Interviewer effects on cooperation during initial and refusal conversion fieldwork phases in telephone panel surveys

Oliver Lipps (oliver.lipps@fors.unil.ch)

FORS and University of Bern, c/o Bâtiment Géopolis, 1015 Lausanne

This is a pre-copyedited, author-produced PDF of an article accepted for publication in *Field Methods* following peer review. The version of record (published version) [Interviewer effects on cooperation during initial and refusal conversion fieldwork phases in telephone panel surveys; Oliver Lipps; Field Methods 2019; doi: 10.1177/1525822X19874285] is available online at: https://journals.sagepub.com/doi/full/10.1177/1525822X19874285
and should be cited as follows:

Lipps, Oliver (2019) 'Interviewer effects on cooperation during initial and refusal conversion fieldwork phases in telephone panel surveys'. *Field Methods,* published online 11-09-2019. DOI: 10.1177/1525822X19874285

Abstract

Specific interviewer characteristics, interviewer continuity, or matching interviewer and household characteristics may increase cooperation, especially for difficult to convince households. In face-to-face surveys, unobserved heterogeneity often makes a proper analysis of interviewer effects impossible. Although surveys conducted in telephone centers usually assign households to interviewers at random, there is less research on interviewer effects on cooperation, probably because telephone surveys produce smaller effects. Using data from a large telephone panel survey, I find interviewer effects only for households that refused to participate in a previous wave. Interviewer continuity or matching interviewers and households on socio-demographic variables has weak effects for any type of household. Interviewer experience has positive effects for previously refusing households only. Telephone survey organizations therefore only need to worry about using specially trained interviewers for refusal conversion calls while specific assignments of interviewers to households are not necessary.

Acknowledgment: This study has used the data collected by the Swiss Household Panel, which is based at the Swiss Centre of Expertise in the Social Sciences (FORS) and financed by the Swiss National Science Foundation. I like to thank Florence Lebert and Jessica Herzing for their helpful comments on an earlier version.

Introduction

Whether using interviewers with specific characteristics, matching interviewers with households on socio-demographic variables, or assigning the same interviewer to the same households across waves increases cooperation rates is not clear. For example, while interviewer experience has been shown to have positive effects in some studies (Lipps and Pollien 2011; Vassallo et al. 2015), Durrant et al. (2010) find a decline in performance when pay grade is controlled for. A worse performance of more experienced interviewers may also be due to being assigned to the more difficult households (West and Blom 2017). While interviewer age, gender, or other personality traits often have no effects (West and Blom 2017), matching the same age, gender and education level is more promising (Groves and Couper 1998, Durrant et al. 2010, Vercruyssen et al. 2017, West and Blom 2017). This supports the theory of liking (Cialdini 1984; Groves et al. 1992; Olson 2007) and status distance, which includes demographic distance (Block and Erskine 2012): Communication is more likely and facilitated and as a consequence potential respondents are more cooperative when approached from interviewers with whom they have some characteristics in common. If the contact with the interviewer was successful in the previous wave (Watson and Wooden 2014), using the same interviewer may increase respondent's trust, establish rapport between respondent and the interviewer, and result in increased cooperation rates (Morton-Williams 1993; Kühne 2018). Finally, positive attitudes about their persuasion power and towards the survey, or high motivation, may have positive effects on cooperation, while beliefs in the importance of the voluntary nature of participating in a survey often have negative effects (Groves et al. 1992; Hox and De Leeuw 2002; Durrant et al. 2010, West and Blom 2017).

While most findings on interviewer effects on cooperation come from face-to-face surveys (Dutwin et al. 2014; West and Blom 2017), scholars wonder whether findings from face-to-face surveys hold for telephone surveys in a similar way (Vercruyssen et al. 2017). One problem with face-to-face surveys is that it may be difficult or costly to assign interviewers to households at random (Campanelli and O'Muircheartaigh 2002; Lynn et al. 2014). As a result, interviewers vary in the types of areas with different unknown cooperation propensities of households. This makes a proper analysis of interviewer effects impossible (Durrant et al. 2010). Studying interviewer effects in telephone centers where fresh numbers are usually assigned to the next free interviewer helps to minimize concerns about unobserved heterogeneity. This guarantees a random assignment of households to interviewers *currently working*. In this "quasi-interpenetrated" (West and Blom 2017) design, interviewer effects need not be disentangled from (often unobserved) co-varying household characteristics. Notwithstanding these analytical advantages, there is less research on interviewer effects on cooperation in telephone surveys, probably because there is only the auditory channel of communication available (Oksenberg and Cannell 1988; Maynard and Schaeffer 1997; Hox et al. 1998; van der Vaart et al. 2006), much less time to convince a reluctant household (Hox et al. 1998), and a higher level of anonymity (Block and Erskine 2012), which all produce smaller effects (Groves and Kahn 1979). Still, telephone interviewers may learn how to develop skills and tactics to convince households to participate, for example by an adequate tailoring and maintaining interaction (Groves et al. 1992).

