UNIVERSITY OF Southampton
Southampton Statistical Sciences Research Institute

Working Paper M11/03
Methodology

Analysing The Process Leading To Cooperation Or Refusal Using Call Record Data: A Multilevel Multinomial Modelling Approach

Julia D'arrigo, Gabriele B. Durrant, Fiona Steele

Abstract

In recent years, survey agencies have started to collect detailed call record data, including information on the timing and outcome of each interviewer call to a household. In interviewbased household surveys, effective interviewer calling behaviours are critical in achieving cooperation and reducing the likelihood of refusal. This paper aims to analyze interviewer call record data to inform the process leading to cooperation or refusal in face-to-face surveys. Of particular interest are the influences on the outcome of a call of interactions between the interviewer and householder and of time-varying characteristics of the call. A multilevel multinomial logistic regression approach is used in which the different possible outcomes at each call are modelled jointly.

Analysing the Process Leading to Cooperation or Refusal Using Call Record Data: A Multilevel Multinomial Modelling Approach

Julia D'Arrigo[1], Gabriele B. Durrant[1] and Fiona Steele[2]

[1]Southampton Statistical Sciences Research Institute (S3RI)
University of Southampton, UK

[2]Centre for Multilevel Modelling
University of Bristol, UK

Summary

In recent years, survey agencies have started to collect detailed call record data, including information on the timing and outcome of each interviewer call to a household. In interview-based household surveys, effective interviewer calling behaviours are critical in achieving cooperation and reducing the likelihood of refusal. This paper aims to analyze interviewer call record data to inform the process leading to cooperation or refusal in face-to-face surveys. Of particular interest are the influences on the outcome of a call of interactions between the interviewer and householder and of time-varying characteristics of the call. A multilevel multinomial logistic regression approach is used in which the different possible outcomes at each call are modelled jointly.

Key Words: event history analysis, interviewer call record data, multilevel multinomial logistic regression, paradata, survey cooperation.

1. Introduction

In recent years, many survey agencies have started to routinely collect call record data in interviewer administered surveys, which include both telephone and face-to-face surveys (Bates et al., 2008; LaFlamme, 2008; Blom et al. 2010). Such call record data contain, at a minimum, information about the day and time of the call, the outcome of the call, and household or sample member and interviewer identifiers which enable linkage to the main survey. They may also include further information, for example who the interviewer talked to and any interaction between the interviewer and the household member. In face-to-face surveys the interviewer may also observe certain characteristics about the household or neighbourhood at each visit, such as the type of accommodation. Such call record and interviewer observation variables constitute a form of paradata (Couper, 1998) since they contain information about the survey data collection process. Survey agencies hope that analysis of call record data may inform best interviewer calling practices. In particular, in interview-based household surveys effective interviewer calling behaviours are crucial in reducing nonresponse - an increasing problem in survey research (Bethlehem et al. 2011). An advantage of paradata is that they are available for *both* respondents and nonrespondents. If variables on the call history are useful for predicting nonresponse outcomes they may represent good candidates for nonresponse modelling and adjustment (Bates et al. 2008; Kreuter and Kohler, 2009).

This paper aims to analyze the process leading to cooperation or refusal in several face-to-face household surveys using call record data. Previous analysis on cooperation has primarily focused on the final response to the survey request, i.e. refusal or cooperation at the end of the data collection process (Groves et al. 1992; Groves and Couper, 1998; O'Muircheartaigh and Campanelli, 1999; Durrant and Steele, 2009). Here, the focus is on the process leading to this final outcome, and we model the response at each visit, conditional on contact being made with the household. The following possible outcomes at each call are

considered: cooperation, appointment made, other forms of postponement (for example where the interviewer withdrew to try again later) and refusal. As identified in Groves and Couper (1996), Groves and Heeringa (2006) and Bates et al. (2008), the interaction between the interviewer and householder and time-varying factors are important determinants of the response process, and our model explores their influences on the call outcome. We also analyse how the call history affects the outcome of future calls. This is of relevance since such models may be used in responsive survey designs (Groves and Heeringa, 2006; LaFlamme, 2008), where survey data collection procedures are continuously monitored allowing early intervention and the alteration of the survey design.

