Journal of Official Statistics, Vol. 30, No. 3, 2014, pp. 443–457, http://dx.doi.org/10.2478/JOS-2014-0028

Panel Attrition: How Important is Interviewer Continuity?

Peter Lynn[1], *Olena Kaminska*[1], *and Harvey Goldstein*[2]

We assess whether the probability of a sample member cooperating 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 nonexperimental data. Consequently, a) interviewer change is generally nonrandom, and b) continuing interviewers are more experienced by the time of the next wave. Our study is based on 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 as well as interactions of interviewer continuity with other variables. We find that continuity reduces refusal propensity for younger respondents but not for older respondents, and that this effect depends on the age of the interviewer. This supports the notion that interviewer continuity may be beneficial in some situations, but not necessarily in others.

Key words: Longitudinal survey; multiple membership multilevel model; nonresponse; refusal.

1. Introduction: Interviewer Continuity

For longitudinal surveys, the perceived benefit of having the same interviewer assigned to sample members 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 cooperation, though possibly also in terms of improved measurement. Consequently, they may sometimes be willing to prioritise the 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 he or she now has to travel 30 km to the address, rather than a different interviewer who lives only 5 km 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

[1] Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, UK. Email: plynn@essex.ac.uk and olena@essex.ac.uk
[2] Centre for Multilevel Modelling, University of Bristol. Tyndall Avenue, Bristol, BS8 1TH, UK. Email: h.goldstein@bristol.ac.uk
Acknowledgments: The contribution of the first author was funded by UK Economic and Social Research Council (ESRC) Award no. ES/H029745/1, "Understanding Society and the UK Longitudinal Studies Centre" (Principal Investigator: Prof. Nick Buck). The contributions of the second and third authors were 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". The third author is partly funded by the LEMMA node of the ESRC National Centre for Research Methods at the University of Bristol. We are grateful to NatCen Social Research for implementing the survey field work and providing the data.

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 reduce refusal propensity. These reasons relate to trust, tailoring and consistency.

Trust in the survey interviewer on the part of the sample member is an important influence on whether or not the sample member chooses to cooperate (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 reduced refusal propensity, would thus be associated with interviewer continuity.

Tailoring of communication and tactics by interviewers reduces the chances of a refusal (Groves et al. 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 reduced refusal propensity (though the additional ability to tailor will be reduced if the survey organisation makes effective efforts to feed forward to the interviewer relevant information about the contact and persuasion attempts from previous waves).

Consistency is generally seen as a desirable personal trait (Cialdini 2008, chap. 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 reduced refusal propensity being associated with continuing interviewers.

However, although it is plausible that interviewer continuity might have the effect of reducing refusal rates, other things being equal, there is very little empirical evidence on this point. A number of longitudinal surveys observe that reinterview rates are higher amongst cases where the same interviewer makes the approach at a subsequent wave (e.g., Rendtel 1990; Schräpler 2001; Waterton and Lievesley 1987). But such an association does not imply causality. In particular, in face-to-face surveys where interviewers tend to work in specific geographic areas, it is quite possible that interviewer continuity and respondent cooperation 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 reinterview rate on a face-to-face survey. This study involved an interpenetrated design at

Wave 2 of the British Household Panel Survey in 1992. No effect of interviewer continuity on reinterview rate was found either at Wave 2 (Campanelli and O'Muircheartaigh 1999) or at Waves 3 and 4 (Campanelli and 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 reinterview propensity and should therefore be controlled in any study of the effect of interviewer continuity;
- The effect of interviewer continuity on reinterview 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 article we examine the effect of interviewer continuity on refusal 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 NatCen Social Research 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 recontacted for further research ($n = 1{,}188$) formed the sample for the study reported here. (Response rate was 55% to the Wave 1 survey and 78% of respondents agreed to be recontacted. However, we would note that inference in our study relies on random allocation within the sample who agreed to be recontacted, so we are not reliant on sampling-based inference.) Ample 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). Grade indicates the position of an interviewer on the NatCen pay scale and therefore, as with any pay scale, tends to reflect a combination of competence and experience. 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 refusal propensity (organisational skills, ability to perceive the concerns and circumstances of respondents, ability to persuade). A low-grade interviewer is likely to have little experience, or could alternatively have more experience but not have performed very well. Of course, any association between interviewer experience and refusal 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 refusal 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 for three primary sampling units (very rural areas) where assignment to random subsets of respondents would have been prohibitively expensive. In these cases, respondents were chosen to be allocated to the same interviewer 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. In total, 181 interviewers worked on Wave 1 of the survey, of whom 69 also worked on Wave 2. A further 136 interviewers worked only on Wave 2, meaning that overall 317 worked on one or both waves. Of these, 51% were female, 43% were aged over 60, 29% had no more than two years of experience as a NatCen interviewer, 52% had between two and ten years' experience, and 18% had more than ten years' experience.

