses, we compared mean evaluation scores by participation pattern and by next-wave participation and used Chi-square tests to test whether participation differed by interviewer’s assessment of the respondent’s ability and reluctance. The logistic multinomial model predicting participation patterns used participation in all eligible waves after 2004 as the reference category. For the model predicting next wave participation, the reference category was participation in the next wave. We included covariates (see measures) from 2004 to predict the participation pattern afterwards, and wave-specific measures from 2004–2016 to predict participation in the next wave. As we had multiple observations for each household, we employed a logistic multinomial random intercept model in the fourth analysis. Although households were crossed with interviewers, we did not include interviewers as a level of analysis in any of the models, because the actual outcome depends on a random interviewer in the *next* wave (short-term model) or on several random interviewers in subsequent waves (long-term model). Although the interviewer-respondent assignment is random, there may be very small selection effects due to different shifts worked by interviewers. We tested if interviewer evaluation heterogeneity (e.g., by systematically providing better evaluations) provided different results by using interviewer-centered evaluations as alternative predictors. Differences compared with our original evaluation variables were only marginal.

The equation of the model is presented below (see Haynes et al. 2005, 9–10). Suppose the outcome variable Y_{it} has $J = 3$ categories ($1 = \text{response} = \text{reference category}$, $2 = \text{non-contact}$, $3 = \text{refusal}$), then the probability for household i in wave t to not being contacted ($j = 2$) or to refuse ($j = 3$) rather than to respond ($j = 1$) given a set of control variables X_{it} can be estimated as:

$$\pi_{ij} = \Pr(Y_{it} = j | X_{it}) = \frac{e^{X_{it}\beta_j}}{\sum_{k=1}^J e^{X_{it}\beta_k}}, \quad j = 2, 3$$

corresponding to the following multinomial logit model:

$$\log\left(\frac{\pi_{ij}}{\pi_{i1}}\right) = X'_{it}\beta_j, \quad j = 2, 3$$

If we allow for household-specific random effects α_{ij} and let Z_{ij} denote a vector of

coefficients for the random effects, then the model extends to:

$$\log \left(\frac{\pi_{ijt}}{\pi_{it1}} \right) = X'_{it}\beta_j + Z'_{ij}\alpha_{ij}, \quad j = 2, 3$$

The random effects α_{ij} capture non-observable household effects that are assumed to come from a multivariate normal distribution with mean zero and variance-covariance matrix Σ . We first estimated the model for participation patterns before turning to next wave participation status. The next section presents the results of our analyses.

4. Results

4.1. Interviewer Evaluations and Participation Patterns

Our first research question concerned the extent to which interviewer evaluations in 2004 were able to predict five different subsequent participation *patterns* in the panel (see definition in Subsection 3.2). We first note that the participation patterns are different for all four (dichotomized) evaluation criteria, with significant (1% level) chi²-values ranging between 61.0 for the ‘easy to convince’ criterion, and between 16.7 and 19.3 for the other three criteria (cross-tabulations not shown). [Table 2](#) presents for each evaluation criterion the mean score by participation pattern.

The majority of the participants, regardless of participation pattern, were evaluated as able and willing to complete the survey task. Respondents who participated in every wave in which they were eligible, but also respondents who could not be contacted at least once but re-entered later, had the most positive evaluations. Respondents who dropped out immediately received the lowest evaluation for “respondent will participate next wave”, but not for the other items. Respondents who refused but re-entered were least likely to be evaluated as friendly and easily convinced. This showed that a less positive evaluation was predictive of refusal at a later wave, but these respondents were not necessarily lost to the study. Overall, the interviewers gave ratings that were more positive to respondents with full participation, or non-contact and re-entry, and more negative ratings to respondents who refused in subsequent waves.

Table 2. Mean scores (in 2004) by participation pattern (2004–2017). Data: SHP 2004–2017.