Previous work on the analysis of call record data has focused primarily on the process leading to contact (Weeks et al., 1980; Kulka and Weeks, 1988; Greenberg and Stokes, 1990; Purdon et al., 1999; Durrant et al., 2011). Although time consuming and expensive, the extent of non-contact can often be minimised by changing calling practices, for example by increasing the number of calls and varying the timing of the calls. For the majority of surveys considered in this paper, the non-contact rate is around 3%. However, survey agencies are facing a much more serious problem due to increasingly high refusal rates (Steeh et al, 2001; Bethlehem et al. 2011). For example, for the surveys considered here, the refusal rate ranges from 15% to 30%. Refusal rates may have serious consequences for the quality of the resulting survey data and survey agencies are faced with the challenge of improving interviewer calling behaviours to increase cooperation. The analysis presented here builds on earlier work by Durrant et al. (2011) which investigates the process leading to contact, in particular the best times to achieve contact. We extend this previous research by focussing on the more challenging issue of cooperation and refusal.

This paper also aims to contribute to the development of statistical models for the analysis of call record data. Previous work in this area modelled the odds of obtaining an interview at a call but did not distinguish between other outcomes (Groves and Couper, 1996;

Groves and Heeringa, 2006). The split here into four different outcome types extends the commonly used dichotomy of cooperation versus other outcomes (see also Purdon et al. 1999). In particular, our model also allows the investigation of the characteristics of households who prefer making an appointment. Earlier work mostly used descriptive analysis techniques and regression models that ignored the hierarchical structure of the data, such as the nesting of sample units within interviewers (Groves and Couper, 1996; Purdon et al., 1999; Sangster and Meekins, 2004; Groves and Heeringa, 2006; Bates et al. 2008). If multilevel models were employed the final response outcome was modelled rather than the response process across calls (O'Muircheartaigh and Campanelli, 1999; Pickery et al., 2001; Durrant and Steele, 2009). More recent work in the area of call record and interviewer observation variables has focused on the use of such data for nonresponse adjustment which implicitly includes the specification of response propensity models (Wood and White, 2006; Peytchev and Olson, 2007; Kreuter and Kohler, 2009; Biemer et al., 2010; Kreuter et al. 2010). In this paper, we use multilevel event history analysis to model the outcome of each call made by an interviewer to a sample unit as a function of household covariates and random effects representing unmeasured characteristics of households and interviewers. A multilevel multinomial logistic discrete-time hazard regression model is specified which jointly models the different types of outcomes at each call to predict the probability of interview or refusal, conditional on contact being made with the household. The effects of both time-varying and time-invariant covariates are considered.

The analysis benefits from an unusually rich dataset, the UK Census Link Study, which combines paradata from six UK face-to-face household surveys, including detailed call record data, interviewer observations about the household and information about the interviewer-household interaction. These data were linked to information about the household from the UK Census. A key advantage of the study is that all of this information is available for both responding and nonresponding households.

Recent developments in the area of paradata have raised issues on the usefulness of paradata for understanding and adjusting for nonresponse, and how best to model such data (Couper and Lyberg, 2005; Groves and Heeringa, 2006; Kreuter and Casas-Cordero, 2010). This paper illustrates the use of a particular type of paradata - interviewer call record and interviewer observation data - which are increasingly collected and used by survey organisations. The findings may have important implications for survey practice, for example what type of paradata to collect, and may inform effective interviewer calling behaviours.

The remainder of the paper is structured as follows. The data, including both the paradata and the linked census data, are described in Section 2. The multilevel multinomial model is outlined in section 3. Section 4 discusses the results from both descriptive analysis and multilevel modelling. The paper concludes with a summary of the main findings, limitations of the study and potential implications for survey practice.

2. Data

2.1 Paradata: Call Record Data And Interviewer Observation Data

This study benefits from the availability of relatively rich paradata from six UK face-to-face household surveys conducted in 2001 which have been linked to records from the UK 2001 Census. The study was designed to coincide with the last UK 2001 Census. The key advantage of this data source is that all of the variables are available for both responding and nonresponding units. The paradata consists of call record data and interviewer observation variables. The call data, the primary focus here, contains basic information recorded by the interviewer at each call, such as the day and time of the call and the outcome of the call. The main outcome of interest is whether cooperation was established with the household at a particular call, defined as at least one household member agreeing to respond to the survey. (We do not distinguish between full cooperation, where the whole household responds, and

partial cooperation where only some household members respond.) Other possible outcomes at each call are refusal, appointment made with the interviewer to come back at a more convenient time and other forms of 'postponement'. The latter category includes broken appointments or the interviewer withdrawing to try again later, for example if the interviewer is unable to make contact with a responsible resident or feels threatened. A call that leads into contact with at least one member of the household is referred to as a *contact call*.