Our key analysis variable is an indicator of interviewer change. We use two forms of this variable, a nine-category version and a three-category version (see results sections

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

	Different Interviewer			Same Interviewer	
Number of assigned Wave 2 cases	Lowest grade 2009	Middle grade 2009	Highest grade 2009	All grades 2009	Total
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

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

below for details of how these are used). The nine-category version is based on the twelve categories in Table 1, but a) combining to single categories all cases with a different interviewer of higher grade and all cases with a different interviewer of lower grade, and b) creating an additional category for cases with the same interviewer but of a higher grade (i.e., an interviewer who had received a promotion in the interim). The nine categories are listed in Table 4.

In the three-category version, the first category consists of all cases involving a different, lower grade, interviewer at Wave 2. The second category consists of cases involving a different interviewer of the same or higher grade. The third category consists of all cases allocated to the same interviewer at Wave 2. Comparison of the second and third categories will allow us to identify the effect of interviewer change, controlling for change in grade.

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. Sociodemographic and classificatory questions were also asked. Mean interview length was 21 minutes. Of the 1,188 issued sample cases, eleven were found to be ineligible for reinterview (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 nonresponse accounted for the remaining 35 cases. Thus, amongst eligible cases, Wave 2 contact rate was 90% and cooperation rate was 80%, giving an overall conditional wave response rate of 72%.

3. Analysis Methods

Our analysis of refusal propensity is restricted to the 1,058 sample members who were successfully contacted at Wave 2, amongst whom the refusal rate was 17%. We use multiple membership multilevel logistic models of propensity to refuse conditional on contact. The dependent variable is coded 1 if the sample member refused the interview at Wave 2 and 0 otherwise. Thus, positive coefficients indicate an increased propensity for the undesirable outcome.

A formal statement of the basic model is as follows:

$$\operatorname{logit}\left(\pi_{i(j_1,j_2)}\right) = X_{i(j_1,j_2)}\beta + w_{j_1}u_{j_1} + w_{j_2}u_{j_2}; \quad w_1 + w_2 = 1 \tag{1}$$

where $\pi_{i(j_1,j_2)}$ is the probability of a refusal 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 (2011, chap. 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 1 and 2 respectively, namely u_{j_1}, u_{j_2}. The corresponding weights reflect the relative importance of the Wave 1 and Wave 2 interviewers. 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).

The multiple membership structure of the data arises from treating the interview occasions as Level 1 units and the interviewers as Level 2 units. This is not a standard two-level model since the Level 1 units, rather than being fully nested within each Level 2 unit (interviewer) with an associated effect from that interviewer, are influenced by a weighted average of the effects associated with both (if they are different) of the interviewers assigned to them. This is reflected in Model (1). The multiple membership model also differs from a cross-classified model where there are two sets of unrelated units (at occasion one and occasion two): treating our data that way would provide no way to differentiate cases where it is actually the same interviewer and where it is a different one at each occasion.

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).

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 by selecting the model with best fit among the models with different weights. Our best fit criterion is to select the model with the smallest Deviance Information Criterion (DIC) (Spiegelhalter et al. 2002).

As the random effect of interviewers turns out not to be significant (see Section 4 below), we do not test for fixed effects of interviewer change between waves or of any other interviewer characteristics. Instead, using the initial weights, we proceed to test random effects of twelve characteristics of respondents in order to establish whether interviewers vary in their relative success with different sample subgroups. These twelve characteristics represent all the sociodemographic variables available in the Wave 1 data for the full sample.

