ISER Working Paper Series

Panel Attrition: How Important is it to Keep the Same Interviewer ?

Peter Lynn
Institute for Social and Economic Research
University of Essex

Olena Kaminska
Institute for Social and Economic Research
University of Essex

Harvey Goldstein
Centre for Multilevel Modelling
University of Bristol

No. 2011-02
January 2011

INSTITUTE FOR SOCIAL
& ECONOMIC RESEARCH

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

www.iser.essex.ac.uk

Non-Technical Summary

Panel surveys involve interviewing the same set of people repeatedly over time. Researchers who carry out such surveys often discuss whether it is beneficial to assign to each sample member the same interviewer who interviewed them on the previous occasion. Furthermore, they would like to know *how* beneficial, as there are often costs associated with assigning the same interviewer.

One possible beneficial effect of assigning the same interviewer is that the sample members might be more willing to be interviewed again. In this paper, we explore the extent to which that is true.

There have been previous studies of this topic, but most of those have used non-experimental data, in which a changed interviewer is associated with the previous interviewer having dropped out for a particular reason such as moving home or finding a different job. In such circumstances, comparing sample members assigned a different interviewer with sample members assigned the same interviewer does not provide a fair comparison, as we are not comparing like with like. In this paper we overcome that problem by using data from a carefully designed experiment. In our experiment, sample members were randomly assigned either the same interviewer or a different one. Additionally, we took account in the design of differences in experience between continuing interviewers and replacement interviewers.

We study the extent to which assigning a different interviewer affects the chances of a sample member refusing an interview or taking part. We also examine whether this effect differs between subgroups of the population.

Panel Attrition: How Important is it to Keep the Same Interviewer?

Peter Lynn
Institute for Social and Economic Research, University of Essex

Olena Kaminska
Institute for Social and Economic Research, University of Essex

Harvey Goldstein
Centre for Multilevel Modelling, University of Bristol

Abstract: We assess whether the probability of a sample member participating at a particular wave of a panel survey is greater if the same interviewer is deployed as at the previous wave. Previous research on this topic mainly uses non-experimental data. Consequently, a) interviewer change is generally non-random, and b) continuing interviewers are more experienced by the time of the next wave. We designed a balanced experiment in which both interviewer continuity and experience are controlled. Multilevel multiple membership models are used to explore the effects of interviewer continuity on refusal rate and response rate as well as interactions of interviewer continuity with other variables.

Keywords: longitudinal survey, multiple membership multi-level model, non-response, refusal

JEL Codes: C81, C83

Acknowledgments: The contribution of the first author is funded by the ESRC via resource centre funding for the UK Longitudinal Studies Centre. The contributions of the second and third authors are funded by the ESRC Survey Design and Measurement Initiative via a research grant to Prof. John Bynner for the project "Solving the problem of attrition". We are grateful to the National Centre for Social Research (NatCen) for implementing the survey field work and providing the data.

Contact: plynn@essex.ac.uk

1. Introduction: Interviewer Continuity

For longitudinal surveys, the perceived benefits of having the same interviewer assigned to a sample member at each wave is a factor that can drive important aspects of survey planning and design. Many survey researchers believe that interviewer continuity – particularly for face-to-face surveys - brings benefits, primarily in terms of continued co-operation, though possibly also in terms of improved measurement. Consequently, they may sometimes be willing to prioritise assignment of the same interviewer as at the previous wave, even when alternative strategies may be less costly or more convenient. For example, when a respondent moves home between waves the researcher may prefer to deploy the same interviewer even if she now has to travel 30km to the address, rather than a different interviewer who lives only 5km away. Considerations of interviewer continuity can also influence decisions about whether to award a survey data collection contract to the existing contractor or to an alternative bidder, as the latter scenario will typically result in considerably less, if any, interviewer continuity at the next wave. Therefore it is important for survey managers and survey commissioners to understand the value of interviewer continuity, in order to make cost-effective decisions.

There are plausible theoretical reasons why interviewer continuity may increase re-interview propensity. These reasons relate to trust, tailoring and consistency.

Trust of the survey interviewer on the part of the sample member is an important influence on whether or not the sample member chooses to co-operate (Beerten and McConaghy 2003, Hox and De Leeuw 2002, Morton-Williams 1993). It is plausible that a sample member will, on average, trust a continuing interviewer more than a replacement one. This should occur if the sample member has experienced no negative consequences (such as crime or unwanted sales calls) of having previously invited this person into their home to interview them. Heightened trust, and therefore greater co-operation propensity, would thus be associated with interviewer continuity.

Tailoring of communication and tactics by interviewers increases the chances of participation (Groves, Cialdini and Couper, 1992). A continuing interviewer is potentially able to draw upon prior knowledge of relevant characteristics of the sample member and his or her household that would not be available to a replacement interviewer. This additional knowledge could make the continuing interviewer better at tailoring both his or her calling

patterns and the arguments that he or she uses to persuade the sample member to take part. This additional ability to tailor could therefore lead to continuing interviewers achieving both greater contact propensity and greater co-operation propensity.