An advantage of the call data is that information on the interaction between the interviewer and the householder have also been recorded, comprising for example information about the way the interviewer made contact, characteristics of the main person the interviewer talked to on the doorstep and whether the person made any positive or negative comments or asked any questions. Further call variables were derived for our analysis such as the number of non-contact calls (both prior to the first contact and in between two contact calls) and the number of previous contacts. Such variables are call dependent (time varying) and are measured at the call level. The interviewer observation data include information about each household, such as type of accommodation, indications of the presence of children and information about the immediate neighbourhood, such as the condition of the house relative to others in the area and how safe the interviewer would feel walking in the area after dark. These variables would not be expected to change across calls and are therefore time invariant. They are collected only once, if possible, at the first call. Both the call record data and the interviewer observations were collected via an interviewer observation questionnaire.

The paradata were linked to demographic and socio-economic household characteristics from the UK 2001 Census, such as type of household, presence of children and an employment status of adults in the household. Deterministic linkage methods were used to link the various data sources for each survey and in total about 95% of cases where linked to their census records. Linkage errors and possible consequences for analysis are assumed to be small as outlined in Durrant et al. (2011). The six surveys included in the study, carried out

around the time of the 2001 UK Census, are: the Expenditure and Food Survey (EFS), the Family Resources Survey (FRS), the General Household Survey (GHS), the Omnibus Survey (OMN), the National Travel Survey (NTS) and the Labour Force Survey (LFS). The six surveys vary in their design and subject matter. Further details about these surveys and the study as a whole can be found in Durrant and Steele (2009) and Durrant et al. (2011).

The analysis sample contains 38,816 contact calls. Calls that did not lead to contact with the household are not considered in this analysis (see Durrant et al. (2011) for an investigation of the contact process). However, the model controls for the number of previous unsuccessful contact attempts made by the interviewer to a household. Households that were never contacted, vacant and non-residential addresses, re-issues and unusable records were all excluded from the analysis (for further details see Durrant and Steele, 2009). The analysis sample includes a total of 15,782 households, nested within 565 interviewers. The median number of contact calls made by an interviewer (after first contact was established and excluding any intermediate non-contact calls) is 2 (average is 2.5). The maximum number of contact calls made to a household is 13, which increases to 15 when non-contact calls are included. The survey organisation provides some guidelines to interviewers on good calling practices. In terms of the frequency and timing of calls, this guidance mostly refers to the best ways of establishing contact. For example, the interviewer is advised to make a minimum of four calls of which at least two should be made in the evening or at the weekend. Some general guidelines are provided to interviewers on how to avoid or deal with a refusal at the doorstep. The interviewer is strongly advised to call back at least once after a refusal.

It should be noted that for the surveys of the Census Link Study, it was at the interviewer's discretion when a call was made to a household and therefore calling times are unlikely to be determined at random. Although an experimental design where calling times of interviewers are allocated at random to households may be to some extent feasable for telephone surveys (see for example West and Olson, 2010) it is impractical, if not impossible,

for face-to-face surveys. However, face-to-face surveys still represent an important mode of data collection and provide rich interviewer observation data about the household and the call. As discussed by Durrant et al. (2011), Purdon et al. (1999), Groves and Couper (1998) and Kulka and Weeks (1988) the data from face-to-face surveys represent *observed* calling times, i.e. the times that interviewers choose to call on a household. To the extent that the interviewer's decision to call at a particular time is independent of the characteristics of the household, a departure from non-randomised calling times should not be important. It may be reasonable to assume this to hold for the first call (or the first calls until first contact) since the interviewer would be expected to have little (or no) prior knowledge about the household. In subsequent calls, however, the interviewer may obtain further information about the household in various ways – for example from interviewer observations, talking to neighbours or having established an initial contact with the household - and this may influence the decision when and how best to call in the future. We attempt to control for this by including information on the call history, interviewer observation variables and household information (here primarily from the census), which extends previous work on the analysis of call record data that did not include such controls (e.g. Bates et al., 2008). In practice, the decision on when best to call may also depend on interviewer characteristics, such as experience of the interviewer. These concerns led us to consider models that included a range of interviewer characteristics, including measures of their attitudes and calling strategies. As the direction and magnitude of the coefficients of the call history variables were unaffected by their inclusion, we focus on a simpler model here which does not include interviewer characteristics.

