

Interviewer Effects on Wave Non-response in a Longitudinal Survey: A Multilevel Analysis

Vassallo, Rebecca¹; Durrant, Gabriele¹; Smith, Peter¹; and Goldstein, Harvey²

¹ Southampton Statistical Sciences Research Institute, University of Southampton

² University of Bristol

E-mail: rv1g09@soton.ac.uk

Introduction

The continuous decline in survey participation rates focuses attention on the efficiency losses and possible bias in survey estimates arising from non-response, and motivates studies aiming to address these problems. Prominent among such studies are those analysing interviewer effects, which aim to reduce non-response at the design stage or during data collection. In their theoretical framework for household non-response, Groves & Couper (1998) identify five factors that influence the process of refusal, of which interviewer attributes and the interviewer-respondent interaction represent two out of only three factors which the researcher has some control over. Studies focused on interviewer effects reflect the understanding that interviewers play an important role in introducing the survey concept, engaging the respondent, addressing any queries, and ultimately gaining response. They also acknowledge the possible influence the research agency can have in minimising negative interviewer effects through effective policies and management strategies. These studies have confirmed the presence of significant interviewer effects on non-response in both cross-sectional and longitudinal surveys. However, very few consistent interviewer attributes associated with higher success rates have been found. This partly reflects the lack of detailed information on interviewers available in many of these studies.

The purpose of this study is the analysis of interviewer effects on wave non-response at a later wave in the life of a longitudinal study. This study aims to identify interviewer characteristics which explain some of the interviewer variation in non-response as well as interaction effects between sample members and interviewers. It will also consider the relative importance of interviewers across two waves. The identification of significant interviewer socio-demographic characteristics, work history, personality traits and job attitudes may provide important guidelines for more effective interviewer recruitment, training, appraisal and work allocation. Just as there is some evidence, albeit weak, that a change for a more experienced interviewer may be beneficial (Lynn et al., 2011), significant interaction effects between respondent and interviewer characteristics may suggest other criteria, such as matching of interviewer and respondent on certain characteristics, by which to determine successful interviewer changes at later waves. Whilst household-level variables will be considered, the main purpose of their inclusion in the model will be as control variables, in the exploration of interactions between householders and interviewers which are predictive of non-response, and to offer some control for possible area confounding, to the extent that contextual effects are area averages of household-level characteristics.

Literature Review

Studies considering the effect of interviewer changes across waves of a longitudinal study using observational data seem to confirm the common belief in survey administration that interviewer continuity is conducive to higher contact and response rates (e.g. Schatteman, 2000). As Campanelli & O'Muircheartaigh (2002) emphasise, the main limitation in non-experimental studies of interviewer continuity effects is that a change in interviewer may reflect non-random processes, such as a respondent move or an interviewer resignation, which may be differential across panel response categories. The study by Lynn et al. (2011), which uses experimental data providing control for both interviewer continuity and experience, shows that in all situations except one – when the wave 1 mid-grade interviewer is replaced by a low-grade interviewer in the subsequent wave – interviewer change either has no impact or reduces the probability of household refusal.

While some studies have found significant effects for demographic or socio-economic characteristics of interviewers as predictors of non-response – such as age (Hansen, 2006), gender (Hox & de Leeuw, 2002) and educational attainment (Haunberger, 2009), other studies including these same variables did not corroborate these relationships. The positive influence of the interviewer's experience (Hox & de Leeuw, 2002) on the probability of a household responding has been confirmed across various studies. Other variants of experience, such as duration in employment on the specific survey considered (Hansen, 2006), years working with current survey agency (Campanelli et al., 1997), and pay grade with current survey agency (Durrant et al., 2010), were considered in the literature, showing the same relationship as that outlined for total experience. However, as Durrant et al. (2010) highlight, this consistent finding does not necessarily confirm a causal effect, as job/profession tenure may simply reflect a selection bias, with better performing staff choosing to remain in the profession or simply being offered work for the long term. There is evidence that interviewers who are confident in their ability to obtain a response and who perceive reluctance as a changeable state of mind achieve better response rates (e.g. Durrant et al., 2010). However, Sinibaldi et al. (2009), who use data from the same NatCen interviewer survey in their analysis, find little evidence of the impact of attitudes on the persuasion of reluctant respondents on respondent refusal. No consistent pattern has yet emerged on the relationship between the interviewer's personality traits and the propensity of the contacted household to refuse. This may either indicate that the measurement of personality traits may be fraught with error, and that the tools used for the general population are not adequate for the analysis of interviewers, or simply that fixed personality traits are not predictive of the success of the doorstep interaction.