Consistency is generally seen as a desirable personal trait (Cialdini, 2008, chapter 3). After committing oneself to a position one should be more willing to comply with requests for behaviours that are consistent with that position. This is a likely explanation for the foot in the door effect in surveys (Freedman and Fraser 1966; Groves and Couper 1998). A sample member who has previously agreed to an interview may be more likely to agree to a similar request in order to appear consistent if it is the same interviewer making the request. Thus, a greater influence of the norm of consistency could result in greater co-operation propensity being associated with continuing interviewers.

However, although it is plausible that interviewer continuity might have a positive effect on response rates, other things being equal, there is very little empirical evidence on this point. A number of longitudinal surveys observe that re-interview rates are higher amongst cases where the same interviewer makes the approach at a subsequent wave (e.g. Rendtel 1990; Schräpler 2001; Waterton & Lievesley 1987). But such association does not imply causality. In particular, on face-to-face surveys where interviewers tend to work in specific geographic areas, it is quite possible that interviewer continuity and respondent co-operation rates have some common causes. For example, these may be associated with geographical mobility or employment mobility in the local area. A study which used more sophisticated analysis techniques found no effect of interviewer continuity on refusal rate (Pickery et al 2001). To our knowledge, only one previous study has used a randomised design to attempt to assess the effect of interviewer continuity on re-interview rate on a face-to-face survey. This study involved an inter-penetrated design at wave 2 of the British Household Panel Survey in 1992. No effect of interviewer continuity on re-interview rate was found either at wave 2 (Campanelli & O'Muircheartaigh 1999) or at waves 3 or 4 (Campanelli & O'Muircheartaigh 2002).

Aside from confounding effects of interviewer continuity with area effects, we note two additional limitations of previous studies of interviewer continuity. As far as we are aware, neither have been noted in the literature:

- Interviewer continuity is, by definition, associated with increasing interview experience. For example, those interviewers who interview the same respondents

over three waves of an annual panel survey all have two years more interviewing experience at the time of wave 3 than they had at the time of wave 1. In cases where there is no interviewer continuity, replacement interviewers are therefore likely to be less experienced, on average, than continuing interviewers. Experience is known to be associated with re-interview propensity and should therefore be controlled in any study of the effect of interviewer continuity;

- The effect of interviewer continuity on re-interview propensity could be positive for some respondents (those who have a good rapport with their interviewer, perhaps) and negative for others (those with a poor rapport). Thus, regardless of whether or not there is a main effect of interviewer continuity, there may be an interaction of interviewer continuity with variables associated with rapport or 'liking' the interviewer. Identification of such interactions could be helpful for survey organisations faced with practical decisions about allocation of panel survey cases to interviewers.

In this paper we examine the effect of interviewer continuity on re-interview propensity using new experimental data. Our experimental design simultaneously controls continuity and interviewer experience. Additionally, our analysis considers interactions of respondent characteristics with interviewer continuity. We believe that these are two original contributions to the literature.

2. Study Design

The March-April 2008 round of the National Centre for Social Research (NatCen) Omnibus Survey involved interviewing a random sample of the population aged 16 and over living in the United Kingdom. We shall refer to this survey as "wave 1". Respondents who agreed to be re-contacted for further research (n=1,188) formed the sample for the study reported here. Sample respondents were allocated to one of four treatment groups for a follow-up interview in March-May 2009 ("wave 2"). The four treatment groups were:

- Same interviewer
- Different interviewer of the same grade
- Different interviewer of each of two different grades (grade was defined as a 3-category variable)

Thus the two control variables are interviewer continuity (whether or not the same interviewer is assigned to the sample case at both waves) and interviewer grade (in three categories). We believe that interviewer grade is a good measure of the relevant characteristics that can differ between continuing and different interviewers in non-

experimental studies, namely those aspects of ability that are associated with length of time working as an interviewer. This is because NatCen interviewers are promoted to higher grades based on a number of criteria, some of which are related to experience *per se* and others of which are related to performance. Thus, grade would seem to capture the aspects of interviewer experience that are relevant to response propensity (organisational skills, ability to perceive the concerns and circumstances of respondents, ability to persuade). Of course, any association between interviewer experience and response rates could be due to either a selection effect (less successful interviewers quit interviewing) or a learning effect (interviewers become more successful over time as they gain new skills). Carton and Pickery (2010) find support for dominance of the selection effect. We do not address the cause of any association. Our intention is simply to control differences between continuing and different interviewers in characteristics that influence response propensity, regardless of the cause of those differences.

Allocation to treatment began by allocating each continuing interviewer to one quarter of his or her wave 1 respondents. This was done at random except in a few cases (mainly very rural areas) where assignment of a different interviewer for a random subset of cases would have been prohibitively expensive, in which case purposive allocation of the appropriate number of wave 1 respondents was carried out, based on geographical location. Remaining respondents were then allocated to other interviewers of different grades, producing the distribution in table 1. The effect of interviewer promotion between waves is shown in table 2 and illustrates the importance of controlling interviewer grade.