As for the little evidence of interviewer effects on cooperation in telephone surveys, Singer et al. (1983) report positive effects from expectations about the response rate, while negative effects are associated with young interviewers, and with experience. Hansen (2006) finds that male interviewers are more successful than female interviewers and that interviewer age and experience have a positive effect on the response rate. Groves and Fultz (1985) find no effects of the interviewer's gender on response rates, while Verger et al. (2001) report higher refusal rates from women and Lipps (2010) finds that older men perform slightly better than younger men and women. Effects from interviewer continuity have not been analyzed so far in telephone panel surveys (but see Lipps 2009, who uses two-wave panels), although some scholars suggest that interviewer continuity may increase response rates also in telephone panels (Erikson et al. 2006). To the best of our knowledge, effects from interviewer-household *matching* of socio-demographic variables have not yet been studied in telephone surveys (but see de Jong 2016 for the importance to match gender in countries with strong gender norms). One exception is the research from Lipps (2010), who finds no age or gender matching effects in a panel with an only small number of waves.

Whether interviewer effects depend on the difficulty of the sample is not clear with the exception of the finding that more experienced interviewers handle more difficult cases better (Durrant and D'Arrigo 2014). Dutwin et al. (2014) state that "anecdotal evidence suggests there is considerable variation across interviewers in their abilities to convert refusals" (p.84). And "some members [of the AAPOR Task Force on Survey Refusals] believe that it is best to match the demographic characteristics of interviewers to refusing cases, but no valid empirical evidence has been reported to support this hypothesis" (op. cit., p.84). Effects from interviewer attitudes such as the expectation about the ease of persuading respondents may well have similar effects on refusal conversions as on households during the initial fieldwork phase but also here no reported evidence exists (op. cit., p.84).

My hypothesis is that some interviewer characteristics, such as experience, have stronger effects for more difficult households. Since not all interviewers have favorable characteristics or attitudes, this finding would suggest optimizing the assignment of the available interviewers to difficult fieldwork situations such as the refusal conversion (in the following RC) fieldwork phase. A systematic comparison of interviewer effects between initial and RC fieldwork phases has not yet been investigated (Dutwin et al. 2014).

In this article, I analyze the effect on cooperation of interviewers' socio-demographic characteristics and attitudes, interviewer continuity, and interviewer and household matching of socio-demographic variables. I use a large nationally representative telephone panel survey which uses a random interviewer-household assignment during the initial as well as the RC fieldwork phase. I hypothesize positive effects from positive interviewer attitudes, interviewer-household matching, and interviewer continuity, which I expect to be stronger during the RC fieldwork phase compared to the initial fieldwork phase. While interviewer conversation skills to convince households to participate in telephone surveys (Broome 2015) may have larger effects on cooperation, recording the conversation often needs the household's consent which may be considered too risky in panel surveys (Groves et al. 2011, ch. 11.7.1).

Data and Methods

I use data from the annually conducted telephone Swiss Household Panel (SHP) between 2005 and 2017 (Tillmann et al. 2016). The SHP sample is representative of the Swiss residential population aged 14 and older. It contains three subsequent questionnaires: The grid questionnaire and the household questionnaire (completed by the household reference person), and for each age eligible individual finally the individual questionnaire. I define characteristics of the household reference person measured in the grid questionnaire to represent the household as a whole (Durrant and Steele 2009; Durrant et al. 2010). I use grid data from all three SHP samples (drawn in 1999, 2004, and 2013, respectively), data from the CATI call records, as well as data from the annually conducted paper and pencil interviewer surveys.