Interviewer evaluations:	Full participation	Refusal and re-entry	Non-contact and re-entry	Participation and dropout	Immediate dropout	Total
Respondent is friendly	0.956	0.910	0.965	0.939	0.921	0.945
Respondent understands questions	0.912	0.905	0.939	0.871	0.884	0.899
Respondent is easy to convince	0.957	0.875	0.948	0.914	0.878	0.933
Respondent will repeat survey	0.680	0.653	0.748	0.625	0.619	0.660
N	2369	400	115	1321	189	4394

In the next step, we analyzed whether interviewer evaluations were predictive of participation patterns after controlling for common covariates. We modeled the same long-term patterns using a multinomial logistic model controlling for all covariates. The first part of Table 3 presents the beta coefficients of the evaluation covariates of the resulting model (see Table 3a). We listed the full model in the Appendix, Section 6 (Table A3).

Since our main interests are the *ceteris paribus* differences in participation pattern by evaluation, we calculated average marginal effects of a more positive evaluation for the different patterns (see Table 3b). Average marginal effects show the *ceteris paribus* increase of the probability of a participation pattern for a positive evaluation compared with a less positive evaluation. For example, the average marginal effect of .142 for the ‘easy to convince’ item indicates that the likelihood of full participation was 14.2 percentage points higher for the easier to convince than for the difficult to convince, of refusal and re-entry 9.4 percentage points lower, of non-contact and re-entry (an insignificant) 0.6 percentage points lower, and of immediate dropout (an insignificant) 2 percentage points lower, holding all other variables constant. Irrespective of the ‘friendliness’ of the respondent or his or her evaluation of the likelihood to repeat the next wave, the patterns do not change. For the ‘understands questions’ item the predicted probability to refuse and re-enter was 3 percentage points higher if the interviewer gave a positive account of the respondent’s understanding, relative to a bad understanding. To assess the part of the marginal effects that is due to confounding predictors, we compared the marginal effects to those calculated with no covariates except the evaluations (see Table 3c). As it turns out, we found somewhat larger effects of the ‘easy to convince’ item and a (5%) significant effect of the ‘understands questions’ item on participation and dropout. However, overall, the interpretation remains largely the same. The results in this section show that the interviewer’s assessment of respondents’ understanding and, in particular, reluctance, added to the prediction of participation patterns in subsequent waves, even after controlling for other common predictors of survey participation.

4.2. Interviewer Evaluations and Short-Term Participation Status

In the last part of our analysis, we investigated the association between the interviewer’s evaluation of the respondent *in each wave* (i.e., across all years and not only in 2004) and the next wave participation *status*. Thus, we focused on short-term participation only, and examined the extent to which interviewer evaluations of separate attributes contributed to predicting nonresponse. Similar to the participation patterns, we first checked bivariate distributions of the interviewer evaluations and the three response outcomes participation, refusal, and non-contact in the next wave (cross-tabulations not shown). Again, all χ^2 -values were significant (1% level) for all four evaluation criteria, ranging between 463.8 for the ‘easy to convince’ criterion, 360.9 for the ‘repeat next wave’ criterion, 140.0 for the ‘friendliness’ criterion and 75.7 for the ‘understands’ criterion. Table 4 presents the mean score by participation outcome for each evaluation criterion.

Again, the majority of the participants were evaluated as able and willing to complete the survey task. Next-wave participants had the most positive evaluations, followed by not contacted respondents. Refusing respondents were least likely to receive positive

Table 3. Multinomial logistic model including all control variables (a, b), without control variables (c), modeling long-term response patterns (base category: full participation).

a) Beta coefficients: full model				
Interviewer evaluations:	Refusal, and re-entry	Noncontact, and re-entry	Participation, and dropout	Immediate dropout
Respondent is friendly	−0.65*	−0.01	−0.18	−0.45
Respondent understands questions	0.46*	0.18	0.04	0.35
Respondent is easy to convince	−1.09**	−0.55	−0.40*	−0.78**
Respondent will repeat survey	0.20	0.39	−0.04	0.06
Constant	−0.20	−1.94	1.28**	0.53
b) Average marginal effects (dy/dx w.r.t. interviewer evaluations): full model				
Interviewer evaluations:	Refusal, and re-entry	Noncontact, and re-entry	Participation, and dropout	Immediate dropout
Respondent is friendly	−.054	.004	−.007	−.013
Respondent understands q.	.030*	.003	−.011	.011
Respondent easy to convince	−.094**	−.006	−.020	−.022
Respondent repeats survey	.016	.009	−.018	.002
c) Average marginal effects (dy/dx w.r.t. interviewer evaluations): only evaluations				
Interviewer evaluations:	Refusal, and re-entry	Noncontact, and re-entry	Participation, and dropout	Immediate dropout
Respondent is friendly	.037	.005	.015	−.008
Respondent understands q.	.025	.010	−.067*	.004
Respondent easy to convince	.190**	−.004	−.051	−.035*
Respondent repeats survey	.009	.009	−.029	−.002