At present there is very little research exploring the potential benefits of matching respondents to interviewers on the basis of specific socio-economic, demographic or personality attributes. In his analysis of respondent refusal in the second wave of a Dutch survey on the transition from education to the labour market, Schatteman (2000) considered whether matching of respondents and interviewers on socio-economic characteristics contributed to higher response rates. In both the bivariate and full logistic models for respondent refusal, variables indicating whether the respondent and interviewer were of the same gender, educational level and occupational category were not found significant for $\alpha=0.05$. On the other hand, in the study by Moorman et al. (1999), descriptive statistics showing the cooperation rates for a case-control study of breast cancer for sample members allocated an interviewer of the same ethnicity compared with those sample members discordant with their interviewer on ethnicity indicate positive effects of matching on ethnicity. Some support for gender and educational achievement cross-level interaction effects comes from the multilevel cross-classified analysis of refusal in six household surveys in the UK by Durrant et al. (2010).

Pickery et al. (2001) attempt to identify whether the interviewer at the current wave or the interviewer at the previous wave has the greatest impact on non-response at the wave being considered. When analysing non-response at the second wave conditional on contact using a cross-classified logistic model where respondents are classified by the interviewer at first wave and the interviewer at the second wave, these authors found the first wave interviewer variance to be significant, whilst the random effect for the second wave interviewer was not. The recent work by Lynn et al. (2011) uses multiple membership models for investigating the relative importance of interviewers from different waves. Their analysis showed the more recent interviewer to have the biggest influence on the propensity to respond.

Data

The data for this study comes from the longitudinal Family and Children Study (FACS), which gathers information on the health and socio-economic status of households with children in the United Kingdom (Lyon et al., 2007). This dataset benefits from linkage to administrative data and detailed information on interviewers from a survey of both waves 7 and 8 interviewers which, in addition to survey data from wave 7 and some participation history variables, provide a varied selection of potential explanatory variables for wave 8 non-response. The interviewer survey, a postal self-completion survey administered in May 2008 addressed to all interviewers who had worked for NatCen at some point since the start of 2006 (equivalent to FACS wave 10), provides rich data for explaining variation in non-response across interviewers. The survey topics include interviewing experience, job expectations and appraisal, flexibility in working hours, personality traits, interpersonal skills, and views on the persuasion of hesitant sample members. Four fifths of eligible interviewers completed the interviewer survey.

The initial dataset includes all 7285 wave 8 cases that had participated in wave 7. The conditioning on wave 7 is to allow information on the respondents to be obtained from the previous wave for all cases. This restriction

excludes cases which had responded to the survey during waves prior to wave 7 and were then maintained or re-introduced in the sample at wave 8. New cases included in the sample for the first time at wave 8 of the study are also excluded since no information would be available on such cases. A wave number represents the number of survey episodes since its incipience in 1999. Due to the varying initial wave and number of interviews for cases considered in the dataset, participation history variables will be included as controls in the multilevel models. Non-response may arise from either non-contact or refusal at a later stage. The small number of cases with the former outcome, and the added model complexity resulting from a distinction between these two outcomes, led to the decision to analyse response conditional on contact. A complete case analysis was carried out, with cases with missing values for interviewer variables being dropped. The main reduction in the number of cases was due to the interviewers with unit non-response in the interviewer survey. The final dataset included 5932 cases nested within 275 wave 8 interviewers.