Table 1: Balanced sample design: interviewer continuity and interviewer grade

Number of assigned wave 2 cases	**Different Interviewer**			**Same Interviewer**	**Total**
	Lowest grade 2009	Middle grade 2009	Highest grade 2009	All grades 2009	
Lowest grade 2008	97	117	131	115	460
Middle grade 2008	114	100	105	115	434
Highest grade 2008	73	75	69	77	294

Table 2: Grades at each wave amongst continuing interviewers

Number of assigned wave 2 cases	**Same Interviewer**			**Total**
	Lowest grade 2009	Middle grade 2009	Highest grade 2009	
Lowest grade 2008	57	58	0	115
Middle grade 2008	0	98	17	115
Highest grade 2008	0	0	77	77

The wave 2 interview was introduced as a survey about safety on public transport, consisting primarily of a module of questions on this topic that had been asked also at wave 1. Socio-demographic and classificatory questions were also asked. Mean interview length was 21 minutes. Of the 1,188 issued sample cases, 11 were found to be ineligible for re-interview (deceased or moved out of the UK). Of the remainder, 844 were successfully interviewed, 119 were not contacted and 179 refused the wave 2 interview. Other reasons for non-response accounted for the remaining 35 cases. Thus, amongst eligible cases, wave 2 contact rate was 90% and co-operation rate was 80%, giving an overall conditional wave response rate of 72%. In our analysis, we exclude 73 cases who had moved and for whom a new address could not be obtained, on the grounds that the interviewer is not able to influence the response propensity in such cases.

3. Analysis

We carry out two sets of analyses. For the first, the dependent variable of interest is propensity not to participate in the survey at wave 2, in other words the propensity to be a non-respondent. Analysis is restricted to the 1,104 cases assumed to be eligible for the wave 2 interview and for whom the known address is assumed to be correct. The second set of analyses is further restricted to the 1,058 contacted cases and the dependent variable of interest is propensity to refuse. We use multiple membership multilevel logistic models of propensity not to participate conditional on eligibility and of propensity to refuse

conditional on contact. For the models of non-response, the dependent variable is coded 1 if the sample member is a non-respondent at wave 2 and 0 if they participated. For the models of refusal, the dependent variable is coded 1 if the sample member refused at wave 2 and 0 if they co-operated. Thus, for both models positive coefficients indicate an increased propensity for the undesirable outcome.

A formal statement of the basic model is as follows:

$$logit\left(\pi_{i(j_1,j_2)}\right) = X_{i(j_1,j_2)}\beta + w_1 u_{j_1} + w_2 u_{j_2} \qquad (1)$$

$$w_1 + w_2 = 1$$

where $\pi_{i(j_1,j_2)}$ is the probability of a refusal (or non-response) for sample member i interviewed by interviewers j_1,j_2 respectively at waves 1 and 2 and the random effects are assumed normal. Further details for such models are given by Goldstein (2010, Chapter 13). In this model, conditional on the fixed effects in the model denoted by $X_{i(j_1,j_2)}\beta$, there are two random interviewer effects contributing to the response from waves one and 2 respectively, namely u_{j_1},u_{j_2}. The corresponding weights reflect the relative importance of the wave 1 and wave 2 interviewer. The overall interviewer effect is thus a weighted average of the two interviewers, or where there is no change in interviewer, simply the effect of that interviewer. We have chosen to assign the same wave 1 weights to each wave 1 interviewer and likewise for wave 2. One of the aims of our analysis is to determine the relative weights which result in the best fitting model (see below). We note that an alternative design that uses a cross-classification of wave 1 by wave 2 interviewer is not appropriate for our data since such a design does not distinguish the case where it is the same interviewer at both waves.

For model estimation we use Markov chain Monte Carlo (MCMC) estimation with orthogonal parameterisation and hierarchical centering with a burn-in length of 5,000 and 20,000 iterations implemented in MLwin 2.19 (Browne, 2009, Rasbash et al., 2009).

The key predictor variable is a 13-category variable that we refer to as "interviewer change." This variable has categories defined by the combination of interviewer grade at each wave and whether or not the same interviewer was assigned at each wave. The categories are defined by the fourteen non-zero cells in tables 1 and 2, with the exception that the two cells

in which a continuing interviewer was promoted to a higher grade between waves 1 and 2 are combined to form a single category. We tested all 78 pairwise contrasts between categories, with no formal correction for multiple comparisons. The large number of comparisons should therefore be borne in mind when interpreting the results.