Data: SHP 2004–2017, N = 4,394 households in 2004.
p < 0.01 (**), p < 0.05 (*).

Table 4. Mean evaluation scores (2004–2016) by participation (2005–2017). Data: SHP 2004–2017.

Interviewer evaluations:	Participation	Refusal	Noncontact	Total
Respondent is friendly	0.966	0.924	0.963	0.964
Respondent understands questions	0.931	0.889	0.928	0.929
Respondent is easy to convince	0.970	0.899	0.940	0.966
Respondent will repeat survey	0.850	0.723	0.802	0.843
N	56149	2990	1159	60298

evaluations on all items. This showed that a less positive evaluation was associated with next wave non-contact and especially refusal.

Next, we used multivariate logistic random intercept models to control for commonly used covariates and to accommodate the clustering of the data on the household level. The first part of [Table 5](#) presents the beta coefficients of the evaluation covariates

Table 5. Multinomial logistic random intercept model including all control variables (a, b), without control variables (c), modeling wave-to-wave participation (participation in next wave (base category), refusal, noncontact).

a) Beta coefficients: full model

Interviewer evaluations:	Refusal	Noncontact
Respondent is friendly	− 0.18	0.36
Respondent understands questions	0.22*	0.09
Respondent is easy to convince	− 0.75**	− 0.77**
Respondent will repeat survey	− 0.37**	− 0.21*
Constant	− .069**	− 1.27**
Variance (observation level)		1.42**
Variance (household level)		2.20**

b) Average marginal effects (dy/dx w.r.t. interviewer evaluations): full model

Interviewer evaluations:	Participation	Refusal	Noncontact
Respondent is friendly	.003	− .009	.006*
Respondent understands questions	− .011*	.009**	.001
Respondent is easy to convince	.060**	− .042**	− .017**
Respondent will repeat survey	.022**	− .018**	− .004

c) Average marginal effects (dy/dx w.r.t. interviewer evaluations): only evaluations (and survey year)

Interviewer evaluations:	Participation	Refusal	Noncontact
Respondent is friendly	.005	− .010	.006
Respondent understands questions	− .004	.000	.004
Respondent is easy to convince	.069**	− .052**	− .017**
Respondent will repeat survey	.024**	− .021**	− .003

Data: SHP 2004–2017, N = 10,185 households, 60,298 observations.

p < 0.01 (**), p < 0.05 (*).

of the results (see [Table 5a](#)). Again, we included the complete model in the [Appendix \(Table A5\)](#).

When looking at the coefficients, all evaluation items, except for whether or not the respondent was friendly, were associated with next-wave participation, refusal or non-contact.

Also for the short-term model, we calculated average marginal effects of participating in the next wave based on the full model for changes of the four interviewer evaluations (see [Table 5b](#)). The average marginal effects showed a significant effect for friendliness: a higher respondent friendliness slightly (at a 5% significance level) increased the predicted probability for next-wave non-contact. A better question understanding meant a 1.1 percentage point *lower* participation on the 5% significance level, and a 0.9 percentage point *higher* refusal, net of other covariates in the model. With regard to difficulty to convince the respondent, the probabilities varied the most: Respondents who were easy to convince exhibited a 6.0 percentage points higher participation rate, a 4.2 percentage points lower refusal rate, and a 1.7 percentage points lower non-contact rate than more reluctant respondents. Again, to assess the part of these marginal effects that is due to confounding predictors, we compared them to average marginal effects calculated with no covariates except the evaluations and the survey years (see [Table 5c](#)). In particular, we investigated the counterintuitive effect in which respondents who understood the questions well were, surprisingly, slightly more likely to refuse in the next wave when all variables were included in the model. Looking at the average marginal effects with no covariates except the evaluations and the survey year, the coefficient of question understanding lost significance. It became significantly positive on refusal only after adding the other covariates. For the other evaluation criteria, we did not observe such a change in the size or direction of the coefficients when covariates were dropped from the model.