3. Methodology

Multilevel multinomial logistic discrete-time event history analysis (Steele et al., 1996) was used to model the response outcome at call t, conditional on contact having been made with the

household at that call. A multilevel model approach is used to allow for the clustering of outcomes by household and interviewer due to unobserved household and interviewer characteristics. Multilevel multinomial models have been employed elsewhere to distinguish non-contact and refusal in analyses of nonresponse (e.g. O'Muircheartaigh and Campanelli, 1999; Durrant and Steele, 2009). However, these studies modelled the final response outcome rather than the process leading to it. We therefore extend this previous work by using a discrete-time hazard model to analyse call outcomes longitudinally, as proposed by Groves and Heeringa (2006). However, their models did not account for any clustering, neither of calls within households nor of households within interviewers. A multilevel model with household random effects allows for the possibility that the events of interest occur more than once to a household; for example during the course of the data collection process a household may make several appointments, an interviewer may withraw several times to come back at a later stage, or different household members may refuse to participate at different calls. Further levels, such as the nesting of households within interviewers to account for the role of the interviewer can be incorporated, and this is proposed here.

Denote by y_{tij} the outcome of call t $(t = 1,...,T_i)$ made to household i $(i = 1,...,n_j)$ by interviewer j $(j = 1,...,J)$ conditional on contact being achieved at t. The outcome of each call is coded as:

$$y_{tij} = \begin{cases} 1 & \text{refusal} \\ 2 & \text{appointment made} \\ 3 & \text{other form of postponement} \\ 4 & \text{full or partial cooperation} \end{cases}$$

A multilevel multinomial logit model for the log-odds of outcome s $(s = 1,2,3)$ relative to outcome 4 (cooperation) may be written

$$\log\left(\frac{\pi_{tij}^{(s)}}{\pi_{tij}^{(4)}}\right) = \boldsymbol{\beta}^{(s)\prime}\boldsymbol{x}_{tij}^{(s)} + \boldsymbol{\delta}^{(s)\prime}\boldsymbol{z}_{ij}^{(s)} + \lambda^{(s)}u_{ij} + \gamma^{(s)}v_j \qquad (1)$$

where $\boldsymbol{x}_{tij}^{(s)}$ is a vector of time-varying covariates, with coefficient vector $\boldsymbol{\beta}^{(s)}$, including indicators of the household's call history prior to t, the time and day of call t, information about the interaction between householder and interviewer and two-way interactions between call level variables. The call history indicators include the number of calls made to the household until first contact and the number of intermediate non-contacts after first contact (i.e. some function of t), which are derived from *all* calls regardless of whether contact was made. Another time-varying covariate is an indicator of whether an appointment was made with the household at the previous call, which allows estimation of transition rates, e.g. the probability that an appointment made at call $t-1$ is converted to cooperation at t. The vector $\boldsymbol{z}_{ij}^{(s)}$ includes time-invariant characteristics of the household, such as those from interviewer observations and the census, with coefficient vector $\boldsymbol{\delta}^{(s)}$. Unobserved household and interviewer characteristics are represented respectively by normally distributed random effects u_{ij} and v_j: $u_{ij} \sim N(\mathbf{0}, \sigma_u^2)$ and $v_j \sim N(\mathbf{0}, \sigma_v^2)$. Both the household and the interviewer random effects have outcome-specific coefficients or 'loadings', $\lambda^{(s)}$ and $\gamma^{(s)}$ respectively (with $\lambda^{(1)}$ and $\gamma^{(1)}$ fixed at 1 for identification). It is assumed that the odds of all non-participation outcomes are influenced by common sets of unmeasured household and interviewer characteristics, but their effects may differ across the three different survey outcomes. Outcome-specific loadings also allow the between-household variance in the log-odds of non-participation to differ across outcomes.