Multilevel multiple membership models allow us to assign different relative weights to interviewers at wave 1 and wave 2. However, we are unable to determine the weights on *a priori* grounds. We are only aware of one previous study that considered the relative influence on wave 2 participation of the wave 1 interviewer and the wave 2 interviewer. Pickery et al (2001) found that the wave 1 interviewer had a stronger influence on wave 2 refusal propensity than the wave 2 interviewer, though this conclusion was based solely on a comparison of coefficients from separate models, without any formal test. We therefore use empirical methods to select appropriate importance weights. We fit models with alternative combinations of importance weights and choose the weights that produce the smallest Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002), i.e. the best-fitting model. As a first step, we identify the most appropriate weights by fitting 2-level models with interviewer change as the sole predictor variable. We then use those weights during subsequent model-fitting, until we have identified the final model, at which point we fit that model with alternative combinations of weights and again select the combination that results in the smallest DIC. The weights chosen at the first step for the models of refusal are 0.80 for 2009 interviewer and 0.20 for 2008 interviewer. For the models of non-response, the weights were 0.85 for 2009 and 0.15 for 2008.

Using the initial weights, we test all pairwise comparisons of categories of interviewer change. We then extend the model by testing for interactions of interviewer change with thirteen other potential predictor variables. We aim to include all categorical predictor variables (other than interviewer change) as dichotomies, as the model otherwise becomes over-parameterised when we include interactions with interviewer change. Few of the variables are naturally dichotomous so combination of categories is necessary. This is done by fitting simple logistic regression models of survey response with the variable in question (full version) as the sole predictor variable, first combining categories with estimated coefficients that were not significantly different from one another ($P > 0.10$) and subsequently, if necessary, combining categories with the smallest absolute difference in estimated β-coefficients until only two categories remain. In addition to the dichotomous

predictors, we have one continuous predictor, age. The thirteen resultant predictor variables are listed in table 3.

Before beginning the process of fitting models with both the interviewer change variable and other predictors, the interviewer change variable is reduced to nine categories in order to gain power. This is done by combining the three categories involving a different interviewer at wave 2 with a higher grade into a single category and similarly combining the three categories involving a different interviewer at wave 2 with a lower grade into a single category. This reduction is justified empirically by joint χ-square tests for all cases with a different interviewer of lower grade (2df, χ^2=0.58; p=0.75) and for all cases with a different interviewer of higher grade (2df, χ^2=1.86; p=0.39).

Table 3: Predictor variables tested for interaction with interviewer change

Variable name	Description	Coding (Ref = 0)
sex	Sex	1 = Female
age	Age	Continuous
edu	Education level	1 = Lower than first degree
rdwell	Dwelling type	1 = Flat (0 = house)
rarea	Interviewer assessment of condition of houses in the	1 = Mainly good (0 = mixed or mainly poor)
rhouse	Interviewer assessment of condition of house relative to other houses in the area	1 = Same as or worse than other houses in the area (0 = Better than others)
rmarried	Marital status	1 = Single
rnumadl	Number of adults in the household	1 = 4 or more
kids	Number of children in the household	1 = 1 or more
work	Whether respondent currently in employment	1 = not working
rent	Housing tenure	1 = renting (0 = own outright or buying on a mortgage)
disab	Whether respondent has a disability	1 = no
health	Self-assessed health status	1 = good or very good

For each of the thirteen predictor variables listed in table 3, we first test the variable as a fixed effect, with interviewer change as the only other predictor in the model. Then we select the significant ones and develop a full fixed effects model through backwards elimination. Using the resultant model, we then test separately a random coefficient (effect) for each variable. Again, we then include all significant random effects in a combined model, eliminating variables through backwards elimination. Only if the random effect remains significant do we then test the interaction of the two variables.

The final step is to fit a combined model with all the random effects and interactions that were significant in the separate models. Any that become non-significant (P>0.10) when included simultaneously with the other predictors are dropped from the final model. The level 2 weights are then re-estimated based on the final model and we again select those that produce the smallest DIC.

4. Results: Effect of Interviewer Change

Table 4 presents all possible contrasts from the estimated model of refusal propensity with interviewer change as the sole predictor variable. The results of the model of non-response are similar. This is not surprising considering that refusals constitute more than half of the non-response and that most of the hypotheses regarding possible effects of interviewer continuity on response rates involve co-operation rather than contact. Consequently, we do not present results in full for the non-response model but we describe findings in the text.

It should be noted that 18 of the contrasts presented in table 4 (shaded yellow) relate to assignment decisions that can be made at wave 2 conditional upon the earlier choice of wave 1 interviewer such that at wave 2 a supervisor can choose between the alternatives. Of the remaining 60 contrasts, 51 relate to unconditional combinations of wave 1 and wave 2 decisions (the decision of the 2-wave interviewer combination should be made at wave 1), while nine (those in the fourth column, where the reference category is the assignment of the same interviewer, who has moved to a higher grade in the meanwhile) are averages of two specific contrasts. To understand this point, we first describe one example of each type of contrast. We then discuss the findings in turn for each type of contrast.