Methodology

The lack of an interpenetrated sample design, which can be attained through the random allocation of interviewers to spatial areas, may have led to confounding between interviewer and area effects. As Von Saden (2004) points out, partial interpenetration still allows the estimation of these parameters, albeit with higher uncertainty. For wave 8 cases there are no primary sampling units (PSUs) in which only one interviewer was allocated work, and approximately 82% of interviewers were allocated households provenance from at least two different PSUs. With this data structure, a multilevel cross-classified model specification which considers both interviewer and area random terms can allow for a distinction between interviewer and area effects. A cross-classified multilevel logistic model was therefore fitted, modelling non-response at wave 8, with random effects for wave 7 and wave 8 interviewers and PSUs. The models were estimated using MLwiN. For the exploratory modeling, for example when considering which fixed effects to include in the model, 2nd order PQL was used to speed up the analysis and MCMC methods were employed for exploring the random structure.

Potential predictors were chosen on the basis of theoretical grounds and a review of significant predictors in the literature. Logistic regressions with only one explanatory variable at a time guided the reduction of the list of potential predictors. A forward selection strategy was used. The first step in the model construction was the specification of the hierarchical levels. Next, groups of explanatory variables were considered in the following order: participation history, household, area and interviewer variables. The order in which the predictor variables for each group were added was determined by the significance of each variable in a separate model with the interviewer 8 random effect and the between-waves 'Interviewer Change' variable (defined below) for the Wald test for continuous and binary variables and the joint chi-squared test for other categorical variables. For all groups excluding interviewer variables, the decision regarding which variables to include in the model was made on the full dataset (cases with unit non-response for the interviewer survey included). These variables were then forced in the restricted dataset irrespective of whether they had become non-significant. In cases of discrepancy, descriptive statistics for the variables in question indicated a similar frequency distribution in the full and restricted dataset, suggesting that non-significance had simply resulted from reduced power. Interviewer sex and the grade/experience variables, predictive of non-response in the interviewer survey, were to be included as control variables in the models for the restricted dataset irrespective of their significance.

Results and Discussion

A multilevel model with the most complex random structure specification, a cross-classification of areas, interviewers at wave 7, and interviewers at wave 8, failed to converge. A simplified version of the multilevel cross-classified model with both interviewers at wave 7 and 8 but excluding area effects yielded numerically unstable results, which suggests that the two interviewer random effects are near non-identifiable. This may reflect a potential misspecification of the interviewer-level structure, since the two models cannot correctly distinguish between and estimate the effects for cases allocated to the same interviewer and others who experience a change of interviewer. To include both random interviewer-level effects simultaneously and estimate the relative effect of each wave interviewer a multilevel multiple membership model may be used (Lynn et al., 2011) and further research will explore this option. In this paper, for simplicity, we decided to use a multilevel cross-classified model including only one of the interviewer-level random effects. The wave 8 interviewer random effect was retained as the initial modeling indicated that this was the more important interviewer. Also this is the interviewer corresponding to the wave when the non-response occurs and for whom more reliable interviewer-level information from the survey is available, which is the focus of this paper. A change in the interviewer between waves 7 and 8 is represented by the indicator variable 'Interviewer Change'

included as a fixed effect in the model. Exploring a multilevel cross-classified logistic model with wave 8 interviewers and areas at the higher level, the area random variance is not significant any more (the same holds in a cross-classified model with wave 7 interviewer and area random effects). We therefore excluded the area random effects and only accounted for interviewer at wave 8 in the multilevel model.

Table 1 presents estimates of the variance and standard error of the interviewer 8 random effect at each stage of the modelling, up until the inclusion of the interviewing history variables, at which point the random effect of the wave 8 interviewer becomes non-significant and is subsequently removed from the model. For the first model specified, not including any fixed effects, the interviewer variance accounts for around 8% [$0.302/(0.302 + 3.29)$] of the total variation in refusal at wave 8. As can be noticed from Table 1, although the interviewer variance is reduced slightly when including the participant history and the household-level variables, the more substantial decreases come from the interviewer-level fixed effects. The variables included in the model explain all of the interviewer random effect. This compares to findings in O'Muircheartaigh & Campanelli (1999) who found interviewer effects were no longer significant once interviewer-level variables were controlled for.