We test all categorical predictor variables (other than interviewer change) as dichotomies, as the model otherwise becomes overparameterised 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 refusal with the variable in question (full version) as the sole predictor variable, first combining categories with estimated coefficients that are 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 twelve resultant predictor variables are listed in Table 3.

For each predictor variable listed in Table 3, we first tested whether the variable had a random coefficient at interviewer level. Significance was judged in terms of whether the 95% interval estimate for a single parameter included zero. More generally, the DIC statistic was used to compare models where models differed in terms of two or more parameters. Retaining each significant variable, our intention was then to develop a full random effects model through backwards elimination, retaining only those predictors and their random coefficients which remain significant. However, as it turned out (see below) this step was not necessary as only one predictor variable showed significance.

When testing the significance of random slopes we use initial Level 2 weights of 0.5 for each wave, until we have identified the final model. We then fit that model with alternative combinations of weights and select the combination that results in the smallest DIC. Finally, we test interactions with the three-category interviewer change variable of each variable for which there is a significant random effect. We use the three-category version in order to retain sufficient statistical power to detect effects. Each of the interactions that is significant in these one-interaction models is then included in a combined model.

Table 3. Predictor variables tested for interaction with interviewer change

Variable	Description	Coding (Ref = 0)	Number of respondents in category 1
Sex	Sex	1 = Female	599
Age	Age	Continuous	
Edu	Education level	1 = Lower than first degree	164
Rdwell	Dwelling type	1 = Flat (0 = house)	168
Rarea	Interviewer assessment of condition of houses in the area	1 = Mainly good (0 = mixed or mainly poor)	530
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)	942
Rmarried	Marital status	1 = Single	209
Rnumadl	Number of adults in the household	1 = 4 or more	52
Kids	Number of children in the household	1 = 1 or more	250
Work	Whether respondent currently in employment	1 = not working	494
Rent	Housing tenure	1 = renting (0 = own outright or buying on a mortgage)	294
Disab	Whether respondent has a disability	1 = no	770

Note: Total number of respondents in the analysis is 1,058. Predictor variables were all collected at Wave 1 of the survey (and are therefore available for both respondents and nonrespondents at Wave 2).

4. Results: Interviewer Effect

We first fit a null model to test for a random intercept for interviewer combinations. The fit of this model is almost identical whether we specify the weights to be 1.0 for Wave 1 and 0.0 for Wave 2 (DIC = 873.0), 0.5 for each wave (DIC = 873.7), or 0.0 for Wave 1 and 1.0 for Wave 2 (873.3). By comparing the above models to a base model containing only a fixed-effect intercept (Model 1 in Table 5, DIC = 872.8), we note that adding a random interviewer combination effect does not improve the model fit. Also, the random effect (in each of the three above weight specifications) is not significant.

We therefore find no evidence of variation between interviewer combinations in propensity for a sample member to refuse. There is therefore no variation that can be explained by fixed characteristics of interviewers. To confirm this we fit a model in which the sole fixed effect predictor is the nine-category interviewer change variable. The fit of the model is slightly worse (DIC = 879.5) than the null model with only a fixed intercept (DIC = 872.8), and none of the coefficients for interviewer change reach significance (we tested all pairwise combinations of interviewer change and none was associated with a significantly different refusal propensity). The unweighted refusal rates for each interviewer combination are presented in Table 4.

5. Results: Random Effects of Respondent Characteristics

Though we found no evidence that interviewer combinations vary in their propensity to elicit a refusal, on average, it is possible that they may differ in the extent to which this propensity varies between sample members with different characteristics. We therefore test whether there is random slope variance associated with each of the twelve respondent characteristics listed in Table 3. We add each random slope in turn to the model which otherwise contains only the fixed intercept. For all respondents' characteristics other than age, the random slope variance is not significant (the mean of 20,000 MCMC parameter estimates is not significantly different from zero and the mode is zero to five decimal places). The only variable for which the random slope variance achieves significance is respondent age. DIC actually increases when the random effect of age is added to the model, but the covariance of age with the intercept is estimated to be 0.00, so we fix the covariance to zero, thereby reducing the number of parameters to be estimated. With the covariance removed, the random effect of age remains significant and DIC reduces.