5. Conclusion

We set out to assess the extent to which interviewers' evaluations are predictive of response patterns and dropout in longitudinal telephone surveys. Prior studies have shown that interviewer assessments are associated with continued participation in the context of face-to-face interviews ([Plewis et al. 2017](#); [Kalton et al. 1990](#); [Lepkowski and Couper 2002](#)). Our study adds to this knowledge by extending it to telephone interviews, a setting in which the interviewer has less information on which to judge the respondent. We can draw the following conclusions.

First, our study showed that even in the absence of face-to-face contact, interviewers' assessments of respondents were predictive of subsequent response patterns, which is in line with the findings from [Kirchner et al. \(2017\)](#). We found that when the interviewer evaluated the respondent as capable with minimal levels of reluctance, respondents were thereafter more likely to become loyal participants, with possible non-contacts in between. In particular, when the interviewer judged the respondent as easy to convince, the respondent was more likely to participate in subsequent waves, rather than to refuse or drop out altogether.

Second, the interviewers' evaluations helped to predict short-term participation, distinguishing participation, refusal, and non-contact. As expected, interviewers were

better able to predict refusal than non-contact. If the respondent was easy to convince and the interviewer judged it likely he or she would return in the next wave, respondents were more likely to participate and less likely to refuse.

An important finding was that the evaluation of the interviewer added to the explanation of next wave or later nonresponse, even if a large set of commonly used predictors of survey participation were taken into account. The ability and reluctance of the respondent as assessed by the interviewer was not fully captured by characteristics such as educational attainment, political interest or civic engagement, nor by past or current survey behavior.

Our study suggests, in line with [Peytchev and Olsen \(2007\)](#), that the use of information provided by interviewers may help to improve longitudinal weights designed to reduce bias from selective attrition. However, for interviewer evaluations to make good adjustment weights, they should not only be able to predict attrition, but also potential research variables ([Little and Vartivarian 2005](#)). Future studies on weighting should explore this further.

Our study is also relevant for responsive design development. Interviewers' assessments may be one of the criteria on which to base decisions on how to allocate fieldwork effort to minimize attrition. For example, respondents who are evaluated as difficult to convince in a given wave can, in the next wave, be offered a higher incentive, a specially tailored newsletter, or be assigned an interviewer who is experienced in refusal conversion. Also, interviewers could be incentivized with additional bonuses for difficult cases. The results of our study can be used to identify respondents for whom special treatment would be most beneficial to improve continued participation in the panel study. Since non-contacts are difficult to anticipate, other measures such as a better timing of the call should be envisaged for this group ([Lipps 2012a, 2012b](#)).

There were some limitations to this study that can be addressed in future research. For example, we had no information on whether households used a landline or a mobile telephone to answer the survey request. Since there are differences between landline and mobile telephone surveys with regard to the mechanism generating nonresponse, as well as conducting the interview ([Kennedy 2010](#)), future studies should distinguish mobile and landline devices, where possible. There are probably additional measurement errors due to inter-interviewer variability, since interviewer evaluations vary in the extent to which they accurately measure objective characteristics such as gender, ethnicity, and dwelling (e.g., [Casas-Cordero et al. 2013](#); [Sinibaldi et al. 2013](#)). In addition, the four evaluation items were significantly skewed towards positive evaluations. We encourage survey methodologists to design and test more elaborated items. To improve the prediction of subsequent response behavior, these items should be based primarily on the evaluated difficulty to convince the respondent to participate, and the evaluated likelihood that the respondent will repeat the survey, and less on the friendliness or degree of question understanding. As our analyses show, these latter items measure respondent characteristics that may not directly predict subsequent participation. However, for the moment, we showed that at least two of the four interviewer evaluations that were considered provide useful additional information at very low cost. A well-designed battery of interviewer evaluations should become an integral part of at least the first wave of every large-scale panel survey. Designing and conducting appropriate experiments remains to be done in future research.