Since the success of a contact with a household depends on the outcome of previous contacts with the household (Durrant et al. 2010) and interviewer effects are highest at first contacts (Lipps 2009), I only use the first contact with households within each wave (that is, the first call where someone in the household was reached) as dependent variable. I define first contacts as successful (=1) if the household finally completes the grid questionnaire in this wave, and unsuccessful (=0) else (Lipps 2008). I distinguish cooperation between the initial and the RC fieldwork phase. During the RC phase, especially trained interviewers start calling refusing households from previous waves about one to two months after the initial fieldwork phase started. Also refusing households from the initial fieldwork phase are tried to be converted during the RC phase but these households are analyzed during the initial phase at their first contact.

I impute missing values for interviewers (2.3% of all values) using within-information from three adjacent waves. In the case of still missing interviewer evaluations, I impute them by the overall mean values (1.8% of all values). Information for the noncontacted households are taken from the last responding wave. Keeping only households with at least one successfully conducted grid questionnaire, this data includes 78,438 observations. For about 13% of household data no interviewer information is available, leaving us with 67,856 observations. To check if missing interviewer information is at random with respect to cooperation, I find an insignificant (1%) difference of 88.6% (finally) completed grids in the full data (ranging between of 81.2% in 2006 and 93.0% in 2009) and 88.5% completed grids in the matched data (from 81.2% in 2006 to 92.6% in 2009).

The analysis data with valid information for all household and interviewer variables between 2005 and 2016 comprises 66,074 first contacts with 13,877 households, carried out by 671 interviewers. To test the effect of using the same interviewer across waves, I use information from the interviewer with whom the household reference person completed the grid questionnaire in wave t-1, and who does the first contact with the household in wave t. Mostly because contact records for 2004 are not yet available, only for 48,622 first contacts an interviewer from the previous wave can be identified. 2.1% of these were conducted by the same interviewer who asked the grid interview in the previous wave. Using the same interviewer had a small negative effect on final grid completion (1% level), which is even smaller (10% level) for RC contacts. As a consequence from the small effect and not to lose 17,452 (=66,074-48,622) first contacts, I drop the variable 'same interviewer' in the following.

Because households and interviewers are crossed, I use interviewer-household crossed multilevel models (Fielding and Goldstein 2006). Specifically, I use the runmlwin (Leckie and Charlton 2013) command in Stata version 14, which estimates the models by using the MLwiN software. The models use Markov Chain Monte Carlo (MCMC) estimation (Browne 2012). Variance components and coefficients from nonlinear models are tricky to interpret and do not allow a comparison across different samples or models, even when they are nested (Mood 2010; Best and Wolf 2015). In addition there are often convergence problems in such nonlinear complex models (Davidian 2017). Therefore, I use linear probability models which allow interpret the coefficients as in usual linear models. The crossed linear probability model can be written as:

$$E(y_{t(h,i)}) = X_{t(h,j)}\beta + u_{0h} + v_{0i} + \varepsilon_{0(h,i)} ,$$

where $X_{t(h,i)}$ denote the covariates and interactions measured at observation t within the crossed second levels i (interviewers) and j (households), β the regression coefficients, u_{0h} the variance on the household level, v_{0i} the variance on the interviewer level, and $\varepsilon_{0(h,i)}$ the variance on the observation level (Fielding and Goldstein 2006; Best and Wolf 2015). I first decompose the total variance into interviewer, household, and observation variance by estimating an empty ('variance-components'-) model (M0). Then I include all other independent variables to estimate the full model (M1), before I finally add interactions of RC with interviewer variables (M2).

Results

I list regression coefficients of the independents variables (year and wave dummy coefficients controlled but not reported) of the three models (M0: multilevel crossed variance components, M1: full multilevel crossed, M2: full multilevel crossed including interactions of RC with the interviewer variables) in table 1:

Table 1: Grid questionnaire completion linear probability model coefficients.

