National Centre for Research Methods Working Paper

1/18

Do interviewers moderate the effect of monetary incentives on response rates in household interview surveys?

Eliud Kibuchi, Patrick Sturgis, Gabriele B. Durrant, Olga Maslovskaya

NCRM National Centre for Research Methods

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

Do interviewers moderate the effect of monetary incentives on response rates in household interview surveys?

Eliud Kibuchi, *University of Southampton*
Patrick Sturgis, *University of Southampton*
Gabriele B. Durrant, *University of Southampton*
Olga Maslovskaya, *University of Southampton*

DRAFT MARCH 2018

Abstract

As citizens around the world become ever more reluctant to respond to survey interview requests, incentives are playing an increasingly important role in maintaining response rates. In face-to-face surveys, interviewers are the key conduit of information about the existence and level of any incentive offered and, therefore, potentially moderate the effectiveness with which an incentive translates non-productive addresses into interviews. Yet, while the existing literature on the effects of incentives on response rates is substantial, little is currently known about the role of interviewers in determining whether or not incentives are effective. In this paper, we apply multilevel models to three different face-to-face interview surveys from the UK, which vary in their sample designs and incentive levels, to assess whether some interviewers are more successful than others in using incentives to leverage cooperation. Additionally, we link the response outcome data to measures of interviewer characteristics to investigate whether interviewer variability on this dimension is systematically related to level of experience and demographic characteristics. Our results show significant and substantial variability between interviewers in the effectiveness of monetary incentives on the probability of cooperation across all three surveys. However, none of the interviewer characteristics considered are significantly associated with more or less successful interviewers.

Introduction

It is widely acknowledged that low and declining response rates pose an existential threat to conventional approaches to data collection in survey research (Brick & Williams, 2013; Couper, 2013; Meyer, Mok, & Sullivan, 2015; Miller, 2017). In response to this pressing challenge, survey methodologists have invested considerable time and resources investigating features of the survey process which can be leveraged to increase the probability of cooperation amongst sampled units (Groves & Couper, 1998; Groves & Heeringa, 2006). As the primary interface between survey organisations and sample members, interviewers are key to this endeavour (Campanelli, Sturgis, & Purdon, 1997; Morton-Williams, 1993; West & Blom, 2017). They are the main reason that response rates for household interview surveys remain substantially higher than all other available modes, albeit that they also come at a commensurately higher cost. A large number of studies in a broad range of contexts have now established that demographic, attitudinal, and behavioural differences between interviewers can account for substantial variability in response rates (Hansen 2006; Hox and Durrant et al. 2010; de Leeuw 2002). For example, Campanelli and O'Muircheartaigh (1999) found that more experienced interviewers were more successful at obtaining contact and cooperation, due to more effective calling patterns and an ability to tailor the survey request to sample members' motivations and concerns.

In addition to interviewers, monetary incentives of various kinds have played a central role in strategies for maximising survey cooperation (Singer, 2002; Singer, Groves, & Corning, 1999; Singer, Hoewyk, Gebler, Raghunathan, & Mcgonagle, 1999). Monetary incentives are considered to operate by acting as a replacement for other non-pecuniary motivations for survey participation, such as interest in the survey topic, enjoyment of social interaction, or a sense of civic duty (Groves, Singer, & Corning, 2000; Singer, Hoewyk, et al., 1999). A large body of evidence, predominantly based on randomized experiments, has established that monetary incentives exert a small to moderate positive effect on response rates and that larger incentives tend to produce more substantial effects but with diminishing marginal returns (Cantor, O'Hare, and O'Connor 2008; Singer, Groves, and Corning 1999; Church, 1993; Singer and Ye, 2013).

Given the sustained focus on the role of interviewers and monetary incentives in the existing survey methodological literature, it is surprising that their potential *joint* influence has seldom been considered. Because interviewers play such a key role in making contact with and persuading sample members to participate, It is *prima facie* plausible that interviewers vary in how effective they are at leveraging incentives to persuade sample members to provide an interview. For example, some interviewers may tailor their doorstep introductions to highlight the availability of a monetary

incentive at households that are most likely to be sensitive to them (Campanelli et al., 1997; Groves & Couper, 1998). Similarly, interviewers may feel more confident in their doorstep approach when they know an incentive is available which may positively affect their persuasive efforts (Singer & Ye, 2013) This joint influence is our focus in this paper. We analyse data from three different face-to-face interview surveys that included a randomized incentive experiment to identify interviewer influences on the effectiveness of monetary incentives in promoting survey cooperation.

The remainder of the paper is structured as follows. We first provide short reviews of the respective literatures on how interviewers and monetary incentives influence survey response, before setting out our expectations regarding the moderating effect of interviewers on the effectiveness of incentives. We then describe the three surveys that form the basis of our analysis and the administrative data on interviewers and areas to which they are linked. This is followed by an exposition of our analysis strategy and presentation of our key findings. We conclude with a consideration of the limitations of our study, a discussion of the implications of our findings for improving survey practice and suggestions on how future research in this area might usefully proceed.

The effect of interviewers on response rates

Face-to-face surveys consistently achieve higher response rates than those undertaken by self-administration or by telephone, a difference that is largely attributable to the role of interviewers. Interviewers locate and make repeated calls at sampled addresses, thereby keeping non-contacts to a minimum (Campanelli et al., 1997). Having made contact with a household, they undertake a number of additional tasks including respondent selection within households, conveying information about the survey such as the topic, sponsor, likely duration of the interview, and the availability of incentives (Couper & Schlegel, 1998). They also often provide accompanying information about the survey, in the form of copies of advanced letters (which will not have been read by all sample members), as well as providing reassurance about the bona fides of the survey and showing identity documentation (Groves & Couper, 1998; Groves et al., 2000).

Interviewers also persuade reluctant respondents to provide an interview, thereby minimizing refusals. A range of dispositional factors and behavioural styles have been identified as important in determining how successful interviewers are at preventing refusals. These include an ability to maintain an interaction, rather than accept a refusal, and to 'tailor' their approach on the doorstep to specific characteristics of sample units, by identifying and presenting aspects of the survey that they judge are likely to be positively valued (Morton-Williams 1993; Campanelli et al 1997; Groves and

Couper 1998). For example, an interviewer may remark upon the respondent's garden if they perceive that gardening is likely to be a hobby of the householder, or they might highlight the topic of the survey if they judge from the observable characteristics of the sample member that it is likely to be of interest. Studies which have examined the causes of non-contact and refusal have consistently found significant interviewer effects across a range of sample designs and international contexts (Campanelli et al., 1997; Durrant, Groves, & Steele, 2010; Durrant & Steele, 2009; Hox & de Leeuw, 2002). For example, Blom, Leeuw, and Hox (2011) found interviewer intra-class correlation coefficients of 0.27 for non-contact and 0.08 for cooperation across ten countries in the 2008 European Social Survey.