A full multinomial model with outcome-specific household and interviewer random effects, $u_{ij}^{(s)}$ and $v_j^{(s)}$, was initially considered (see, for example, Steele et al. 1996). However, the relatively small number of households with repeated outcomes of the same type caused estimation problems of the household- and the interviewer-level variances and covariances. It was therefore decided to employ a simplified model with common random effects but with outcome-specific loadings.

The analysis file contains a record for each call that resulted in contact being made with the household. Each household may therefore contribute multiple records, up to a maximum of T_i, with their sequence of calls terminating in refusal, cooperation or the interviewer giving up (right-censored histories). Estimation of (1) is carried out using maximum likelihood as implemented in the aML software package (Lillard and Panis, 2003).

To aid interpretation of the fitted model, predicted probabilities of each type of response outcome are calculated for each value of a given covariate, holding constant the values of all other covariates in the model at their sample means. We obtain population-averaged probabilities as follows: (i) a large number M of random draws are taken from the household and interviewer random effect distributions (based on the estimated random effect variances); (ii) M predicted probabilities are calculated for each response outcome, based on the generated household and interviewer random effect values and the estimated coefficients; and (iii) for each outcome s the mean of the predicted probabilities $\hat{\pi}_{tij}^{(s)}$ across the M random effect values is calculated.

4. Results

4.1 Results from descriptive statistics

We first present results from descriptive analysis and some preliminary modelling before discussing the final selected model. Table 1 shows the observed probability of each of the four possible outcomes at the first contact by time of day and day of the week. At first contact, it may be assumed that the interviewer has little, if any, information about the household that might influence his/her calling behaviour. In particular, there can be no appointments made before the first contact to influence the timing of the call. We can see that most first contacts are made on weekday afternoons, followed by weekday evenings and weekday mornings, with a clear decline in the number of contacts from the beginning to the end of the week for all

times of the day. (However, as reported in Durrant et al. (2011), the contact rate at the first call is highest for evening and weekend calls.) Overall, 26% of all households cooperate straightaway at the first contact, 8% refuse, 43% make an appointment and the remaining 24% result in another form of postponement. The chance of immediate cooperation at the first call is highest (above 30%) for calls made during mornings and afternoons at the beginning of the week (Monday and Tuesday) with a clear decline thereafter for all days of the week. The lowest cooperation rates are in the evenings, in particular towards the end of the week. On the other hand, the chances of making an appointment are highest for evening calls with above 45% for all days of the week but especially at the weekend; similar findings are reported in Purdon et al. (1999). The probability of other forms of postponement and refusal are relatively stable at around 25% and 8% respectively for most days and times of the week. It should be noted that only a few first contact calls are made at the weekend and Sunday calls are especially rare.

[Table 1 about here]

Since the first contact call is only indicative of the chances of achieving cooperation with a household we now examine changes in the probabilities of the different outcomes across calls. Figure 1 shows the observed probabilities of each outcome for the first seven contact calls (few calls were made after this point). The probability of making an appointment is over 40% at the first call, declines substantially to about 17% for the second call and then stabilises at around 10% for all subsequent calls. The cooperation rate is lowest at the first call (26%), increases sharply at the second call to about 60% and then stabilises at just above 70% at the fourth and subsequent calls. (The probability of cooperation in fact remains high even after 7 contact calls – results not shown). The rise in the cooperation rate for calls 2 to 4 may be explained by the large number of appointments that were made at the early calls, in particular at the first call. It may be speculated that prior appointments are usually turned into successful interviews at the next call. (This is further investigated in the next section and in

Table 3.) The probability of refusal is highest at calls 1 and 2 (at around 8%), then drops quickly towards zero. It seems that people that are inclined to refuse do so early on. Other forms of postponement are relatively high at the first call (25%), then fall to just over 10% and continue to rise again steadily from call 4 onwards. Taken together, these patterns suggest that for later calls (from about call 4 onwards) the household either cooperates or postponements occur (e.g. appointments are broken; the interviewer decides to postpone to another time), rather than the interviewer receiving a refusal.

[Figure 1 about here]

4.2 Results of final model

To investigate the joint effects of household characteristics, interviewer observations, characteristics of the current call and the call history on the outcome of the call various specifications of the multilevel multinomial discrete-time logistic models were explored. We now turn to the discussion of the final model of the process leading to cooperation or refusal across calls. Characteristics of households who prefer making an appointment are also described. Table 2 shows parameter estimates of the multilevel multinomial model with time-varying call characteristics, fixed interviewer observations and household characteristics, and household and interviewer random effects.