Example 1: Consider the third entry in the second column of the table, an estimated coefficient of +1.328. This relates to a reference category of assigning the same mid-grade

interviewer at both waves. The contrast is with assigning a different interviewer at wave 2, who is of low grade. In other words, conditional on the wave 1 interviewer being of mid-grade, assigning a different, low grade, interviewer at wave 2, rather than the same interviewer with no change in grade, increases the log-odds of refusal by a factor of 1.328, an estimate which is significantly different from 1.000 at the 0.05 level. So we conclude that a survey manager, faced with the decision of whether to assign the same mid-grade interviewer at wave 2 or a different low-grade interviewer, should recognise that on average the latter decision is likely to increase the chances of a refusal at wave 2 (by around 33%).

Example 2: Consider the first entry in the first column, an estimated coefficient of -1.171. This relates to a contrast between assigning the same low-grade interviewer at each wave and assigning the same mid-grade interviewer at each wave. Clearly, once a low grade interviewer has been assigned at wave 1 it is not possible at wave 2 to decide to assign a mid-grade interviewer at both waves. This contrast does not therefore correspond to a conditional decision that can be made at wave 2. Instead, it contrasts one possible pair of decisions (at waves 1 and 2) with another pair. In fact, this particular contrast can be thought of as corresponding to a main effect of grade, conditional on assigning the same interviewer at each wave. Conditional on allocation of the same interviewer who obtained participation at wave 1, a mid-grade interviewer is less likely ($0.05 < P < 0.10$) to obtain a refusal at wave 2 than a low-grade interviewer. A survey manager, knowing it is likely that the interviewer assigned at wave 1 will also work the case at wave 2, should conclude that a wave 2 refusal is more likely if a low-grade rather than mid-grade interviewer is assigned.

Example 3: Consider the first entry in the fourth column, a positive but non-significant estimated coefficient. The reference category here is the situation in which there is interviewer continuity, but the interviewer has been promoted between waves. Table 2 shows us that this category consists both of interviewers who have moved from mid-grade to high grade and interviewers who have moved from low-grade to mid-grade. This entry in the table contrasts this situation with the situation where a mid-grade interviewer is assigned at wave 1 and a different, low-grade, interviewer is assigned at wave 2. Thus, this contrast is a combination of two contrasts, one of which corresponds to an assignment decision that can be made at wave 2 conditional upon the wave 1 assignment (like example 1), while the other does not (like example 2).

Example 1 is one of the 18 contrasts that correspond to alternative decisions that can be made at wave 2, conditional on the wave 1 assignment decision (shaded yellow in table 4). These are in some ways the most interesting contrasts, as real field work assignment decisions are typically made independently at each wave of a longitudinal survey, with little or no consideration of the implications for future waves. Of these 18 contrasts, nine relate to a decision between assigning the wave 1 interviewer or a different one, while nine relate to a decision between two alternative different interviewers. Of the nine possible choices between assigning the same interviewer or a different one, only one displays a significant difference in the probability of receiving a refusal: given a mid-grade interviewer at wave 1, changing to a low-grade interviewer is more likely to result in a refusal than re-assigning the same interviewer at wave 2. The same contrast is significant for non-response. Two additional contrasts also exhibit a significant difference in the probability of non-response: given a high-grade interviewer at wave 1, changing to a low-grade interviewer or to a different high-grade interviewer are both more likely to result in non-response than re-assigning the same interviewer. However, the direction of the estimated effects shows a consistent pattern for both dependent variables: all six of the estimated contrasts involving a mid-grade or high grade interviewer at wave 1 suggest a higher refusal propensity and a higher non-response propensity if a different interviewer is assigned at wave 2, while all three of the estimated contrasts involving a low-grade wave 1 interviewer suggest the opposite.

Of the nine contrasts that compare alternative choices of a different interviewer at wave 2, again only one choice is estimated to have a significant impact on the probability of a refusal: given a mid-grade interviewer at wave 1, changing to a low-grade interviewer is more likely to result in a refusal than changing to a different mid-grade interviewer. The same contrast is significant for non-response, in addition to which changing from a mid-grade to a low-grade interviewer is also more likely to result in non-response than changing from a mid-grade to a high-grade interviewer. Again the direction of estimated effects shows a consistent pattern: regardless of the grade of the wave 1 interviewer and conditional upon assignment of a different interviewer at wave 2, refusal and non-response at wave 2 both appear less likely with a high grade interviewer than with a low grade interviewer and less likely with a mid-grade interviewer than with either a low or high grade interviewer.

Example 2 is one of 51 contrasts between alternative pairs of assignment decisions in which the wave 1 assignment differs. Four of these contrasts reach significance ($P < 0.10$) for refusal

and six reach significance for non-response. Assigning the same mid-grade interviewer at both waves results in a lower refusal propensity than assigning the same low grade interviewer at both waves, than assigning a different low grade interviewer at each wave, and than assigning a high grade interviewer at wave 1 and a low grade interviewer at wave 2. Additionally, assigning a different mid-grade interviewer at each wave results in a lower refusal propensity (but not lower overall non-response propensity) than assigning a high grade interviewer at wave 1 and a low grade interviewer at wave 2. Non-response propensity appears also to be lower, a) if the same high-grade interviewer is assigned at both waves, compared to assigning the same low grade interviewer at both waves or a mid-grade interviewer at wave 1 followed by a low grade interviewer at wave 2, b) if different mid-grade interviewers are assigned at each wave, compared to different low grade interviewers at each wave, c) if the same mid-grade interviewer is assigned at both waves, compared to assigning a different high grade interviewer at each wave. The directions of the other estimated contrasts are generally consistent with a change of interviewer being associated with increased refusal or non-response propensity only when compared to interviewer continuity with mid or high grade interviewers.