Existing research has also considered which characteristics of interviewers are important in producing these effects (Blom & Korbmacher, 2013). This has found that experienced interviewers tend to be better at tailoring their approaches to household idiosyncrasies and concerns (Groves and Couper 1998; Lemay and Durand 2002). More experienced interviewers, both in terms of experience on the particular survey and of interviewing more generally, have also been found to obtain higher response rates, even though they are often allocated to more difficult areas (Purdon, Campanelli, & Sturgis, 1999; West & Blom, 2017). Other studies have found that interviewers with higher levels of self-confidence and more positive appraisals of the likelihood of achieving interviews also obtain higher cooperation rates, an effect which is thought to arise from the positive effect of confidence on the quality of doorstep interactions (Singer and Kohnke-Aguirre 1977); Groves and Couper 1998; Hox and de Leeuw 2002). The existing evidence, then, suggests that interviewer skills and experience in recognising, interpreting, and addressing visual cues and the confidence and self-belief with which interviewers approach the task of obtaining cooperation on the doorstep are the key mechanisms through which interviewers influence individual cooperation decisions.

Using incentives to increase response rates

Under the influential 'Leverage-salience' theory of survey cooperation (Groves et al., 2000), incentives are postulated to work by acting as a replacement for non-financial motivating factors, such as engagement in the topic of the survey, enjoyment of social interaction, and a sense of civic or moral obligation. Incentives may also invoke norms of reciprocity, such that respondents feel a sense of obligation to provide an interview, when they are offered or receive an incentive before the interview request is made (Dillman, Smyth, & Christian, 2009).

The field of survey research has benefited from a wealth of systematic reviews and meta-analyses of the effects of survey incentives which have yielded a robust set of conclusions. We know from this

body of evidence that monetary incentives are more effective in motivating participation than non-monetary incentives such as pens, calendars, diaries and so on (Cantor, O'Hare, and O'Connor 2008; Church 1993; Singer and Ye, 2013). It is also well established that pre-paid (or unconditional) incentives tend to produce more substantial effects on response rates than those that are promised (or conditional) on completion of the survey (Cantor et al., 2008; Church, 1993; Lavrakas, 2008; Singer, Hoewyk, et al., 1999),though it does not follow from this that they are necessarily more *cost-effective* (Brick, Montaquila, Hagedorn, Roth, & Chapman, 2005). It is also apparent from these studies that the effect of incentives is greater for surveys that have a low response rate when no incentive is offered, presumably because there is more scope for the incentive to act as a replacement for non-monetary motivations amongst a larger pool of potential nonrespondents(Singer, Hoewyk, et al., 1999).

Researchers have also established that the magnitude of the effect of incentives on response rates increases with the size of the incentive. For instance, in a meta-analysis of 39 experimental studies Singer, Groves, and Corning (1999) found that each dollar of incentive paid resulted in one third of a percentage point increase in response rate, compared to the no incentive condition. However, other studies have found that this 'dose-response' relationship is curvilinear, with the size of the increase in the response rate declining with additional increases in the value of the monetary incentive(Cantor et al., 2008; Gelman, Stevens, & Chan, 2002). In sum, the existing evidence demonstrates that monetary incentives have a robust, positive effect on the probability of survey cooperation.

The joint effect of interviewers and incentives on response rates

We know, then, that interviewers and incentives have a positive influence on response rates, what though of their joint effect? It seems plausible that interviewers might moderate the effect of incentives on cooperation probability for three inter-related reasons. First, interviewers are the primary conduit of information between survey organisation and sample members and are, therefore, essential to ensuring that potential respondents are aware that an incentive is available. While most surveys will highlight incentives in an advanced letter, many respondents do not open, let alone read them (Stoop, 2005). Furthermore, it seems reasonable to assume that the sorts of people who do not read advanced letters – those who are busy and/or uninterested in the survey topic - are also more likely to be susceptible to monetary incentives. Second, interviewers may have more confidence in the likelihood of obtaining an interview when a monetary incentive is offered. This might exert an additional positive effect on cooperation over and above the influence of the incentive on respondents, because higher levels of confidence improve the quality of interviewer approaches

(Groves & Couper, 1996; Singer, Hoewyk, et al., 1999; Singer & Ye, 2013). Third, interviewers may vary in the extent to which they tailor their doorstep introductions by highlighting the availability of the incentive at addresses where they believe it is likely to be effective. For example, some interviewers might ask sample members whether they received the letter with information about the payment at an early stage of the interaction, while others do not mention it at all.

Existing research, however, offers little in the way of hard evidence on the question of whether or not interviewers moderate the effects of monetary incentives on cooperation probability. An exception is Singer, Hoewyk, and Maher (2000), who investigated the influence of interviewer expectations on the effect of incentives on cooperation rates, using data from the Survey of Consumer Attitudes, a telephone survey of the American public. Singer and colleagues randomly assigned interviewers and respondents to three groups: in groups 1 and 2 respondents received an advance letter and a $5 unconditional incentive, while respondents in group 3 received an advance letter but no incentive. Interviewers in group 1 were unaware of the incentive but interviewers in groups 2 and 3 were made aware of the incentive level via messages on their computers. Interviewers in groups 1 and 2 achieved response rates of 76% and 75%, respectively, compared to 62% for interviewers in group 3. Singer, Hoewyk, and Maher (2000) concluded that, although the unconditional incentive boosted response, interviewer expectations about the likely cooperativeness of sample members had no additional effect. Lynn (2001) found similar evidence from a focus group of interviewers that expectations about the likely impact of incentives on cooperation bore little resemblance to actual response outcomes. While these studies support the conclusion that incentives operate primarily or exclusively via their effects on respondents rather than on interviewers, they do not rule out the possibility that interviewers vary in the effectiveness with which they deploy incentives. We turn next to a direct empirical assessment of this question.