6. Appendix

Table A3. Multinomial logistic model including all control variables modeling long-term response patterns (base category: full participation).

Interviewer evaluations	Refusal, and re-entry	Noncontact, and re-entry	Participation, and dropout	Immediate dropout
Respondent is friendly	− 0.65*	− 0.01	− 0.18	− 0.45
Respondent understands questions	0.46*	0.18	0.04	0.35
Respondent is difficult to convince	− 1.09**	− 0.55	− 0.40*	− 0.78**
Respondent will repeat survey	0.20	0.39	− 0.04	0.06
Owner of house	0.14	− 0.07	0.05	− 0.19
Degree of willingness to move [0..10]	− 0.02	− 0.02	− 0.01	0.03
Swiss (ref: from another country)	0.23	− 0.03	− 0.50*	− 0.30
From a neighbouring country (ref: from another country)	0.21	0.25	− 0.64**	− 0.46
Lives in Switzerland for 14 years or more	0.30	0.35	0.30*	− 0.41
Age: 40–49 (ref: 14–39)	0.22	− 0.58*	− 0.12	− 0.59**
Age: 50–59 (ref: 14–39)	0.25	− 0.92**	− 0.30**	− 0.41
Age: 60–69 (ref: 14–39)	0.17	− 1.67**	− 0.25	− 0.41
Age: 70+ (ref: 14–39)	0.26	− 1.65**	− 0.01	− 0.31
Survey language is second language (ref: first)	0.17	0.31	− 0.05	− 0.53
Survey language is neither first nor second	− 0.34	− 13.62**	0.34	0.55
Education equivalent to high school (ref: low)	− 0.08	− 0.08	− 0.16	− 0.16
Education more than high school (ref: low)	− 0.48**	− 0.05	− 0.26*	− 0.36
Male (ref: female)	0.13	0.11	0.23**	0.22
Lives with partner (ref: no partner)	− 0.13	0.34	0.05	0.43
Partner but not living together (ref: no partner)	− 0.32*	0.37	0.18*	0.15
Child under 7 in the household (ref: no child)	0.08	− 0.82	− 0.29*	− 0.31
Number interview eligible household members	0.09	0.06	0.13**	0.12
Member of a club	− 0.14	− 0.10	− 0.17*	− 0.01
Political interest [0..10]	− 0.06**	− 0.11**	− 0.06**	− 0.12**
Trust in people [0..10]	− 0.02	0.03	− 0.04**	− 0.03
Number of waves in the panel	− 0.17**	− 0.07	− 0.15**	− 0.26**
Equivalised household income	− 0.00	− 0.00	− 0.00*	0.00
Proportion of refused answers	− 2.66	− 6.86	6.66*	14.78**
Proportion of don't know answers	10.06	4.79	3.50	6.25
Constant	− 0.20	− 1.94	1.28**	0.53

Data: SHP 2004–2017, N = 4,394 households in 2004, $r^2 = .055$.

$p < 0.01$ (**), $p < 0.05$ (*).

Table A5. Multinomial logistic random intercept model including all control variables modeling wave-to-wave participation (participation in next wave (base category), refusal, noncontact).