Fixed Effects Parameters (no. of parameters)	Variance	S.E.	Wald Chi	df	p-value	1/2 p-value
Null	0.302	0.081	13.899	1	0.000	0.000
Added Interviewer Change (1)	0.229	0.073	9.845	1	0.002	0.001
Added Participation History Variable (6)	0.213	0.071	8.903	1	0.003	0.001
Added Respondent/Household Variables (12)	0.197	0.070	7.924	1	0.005	0.002
Added Interviewer 8 Experience/Grade & Sex (10)	0.118	0.060	3.824	1	0.051	0.025
Added Interviewer 8 Work History Variable (6)	0.073	0.056	1.746	1	0.186	0.093

Table 2 presents the estimated coefficients of the final logistic regression model. For space reasons, the results of household-level characteristics are not discussed. The variables describing the geographical area of the household, such as indicator for the UK regions, the London indicator, and various respondent neighbourhood perception variables were found not to be significant, after controlling for other household-level variables. This result supports the conclusion that, after controlling for household and interviewer effects, area effects are negligible. Prior to discussing significant fixed interviewer-level effects, an overview of non-significant interviewer variables will be presented. The variable working field area, indicating the geographical area within which the interviewer may be allocated work, was not found significant. This provides further evidence that there are no area effects to non-response. Demographic variables, such as gender and age, were also not found significant. Variables indicating the interviewer satisfaction rating with various aspect of the survey agency, the importance they allocate to various aspects of a job such as monetary compensation and flexibility, and indexes of time availability for this job all were not significant. Similarly to the analysis by Sinibaldi, et al. (2009) using data from the same NatCen interviewer survey, interviewer attitudes on the persuasion of reluctant respondents were not found to be predictive of respondent refusal.

In agreement with previous observational studies, an interviewer change between waves was found to be positively associated with refusal. However, a causal relationship cannot be inferred from this finding. Rather than a random allocation, an interviewer change may reflect a respondent move or the resignation of an interviewer. For example, if in survey administration an interviewer change occurs for respondents with a change in their address, to the extent that respondents who have just experienced a house move represent a type of individual who is more or less likely to respond to the survey request, then using results from observational data to analyse the effect of a 'random' interviewer change will be erroneous.

Both interviewer grade and years of experience were highly significant predictors when included one at a time in the model. While years of experience simply constitute a consistent measure of exposure in the profession, the grade held within the organisation is a more subjective construct and may reflect experience, educational background and skill. In fact, while generally interviewers with higher grades were more likely to have many years of experience, for a particular grade there were interviewers with a range of years of experience. However, the variables are positively correlated and their simultaneous inclusion would create problems with multicollinearity. Consequently, a variable distinguishing between different years of experience bands for the same grade was created. All categories of this variable show a higher propensity of refusal compared to the highest grade – T – interviewers and this higher prediction is significant for $\alpha=0.05$ for all categories except grade R. As hypothesised, the predicted impact of being in the lower experience group for grades C, D and S on

refusal is greater than the impact of being in the higher experience group for the same grade. The difference between these groups of dummy variables is significant for grade C for $\alpha=0.05$ and for grade S for $\alpha=0.1$. These results confirm the positive influence of longer years of experience and higher grade on household response.