Independent variables	M0 Variance components crossed	M1 full crossed	M2 full crossed + RC interact
Household characteristics			
Sample 2004 (Reference 1999)		-.064 **	-.064 **
Sample 2013 (Reference 1999)		.029 **	.030 **
Household lives in Swiss-German speaking part (reference: French)		-.009 *	-.008
Household lives in Italian speaking part (reference: French)		.015 *	.015 *
Household single		-.048 **	-.047 **
Number of children (<18 years) in household		-.011 **	-.011 **
Household male		.010 **	.010 **
Household old [>50 years in first wave])		-.017 **	-.017 **
Household has higher education in first wave		.033 **	.034 **
Household works full time [>=36 h/week]		-.020 **	-.020 **
Household works part time [4/<36 h/week]		.009 *	.009 *
Household has Swiss nationality (reference: other than Swiss or neighbor)		.074 **	.073 **
Household has nationality from a neighboring country		.049 **	.048 **
Fieldwork characteristics			
First contact Saturday (reference: Working day evening 19h and later)		-.027 **	-.030 **
First contact working day morning (8h – 13h59h)		.021 **	.005
First contact working day afternoon (14h – 18h59)		.002	.003
Interviewer characteristics			
Interviewer experience [years at survey firm]		-.002 **	-.001
Interviewer male		.000	.004
Interviewer old [>25 years]		.002	-.005
Interviewer has higher education		-.011 **	-.009 **
Interviewer evaluation: survey is voluntary [0no .. 10yes]		.001	.001
Interviewer evaluation: I can convince majority [0no .. 10yes]		.000	.001
Interviewer evaluation: I must accept refusal [0no .. 10yes]		-.000	-.000
Interviewer evaluation: SHP is interesting [0no .. 10yes]		-.001	-.001
Interviewer evaluation: SHP is important for society [0no .. 10yes]		.001	.000
Interviewer eval.: I would participate in SHP survey [0no, 1perhaps, 2yes]		.000	-.001
Interviewer evaluation: I would report income [0no, 1perhaps, 2yes]		-.006 *	-.005 *
Interviewer evaluation: I am satisfied with working conditions [0no .. 10yes]		.000	.001

Interviewer evaluation: I am able to adapt to various situations [0no .. 10yes]		-.000	-.000
Interviewer evaluation: I am a good actor [0no .. 10yes]		.000	.000
Matching characteristics			
Interviewer and Household male		-.007	-.009 *
Interviewer and Household old		.006	.006
Interviewer and Household higher education		.008 *	.007
RC characteristics			
Refusal conversion phase (RC)		-.286 **	-.188 **
RC X First contact Saturday (reference: Working day evening 19h and later)			-.017
RC X First contact working day morning (8h – 13h59h)			.244 **
RC X First contact working day afternoon (14h – 18h59)			.028 **
RC X Interviewer characteristics			
RC X Interviewer experience [years at survey firm]			-.013 **
RC and Interviewer male			-.037 **
RC and Interviewer old [>25 years]			.057 **
RC and Interviewer has higher education			-.007
RC X Iwer eval.: survey is voluntary [0no .. 10yes]			-.001
RC X Iwer eval.: I can convince majority [0no .. 10yes]			.013 **
RC X Iwer eval.: I must accept refusal [0no .. 10yes]			-.006 **
RC X Iwer eval.: SHP is interesting [0no .. 10yes]			-.018 **
RC X Iwer eval.: SHP is important for society [0no .. 10yes]			-.001
RC X Iwer eval.: I would participate in SHP survey [0no, 1perhaps, 2yes]			.020 *
RC X Iwer eval.: I would report income [0no, 1perhaps, 2yes]			-.015 *
RC X Iwer eval.: I am satisfied with working conditions [0no .. 10yes]			-.006 *
RC X Iwer eval.: I am able to adapt to various situations [0no .. 10yes]			.004 **
Iwer eval.: I am a good actor [0no .. 10yes]			-.001
RC X Matching characteristics			
RC and Interviewer and Household male			.015
RC and Interviewer and Household old			-.004
RC and Interviewer and Household higher education			.004
Constant	.857 **	.981 **	.968 **
Interviewer variance	.0013 **	.0006 **	.0005 **
Household variance	.0436 **	.0203 **	.0203 **
Observation variance	.0669 **	.0665 **	.0658 **
Bayesian DIC	18,571	15,891	15,267

Note: **(*)=significant on 1(5)% level. N=66,074 observations, 13,877 households, 671 interviewers. Source: SHP 2005-2017.

From model M0, the variance share on the interviewer level amounts to 1.2% (.0013/(.0013+.0436+.0669)) of the total variance, on the household level to 39.2%, and the observation level (within household and interviewer) to 59.7%. The interviewer variance share is in line with findings in the literature (Hox et al. 1991; Lipps 2009). The observation level promises most opportunities to improve cooperation, possibly by a more efficient assignment of interviewers to households.