Interviewer evaluations	Refusal	Noncontact
Respondent is friendly	– 0.18	0.36
Respondent understands questions	0.22*	0.09
Respondent is difficult to convince	– 0.75**	– 0.77**
Respondent will repeat survey	– 0.37**	– 0.21*
Owner of house	0.11*	– 0.27**
Degree of willingness to move [0..10]	0.02*	0.04**
Swiss (ref: from another country)	– 0.07	– 0.45**
From a neighbouring country (ref: from another country)	– 0.23	– 0.12
Lives in Switzerland for 14 years or more	– 0.13	– 0.28
Age: 40–49 (ref: 14–39)	0.15*	– 0.33**
Age: 50–59 (ref: 14–39)	0.30**	– 0.81**
Age: 60–69 (ref: 14–39)	0.10	– 1.56**
Age: 70+ (ref: 14–39)	0.54**	– 2.34**
Survey language is second language (ref: first)	0.13	– 0.22
Survey language is neither first nor second	0.73*	0.80
Education equivalent to high school (ref: low)	– 0.15*	– 0.18
Education more than high school (ref: low)	– 0.40**	– 0.40**
Male (ref: female)	0.17**	0.39**
Lives with partner (ref: no partner)	– 0.22*	0.40**
Partner but not living together (ref: no partner)	– 0.28**	0.42**
Child under seven in the household (ref: no child)	– 0.23**	– 0.25*
Number interview eligible household members	0.06	– 0.13**
Member of a club	– 0.19**	– 0.31**
Political interest [0..10]	– 0.05**	– 0.04**
Trust in people [0..10]	– 0.03**	– 0.06**
Number of waves in the panel	– 0.09**	– 0.06**
Equivalised household income	– 0.00	0.00
Proportion of refused answers	7.31**	2.80
Proportion of don't know answers	8.83**	2.68
Survey year 2004	– 0.24*	– 0.79**
Survey year 2005	– 0.87**	– 0.65**
Survey year 2006	– 0.37**	– 0.53**
Survey year 2007	– 0.50**	– 0.85**
Survey year 2008	– 0.88**	– 0.67**
Survey year 2009	– 1.58**	– 0.60**
Survey year 2010	– 1.23**	– 0.47**
Survey year 2011	– 1.29**	– 0.36*
Survey year 2012	– 0.70**	– 0.24
Survey year 2013	– 0.37**	0.01
Survey year 2014	– 0.38**	– 0.25
Survey year 2015	– 0.18*	0.31*
Constant	– .069**	– 1.27**
Variance (observation level)	1.42**	
Variance (household level)	2.20**	

Data: SHP 2004–2017, N = 10,185 households, 60,298 observations. $p < 0.01$ (**), $p < 0.05$ (*).

7. References

- Barrett, K., M. Sloan, and D. Wright. 2006. "Interviewer Perceptions of Interview Quality". In *Proceedings of the American Statistical Association, Survey Research Methodology Section*, 4026–4033. Alexandria, VA: American Statistical Association. Available at: <http://www.asasrms.org/Proceedings/y2006/Files/JSM2006-000644.pdf> (accessed February 2020).
- Block, E.S. and L. Erskine. 2012 "Interviewing by Telephone: Specific Considerations, Opportunities, and Challenges." *International Journal of Qualitative Methods* 11(4): 428–445. DOI: <https://doi.org/10.1177/160940691201100409>.
- Casas-Cordero, C., F. Kreuter, Y. Wang, and S. Babey. 2013. "Assessing the Measurement Error Properties of Interviewer Observations of Neighbourhood Characteristics", *Journal of the Royal Statistical Society: Series A* 176: 229–249. DOI: <https://doi.org/10.1111/j.1467-985X.2012.01065.x>.
- Chun, A., B. Schouten and J. Wagner. 2017. JOS Special Issue on Responsive and Adaptive Survey Design: "Looking Back to See Forward-Editorial". *Journal of Official Statistics* 33(3): 571–577. DOI: <http://dx.doi.org/10.1515/JOS-2017-0027>.
- Couper, M.P. and M.B. Ofstedal. 2009 "Keeping in Contact with mobile Sample Members." In *Methodology of Longitudinal Surveys*, edited by P. Lynn, 183–203. New York: Wiley.
- Eckman, S., J. Sinibaldi, and A. Möntmann-Hertz. 2013. "Can Interviewers effectively rate the Likelihood of Cases to cooperate?" *Public Opinion Quarterly* 77(2): 561–573. DOI: <https://doi.org/10.1093/poq/nft012>.
- Feldman, J.J., H. Hyman, and C.W. Hart. 1951. "A Field Study of Interviewer Effects on the Quality of Survey Data." *Public Opinion Quarterly* 15(4): 734–761. DOI: <https://doi.org/10.1086/266357>.
- Groves, R.M., F.J. Fowler, M.P. Couper, J.M. Lepkowski, E. Singer, and R. Tourangeau. 2009. *Survey Methodology*. 2nd Ed., Wiley Series in Survey Methods. Hoboken, NJ: John Wiley & Sons.
- Groves, R.M. and S.G. Heeringa. 2006. "Responsive Design for Household Surveys: Tools for actively controlling Survey Errors and Costs." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 169(3): 439–457. DOI: <https://doi.org/10.1111/j.1467-985X.2006.00423.x>.
- Haynes, M., M. Western, and M. Spallek. 2005. "Methods for categorical longitudinal Survey Data: Understanding Employment Status of Australian Women." Paper prepared for the HILDA Survey Research Conference, University of Melbourne, Australia, 29–30 September, 2005. Available at: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&uact=8&ved=2ahUKEwjfrIGR1sbnAhUkShUIHf1PAQgQFjABegQIBRAB&url=https%3A%2F%2Fwww.researchgate.net%2Fpublication%2F43458841_Methods_for_Categorical_Longitudinal_Survey_Data_Understanding_Employment_Status_of_Australian_Women&usg=AOvVaw26JHyUI5MCUOF5FqgSIINc (accessed February 2020).
- Kennedy, C.K. 2010. *Nonresponse and Measurement Error in Mobile Phone Surveys*. Doctoral thesis, University of Michigan, Ann Arbor. Available at: <https://deepblue.lib.umich.edu/handle/2027.42/75977> (accessed February 2020).