Data

We use data from three different UK face-to-face interview surveys. These are the 2015 National Survey for Wales Field Test (NSW2015), the 2016 National Survey for Wales Incentive Experiment (NSW2016), and Wave 1 of the UK Household Longitudinal Study Innovation Panel (UKHLS-IP). All three surveys use stratified random sampling, with addresses selected from the Postcode Address File. The two Welsh surveys randomly select one eligible adult (aged 16 and over), while UKHLS-IP attempts interviews with all eligible adults (aged 18 and over) in the household. For UKHLS-IP a cooperating household is defined as one in which at least one eligible adult provided an interview. NSW2015

randomly allocated 50% of addresses to receive no incentive and 50% to receive £10, NSW2016 also used a 50/50 allocation but with a treatment condition of £5 and a control condition of no incentive. The UKHSL-IP randomly allocated a third of addresses to receive a £10 incentive and the remainder £5. Incentives in all three surveys were offered conditional on completion of the questionnaire and allocation of addresses to experimental conditions was implemented within interviewer workloads. More detailed information about the design of each survey is provided in the Appendix.

Each survey was linked to administrative data held on interviewers by the respective survey agencies. These were: age, sex, and experience (number of years working for the agency). We use these variables to assess whether interviewer characteristics are associated with variability in the effectiveness of deploying incentives. For the UKHLS-IP, we also link aggregate census variables from the 2011 census to the sample file. A total of 21 census count variables were combined using a factorial ecology model (Rees, 1971), with a total of five neighbourhood indices extracted. These measures cover the extent of *concentrated disadvantage* (areas with a higher number of single parent families, those on income support and unemployed, fewer people in managerial and professional occupations, and less owner occupiers), *urbanicity* (high population density and domestic properties, and relatively little green space) and *population mobility* (higher levels of in- and out-migration and more single person households). We also account for differences in the neighborhood *age structure* (with higher scores for areas with a younger population), *housing structure* (higher scores for areas with more terraced and vacant properties), and the police recorded crime rate.

Response outcomes for the three surveys are presented in Table 1. The cooperation and response rates were higher in the incentive condition for all three surveys, with the UKHLS-IP and NSW2016 having a 2 percentage point, and the NSW2015 a 5 percentage point higher cooperation rate for the incentivised households. The difference is statistically significant at the 95% level of confidence (using a Chi Square test) for NSW2015 but not for UKHLS-IP or NSW2016. For UKHLS-IP, the response rate is 4 percentage points higher in the incentive condition, due to a slightly lower rate of non-contacts and ineligible addresses in the higher incentive group.

Table 1, then, demonstrates the cooperation rate was higher in the incentivised condition for all three surveys, though the difference was statistically significant in only one. Next we proceed to a multivariate analysis to assesses whether these average differences in cooperation rates are constant across interviewers, or whether some interviewers are more successful at using the incentive to convert refusals into interviews.

Table 1: Incentives and& fieldwork outcomes for the three surveys

	NSW2015		NSW2016		UKHLS-IP	
	£10	£0	£5	£0	£10	£5
Interviews	1,504	1,319	1,801	1,693	1,020	469
Refusals	591	649	972	978	459	233
Non-contact	221	224	225	232	65	50
Other nonresponse	247	246	207	204	50	44
Ineligible	402	392	392	367	256	135
Cooperation Rate	72%	67%	65%	63%	69%	67%
Response Rate	65%	60%	60%	58%	66%	62%
Total issued sample	2,965	2,830	3,604	3,467	1,850	931

Analysis

The influence of interviewers on the effectiveness of incentives on survey cooperation is assessed using multilevel logistic regression models (Goldstein 2010; Durrant and Steele 2009; Hox and de Leeuw 2002). The model applied here has the following form. Let y_{ij} denote the binary response for household i $(i = 1, \dots, i)$, interviewed by interviewer j $(j = 1, \dots, j)$ where

$$y_{ij} = \begin{cases} 1 & \text{cooperation} \\ 0 & \text{refusal} \end{cases}$$

y_{ij} is assumed to follow a Bernoulli distribution, with conditional response probabilities $\pi_{ij} = Pr(y_{ij} = 1)$ and $1 - \pi_{ij} = Pr(y_{ij} = 0)$. The multilevel logistic regression model accounting for interviewer effects takes the form

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + \beta_1 x_{1ij} + \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_j\boldsymbol{\alpha} + \mu_{oj} + \mu_{1j} \qquad (1)$$

where x_{1ij} is a dummy indicator of the incentive group for household *i* within the assignment of interviewer *j*, $\mathbf{x}'_{ij}$ is a vector of household-level characteristics with coefficient vector $\boldsymbol{\beta}$, $\mathbf{z}'_j$ is a vector of interviewer-level covariates with coefficient vector $\boldsymbol{\alpha}$, μ_{0j} is a random intercept and μ_{1j} is a random coefficient for the incentive dummy. The random intercept and slope variances are assumed

to follow a normal distribution with zero mean and constant variances: $\mu_{oj} \sim N(0,\ \sigma_{\mu 0}^2)$, $\mu_{1j} \sim N(0,\ \sigma_{\mu 1}^2)$. The random coefficient on the dummy variable for incentive allows an interviewer influence on the effect of incentive on cooperation. This random coefficient introduces a covariance between μ_{oj} and μ_{1j} which is denoted σ_{u01}. Positive values of σ_{u01} indicate that the effect of the incentive is greater for interviewers with higher response rates, negative values indicate the opposite. Cross-level interactions between interviewer characteristic variables and the incentive variable are included to test whether observable characteristics of interviewers are associated with variability in the effectiveness of deploying incentives.

In standard face-to-face survey designs such as those considered here, identification of interviewer effects is complicated by the confounding of interviewer assignments and areas (Campanelli & O'Muircheartaigh, 1999; Durrant, Groves, Staetsky, et al., 2010). Failure to account for differences in the area-level composition of interviewer assignments can result in over-estimation of the magnitude of interviewer effects (O'Muircheartaigh and Campanelli, 1998). Where there is an overlap between interviewer assignments and areas, this can be mitigated using a cross-classified multi-level model (Durrant and Steele, 2009). However, this could not be done for the three datasets analysed here, because it was not possible to obtain geographic identifiers for the two Welsh surveys and the UKHLS-IP did not contain sufficient crossing of interviewers and areas to implement a cross-classified model. We therefore control for area characteristics as fixed effects in the models for the UKHLS-IP data and assess the impact this has on the interviewer random effects.

Models are estimated using Markov Chain Monte Carlo (MCMC) methods using MLwiN software (Browne, Kelly, Charlton, & Pillinger, 2016; Gelman, Carlin, Stern, & Rubin, 2004). The starting values for the random parameters are vague priors and second-order penalised quasi-likelihood (PQL) estimates for fixed effects. Priors for the variance matrix are assumed to follow an inverse Wishart distribution $p(\Omega^{-1}) \sim Wishart_n(n,)$, where n is the number of rows in the variance matrix and is an entire estimate for the true value of Ω (Browne et al., 2016). Because we are using MCMC we also assess significance of coefficient estimates using the change in model Deviance Information Criterion (DIC) (Spiegelhalter, Best, Carlin, & van der Linde, 2002). DIC balances model fit and model complexity by taking the sum of the posterior expectation (mean) of the deviance function ($\overline{D}$) and the effective number of parameters (pD). When comparing DIC values, a model with a DIC value of at least 3 points lower than the previous model is considered to have a significantly better fit (Rasbash et al., 2012; Spiegelhalter et al., 2002).