In model M1, the Bayesian DIC[1] is much lower than that of model M0 which shows that model M1 fits the data better. This is mostly due to the high explanation power of the household variables and in particular the fieldwork phase: including RC alone (thus dropping household and interviewer characteristics) would result in an only slightly higher household variance (.222) and an interviewer variance between that of model M0 and model M1 (.0009). Both the household and the interviewer variance drop to less than half their values from model M0, while the observation variance remains unchanged. Three interviewer variables are significant, of the matching variables, only education is significant (5% level).

Since there is a selection of interviewers with specific characteristics into the RC fieldwork phase such as more experienced or older interviewers, the coefficients of these interviewer characteristics are probably biased because they work on average a more difficult sample. I check the proportion of male, older, and higher educated interviewers, and the mean number of years of experience by initial versus RC fieldwork phase in table 2, and compare the cooperation rate of these interviewer groups.

Table 2: Characteristics of interviewers and cooperation differences by fieldwork phase.

	Initial phase [%]	RC phase [%]	Cooperation rate initial phase	Cooperation rate RC phase
Interviewer male	35.8	37.3 **	F: .935, M: .936	F: .546, M: .533
Interviewer Old [>25 years]	42.1	58.0 **	Y: .934, O: .938	Y: .511, O: .564 **
Interviewer higher education	53.6	55.1 *	L: .940, H: .931 **	L: .550, H: .534
Interviewer experience [mean years]	2.85	3.57 **		

Note: **(*)=significant on 1(5)% level. N=66,074 observations, 13,877 households, 671 interviewers. Source: SHP 2005-2017.

Interviewers who work during the RC fieldwork phase are rather men, more experienced, and in particular older. Also for the other interviewer characteristics (not shown), I find differences between interviewers working during RC phase or not. While older interviewers perform better than younger interviewers to convince households to cooperate during the RC

[1] The Bayesian Deviance Information Criterion (DIC) is an MCMC penalised goodness of fit measure and is equivalent to the Akaike Information Criterion (AIC) used in maximum likelihood estimation. Given a data set, of several competing models the one with the lowest AIC is the best.

phase, lower educated are better than higher educated interviewers during the initial phase. Including interaction effects of interviewer characteristics with the fieldwork phase should thus improve the proportion of explained variance and provide insights how to optimize interviewer-household assignments.

Coefficients of the model which adds these interactions to model M1 are shown in table 1 (model M2). From the lower DIC this model fits the data better than model M1. Compared to model M1, there is a strong (absolute) drop of the RC main effect from -.286 to -.188. Interestingly, I find that the positive effect of first contacting households during mornings (model M1) holds only during the RC fieldwork phase (model M2), where it turns out to be very strong. All three significant interviewer coefficients from model M1 reduce when limited to the initial phase and there are only two significant variables left in model M2. Model M2 shows that while most interviewer variables have no effect during the initial phase, the majority of them (10 out of 14) are effective during the RC phase. The strongest effect comes from using older interviewers, which increases the RC cooperation rate by ceteris paribus 5.7% points, followed by using women (+3.7% points), and less experienced interviewers (-1.3% point per year of experience). Also interviewer attitudes towards the surveys show much stronger effects during the RC phase. With the exception of the negative effect of the belief that the SHP is interesting and the negative effect of experience, the findings are in line with previous research: a positive thinking about one’s own persuasion capacities, a more rigorous treatment of reluctant households, and a positive opinion about the survey are associated with higher cooperation rates.

The strong positive effect of RC first contacts during mornings are based on relatively few contacts (N=624). Bivariate statistics indicate (results not shown) that a slightly different sample was reached during this time (less singles and Swiss-German speakers, more women, etc.). Nevertheless I cannot explain this strong effect by co-varying household (or interviewer) variables such that doing more RC calls during the morning may be a good idea to improve cooperation rates.

As four robustness checks, I first re-estimate the model M2 from table 1 excluding observations where interviewer information was imputed, leaving us with 59,434 complete case observations. The coefficients of both models are very similar, with the same number of significant interviewer variables for both fieldwork phases.