Table 2: Estimated Coefficients for the Final Logistic Model (Wave 8 interviewer effect was no longer significant)									
Variable	Reference	Dummy	β	S.E.	β /S.E.	Wald Chi	df	p-value	Odds
Interviewer Change	Same	Change	0.385	0.110	3.500	12.357	1	0.000	1.000
Participation History Variables									
First Wave for Respondent	Wave 7	Wave 1	-0.814	0.165	-4.933	39.381	1	0.000	1.000
		Wave 2	-0.820	0.197	-4.162				1.000
		Wave 3	-0.894	0.168	-5.321				1.000
		Wave 4	-1.026	0.233	-4.403				1.000
		Wave 5	-0.439	0.199	-2.206				1.000
		Wave 6	-0.453	0.196	-2.311				1.000
Respondent/Household Variables									
Ethnicity	Nonwhite & Missing	White	-0.596	0.151	-3.947	15.522	1	0.000	1.000
Any Vocational or Academic Qualifications	Yes	No	0.285	0.143	1.993	3.975	1	0.046	1.047
Age of Youngest Child	No dependent children & 16-18 year olds	0-4 years	-0.529	0.173	-3.058	12.263	3	0.007	1.007
		5-10 years	-0.441	0.170	-2.594				1.000
		11-15 years	-0.188	0.173	-1.087				1.000
Heating Problems in the Dwelling	No & Don't know	Yes	0.265	0.213	1.244	1.548	1	0.213	1.238
Gender	Female	Male	-1.228	0.521	-2.357	5.643	1	0.018	1.018
		Accommodation Type	Detached house	Semi-detached house	-0.266	0.126	-2.111	5.357	4
		Terraced house	-0.269	0.139	-1.935				1.000
		Flat or maisonette - purpose built & Other	-0.190	0.210	-0.905				1.000
		Flat or maisonette - conversion	-0.398	0.498	-0.799				1.000
Household Size			0.088	0.045	1.956	3.784	1	0.052	1.053
Interviewer 8 Administrative Variables									
Grade Experience	Grade T (Highest grade)	Grade A (Lowest Grade)	2.155	0.847	2.544	34.722	9	0.000	1.000
		Grade B	2.292	0.741	3.093				1.000
		Grade C, 0-4 years experience	2.260	0.720	3.139				1.000
		Grade C, 5+ years experience	1.704	0.727	2.344				1.000
		Grade D, 0-4 years experience	2.048	0.733	2.794				1.000
		Grade D, 5+ years experience	1.812	0.733	2.472				1.000
		Grade R	0.661	1.257	0.526				1.000
		Grade S, 0-4 years experience	1.974	0.741	2.664				1.000
		Grade S, 5+ years experience	1.451	0.743	1.953				1.000
Sex	Female	Male	0.047	0.106	0.443	0.201	1	0.654	1.923
Interviewer 8 Work History, Time Availability, Attitudes towards Refusal, Work Priorities, Satisfaction with Job Variables									
Interviewing Work History - Work status with another survey agency & Experience with other (phone, marketing) survey interviewing	Never worked for another survey agency	Currently working & Done other survey interviewing	-0.104	0.186	-0.559	18.317	6	0.005	1.006
		Working since 01/01/06 & Done other survey interviewing	0.000	0.300	0.000				1.000
		Worked prior to 01/01/06 & Done other survey interviewing	0.447	0.124	3.605				1.000
		Currently working & Never done other survey interviewing	-0.274	0.334	-0.820				1.000
		Working since 01/01/06 & Never done other survey interviewing	-0.700	0.498	-1.406				1.000
		Worked prior to 01/01/06 & Never done other survey interviewing	0.187	0.177	1.056				1.000
Interviewer 8 Personality Traits Variables									
Is talkative	7 (Applies perfectly to me)	1 (Does not apply to me at all), 2, 3	0.026	0.170	0.153	8.559	3	0.036	1.036
		4, 5	-0.152	0.145	-1.048				1.000
		6	-0.434	0.178	-2.438				1.000
Worries a lot	1 (Does not apply to me at all)	2, 3	0.498	0.194	2.567	11.252	4	0.024	1.024
		4	0.224	0.210	1.067				1.000
		5, 6	0.541	0.204	2.652				1.000
		7 (Applies perfectly to me)	0.515	0.333	1.547				1.000

Variables indicating whether the interviewer had done any other type of work – including other (phone or marketing) types of interviewing, interviewing for another survey agency, and non-interviewing work were all found significant when considered one at a time. Only one of these variables could be retained in the model because of the high correlation between these variables. Those who have never done other types of interviewing are more likely to have never undertaken interviewing with another survey agency. Contrary to what might be expected, those who have another job (other than interviewing) are also more likely to be working with another survey agency. This could indicate some problems with the interpretation of the question relating to whether they had another non-interviewing job, and consequently this variable was excluded. With the other two variables, an interaction variable was created summarising the interviewer work history. This interaction variable shows higher predicted odds of refusal for interviewers with experience with other types of interviewing compared to those with no such experience, and higher odds for interviewers who had worked for another survey agency prior to 2006 and done other survey interviewing compared with interviewers with no work history with any other survey agency. Experience in different interviewing modes or research areas may have a negative influence on the performance of face-to-face interviewing, simply because different techniques apply. The

explanation for the negative effect of previous work with other survey agencies is unclear and more data on interviewing work history is necessary to explore this relationship further.