Results

Table 2 presents the coefficient estimates, their standard deviations, and the corresponding 95% credible intervals for the NSW2015 and NSW2016 models. As we saw in Table 1, the coefficients for the incentive fixed effect are positive for both surveys, although only for NSW2015 does the 95% credible interval not include zero. The random coefficient variances of 0.11 and 0.07 are both significant, indicating that interviewers vary in the effectiveness with which they deploy incentives. The DIC decreases by 14.7 for NSW2015 and by 13.3 for NSW2016 when the interviewer random coefficient is introduced, indicating an improvement in model fit.

The cross-level interactions between the three interviewer characteristic variables – age, sex, and experience – and the incentive dummy are all non-significant, indicating that these interviewer characteristics do not explain between interviewer variability in the effectiveness of incentives on cooperation. The DIC change, when these interaction terms are added, are -2 for NSW2015 and 3.5 for NSW2016, indicating a small improvement in model fit after the inclusion of these interactions for NSW2016. The covariance between the random intercept and random coefficient, σ_{u01}, is non-significant for both surveys, with a point estimate of 0.023 for NSW2015 and of -0.031 for NSW2016. This indicates that the effectiveness of incentive deployment between interviewers is not related to the overall response rate an interviewer achieves on their assignment of addresses.

Figure 1 plots the difference in the mean predicted probability of cooperation for each interviewer derived as fitted values from the models in Table 2. Each blue dot in Figure 1 represents an interviewer, with the left Y axis being the difference in the response rate for households in the incentive and non-incentive conditions. The brown triangles show the overall cooperation rate (plotted against the right Y axis) for each interviewer across all eligible households in their assignment. There is substantial variability across interviewers in the effectiveness of the incentive in obtaining cooperation. For NSW2015, the difference ranges from -9% to +13%, with the corresponding values for NSW2016 being -8% and +14%. Not all of this variability is attributable to how skilful interviewers are in deploying incentives and simply reflects random variability in response propensities across interviewer assignments. We can get a better sense of the effect of interviewers on incentive effectiveness by taking the expected response rate for an incentivised household using interviewers from the top and bottom deciles of the random coefficient variance, σ_{1j}^2 while holding all other variables constant. For NSW2015, this shows that interviewers in the top performing decile achieve an expected cooperation rate of 81% for incentivised households compared to 65% for those in the bottom decile and compared to 66% for the median interviewer for non-incentivised households, a quite substantial difference.

Table 2: Estimated coefficients for the final model for NSW2015 and NSW2016 Cooperation)

		NSW2015				*NSW2016*			
Variable	*Category*	β	*SD*	*0.025 Quantile*	*0.975 Quantile*	β	*SD*	*0.025 Quantile*	*0.975 Quantile*
Intercept		0.244	0.276	-0.281	0.814	0.623	0.264	0.105	1.148
Incentive *{no incentive}*	£10 Incentive	0.255	0.080	0.099	0.413	0.067	0.317	-0.534	0.678
Interviewer age *{young }*	Lower middle	0.222	0.201	-0.165	0.625				
	Upper Middle	0.289	0.197	-0.102	0.682	0.189	0.134	-0.074	0.453
	Old	0.331	0.242	-0.139	0.810				
Interviewer Experience {*less }*	Lower middle	0.089	0.216	-0.342	0.511	0.017	0.262	-0.503	-0.527
	Upper middle	0.353	0.242	-0.216	0.829	-0.294	0.301	-0.893	0.290
	Highest	0.431	0.230	-0.024	0.887	-0.364	0.319	-1.000	0.256
Interviewer Sex *{Female}*	Male	-0.060	0.137	-0.330	0.209	-0.155	0.133	-0.376	0.151
Incentive *{£10 per adult}*Gender *{Female}*	£10 per adult *Male	-0.020	0.184	-0.376	0.343	-0.281	0.145	-0.566	0.004
Incentive *{£10 per adult}* * Age *{young}*	£10* Lower middle	-0.089	0.275	-0.634	0.447	0.057	0.315	-0.554	0.649
	£10* Upper Middle	-0.035	0.272	-0.570	0.498	0.584	0.361	-0.125	1.282
	£10* Old	-0.362	0.328	-1.012	0.281	0.065	0.371	-0.671	0.782
Incentive *{£5}* * Experience *{less}*	£10*Lower Middle	-0.160	0.284	-0.721	0.394	0.057	0.315	-0.554	0.649
	£10*Upper Middle	-0.305	0.317	-0.935	0.323	0.584	0.361	-0.125	1.282
	£10*Highest	0.024	0.304	-0.575	0.619	0.065	0.371	-0.671	0.782
$\sigma_{u0}^2 = var(\mu_{oj})$		0.156	0.055	0.070	0.285	0.145	0.048	0.0781	0.257
$\sigma_{u1}^2 = var(\mu_{1j})$		0.109	0.060	0.031	0.260	0.067	0.035	0.021	0.154
$\sigma_{u01} = cov(\mu_{0j}, \mu_{1j})$		0.023	0.041	-0.069	0.095	-0.031	0.036	-0.115	0.025
DIC		4799.673				6683.025			
N=4,063 (for NSW 2015)						N=5,264 (for NSW 2016)			

The corresponding figures for NSW2016 are 63% and 77% for the top and bottom deciles, respectively, and 63% for the median interviewer for non-incentivised households. There is no obvious relationship between the overall response rate and the effectiveness of the incentive within interviewers, so we find no evidence that interviewers who are, on average, better at obtaining cooperation are also more effective in deploying the incentive.