Second, to be able to use the households' socio-demographic characteristics from the last responding wave, the household must have participated at least once before. This leads to a positive selection. To test a less selective sample, I use the SHP 2013 sample in the year 2013, which comprises 5,302 households (4,500 successfully matched with interviewers), and a cooperation rate of 60.9% (60.4%). The advantage of this sample is that it was drawn from the population register, which contains the socio-demographic variables from model M1 with the exception of education and working status. Some (12.6%) of the households were deemed "difficult" by the survey agency (negative feedback, difficult to find) and treated as RC cases from the onset (cooperation rate 40.7%). I use a hierarchical two-level model (households nested in interviewers) and the variables from model M2. While the socio-demographic variables have the expected coefficients, first calls during mornings are still significantly positive (main effect: .07), especially for RC cases (interaction effect: .28). With two exceptions (positive effects from older interviewers and negative effects from the belief that the SHP is important for the society), neither the interviewer variables nor matching households and interviewers on age and sex is significant, both in the initial and in the RC fieldwork phase.

Third, I estimate the coefficients of the timing and the interviewer variables and their interaction with RC using household fixed effects (FE) models (no household variables included). The reason is – similar to a Hausman test – to compare the coefficients of the multilevel crossed model M2 with those of the household FE model. I assume that the FE model produces unbiased coefficients. While most of the coefficients from model M2 change somewhat, the conclusion from both model M2 and the FE model is the same: I find few interviewer effects during the initial fieldwork phase while many more are effective during the RC fieldwork phase. The coefficients of the FE model are very close to those of model M2.

Fourth, I test alternative measures of interviewer experience (Lipps 2009), using multilevel (ML) and household FE models. Recall that the number of years worked with the survey agency is insignificant during the initial fieldwork phase and has a negative coefficient during the RC phase (ML:-.013, FE:-.013). The number of all first contacts ever (recoded to ten deciles) has a negative coefficient during the initial phase (ML: -.005, FE:-.003) and a positive during the RC phase (ML:.015, FE:.020), which is similar to those of the number of (recoded) successful first contacts, i.e. those with a final completion (initial phase: ML:-.004, FE:-.002; RC phase: ML:.015, FE:.020). The number of (recoded) *un*successful first contacts, i.e. those with a final refusal, have a negative coefficient during the initial phase (ML:-.013, FE:-.007) and a positive during the RC phase (ML:.005, FE:.014). Since the model with experience measured as the number of unsuccessful first contacts has the smallest DIC (15,074 versus values between 15,267 and 15,282) in the ML model and the highest within-R^2 (.067 versus .065) in the FE model, we favor this experience measure.

Summary and Conclusion

I analyze the effect of interviewer characteristics and attitudes towards surveys and interviewer-household matching of socio-demographic variables and interviewer continuity on cooperation in a centralized telephone panel survey. I hypothesize that interviewer effects are stronger during the refusal conversion (RC) fieldwork phase.

First of all, interviewer continuity has no effect on cooperation neither during the initial nor the RC phase. Unlike in face-to-face panel surveys, telephone panel households either do not remember the previous interviewer or do not mind being contacted by another. Telephone survey organizations thus do not have to worry about compromising flexibility by assigning the same interviewer to the same households across waves. Also interviewer-household matching of socio-demographic variables has small effects at best such that also here the survey organizations can be flexible without reducing cooperation probability.

While interviewer variables have small effects during the initial fieldwork phase, they make a difference during the RC phase. It may be that in telephone surveys households do not correctly infer or simply do not mind interviewer characteristics when they are positively inclined towards the survey, but that those who previously refused are more sensitive when being re-approached. That there are no effects of interviewer characteristics on both “normal” and more difficult households (but who did not yet refuse) in a fresh sample supports this interpretation: Households may become sensitive to differences between interviewers in telephone surveys only after a refusal.

Households contacted during the morning show large positive effects on cooperation in both the initial and the RC phase. This is only partially due to a specific sample composition accessible at this time and deserves further investigation. I find negative effects from interviewer experience during the initial phase and mostly positive effects during the RC phase. It may be that previous research confused the effect of age and experience of interviewers due to a lacking interpenetrated design (Lynn et al. 2014) and maybe a lacking distinction of the fieldwork phases. I can only speculate why older interviewers perform better during the critical fieldwork phase only. It may be that they have better tactics to gain cooperation at their disposal (Hox et al. 1998) which pay off during more difficult situations only. Other findings for the RC phase are mostly in line with the literature: men perform worse than women, a positive thinking about one's own persuasion capacities and a more rigorous treatment of reluctant households are associated with higher cooperation rates.