The inclusion of seventeen personality trait items was considered. Initially the possibility of creating a composite measure for these items using a principal component analysis was explored. However, due to the relatively large variance proportion left unexplained, and the lack of congruence between the weights assigned and the five theoretical personality dimensions: openness, extraversion, agreeableness, neuroticism and conscientiousness, an alternative method was chosen. Only items which were significant in a model including only the random effect and the interviewer change dummy would be considered for inclusion in the full model. A less stringent criterion, $\alpha=0.1$, resulted in only four items for further consideration. This result suggests the consideration that rather than a personality dimension, a particular interviewer trait may influence the interaction process and consequently the outcome of the participation request. The personality traits retained in the final model seem to indicate that households are more likely to refuse if they are allocated interviewers with a high self-rating of neuroticism in terms of worrying tendencies or interviewers who perceive themselves as either not talkative or also extremely talkative.

Conclusion

While confirming previous findings on the relationship between interviewer experience and non-response, this paper sheds light on the need for further data on the work history of interviewers and supports the hypothesis that fixed interviewer personality traits are not important predictors of household non-response. The paper presents work in progress. Further work includes the consideration of interaction effects and interviewer skills as additional fixed effects and the use of multiple membership models to include both interviewer 7 and 8 as part of the random structure.

REFERENCES

- Campanelli, P. & O'Muircheartaigh, C. (1999). Interviewers, interviewer continuity, and panel survey nonresponse. *Quality & Quantity*, 33, 59-76.
- Campanelli, P. & O'Muircheartaigh, C. (2002). The importance of experimental control in testing the impact of interviewer continuity on panel survey nonresponse. *Quality & Quantity*, 36, 129-144.
- Campanelli, P., Sturgis, P. & Purdon, S. (1997). *Can you hear me knocking: An investigation into the impact of interviewers on survey response rates*. Final Report for UK ESRC Grant R000235776, London, National Centre for Social Research.
- Durrant, G. B., Groves, R. M., Staetsky, L. & Steele, F. (2010). Effects of interviewer attitudes and behaviors on refusal in household surveys. *Public Opinion Quarterly*, 74, 1-36.
- Groves, R.M. & Couper, M.P. (1998). *Nonresponse in household interview surveys*. New York: Wiley.
- 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, 112-121.
- Haunberger, S. (2009). The effects of interviewer, respondent and area characteristics on cooperation in panel surveys: a multilevel approach. *Quality & Quantity*, 44, 957-969.
- Hox, J. & De Leeuw, E. (2002). The influence of interviewers' attitude and behavior on household survey nonresponse: An international comparison. In Groves, R. M., Dillman, D. A., Eltinge, J. L. & Little, R. J. A., *Survey Nonresponse* (pp.103-119). New York: Wiley.
- Lynn, P., Kaminska, O. & Goldstein, H. (2011). Panel attrition: How important is it to keep the same interviewer? *ISER Working paper Series, 2011-02, Institute for Social & Economic Research, ESRC*.
- Lyon, N., Mangla J., Tait, C. & Scholes, S. (2007). *Families and Children Study (FACS) 2005, Wave 7 Technical Report*. National Centre for Social Research: London.
- Moorman, P. G., Newman, B., Millikan, R. C., Tse, C. J. & Sandler, D. P. (1999). Participation rates in a case-control study: The impact of age, race, and race of interviewer. *Annals of Epidemiology*, 9(3), 188-95.
- Pickery, J., Loosveldt, G. & Carton, A. (2001). The effects of interviewer and respondent characteristics on response behavior in panel surveys - A multilevel approach. *Sociological Methods & Research*, 29, 509-523.
- Schatteman, T. (2000). *Do interviewer characteristics influence respondents participation in panel surveys*. Paper presented at 2004 JRSS/ESRC conference on Statistical methods for attrition and non-response in social surveys, London.
- Sinibaldi, J., Jäckle, A., Tipping, S. & Lynn, P. (2009). Interviewer characteristics, their doorstep behaviour, and survey cooperation. Proceedings of the Survey Research Methods Section of the 2009 Joint *Statistical Meetings*.
- Von Sanden, N. D. (2004). *Interviewer effects in household surveys: estimation and design*. Unpublished PhD thesis, School of Mathematics and Applied Statistics, University of Wollongong.