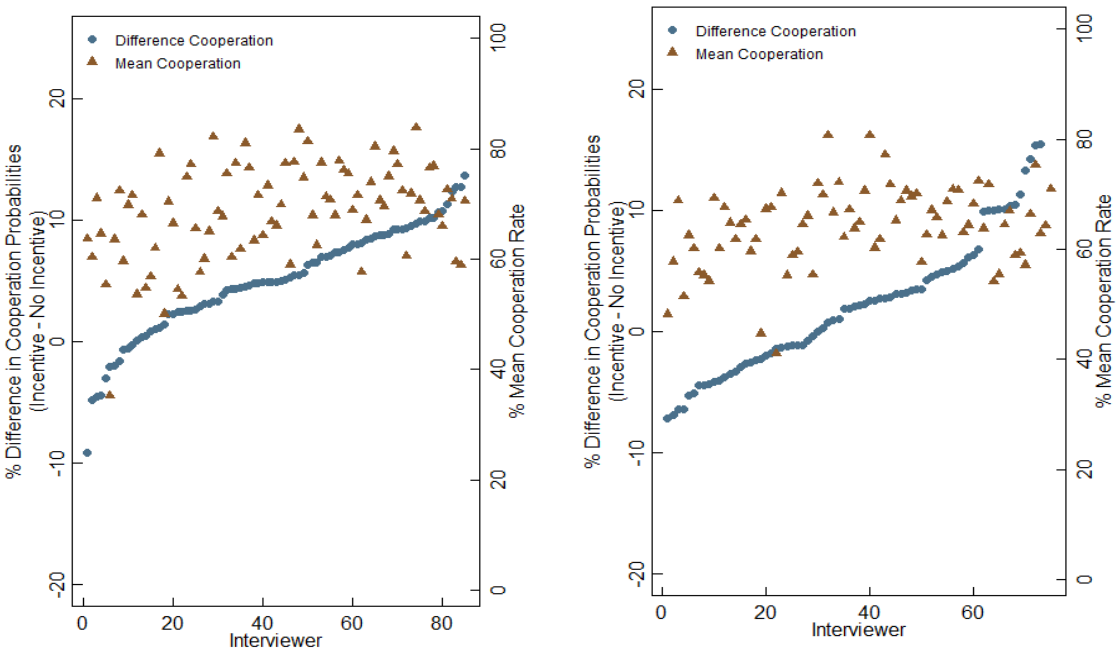

Figure 1: Difference in predicted probability of cooperation for incentive and non-incentive households by interviewer for NSW 2015 (left panel) and NSW 2016 (right panel)

Not all of this variability is attributable to how skilful interviewers are in deploying incentives and simply reflects random variability in response propensities across interviewer assignments. We can get a better sense of the effect of interviewers on incentive effectiveness by taking the expected response rate for an incentivised household using interviewers from the top and bottom deciles of the random coefficient variance, σ_{1j}^2 while holding all other variables constant. For NSW2015, this shows that interviewers in the top performing decile achieve an expected cooperation rate of 81% for incentivised households compared to 65% for those in the bottom decile and compared to 66% for the median interviewer for non-incentivised households, a quite substantial difference. The corresponding figures for NSW2016 are 63% and 77% for the top and bottom deciles, respectively, and 63% for the median interviewer for non-incentivised households. There is no obvious relationship between the overall response rate and the effectiveness of the incentive within interviewers, so we

find no evidence that interviewers who are, on average, better at obtaining cooperation are also more effective in deploying the incentive.

Next we turn to the same analysis of UKHLS-IP, which as a household longitudinal survey, has a rather different design to the Welsh cross-sectional surveys, albeit that we focus here on wave 1 only. Table 3 presents the estimated coefficients, standard deviations, and corresponding 95% credible intervals. There are consistent with those presented in Table 2; the fixed effect for the incentive predicting cooperation is positive but non-significant and the interviewer characteristics - age, gender, and experience - are all non-significant, as are the interactions between these variables and the incentive fixed effect.

Two of the area level variables are significantly associated with cooperation; the higher the crime rate, the lower the level of survey cooperation, while areas with a housing structure comprising more terraced housing and vacant properties have higher levels of cooperation. Even after controlling for these differences in area composition, the random coefficient for the incentive is significant, with a variance of 0.17 (95% credible interval 0.04 – 0.44). This suggests that the between interviewer variability in the effectiveness of the incentive is caused by interviewer behaviour, rather than by differences in the sorts of people they have been allocated to interview. The model DIC decreases by 3.10 with the inclusion of the random coefficient, so, we also find evidence of a between interviewer difference in the effectiveness of the incentive on this alternative measure of statistical significance.

As with the Welsh surveys, the covariance between the random intercept and random slope is positive but with a 95% credible interval that includes zero. We therefore also find no support from UKHLS-IP for the idea that interviewers who, on average, obtain higher cooperation rates might also be more effective in their deployment of incentives.

Figure 2 plots the difference in the mean predicted probability of cooperation for each interviewer derived as fitted values from the models in Table 3. It shows a very similar pattern to what we saw in Figure 1 for the Welsh surveys, with substantial between-interviewer variation in cooperation probabilities between high and low incentive groups with a range of -16% to +23%. Visually, there is more evidence of a positive correlation between percentage difference in cooperation probabilities and the overall response rate for each interviewer, although this difference is not statistically significant.

Table 3: Estimated coefficients for the final model for Innovation Panel Cooperation

Variable *{reference category}*	*Category*	*β*	*SD*	*Posterior 0.025 Quantile*	*Posterior 0.975 Quantile*
Intercept		0.668	0.538	-0.343	1.803
Incentive	£10 per adult	0.512	0.569	-0.630	1.643
Neighbourhood Characteristics					
Housing structure		0.253	0.082	0.095	0.415
Crime rate		-0.320	0.149	-0.616	-0.034
Socio-economic disadvantage		0.155	0.092	-0.023	0.338
Urbanicity		-0.054	0.098	-0.245	0.139
Population Mobility		0.028	0.094	-0.155	0.213
Age Profile		0.084	0.069	-0.052	0.220
Interviewer Characteristics					
Gender {Female}	Male	-0.204	0.255	-0.710	0.287
Age {less than 40 years}	41 to 50 years	-0.344	0.601	-1.599	0.791
	50 to 60 years	0.729	0.543	-0.378	1.776
	> 60 years	0.397	0.558	-0.748	1.485
Experience {less than 2 yrs.)	3 to 6 years	-0.213	0.286	-0.776	0.349
	7 to 9 years	-0.161	0.355	-0.857	0.533
	>10 years	-0.200	0.472	-1.129	0.719
*Incentive * Gender*	£10 per adult *Male	0.017	0.260	-0.491	0.537
*Incentive * Age*	£10 per adult *41 to 50 years	-0.265	0.632	-1.542	0.954
	£10 per adult *50 to 60 years	-0.487	0.576	-1.655	0.645
	£10 per adult *> 60 years	-0.365	0.586	-1.567	0.785
*Incentive * Experience*	£10 per adult *3 to 6 years	-0.138	0.290	-0.708	0.428
	£10 per adult *7 to 9 years	0.329	0.373	-0.400	1.060
	£10 per adult *>10 years	-0.160	0.480	-1.106	0.775
Random effects					
$\sigma^2_{0jk} = var(\mu_{ojk})$		0.493	0.215	0.179	1.008
$\sigma^2_{1k} = var(\mu_{1k})$		0.154	0.107	0.032	0.430
$\mu_{01} = cov(\mu_{ok}, \mu_{1k})$		0.092	0.113	-0.174	0.280
DIC		2116.654			