Table 4. Refusal rates by interviewer combination

	Refusal rate	*n*
Same interviewer: low grade	19.2	52
Same interviewer: medium grade	7.8	90
Same interviewer: high grade	11.3	71
Same interviewer: higher grade	14.7	68
Different interviewer, same grade: low	15.2	79
Different interviewer, same grade: medium	10.5	86
Different interviewer, same grade: high	16.1	56
Different interviewer, lower grade	18.2	236
Different interviewer, higher grade	13.8	320

This suggests that interviewer combinations may differ in the extent to which they are relatively more (or less) likely to elicit a refusal from older (or younger) respondents. It is therefore of interest to know whether this variation can be explained by fixed characteristics of interviewers, notably interviewer change.

For the model containing a fixed intercept and a random slope of respondent age, we compare alternative assignment of weights to the two waves. We find that minimum DIC is achieved with weights of 0.25 for Wave 1 and 0.75 for Wave 2, suggesting that the Wave 2 interviewer has approximately three times as much influence on the Wave 2 outcome as the Wave 1 interviewer (Table 5). We use these weights in subsequent modelling.

6. Results: Interactions Between Interviewer and Respondent Characteristics

We next explore whether the variation between interviewers in the effect of respondent age on refusal propensity (significant random slope for respondent age) can be explained by known characteristics of interviewers, notably interviewer change. We therefore explore fixed-effect interactions between respondent age and interviewer characteristics. The three-category version of the interviewer change variable is used: a different interviewer of a lower grade, a different interviewer of the same or higher grade, and the same interviewer.

The interaction between respondent age and interviewer change does not reach statistical significance, though the model with this term added (including the respective main effects as fixed effects) is a better fit (DIC = 870.4) than the model with only a fixed intercept and a random effect of respondent age (DIC = 893.5). However, we can also explore the possible effects of other known characteristics of interviewers, namely age and sex. Specifically, we hypothesise that the random effect of respondent age may be related to interviewer age. Such an interaction could be driven by liking, whereby respondents are more likely to comply with a survey request from someone they like (Groves et al. 1992) and are more likely to like someone who is similar to themselves (Stotland and Patchen 1961), in this case in terms of age. Alternatively, the effect could be driven by

Table 5. Comparison of models

Model no.	Fixed part	Random part	Weights (Wave 1 : Wave 2)	DIC
1	Intercept	None	0.5 : 0.5	872.8
2a	Intercept	Respage	0.5 : 0.5	867.7
2b	Intercept	Respage	0.25 : 0.75	867.5
3	Intercept Intchg Agedum Intchg*Agedum	Respage	0.25 : 0.75	856.9

Notes: Respage is respondent age in years; Agedum is a binary indicator of whether or not the respondent is aged over 60 (at Wave 2); Intchg is a five-category variable indicating whether the Wave 2 interviewer is a) same as Wave 1, up to 60, b) same as Wave 1, over 60, c) different, same or higher grade, up to 60, d) different, same or higher grade, over 60, e) different, lower grade. All models based on $n = 1{,}058$.

a tendency to show greater respect towards elders, which would suggest that younger respondents should be less likely to refuse to older interviewers.

We create a new five-category variable defined by interviewer change and interviewer age. This variable is created by subdividing both the cases with the same interviewer at Wave 2 and the cases with a different interviewer of same or higher grade into those where the Wave 2 interviewer is aged over 60 and those with a younger interviewer. The cases with a different interviewer of lower grade are not subdivided by interviewer age as this distinction is not of substantive interest (as there is no comparison group of same interviewers of lower grade). We also recode respondent age as a binary variable indicating whether or not the respondent is aged over 60. This is done to gain statistical power, and the cut point is chosen based on previous research that shows people of retirement age to be distinctive in terms of the determinants of survey participation (Lynn 2012 showed that people aged over 60 were more likely to agree to take part in an interviewer-administered survey, more likely to continue participating in a panel, and that their decision to take part was more likely to be sensitive to incentives to do so.) The sample contained 324 respondents aged over 60 and 734 aged 60 or under.