[Table 2 about here]

Time-varying call characteristics

The model controls for any previous calls, i.e. the number of previous contacts, the number of non-contact calls until first contact and the number of intermediate non-contact calls. The inclusion of the previous contact indicator means that the coefficients of number of contact calls are interpreted as the effect on the different forms of non-participation of each

additional call after the first call. We find that the probabilities of refusal, appointment and other forms of postponement are highest for the first call and decrease with each subsequent call, after controlling for the other explanatory variables in the model. This is consistent with the finding from the descriptive analysis that sample members who are inclined to refuse, do so earlier on. The odds of cooperation increase with each additional contact made. This supports the findings of Sangster and Meekins (2004) and Groves in Heeringa (2006) who report a significant positive effect of a prior contact with the household on the likelihood of a main interview. This effect may indicate that an ongoing interaction between the interviewer and the householder may be more likely to lead to a positive outcome, which would support the interaction hypothesis of Groves and Couper (1996 and 1998). It could also indicate that interviewers are persistent in returning to a household if they feel they have a chance of a positive outcome. There is a (small) negative effect of the number of calls made until first contact on the probabilities of refusal, appointment and other postponements, which may imply that households that are more difficult to reach may be more likely to cooperate once contacted, possibly justifying increased costs for survey agencies to follow up difficult to reach households. The effects on non-participation of the number of non-contact calls after first contact are in the opposite direction: the more non-contact calls are made after first contact the more likely it is that the household refuses, makes an appointment or that the interviewer withdraws. This may indicate that a non-contact could in fact be a hidden evasion or refusal (Stoop, 2005).

The 21 possible day and time of the week combinations were reduced to six categories, distinguishing early week (Sun-Tue) and late week (Wed-Sat) and morning, afternoon and evening (see Table 2). This coding is based on the descriptive analysis reported in Table 1 and initial modelling which began with all 21 categories (results not shown). These analyses revealed quite different patterns for Saturday and Sunday with Sunday more like the early part of the week (Mon-Tue) and Saturday more like late week (Wed-Fri), especially Friday. In

addition, the few calls made on Saturdays and Sundays made it necessary to merge these categories with other days of the week.

The effects of day and time of the week may be expected to depend on whether an appointment was made at the previous call, and indeed an interaction term between these two variables was found to be highly significant. For ease of interpretation, predicted probabilities for this interaction are presented in Table 3. When there was no prior appointment, all evening calls have very low probabilities of resulting in an immediate cooperation (below 13%), while the refusal rate is relatively high (12-14%) and the postponement rate is around 30%. However, the probability that a householder books an appointment is considerably higher if the call is made in the evening rather than at any other time of the day. Without an appointment, the immediate cooperation rates for morning and afternoon calls are at 20-30% almost three times higher than for evening calls; conversely, the refusal rate for morning and afternoon calls is at around 2% only a fraction of the evening refusal rate. If the previous call results in an appointment the chances of experiencing cooperation at the next call is very high (around 70%), and this is the case for any time of day including evening calls. That means that appointments are likely to lead to cooperation irrespective of the time of the appointment. The findings may reflect and justify common interviewing practices. They may indicate that without a prior appointment daytime calls are more likely to lead to (immediate) cooperation than evening calls, but evening calls may of course be necessary if no prior contact has been established at other times and may be used to make an appointment – strategies often adhered to by interviewers. This illustrates that good times to achieve cooperation are not necessarily the same as good times to establish contact, which are generally recognised to be evenings and weekends (Weeks et al., 1980; Weeks et al., 1987, Swires-Hennessy and Drake, 1992; Purdon et al. 1999; Durrant et al., 2011); however, good times to establish contact can be used to make

appointments. Similar findings were reported by Purdon et al. (1999) based on descriptive analysis of the first contact call of the UK Family and Resources Survey.