UKHLS-IP, wave 1, N= 1847

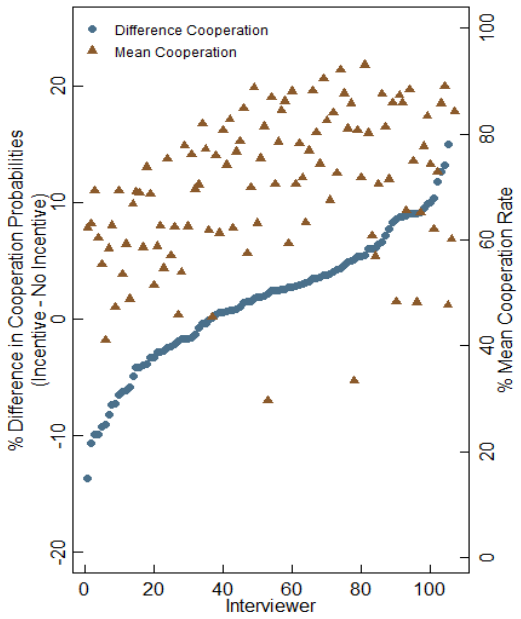

Figure 2: Difference in predicted probability of cooperation for incentive and non-incentive households by interviewer for UKHLS-IP

Discussion

John Wannamaker, the American department store magnate, once (apocryphally) observed that "half the money I spend on advertising is wasted; the trouble is I don't know which half'. The same sentiment might also be applied to monetary incentives in surveys, although in this context considerably more than half of the money is wasted. This is because incentives generally add only a few percentage points or so to the headline response rate. It follows, therefore, that the majority of respondents in any survey using a monetary incentive would have agreed to provide an interview anyway. A small minority, however, are susceptible to being converted from refusal to interview with the provision of an incentive and this, in turn, raises the possibility that interviewers might play an important role in determining the rate of such 'conversions'. While there are, of course, other reasons for providing monetary incentives than boosting the response rate, this remains the primary rationale in most cases. It is, therefore, important to understand how best to maximise the effectiveness of monetary incentives in converting refusals to interviews. This is all the more pressing, given the likely need to place greater reliance on incentives to maintain response rates in the future.

Our findings show that, across three different UK face-to-face surveys, interviewers vary significantly in how effective they are at using incentives to increase rates of cooperation. The effects we observe are substantively as well as statistically significant; our model estimates show that exchanging interviewers from the top to the bottom decile of interviewer performance would yield an expected 14 to 15 percentage point increase in the effect of the incentive relative to the control condition. We have speculated that this heterogeneity results from interviewer expectations and behaviour, particularly the use of 'tailoring' of doorstep interactions (Groves & Couper, 1998) and greater confidence in the probability of obtaining an interview when an incentive is offered (Singer, Frankel, & Glassman, 1983; Singer & Maher, 2000). However, while the between interviewer variability in the effectiveness of incentives was consistent across the three surveys, we found no significant predictor of this variance amongst the covariates considered: interviewer age, sex, and experience. Nor was variability in incentive effectiveness related to the overall response rate an interviewer achieved. The mechanisms underpinning this effect, therefore, remain unclear.

Our focus in this paper has been on the effect of incentives on cooperation because incentives seem likely to exert their primary influence on the cooperation decision. However, it is possible that they also have an effect on contact rates and other categories of nonresponse. We have therefore also carried out the analyses reported here with the dependent variable specified as response/nonresponse. The results are substantively identical to those reported here, so we find no evidence of a differential effect of interviewers on cooperation relative to total nonresponse.[1]

Our findings have implications for survey practice. The approach we have implemented here to identify interviewer effectiveness in deploying incentives could be used as a way of identifying underperforming interviewers. This sort of monitoring is now implemented routinely in many large-scale survey operations, often in real-time, as a way of identifying interviewers who show signs of missing fieldwork targets (Edwards, Maitland, & O'Connor, 2017; Kreuter, 2013). It should be feasible to include 'incentive performance' alongside other forms of paradata to raise flags against particular interviewers on this performance dimension, although how this would be adapted to designs in which all households are offered the same incentive would require further consideration.

Relatedly, the ability to identify interviewers at the top end of the performance distribution offers opportunities for better understanding the sorts of strategies employed by more successful interviewers. Information on successful approaches to incentive use that are identified in this way could be integrated into sections of interviewer briefings which address doorstep approaches, both for generic and survey-specific training. Indeed, simply highlighting to interviewers that the way they

[1] These analyses are available from the corresponding author upon request.

administer incentives can have substantial effects on their response outcomes may, on its own, have some effect on their subsequent behaviour.

While our methodological approach and findings represent an advance in our understanding of how interviewers and incentives interact to promote cooperation, this study is not without limitations and these should be acknowledged. First, the surveys we have considered all use a relatively narrow range of incentive values which are administered to all households in the incentive condition. Caution should therefore be exercised in generalising to contexts where larger incentives are used, or where incentives of varying values are targeted at different sub-groups of the sample based on response propensities (Lavrakas, McPhee, & Jackson, 2016). Our results also have little relevance to the use of incentives in online surveys, which comprise a large and growing proportion of total survey volume, both in the UK and internationally.

We were also able to link the sample file and response outcome data to a limited range of area and interviewer characteristics. It is, therefore, possible that with stronger controls for differences between interviewers in the composition of their allocated addresses the magnitude of the effects we have observed here might be reduced. The paucity of interviewer characteristic data available to us, particularly the absence of variables measuring interviewer attitudes, beliefs, and behaviours means that our ability to explain *why* some interviewers are more effective in deploying monetary incentives than others is weak. These limitations, we contend, represent potentially fruitful avenues for future research.

References

Aumeyr, M., Brown, Z., Doherty, R., Fallows, A., Pegg, T., Perez-dominguez, R., … Shemwell, L. (2017). *National Survey for Wales 2016-17 Technical Report*. Cardiff. Retrieved from http://www.gov.wales/nationalsurvey

Blom, A. G., & Korbmacher, J. M. (2013). Measuring Interviewer Characteristics Pertinent to Social Surveys: A Conceptual Framework. *Survey Methods: Insights from the Field*.