The interaction between respondent age and this five-category measure of interviewer change and age combinations includes significant differences (details in Section 7 below) and the model fit is significantly improved (DIC $=$ 856.9, compared to 867.5 in the model with only a random effect of age). We therefore retain this term in the model and proceed to test the interaction of interviewer sex with respondent age. This interaction is not significant and does not improve model fit. We also test the effects of interactions of respondent age with sex of Wave 1 interviewer and with age of Wave 1 interviewer, both instead of or as well as the interaction with age of Wave 2 interviewer. None of these interactions improve the model. Thus, we retain as our final model the model containing, in the fixed part, the interaction between respondent age (two categories) and the combination of interviewer age and interviewer change (five categories), plus a random effect of respondent age (continuous variable). This model is denoted Model 3 in Table 5.

7. Final Model

The final model is summarised in Table 6. To aid interpretation, Figure 1 displays the model-predicted propensities to refuse for each combination of interviewer continuity and respondent age (different interviewer of a lower grade is not shown, as this is not of relevance to the central theme of this article, as explained earlier). The model suggests that for sample members aged up to 60, interviewer continuity reduces the propensity for refusal if the interviewer is aged over 60 (left-hand panel in Figure 1). For sample members aged over 60, assigning an older interviewer reduces the propensity to refuse, regardless of whether or not it is the same interviewer who carried out the Wave 1 interview (right-hand panel in Figure 1). Specifically, for sample members aged up to 60, assignment of the same interviewer, aged over 60, results in a significantly lower probability of refusing than assignment of a different interviewer aged 60 or under ($p = 0.04$) or assignment of a different interviewer over 60 ($p = 0.03$). For sample members aged over 60, assignment of a different interviewer, aged over 60, results in a significantly lower probability of refusing than assignment of the same interviewer, aged

Table 6. Final model of propensity to refuse

	Coefficient	Standard Error	
Fixed Part			
Intercept	−1.59	0.29	**
respondent age 60+	−0.49	0.60	
same interviewer 61+	−0.83	0.46	
different interviewer <61	0.00	0.34	
different interviewer 61+	0.04	0.37	
different interviewer lower grade in w2	0.23	0.35	
same int 61+* resp age 60+	−1.52	1.50	
different interviewer <61* respondent age 60+	−0.69	0.75	
different interviewer 61+* respondent age 60+	−2.07	1.02	**
different interviewer, lower grade in w2* resp age 60+	−0.50	0.74	
Random Part			
Level: combination of 2008 interviewers (35%) and 2009 interviewers (65%)			
var (intercept)	0.147	0.172	
var (age-gm)	0.00119	0.00068	
Model Fit			
DIC:	856.9		
Units: interviewers (2009)	227		
Units: respondents	1058		

Notes: Dependent variable is an indicator of whether the sample member refused to cooperate at Wave 2. Base is all sample members contacted at Wave 2. Reference category for respondent age is 60 or under. Reference category for interviewer change is the same interviewer, aged 60 or under.

up to 60 ($p = 0.03$). There is also a suggestion that continuity with an interviewer aged over 60 results in a lower probability of refusing than continuity with an interviewer aged 60 or under, though the difference is only of marginal significance ($p = 0.10$ for respondents over 60 and $p = 0.07$ for respondents 60 or under).

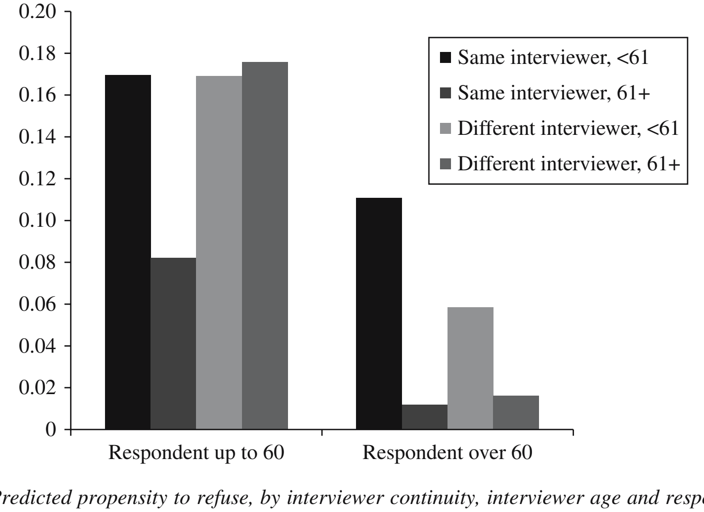

Fig. 1. Predicted propensity to refuse, by interviewer continuity, interviewer age and respondent age