[Table 3 about here]

Of particular interest is the effect on the call outcome of what happens at the doorstep, especially the initial interaction between the householder and the interviewer, also a focus in previous research (e.g. Groves and Couper, 1996; Maynard and Schaeffer, 1997; Campanelli et al., 1997; Sturgis and Campanelli, 1998; Bates et al., 2008). As argued in Bates et al. (2008) such information can greatly improve models predicting nonresponse relative to models that only include basic call history measures, such as the number of contact attempts. Here, the mode of contact appears relevant for the likelihood of gaining immediate cooperation: the chances of a refusal, making an appointment or the interviewer withdrawing to try again later are significantly lower if the contact is face-to-face rather than through an intercom system, a window or a closed door. This effect remains after controlling for potential area effects, such as urban/rural indicator, London indicator and the condition of the house in comparison to others in the area. Non face-to-face contact could indicate a potential fear of crime or a reluctance to talk to strangers which has been shown in other studies to lead to a higher refusal rate (Groves and Couper, 1998). If the householder asks at least one question, the chances of refusal, appointment or postponement are significantly reduced. Likewise, if the householder makes at least one positive or neutral comment as opposed to no comment, the odds of refusal or the interviewer withdrawing are much reduced while the odds of making an appointment increase. As would be expected, people who engage in a positive or neutral way with the interviewer (asking a question or making a comment), potentially expressing some interest in the survey, tend to cooperate more than those who do not. On the other hand, if the householder makes at least one negative comment, refusal, appointment and

postponement are much more likely than if no comment was made – supporting findings by Groves and Couper (1996), Groves and Heeringa (2006) and Bates et al. (2008).

Characteristics of the person the interviewer talked to at the doorstsep (based on interviewer observations) also seem to be useful in predicting the outcome of the call. For example, the older the person at the doorstep the less likely he/she is to refuse, make an appointment or postpone. Particularly high rates of refusal and postponements can be seen for children younger than 16 years. A potentially higher (final) cooperation rate for older householders has been noted in other studies (Durrant and Steele, 2009), although some contradictory effects of the age of the householder have been found (Groves and Couper 1996; Groves and Couper, 1998). If the person at the doorstep is female the call is more likely to result in an appointment or a postponement, which may reflect a greater reluctance to speak to strangers or fear of crime among women. Differences in lifestyles such as looking after children when at home may also contribute to this effect. However there is no gender difference in the immediate refusal behaviour.

Time invariant interviewer observations and household characteristics

We now turn to the effects of time invariant interviewer observations and household characteristics. Interviewer observations on the household and neighbourhood are found to be useful in predicting the outcome of a call. Direct observations about the household as well as interviewer judgements were explored. Compared to householders living in flats, those living in houses have higher chances of immediate refusal, an appointment or the interviewer to withdraw. The interviewer was also asked to judge the condition of the house and area. Living in a house that the interviewer reports to be in a worse condition than others in the area is associated with higher rates of refusal, appointment and postponement, as might be expected since socially deprived households have been found to be less likely to cooperate in other

studies (Goyder, 1987; Groves and Couper, 1998). Leaving a card or message behind was not found to affect the probabilities of any type of nonresponse.

Some of the variables considered in the present study are available from both the census and the interviewer observation questionnaire, for example information on the presence of children and the household type. Census variables, where available, were included in the final model because these data are likely to be of higher quality than interviewer reports. Although it may be regarded as unusual to have access to Census records, it is (at least in principle) possible to obtain information about the households prior and during data collection. Some information may come from the sampling frame -although such information may be limited- or from register or adminstrative data, such as in Scandinavian countries, The Netherlands (Cobben and Schouten, 2007) or Germany (Trappman and Mueller, 2010; Kreuter et al. 2011). Other studies without access to census or administrative variables may be able to include similar information based on interviewer observations.

For households with pre-school children the immediate refusal and the postponement rate are lower. Such households are, however, more likely to request an appointment for a different time. This may be expected since, for example, households with children can be contacted relatively easily, but it may not be convenient to participate in a survey in the presence of children; in which case an appointment for another time may be made. Refusals, appointments and other postponements are more likely outcomes than cooperation in London and urban areas, and for couple households and households with at least one member in employment. Households where the household representative has a high educational attainment are less likely to refuse, to make an appointment or to postpone, leading to a higher cooperation rate (see also Goyder, 1987; Groves and Couper, 1998). Although the analysis above clearly indicates that certain households prefer making an appointment, the probability of appointment depends on the time of day when the household was contacted, as discussed in the previous section. This finding is in contrast to the conclusions drawn in Purdon et al.