Blom, A. G., Leeuw, E. D. De, & Hox, J. J. (2011). Interviewer Effects on Nonresponse in the European Social Survey. *Journal of Official Statistics*, *27*(2), 359–377.

Boreham, A. R., & Constantine, R. (2008). Understanding Society Innovation Panel Wave 1 Technical Report, (7083), 1–14.

Brick, M. J., Montaquila, J. M., Hagedorn, M. C., Roth, S. B., & Chapman, C. (2005). Implications for RDD design from an incentive experiment. *Journal of Official Statistics*, *21*(4), 571–589.

Brick, M. J., & Williams, D. (2013). Explaining Rising Nonresponse Rates in Cross-Sectional Surveys. *The ANNALS of the American Academy of Political and Social Science*, *645*(1), 36–59. http://doi.org/10.1177/0002716212456834

Browne, W. J., Kelly, M., Charlton, C., & Pillinger, R. (2016). *MCMC estimation in MLwiN: Version 2.36*.

Campanelli, P., & O'Muircheartaigh, C. (1999). Interviewers , Interviewer Continuity , and Panel Survey Nonresponse 1. *Quality & Quantity*, *33*, 59–76.

Campanelli, P., Sturgis, P., & Purdon, S. (1997). Can you hear me knocking? and investigation into the impact of interviewers on survey response rates. *National Centre for Social Research*.

Cantor, D., O'Hare, B. C., & O'Connor, K. S. (2008). The use of monetary incentives to reduce nonresponse in random digit dial telephone surveys. In *Advances in telephone survey methodology* (pp. 471–498).

Church, A. H. (1993). Estimating the Effect of Incentives on Mail Survey Response Rates : A Meta-Analysis. *Public Opinion Quarterly*, *57*(1), 62–79.

Couper, M. P. (2013). Is the Sky Falling? New Technology, Changing Media, and the Future of Surveys. *Survey Research Method*, *7*(3), 145–156.

Couper, M. P., & Schlegel, J. (1998). Evaluating the NHIS CAPI instrument using trace files. In *In Annual Meeting of the American Association for Public Opinion Research*. St. Louis. Retrieved from https://ww2.amstat.org/sections/srms/Proceedings/papers/1997_142.pdf

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method* (3rd ed.). New York, NY: John Wiley & Sons.

Durrant, G. B., Groves, R. M., Staetsky, L., & Steele, F. (2010). Effects of interviewer attitudes and behaviors on refusal in household surveys. *Public Opinion Quarterly*, *74*(1), 1–36. http://doi.org/10.1093/poq/nfp098

Durrant, G. B., Groves, R. M., & Steele, F. (2010). EFFECTS OF INTERVIEWER ATTITUDES AND BEHAVIORS ON REFUSAL IN HOUSEHOLD SURVEYS. *Public Opinion Quarterly*, *74*(1), 1–36. http://doi.org/10.1093/poq/nfp098

Durrant, G. B., & Steele, F. (2009). Multilevel modelling of refusal and non-contact in household

surveys: Evidence from six UK Government surveys. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, *172*(2), 361–381. http://doi.org/10.1111/j.1467-985X.2008.00565.x

Edwards, B., Maitland, A., & O'Connor, K. S. (2017). Measurement error in survey operations management: detection, quantification, visualization, and reduction. In P. Biemer, E. de Leeuw, S. Eckman, B. Edwards, F. Kreuter, L. Lyberg, … B. T. West (Eds.), *Total Survey Error in Practice*. Hoboken,NJ: John Wiley & Sons.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian Data Analysis*. *Chapman Texts in Statistical Science Series*. http://doi.org/10.1007/s13398-014-0173-7.2

Gelman, A., Stevens, M., & Chan, V. (2002). Regression modeling and meta-analysis for deci- sion making: A cost-benefit analysis of incentives in telephone surveys. *Journal of Business & Economic Statistics*, *21*(2), 213–225.

Groves, R. M., & Couper, M. P. (1996). contact level influences on cooperation in Face to Face surveys.pdf. *Journal of Official Statistics*, (12), 63–83.

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

Groves, R. M., & Heeringa, S. G. (2006). Responsive design for household surveys: Tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, *169*(3), 439–457. http://doi.org/10.1111/j.1467-985X.2006.00423.x

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-Saliency Theory of Survey Participation Description and an Illustration. *The Public Opinion Quarterly*, *64*(3), 299–308.

Hansen, K. M. (2006). The Effects of Incentives , Interview Length , and Interviewer Characteristics on Response Rates in a CATI-Study. *International Journal of Public Opinion Research*, (19), 112–121. http://doi.org/10.1093/ijpor/edl022

Hanson, T., Sullivan, S., & Mcgowan, C. (2015). *National Survey for Wales 2014-15 Technical Report*. Cardiff.

Hox, J., & de Leeuw, E. (2002). The influence of interviwers' attitude and behavior on household survey nonresponse: An international comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 103–120). New York: Wiley.

Kreuter, F. (2013). *mproving Surveys with Paradata: Analytic Uses of Process Information*.

Hoboken,NJ: John Wiley & Sons.

Lavrakas, P. J. (2008). *Encyclopedia of Survey Research Methods*. Thousand Oaks, Calif: SAGE Publications.

Lavrakas, P. J., McPhee, C., & Jackson, M. (2016). Conceptual Background on Response Propensity Modeling (RPM) for Allocating Differential Survey Incentives: JSM 2016 - Survey Research Methods Section 3280 Purpose, Rationale, and Operationalization. In *Presented at the 71st annual conference of the American Association for Public Opinion Research*. Austin Texas.

Lynn, P. (2001). The impact of incentives on response rates to personal interview surveys: Role and perceptions of interviewers. *International Journal of Public Opinion Research*, *13*(3), 326–336.

Meyer, B. D., Mok, W. K. C., & Sullivan, J. X. (2015). Household Surveys in Crisis. *Journal of Economic Perspectives*, *29*, 199–226.

Miller, P. (2017). Is There a Future for Surveys? *Public Opinion Quarterly*, *81*(1), 205–212. Retrieved from https://doi.org/10.1093/poq/nfx008

Morton-Williams, J. (1993). *Interviewer Approaches*. Aldershot: Dartmouth Publishing Company Limited.

Purdon, S., Campanelli, P., & Sturgis, P. (1999). Interviewers â€™ Calling Strategies on Face-to-Face Interview Surveys. *Journal of Offcial Statistics*, *15*(2), 199–216.

Rasbash, J., Steele, F., Browne, W. J., Goldstein, H., Charlton, C., & Browne, W. J. (2012). A User ' s Guide to MLwiN, v2.26. *Centre for Multilevel Modelling, University of Bristol*.