It is interesting to note that the effect of interviewer continuity for younger sample members would have appeared larger if we had not controlled for interviewer experience. The difference in predicted probability of refusal between the same interviewer over 60 and a different interviewer of lower grade is even greater ($p = 0.01$) than the differences reported in the previous paragraph between the same interviewer and a different interviewer of the same or higher grade (of either age group).

8. Discussion

This experimental study has provided evidence of heterogeneous effects of interviewer continuity on cooperation by panel survey members. We believe it is the first study to find such evidence. Specifically, we find that continuity reduces refusal propensity for one sample subgroup (respondents aged 60 or under) but not for another (respondents aged over 60) and that this effect depends on a characteristic (age) of the interviewer. This supports the notion that interviewer continuity may be beneficial in some situations, but not necessarily in others. Whether interviewer continuity is beneficial may depend on the characteristics of the previous interviewer, the available alternative interviewers, and the respondent. What we conclude from this is that interviewer continuity should neither be blindly pursued in all cases nor completely ignored. Rather, survey organisations would be well advised to attempt to restrict the pursuit of interviewer continuity to situations where it is likely to matter. This can be thought of as an example of targeting of survey design features (Lynn, forthcoming).

We find that for younger respondents, interviewer continuity may only be beneficial if the interviewer is aged over 60. And in the case of older sample members, changing the interviewer may be beneficial if this involves switching from a younger to an older interviewer. The effect for younger respondents is intriguing, though the explanation is unclear. Maybe the trust of younger respondents is more likely to be engendered by older interviewers. Maybe older interviewers are generally better at tailoring but this only matters when the respondent is younger. Maybe younger respondents feel more strongly the need to appear consistent when the interviewer is older. Or maybe a greater positive age difference between interviewer and respondent engenders greater respect. The explanation of this finding requires further research.

Furthermore, we have demonstrated the importance of controlling for interviewer experience in studying interviewer continuity. We would have overestimated the benefits of continuity had we ignored experience, as changing to a less experienced (lower-grade) interviewer tends to increase the probability of a refusal.

It should be remembered that observed main effects of interviewer continuity are likely to mask a range of respondent-specific effects. Thus even if, for example, a switch to a different, lower-grade, interviewer reduces cooperation 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 finding that the effect of interviewer continuity on refusal propensity differed between younger and older sample members is an example of such a nonuniform effect.

Our study is somewhat exploratory and some of the decisions we made in the course of the analysis were data driven rather than theory driven. For this reason, the specific substantive findings should be treated with caution. Furthermore, our complex models require large sample sizes for good estimation. Other interactions between respondent characteristics and interviewer continuity may have become apparent with greater statistical power. Good measures of other relevant characteristics could also reveal other interactions. In particular, 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 in this study we had available only very limited characteristics of the interviewer. Furthermore, the available respondent characteristics may not be the most relevant ones. 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 the 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 reinterviewed by the same interviewer.

We recognise that interviewer grade is not a perfect measure of the relevant concepts of experience or performance capability. There is an opportunity for future studies to benefit from attempting to measure more directly the qualities of an interviewer that determine success at making contact and gaining cooperation. Measures of experience might include numbers of cases worked, the period of time over which these cases were worked, and the variability in characteristics of those cases. Measures of competence might include input-adjusted outcome measures, such as response rates conditional on sample characteristics. Separate identification of experience and competence in future studies might provide insights into the mechanisms by which interviewer grade effects operate. This could assist sample allocation decisions.

This study was designed to identify the effects of interviewer continuity, not to explain the causes of such effects. We posited three possible causes: trust, tailoring and consistency. There is no particular reason why any of these causes should not apply more strongly to younger respondents than older respondents, or to older interviewers rather than younger ones. Thus, the identification of heterogeneous effects cannot assist us to identify the cause of the effects.