- Kalton, G., J. Lepkowski, G.E. Montanari, and D. Maligalig. 1990. Characteristics of second Wave Nonrespondents in a Panel Survey. In *Proceedings of the American Statistical Association, Survey Research Methodology Section*, 462–467. Alexandria, VA: American Statistical Association.
- Kaminska, O., A.L. McCutcheon, and J. Billiet. 2010. “Satisficing among reluctant Respondents in a Cross-National Context.” *Public Opinion Quarterly* 74(5): 956–984. DOI: <https://doi.org/10.1093/poq/nfq062>.
- Kirchner, A., K. Olson, and J.D. Smyth. 2017. “Do Interviewer Postsurvey Evaluations of Respondents’ Engagement measure who Respondents are or what they do?” A Behavior coding Study. *Public Opinion Quarterly*: DOI: <https://doi.org/10.1093/poq/nfx026>.
- Laurie, H., R. Smith, and L. Scott. 1999. “Strategies for reducing Nonresponse in a longitudinal Panel Survey.” *Journal of Official Statistics* 15(2): 169–282. Available at: <https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/strategies-for-reducing-nonresponse-in-a-longitudinal-panel-survey.pdf> (accessed February 2020).
- Lemay, M. 2009. *Understanding the Mechanism of Panel Attrition*. Unpublished Doctoral thesis, University of Maryland, College Park, MD, U.S.A.
- Lepkowski, J.M. and M.P. Couper. 2002. “Nonresponse in the second Wave of longitudinal Household Surveys.” In *Survey nonresponse*, edited by R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J. Little, 259–272. New York: Wiley.
- Lipps, O. 2012a. Using Information from Telephone Panel Surveys to predict Reasons for Refusal. *Methods, data, analyses* 6(1): 3–20. Available at: <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-314544> (accessed August 2019).
- Lipps, O. 2012b. “A Note on improving Contact Times in Panel Surveys.” *Field Methods* 24(1): 95–111. DOI: <https://doi.org/10.1177/1525822X11417966>.
- Little, R.J. and S. Vartivarian. 2005. “Does weighting for Nonresponse increase the Variance of Survey Means?” *Survey Methodology* 31(2): 161–168. Available at: <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&ved=2ahUKEwiTkPGQ2MbnAhXHUBUIHWLjCq4QFjACegQIBRAB&url=http%3A%2F%2Fciteseerx.ist.psu.edu%2Fviewdoc%2Fdownload%3Fdoi%3D10.1.1.692.507%26rep%3Drep1%26type%3Dpdf&usg=AOvVaw1fkEbnvfviCWWInhD3jfs6> (accessed February 2020).
- Lutig, P. 2014. “Panel Attrition: Separating Stayers, fast Attriters, gradual Attriters, and Lurkers.” *Sociological Methods & Research* 43(4): 699–723. DOI: <https://doi.org/10.1177/0049124113520305>.
- Perez, F.P. and B. Baffour. 2018. “Respondent mental Health, mental Disorders and Survey Interview Outcomes.” *Survey Research Methods* 12(2): 161–176. DOI: <https://doi.org/10.18148/srm/2018.v12i2.7225>.
- Peytchev, A. and K. Olson. 2007. “Using Interviewer Observations to improve Nonresponse Adjustments: NES 2004.” In *Proceedings of the American Statistical Association, Survey Research Methodology Section*, 3364–3371. Alexandria, VA: American Statistical Association. Available at: <https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1149&context=sociologyfacpub> (accessed February 2020).
- Plewis, I., L. Calderwood, and T Mostafa. 2017. “Can Interviewer Observations of the Interview Predict future Response?” *Methods, data, analyses* 11(1): 29–44. DOI: <https://doi.org/10.12758/mda.2016.010>.