A practical implementation of our findings may be to perform more first RC fieldwork calls during mornings, a re-assignment of interviewers with a positive thinking about their persuasion capacities and a more rigorous treatment of reluctant households, women, and in particular older and more experienced interviewers to the RC phase.

Our research has three important findings: first, many interviewer characteristics are effective only for households that refused in a previous wave. Secondly, making contact at mornings is very effective, especially for RC calls. Third, interviewer experience, and in particular negative experience (see Lipps 2009) appears to have differential effects in the initial and the RC fieldwork phase in telephone surveys. Interestingly both negative and positive experiences have negative effects during the initial fieldwork phase and generally positive effects during the RC phase. Since also first contacts from the same interviewer who has conducted the previous (grid) interview and interviewer-household socio-demographic matching have no effects, survey organizations need not worry about high interviewer turnover or interviewer-household matching during the initial fieldwork phase. In this phase, survey organizations and their interviewers may be flexible without compromising cooperation rates in telephone panel surveys.

References

Best, H., and C. Wolf. 2015. Logistic regression. In *The SAGE handbook of regression analysis and causal inference,* eds. H. Best and C. Wolf, 153-71. Los Angeles: Sage.

Block E., and L. Erskine. 2012. Interviewing by Telephone: Specific Considerations, Opportunities, and Challenges. *International Journal of Qualitative Methods* 11(4):428-445.

Broome J. 2015. How telephone interviewers' responsiveness impacts their success. *Field Methods* 27:66–81.

Browne, W. 2012. MCMC Estimation in MLwiN, v2.26. Centre for Multilevel Modelling, University of Bristol.

Campanelli, P., and C. O'Muircheartaigh. 2002. The importance of experimental control in testing the impact of interviewer continuity on panel survey nonresponse. *Quality & Quantity*, 36(2):129-44.

Cialdini, R. 1984. Influence: the new Psychology of modern persuasion. New York: Quill.

Davidian, M. 2017. Nonlinear models for repeated measurement data. New York: Routledge.

de Jong, J. 2016. Data collection: Telephone surveys. In *Cross-cultural Survey Guidelines*, Survey Research Center, Institute for Social Research. University of Michigan.

Durrant, G., and J. D'Arrigo**.** 2014. Doorstep interactions and interviewer effects on the process leading to cooperation or refusal. *Sociological Methods and Research*, 43(3):490-518.

Durrant, G., R. Groves, L. Staetsky, and F. Steele 2010. Effects of interviewer attitudes and behaviors on refusal in household surveys. *Public Opinion Quarterly* 74:1-36.

Durrant, G., and F. Steele. 2009. Multilevel modelling of refusal and noncontact nonresponse in household surveys: Evidence from six UK government surveys. *Journal of the Royal Statistical Society*, *Series A* 172(2):361-81.

Dutwin, D., J. Loft, J. Darling, A. Holbrook, T. Johnson, R. Langley, P. Lavrakas, K. Olson, E. Peytcheva, J. Stec, T. Triplett and A. Zukerberg. 2014. *Current knowledge and considerations regarding survey refusals.* Available at https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/RefusalTF_FINAL090814.pdf. (accessed September 6, 2018).

Erikson, R., N. Buck, D. Lievesley, and J. Schupp. 2006. The Swiss Household Panel. Evaluation report, FORS, Lausanne. http://forscenter.ch/wp-content/uploads/2013/11/shp_evaluation_report_2006.pdf (accessed September 6, 2018)

Fielding, A. and H. Goldstein. 2006. Cross-classified and multiple membership structures in multilevel models: An introduction and review. Research Report RR791, department for education and skills, University of Birmingham.

Groves, R., R. Cialdini, and M. Couper. 1992. Understanding the decision to participate in a survey. *Public Opinion Quarterly,* 56:475-95.

Groves, R., and M. Couper. 1998. Household survey nonresponse. New York: Wiley.

Groves, R., F. Fowler Jr, M. Couper, J. Lepkowski, E. Singer, and R. Tourangeau. (2011). Survey methodology (Vol. 561). New York: Wiley.

Groves R., and N. Fultz. 1985. Gender effects among telephone interviewers in a survey of economic attitudes. *Sociological Methods and Research,* 14:31–52.

Groves, R., and R. Kahn. 1979. Surveys by telephone. New York: Academic Press.