We cannot rule out the possibility that interviewer continuity effects are sensitive to the survey context. Our study is based on a request to take part in a relatively short interview (21 minutes) on a particular topic (safety on public transport). Results for a different type of survey request could be different. This issue could warrant investigation.

In conclusion, we have demonstrated that the effect of interviewer continuity on subsequent survey response may be rather more complex than has been implied by previous literature. The effect may depend on the interaction between characteristics of the previous interviewer, of the available alternative interviewers, and of the respondent. We have found examples of such interaction. We have also demonstrated the importance of controlling for the effect of interviewer experience, of appropriate analysis methods, and of capturing interviewer characteristics. We believe there is considerable potential to learn more about the nature of interviewer continuity effects. This knowledge could help to reduce panel survey refusal rates in the future. But to gain this knowledge, further research 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. In addition, randomised designs and appropriate analysis methods are needed.

9. References

Beerten, R. and M. McConaghy. 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, May 2003. Available at: http://www.allacademic.com/meta/p116455_index.html (accessed December 13, 2013).

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

Campanelli, P. and C. O'Muircheartaigh. 1999. "Interviewers, Interviewer Continuity, and Panel Survey Nonresponse." *Quality & Quantity* 33: 59–76. DOI:http://dx.doi.org/10.1023/A:1004357711258.

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: 129–144. DOI:http://dx.doi.org/10.1023/A:1014928107205.

Carton, A. and J. Pickery. 2010. "Interviewer (Non-)Response Performance Over Time." Paper presented at the *21st International Workshop on Household Survey Nonresponse*, Nuremberg, August 30.

Cialdini, R.B. 2008. *Influence: Science and Practice*. 5th ed. Boston BA: Prentice Hall.

DeMaio, T.J. 1984. "Social Desirability and Survey Measurement: A Review." In *Surveying Subjective Phenomena (Vol. 2)*, edited by C.G. Turner and E. Martia, 257–281. New York: Russell Sage Foundation.

Freedman, J.L. and S.C. Fraser. 1966. "Compliance Without Pressure: The Foot-in-the-Door Technique." *Journal of Personality and Social Psychology* 4: 196–202.

Goldstein, H. 2011. *Multilevel Statistical Models*. 4th ed. Chichester: Wiley.

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

Groves, R.M., R.B. Cialdini, and M.P. Couper. 1992. "Understanding the Decision to Participate in a Survey." *Public Opinion Quarterly* 56: 475–495. DOI:http://dx.doi.org/10.1086/269338.

Hox, J. and E. de Leeuw. 2002. "The Influence of Interviewers Attitudes and Behaviour in Household Non-Response." In *Survey Nonresponse*, edited by R. Groves, D. Dillman, J. Eltinge, and R. Little. London: Wiley.

Lynn, P. 2012. "The propensity of older respondents to participate in a general purpose survey." *Understanding Society Working Paper* 2012-03. Colchester: University of

Essex. Available at: www.iser.essex.ac.uk/publications/working-papers/understanding-society/2012-03 (accessed December 13, 2013).

Lynn, P. Forthcoming. "Targeted Response Inducement Strategies on Longitudinal Surveys." In *Improving Survey Methods: Lessons from Recent Research*, edited by U. Engel, B. Jann, P. Lynn, A. Scherpenzeel, and P. Sturgis. Abingdon, UK: Psychology Press.

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

Pickery, J., G. Loosveldt, and A. Carton. 2001. "The Effects of Interviewer and Respondent Characteristics on Response Behaviour in Panel Surveys: a Multilevel Approach." *Sociological Methods and Research* 29: 509–523. DOI:http://dx.doi.org/10.1177/0049124101029004004.

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. Berlin: DIW.

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. DOI:http://dx.doi.org/10.1111/1467-9868.00353.

Stotland, E. and M. Patchen. 1961. "Identification and Change in Prejudice and authoritarianism." *Journal of Abnormal and Social Psychology* 62: 250–256. DOI:http://dx.doi.org/10.1037/h0043040.

Waterton, J. and D. Lievesley. 1987. "Attrition in a Panel Study of Attitudes." *Journal of Official Statistics* 3: 267–282.

Received December 2012
Revised December 2013
Accepted December 2013