Rees, P. H. (1971). Factorial ecology: An extended definition, survey and critique of the field. *Economic Geography*, *47*, 220–233. http://doi.org/10.2753/JEI0021-3624440403

Singer, E. (2002). The Use of Incentives to Reduce Nonresponse in Household Surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey Nonresponse* (pp. 163–177). New York: Wiley.

Singer, E., Frankel, M. R., & Glassman, M. B. (1983). The Effect of Interviewer Characteristics and Expectations on Response. *Public*, *47*, 68–83.

Singer, E., Groves, R. M., & Corning, A. D. (1999). Differential incentives: Beliefs about practices, perceptions of equity, and effects on survey participation. *Public Opinion Quarterly*, *63*(2), 251–260. http://doi.org/10.1086/297714

Singer, E., Hoewyk, J. Van, Gebler, N., Raghunathan, T., & Mcgonagle, K. (1999). The Effect of Incentives on Response Rates in Interviewer-Mediated Surveys, *15*(2), 217–231.

Singer, E., Hoewyk, J. Van, & Maher, M. P. (2000). Experiments with Incentives In Telephone Surveys. *Public Opinion Quarterly*, *64*(2), 171–188.

Singer, E., & Kohnke-Aguirre, L. (1977). Interviewer Expectation Effects : A Replication and Extension. *Public Opinion Quarterly*, *63*, 245–260.

Singer, E., & Maher, M. P. (2000). Experiments with Incentives in Telephone Surveys. *Public Opinion Quarterly*, *64*(2), 171–188.

Singer, E., & Ye, C. (2013). The Use and Effects of Incentives in Surveys. *The ANNALS of the American Academy of Political and Social Science*, *645*(1), 112–141. http://doi.org/10.1177/0002716212458082

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian Measures of Model Complexity anf Fit. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, *64*(4), 583–639. http://doi.org/10.1111/1467-9868.00353

Stoop, I. A. L. (2005). *The hunt for the last respondent*. The Hague, Netherlands.

Vassallo, R., Durrant, G. B., & Smith, P. W. F. (2015). Interviewer effects on non-response propensity in longitudinal surveys : a multilevel modelling approach. *Ournal of the Royal Statistical Society: Series A (Statistics in Society)*, *178*(1), 83–99.

West, B. T., & Blom, A. G. (2017). EXPLAINING INTERVIEWER EFFECTS : A RESEARCH SYNTHESIS. *Journal of Survey Statistics and Methodology*, *5*(November), 175–211. http://doi.org/10.1093/jssam/smw024

Appendix 1 Descriptions of Survey Sample Designs & Methodology

National Survey for Wales Field Test 2015 (NSW 2015)

The sample design of the NSW2015 used a stratified, single-stage random selection of addresses across Wales drawn from the small user Postcode Address File (PAF). Adult aged 16 or over within each sampled household were interviewed face-to-face and each interview lasted for an average of 25 minutes. Where a household contained more than one adult, a single adult was randomly selected. The aim of the incentive experiment was to assess the extent to which response rates improved by offering respondents a £10 gift-card upon completing an interview. The experimental group (N=2,960) received a £10 conditional incentive and the second group received no incentive (N=2,828). The households which were randomly selected to be offered a conditional £10 received advance letters mentioning the incentive, while the other half of households received advance letters that contained no information about incentives. To ensure that any differences in response rates between respondents who were offered £10 and those offered no incentive are not attributed to any interviewer abilities, addresses that were offered incentives were randomly allocated within each interviewer assignment. The survey was implemented by a team of 86 interviewers with the number of households interviewed by each interviewer ranging between 14 and 134. Further details on the NSW2015 sample design can be found in Hanson, Sullivan, and Mcgowan (2015).

National Survey for Wales Incentive Experiment 2016 (NSW 2016)

The Welsh Government commissioned the office for National Office of National Statistics (ONS) to conduct the National Survey for Wales 2016 (NSW 2016) incentive experiment between July and October 2016. The sample was drawn from the Postcode Address File (PAF). The stratification was by Local Authority (LA) using an allocation designed to ensure a minimum effective sample size was achieved in each LA based on estimated response rate. Further details on the sample design can be found in Aumeyr et al. 2017). Half of addresses in each odd numbered quota[2] were offered a £5 incentive conditional on participation (N=3604), and addresses with even quota number were offered no incentive (N=3467). The incentive experiment ran from July to October 2016. Originally, it was intended to run the experiment until December 2016 but it was terminated at the end of October 2016 as both experimental and control groups experienced lower response rates at 55% and 53% respectively which were lower than expected. With an aim of boosting response rates, a new £10 incentive conditional on participation was introduced to the full sample from November 2016. This

[2] Each quota contained between 20 and 30 addresses on average.

study will only considers the experiment sample size from July to October 2016 that consist of 7,071 households across the two conditions. There were 85 interviewers working on the survey with the minimum and maximum number of interviews per interviewer ranging between 1 and 219. Socio-demographic characteristics of 10 (12%) interviewers who conducted interviews on 206 (3%) households were missing because they did not provide consent. The final analysis sample had 6,106 households after excluding 965 (13.6%) ineligible households and those interviewed by interviewers with missing socio-demographic characteristics.

UK Household Longitudinal Survey Innovation Panel Wave 1 (UKHLS-IP)

The sample for the UK Household Longitudinal Survey Innovation Panel Wave 1 (UKHLS-IP) was clustered and stratified, consisting of 2,760 addresses from 120 primary sampling units (PSUs) from the Postcode Address File (PAF). The incentive experiment comprised three conditions, with each condition receiving a different conditional incentive: Group 1 were offered £5 per adult, Group 2 £10 per adult, and Group 3 were offered £5 per adult, rising to £10 per adult if all adults in the household completed interviews. Single person households randomly assigned to the Group 3 received £5 initially that increased to £10 if they participated. For the purposes of our analysis, Groups 2 and 3 are combined. Note that all households were also sent an unconditional £5 incentive with the advance letter. There were 27 households in the UKHLS-IP that did not successfully merge with interviewer data due to lack of common unique identifiers. The neighbourhood characteristic variables are drawn from the census and were available for England only. This resulted in the exclusion of 344 households (13.6%) from Wales and Scotland. In addition, 39 (1.5%) households in 5 MSOAs in England did not successfully merge with Innovation Panel data due to lack of common unique identification codes. Therefore, the final analysis sample contained 2,113 households. The number of interviewers working on the UKHLS-IP was 107, with the number of households interviewed by each interviewer ranging between 2 and 50. Further details about the UKHLS-IP can be found in Boreham and Constantine 2008).