Hansen, K. M. 2006. The effects of incentives, interview length, and interviewer characteristics on response rates in a CATI-study. *International Journal of Public Opinion Research*, 19(1):112-21.

Hox, J., and E. De Leeuw. 2002. The influence of interviewers' attitude and behavior on household survey nonresponse: An international comparison. In *Survey nonresponse,* eds. M. Couper, R. Groves, W. De Heer, V. Kuusela, ... and J. Martin, 103-12. New York: Wiley.

Hox, J., E. De Leeuw, and Kreft, G. 1991. The effect of interviewer and respondent characteristics on the quality of survey data: a multilevel model. In *Measurement Errors in Surveys*, eds. P. Biemer, … and S. Sudman, 439-61. New York: Wiley.

Hox, J., E. De Leeuw, and G. Snijkers. 1998. Fighting nonresponse in telephone interviews: Successful interviewer tactics. In *Nonresponse in Survey Research. Zuma Spezial* 4, eds. A. Koch and R. Porst, 173-85. Mannheim.

Kühne, S. 2018. From strangers to acquaintances? Interviewer continuity and socially desirable responses in panel surveys. *Survey Research Methods*, 12(2):121-46.

Leckie, G., and C. Charlton. 2013. runmlwin - A Program to Run the MLwiN Multilevel Modelling software from within stata. *Journal of Statistical Software*, 52(11):1-40.

Lipps, O. 2008. A note on interviewer performance measures in centralised CATI surveys. *Survey Research Methods,* 2(2):61–73.

Lipps, O. 2009. Cooperation in centralised CATI household panel surveys—A contact-based multilevel analysis to examine interviewer, respondent, and fieldwork process effects. *Journal of Official Statistics,* 25(3):323-38.

Lipps, O. 2010. Does interviewer-respondent socio-demographic matching increase cooperation in centralized CATI household panels? *Survey Practice* August 2010, AAPOR.

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

Lynn, P., O. Kaminska, and H. Goldstein. 2014. Panel attrition: How important is interviewer continuity? *Journal of Official Statistics*, 30(3):443-57.

Maynard, D., and N. Schaeffer. 1997. Keeping the gate: Declinations to the request to participate in a telephone survey interview. *Sociological Methods and Research,* 26:34–79.

Mood, C. 2010. Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European Sociological Review,* 26:67–82.

Morton-Williams, J. 1993. *Interviewer approaches*. Aldershot: Dartmouth.

Oksenberg, L., and C. Cannell. 1988. Effects of interviewer vocal characteristics on nonresponse. In *Telephone Survey Methodology.* eds. R. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls, and J. Waksberg, 257-69. New York: Wiley.

Olson, K. 2007. An investigation of the nonresponse-measurement error nexus. PhD Dissertation, University of Michigan.

Singer, E., M. Frankel, and M. Glassman. 1983. The effect of interviewer characteristics and expectations on response. *Public Opinion Quarterly*, 47(1):68-83.

Tillmann, R., M. Voorpostel, E. Antal, U. Kuhn, F. Lebert, V.-A. Ryser, O. Lipps, and B. Wernli. 2016. The Swiss household panel study: Observing social change since 1999. *Longitudinal and Life Course Studies*, 7(1):64-78.

van der Vaart, W., Y.Ongena, A. Hoogendoorn, and W. Dijkstra. 2006. Do interviewer voice characteristics influence cooperation rates in telephone surveys? *International Journal of Public Opinion Research*, 18(4):488-99.

Vassallo, R., G. Durrant, P. Smith, and H. Goldstein. 2015. Interviewer effects on non-response propensity in longitudinal surveys: a multilevel modelling approach. *Journal of the Royal Statistical Society: Series A*, 178:83–99.

Verger, P., E. Baruffol, and M. Rotily. 2001. Using the telephone for mental health surveys: An analysis of the impact of call rank, non-response and interviewer effect: *Population*, 6:987-1010.

Vercruyssen, A., C. Wuyts, and G. Loosveldt. 2017. The effect of sociodemographic (mis) match between interviewers and respondents on unit and item nonresponse in Belgium. *Social Science Research*, 67:229-38.

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

West, B. and A. Blom. 2017. Explaining Interviewer Effects: A Research Synthesis, *Journal of Survey Statistics and Methodology*, 5(2):175–211.