In summary, the results in table 3 show that:

- *Reduced refusal and non-response propensity is more likely to be associated with interviewer continuity than with change*. There are three significant contrasts in which interviewer continuity is associated with a lower propensity for refusal than interviewer change and eight in which continuity is associated with a lower propensity for non-response; there are no contrasts in which continuity is associated with higher propensity for either refusal or non-response;

- *Any such benefits of continuity may be restricted to situations where the continuing interviewer is mid or high grade*. The direction of the effect for interviewer continuity rather than change indicates both lower refusal propensity and lower non-response propensity for 23 of the 27 contrasts with a continuing interviewer who is mid or high grade at both waves or is promoted between waves; with a continuing interviewer who is low grade at both waves the opposite is true regarding refusals for 7 of the 9 contrasts and regarding non-response for 5 of the 9 contrasts;

- *A change in interviewer may not be detrimental to co-operation or response if the new interviewer is mid or high grade*. There is no situation in which changing interviewer significantly increases either refusal propensity or non-response propensity if the wave 2 interviewer is of higher grade than the wave 1 interviewer (12 contrasts for each outcome). There is no situation in which changing to an interviewer of the same grade increases refusal propensity (3 contrasts) and only one in which changing to an interviewer of the same grade increases non-response propensity;

- *Changing to a lower grade interviewer may be more likely to have a detrimental impact on co-operation or response than changing to an interviewer of the same or higher grade than the previous interviewer*. Out of 36 contrasts between alternative scenarios of interviewer change, two are significant with respect to refusal and four are significant with respect to non-response. Five of these six are situations where changing to a lower grade interviewer produces a poorer outcome than changing to an interviewer of the same grade as the wave 1 interviewer. The sixth indicates that changing from a low grade interviewer to a different low grade interviewer is more likely to result in non-response than changing from a mid-grade interviewer to a different mid-grade interviewer.

Table 4: Multiple membership multilevel logistic model of refusal

		Reference category											
	Interviewer	Same				Different							
Interviewer	Grade	Same (low)	Same (mid)	Same (high)	Higher (all)	Lower (mid-low)	Lower (high-low)	Lower (high-mid)	Same (low)	Same (mid)	Same (high)	Higher (low-mid)	Higher (low-high)
Same	Same (low)												
	Same (mid)	-1.171*											
	Same (high)	-	+										
	Higher (all)	-	+	+									
Different	Lower (mid-low)	+	+1.328**	+	+								
	Lower (high-low)	+	+1.369**	+	+	+							
	Lower (high-mid)	-	+	+	+	-	-						
	Same (low)	-	+0.976*	+	+	-	-	+					
	Same (mid)	-	+	-	-	-0.896*	-0.950*	-	-				
	Same (high)	-	+	+	-	-	-	+	-	+			
	Higher (low-mid)	-	+	+	+	-	-	+	-	+	-		
	Higher (low-high)	-	+	+	+	-	-	+	-	+	-	+	
	Higher (mid-high)	-	+	+	-	-	-	-	-	+	-	-	-

Cell entries are estimates of β in the case of estimates that are significantly different from 0.00 (P<0.10) and are the direction of β (+ or -) in the case of estimates that are not significantly different from 0.00. ** indicates P<0.05; * indicates 0.05<P<0.10. 80% of the level 2 variance is between wave 2 interviewers and 20% is between wave 1 interviewers.

6. Results: Interactions of Interviewer Change with Other Predictors

Amongst the 13 mini-models with just interviewer change (9-category version) and one other fixed effect predictor variable, 6 exhibited a significant association of the predictor variable with refusal propensity, namely *age*, *rarea*, *rhouse*, *kids*, *work* and *disab* (see table 3 for variable definitions). In a combined model, *kids* and *disab* were both eliminated via backward elimination, leaving only four significant predictors: *rarea*, *rhouse*, *work* and *age*.