- Schonlau, M., N. Watson, and M. Kroh. 2010. "Household Survey Panels: How much do following Rules affect Sample Size?" *Survey Research Methods* 5(2): 53–61. DOI: <https://doi.org/10.18148/srm/2011.v5i2.4665>.
- Schouten, B., A. Peytchev, and J. Wagner. 2017. *Adaptive Survey Design*. Boca Raton: CRC Press.
- Sinibaldi, J., G. Durrant, and F. Kreuter. 2013. "Evaluating the Measurement Error of Interviewer observed Paradata." *Public Opinion Quarterly* 77: 173–193. DOI: <https://doi.org/10.1093/poq/nfs062>.
- Stoop, I. 2005. *The Hunt for the Last Respondent, Nonresponse in Sample Surveys*. The Hague: Social and Cultural Planning Office of the Netherlands. Available at: <https://dspace.library.uu.nl/bitstream/handle/1874/2900/full.pdf> (accessed February 2020).
- Tillmann, R., M. Voorpostel, U. Kuhn, F. Lebert, V-A. Ryser, O. Lipps, B. Wernli, and E. Antal. 2016. "The Swiss Household Panel Study: Observing social Change since 1999." *Longitudinal and Life Course Studies* 7(1): 64–78. DOI: <http://dx.doi.org/10.14301/llds.v7i1.360>.
- Uhrig, N.S.C. 2008. *The Nature and Causes of Attrition in the British Household Panel Survey*. ISER Working Paper 2008-05. Colchester: ISER, University of Essex. Available at: https://www.researchgate.net/profile/Sc_Uhrig/publication/242116089_The_Nature_and_Causes_of_Attrition_in_the_British_Household_Panel_Survey/links/0a85e537c6cd161804000000/The-Nature-and-Causes-of-Attrition-in-the-British-Household-Panel-Survey.pdf (accessed February 2020).
- Voorpostel, M. 2010. Attrition Patterns in the Swiss Household Panel by Demographic Characteristics and Social Involvement. *Swiss Journal of Sociology* 36(2): 359–377. Available at: https://serval.unil.ch/resource/serval:BIB_78D83DD99A6F.P001/REF (accessed February 2020).
- Voorpostel, M. and O. Lipps. 2011. Attrition in the Swiss Household Panel: Is change associated with drop-out? *Journal of Official Statistics* 27(2): 301–318. Available at: <https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/attrition-in-the-swiss-household-panel-is-change-associated-with-drop-out.pdf> (accessed February 2020).
- Wagner, James. 2008. *Adaptive Survey Design to Reduce Nonresponse Bias*, Dissertation, University of Michigan.
- Watson, Nicole, and Mark Wooden. 2004. Sample attrition in the HILDA survey. *Australian Journal of Labour Economics* 7(2): 293–308. Available at: <https://business.law.curtin.edu.au/wp-content/uploads/sites/5/2016/05/AJLE-v7n2-watson.pdf> (accessed February 2020).

Received January 2019

Revised July 2019

Accepted September 2019