Six of the 13 random effects were significant at P<0.10 level, but only 4 of these were significant at P<0.05 level. Given computational restrictions, we retained only these four: *rent*, *age*, *sex* and *disab*. After backward elimination, only two random effects remained: *age* and *sex*. However, the *sex* random effect was not well estimated, due to very poorly mixing chains or chains that just keep increasing or decreasing (effective sample size = 27). Thus, we dropped the random effect of *sex*, leaving a model with five fixed effects and one random effect. We tested the interactions with interviewer change of each of the four variables that initially displayed significant random effects. Only one, *rent*, was significant. With this final model we re-tested the weights and found the optimal weight to be 0.90 for wave 2 interviewer and 0.10 for wave 1 interviewer. The final estimated model is presented in table 5, alongside an analogous final model of non-response. The one significant interaction in the refusal model suggests that two contrasts differ between home owners and renters. First, conditional on a mid-grade wave 1 interviewer, a significant reduction in refusal propensity if the same interviewer is assigned at wave 2, compared to assigning a different mid-grade interviewer, may apply only to home owners and not to renters, for whom the direction of the effect is reversed. Second, changing to a higher grade interviewer appears to reduce refusal propensity only for renters. However, given that out of 13 characteristics tested home ownership is the only one for which contrasts differ significantly, the finding should be interpreted with caution. In modelling response, rather than refusal, we again find just one characteristic for which the effect of interviewer change differs between categories. In this case the significant characteristic is whether or not the respondent is in employment, rather than home ownership (table 5).

Table 5: Multiple membership multilevel logistic models of refusal and non-response with fixed and random covariates

	Non-Response Rate			Refusal Rate		
	$\hat{\beta}$	Standard Error		$\hat{\beta}$	Standard Error	
Fixed Part						
constant	-2.310	0.592	**	-3.856	0.815	**
diff / lower	0.384	0.510		1.358	0.730	**
diff / same (low)	0.197	0.651		1.334	0.900	
diff / same (mid)	-0.474	0.681		0.889	0.868	
diff / same (hi)	0.570	0.706		1.040	0.910	
diff / higher	0.075	0.518		1.052	0.727	
same / same (low)	0.704	0.766		0.077	1.100	
same / same (hi)	-0.945	0.811		0.880	0.931	
same / higher	-0.256	0.745		1.076	0.936	
diff / lower.work_2	1.980	1.246				
diff / same (low).work_2	2.753	1.377	**			
diff / same (mid).work_2	2.709	1.408	**			
diff / same (hi).work_2	1.712	1.454				
diff / higher.work_2	1.998	1.236				
same / same (mid).work_2	-0.061	1.540				
same / same (hi).work_2	3.239	1.495	**			
same / higher.work_2	3.349	1.419	**			
diff / lower.rent_1				-2.175	1.236	
diff / same (low).rent_1				-2.412	1.465	
diff / same (mid).rent_1				-3.849	1.858	**
diff / same (hi).rent_1				-1.844	1.710	
diff / higher.rent_1				-2.304	1.223	
same / same (mid).rent_1				-2.185	1.732	
same / same (hi).rent_1				-1.859	1.580	
same / higher.rent_1				-1.830	1.507	
rarea_1	-0.475	0.220	**	-0.671	0.255	**
rent_1	0.389	0.233		2.420	1.149	**
rhouse_2	1.040	0.407	**	1.131	0.475	**
work_2	-2.889	1.196	**	-1.261	0.286	**
edu1_1	-0.496	0.309				
edu1_1.rent_1	0.906	0.637				
(age-gm)	-0.015	0.011		-0.017	0.012	
Random Part						
Level: int09						
(age-gm)/(age-gm)	0.009	0.002	**	0.009	0.002	**
(age-gm)/cons	0.024	0.017		0.019	0.019	
cons/cons	0.966	0.372	**	1.230	0.547	**
-2*loglikelihood:						
DIC:	1099.635			844.727		
Units: int09	181			181		
Units: serial_num	1086			1047		

7. Discussion

The findings from our experimental study suggest that interviewer continuity may be beneficial for co-operation and participation in some situations, but not necessarily in others. Specifically, we find that, conditional on the grade of the wave 1 interviewer, the only situation in which changing the interviewer at wave 2 reduces co-operation propensity is when the wave 1 interviewer is mid-grade and the wave 2 interviewer is low grade. For other situations in which the wave 1 interviewer is either mid or high grade, the direction of the estimated effect is the same, but the effect is not statistically significant. When the wave 1 interviewer is of low grade, the direction of the estimated effect indicates that a change of interviewer is associated with an *increased* co-operation propensity, though these effects are not significant. Thus, whether interviewer continuity is beneficial may depend on the alternatives available. If the alternative interviewer is less experienced or less capable than the continuing interviewer – who was, after all, successful at obtaining an interview at the previous wave – then perhaps continuity should be preferred. But if the alternative interviewer is at least as experienced or capable as the continuing interviewer, then perhaps no effect on co-operation propensity should be expected. There may even be situations in which changing interviewer is, on average, beneficial, for example switching from a very inexperienced interviewer to a very experienced one.

We should also remember that these main effects of certain types of interviewer change are likely to mask a range of respondent-specific effects. Thus, even if a switch from a mid-grade interviewer to a different, low-grade, interviewer reduces co-operation propensity on average, there may be some respondents for whom such a switch is neutral, or even positive. In other words, the effect may not be uniform across respondents. Our study provided only a hint of such non-uniform effects, with the findings that, a) the effect of interviewer continuity on refusal propensity differed between home owners and renters, and b) the effect on response propensity differed between persons in employment and others. Thus, we find only modest evidence that any effect of interviewer continuity, once interviewer grade is controlled, varies between subgroups. However, our failure to find more interactions between respondent characteristics and interviewer change may be caused by limitations in the available measures of respondent characteristics.

We would expect that the effect of interviewer continuity should depend on the rapport between respondent and interviewer and the extent to which the respondent likes the

interviewer. Rapport and liking should depend on the combination of characteristics of respondent and interviewer, not merely the characteristics of the respondent. But the measures available in this study are only characteristics of the respondent. Furthermore, they may not be the most relevant characteristics. We suggest that future studies should consider measuring respondent personality and behavioural traits and preferences or, ideally, aspects of the respondent-interviewer interaction. Direct questions to the respondent regarding how they perceived the interviewer may provide the most powerful indicators of the likely effect of interviewer continuity. There are, of course, issues to be addressed in asking such questions. If they are administered by the interviewer who is the subject of the questions, there will be a risk of social desirability bias affecting the answers given (DeMaio 1984). Thus, a confidential self-completion mode may be preferred for administration of these questions. Aside from the mode in which the questions are asked, there is also work to be done to develop questions that effectively capture the extent to which the respondent is likely to be willing to be re-interviewed by the same interviewer.

A second limitation of our study is that interviewer grade is not a perfect measure of the relevant concepts of experience or performance capability. Future studies would benefit from attempting to measure more directly the qualities of an interviewer that determine success at making contact and gaining co-operation. These might include measures of experience in terms of numbers of cases worked, the period of time over which these cases were worked, and the variability in characteristics of those cases. One might also consider input-adjusted outcome measures, such as response rates conditional on sample characteristics.

In conclusion, we believe that the effect of interviewer continuity on subsequent survey response may be rather more complex than has been implied by previous literature. To better understand the effect, randomised designs and appropriate analysis methods are needed. Our study is a step in the direction of such more informative studies, but there remains work to be done. Further research is needed and would benefit from better measures of both respondent and interviewer characteristics, including interviewer experience and ability, and direct measures of the respondent's perception of his or her interviewer.

References

Beerten, R. and McConaghy, M. (2003) Respondents' confidence in survey taking and their co-operation with government surveys: some evidence from the UK, Paper presented at the *Annual meeting of the American Association for Public Opinion Research,* Sheraton Music City, Nashville, TN. http://www.allacademic.com/meta/p116455_index.html

Browne, W. J. (2009). *MCMC estimation in MLwiN*. Version 2.10. Bristol, Centre for Multilevel Modelling, University of Bristol.

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

Campanelli, P. and 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.

Carton, A. and Pickery, J. (2010) Interviewer (non-)response performance over time, presented at *21st International Workshop on Household Survey Nonresponse*, Nürnberg, August 30.

Cialdini, R.B. (2008) *Influence: Science and Practice* (5th Edition), Prentice Hall, Location.

DeMaio, T. J. (1984), Social desirability and survey measurement: A review, pp. 257–281 in C. G. Turner and E. Martia (ed.s) *Surveying Subjective Phenomena (Vol. 2)*, Russell Sage Foundation, New York

Freedman, J. L., and Fraser, S. C. (1966) Compliance without pressure: The foot-in-the-door technique, *Journal of Personality and Social Psychology* 4, 196-202.

Goldstein, H. (2010). *Multilevel Statistical Models*. Fourth Edition. Wiley, Chichester.

Groves, R.M. and Couper, M.P. (1998) *Nonresponse in Household Interview Surveys*, Wiley, New York.

Groves, R.M., Cialdini, R.B. and Couper, M.P. (1992) Understanding the decision to participate in a survey, *Public Opinion Quarterly* 56, 475-495.

Hox, J. and De Leeuw, E. (2002) The influence of interviewers attitudes and behaviour in household non-response, in Groves, R., Dillman, D., Eltinge, J., and Little, R., *Survey Nonresponse*, Wiley, London.

Morton-Williams, J. (1993) *Interviewer Approaches*, Dartmouth, Aldershot.

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

Rasbash, J., F. Steele, W. Browne and H. Goldstein (2009). *A user's guide to MLwiN version 2.10*. Bristol, Centre for Multilevel Modelling, University of Bristol.

Rendtel, U. (1990) Teilnahmebereitschaft in Panelstudien: Zwischen Beeinflussung, Vertrauen und Sozialer Selektion, *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 42, 280-299.

Schräpler, J.-P. (2001) Respondent Behavior in Panel Studies. A Case Study of the German Socio-Economic Panel (GSOEP). *DIW Discussion Paper* 244, DIW Berlin

Spiegelhalter, D., N. Best, B. P. Carlin and A. Van der Linde (2002) Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B* 64: 583-640.

Waterton, J. and Lievesley. D. (1987) Attrition in a panel study of attitudes, *Journal of Official Statistics* 3, 267-282.