(1999, p. 214), based on initial descriptive analysis of first contact calls, that whether or not an appointment is made seems to be more closely related to the type of respondent than to the time of day.

The model also allows for differences in cooperation and refusal across the six surveys. We find the highest refusal, appointment and postponement rates for the EFS, a survey with a relatively high response burden due to the requirement to keep a diary and a long questionnaire. The lowest rates are achieved for the LFS, a less burdensome survey with a comparatively short interview. Further details on the differences between the surveys and an analysis of survey-dependent effects on ultimate contact and refusal rates can be found in Durrant and Steele (2009).

Random household and interviewer effects

Table 4 presents the estimated household and interviewer random effect parameters from the final multilevel multinomial discrete-time logistic regression model. The results show significant residual variation in the log-odds of a nonresponse outcome between households and between interviewers. The fitted model is a simplification of model (1) with loadings on the interviewer random effect $\gamma^{(s)}$ (s=1,2,3) constrained to be equal across all three outcomes (the likelihood ratio test statistic for a test of the null hypothesis H_0: $\gamma^{(1)} = \gamma^{(2)} = \gamma^{(3)} = 1$ is 2.80 on 2 d.f., p=0.246. We therefore conclude that unmeasured interviewer characteristics, represented by v_j, have the same effect on the log-odds of each of the three forms of non-participation. At the household level, however, there is evidence of differential effects of unmeasured household characteristics u_{ij} across the three outcomes (based on t-tests that the loadings for postponement and appointment are equal to 1: $t = 3.1$, p=0.002 for H_0: $\lambda^{(2)} = 1$ and $t = 5.1$, p=0.000 for H_0: $\lambda^{(3)} = 1$). While there is significant between-household variation

in the log-odds of all forms of non-participation, household effects are strongest for postponement and weakest for appointments.

[Table 4 about here]

5. Summary and Implications for Survey Practice

This paper analyses call record data in interviewer administered face-to-face surveys to inform the probability of achieving cooperation at each call to a household. The aim is to better understand the process leading to cooperation or refusal rather than focussing on predicting final response. In the following, we summarise the main results and indicate potential implications for survey practice:

1. We have found some indication that households that are inclined to refuse do so early on. For later calls (from about contact call 4 onwards) the household either cooperates or the interviewer decides to postpone to another time or to stop calling, rather than that a refusal is reported.
2. Time-varying call record information, such as features of the call history and of the current call, play a key role in predicting the outcome of each call. Characteristics of the interaction process between the interviewer and the householder were of particular relevance, including how contact was established, characteristics of the person who came to the door and whether this person asked questions or made comments.
3. Best times to establish contact with a household may differ from the best times to establish cooperation. For example, calls made in the evening and at weekends are most likely to result in contact. However, without a prior appointment, households contacted at those times are more likely than at other times to refuse, book an appointment or the interviewer feels the need to withdraw. The results indicate that good times to establish

cooperation strongly depend on if an appointment was previously made. Overall, a call made at a previously agreed time is likely to lead to a successful interview.

4. The model identified types of households which prefer making an appointment. For example, householders who are female, younger than 60 (in particular if younger than 16), live in a house or have pre-school children are more likely to make an appointment. The circumstances of the call also influence the probability of appointment: for example, if the call is made in the evening the probability of appointment is significantly higher than for a call during day time.
5. The more contact calls are made the higher the odds of cooperation. This may provide some evidence that keeping in contact with the household may increase the chances of a successful interview. The finding could suppport the hypothesis expressed in Groves and Couper (1996 and 1998) that maintaining the interaction with the household is more likely to lead to cooperation. Rather than pressing for an immediate cooperation the interviewer may be advised to keep the conversation and the contact with the household going, for example by making an appointment for another time.
6. Interviewer observations, for example on the type and condition of the house and the presence of dependent children, proved to be useful for predicting the likelihood of cooperation.

It should be noted that we cannot necessarily infer about possible causal effects since the data available are not from a randomised experiment with calls allocated at random to households. The model, however, controls for relevant information about the household, area and call history that may have been used by the interviewer to decide when best to call. A note of caution needs to be made since the data do not contain many weekend calls and analysis of such calls is therefore limited.

Paradata are often of a complex multilevel structure which needs to be reflected in a statistical model. For example, in the present study we have information at the call, household and interviewer level. This paper contributes to methodological development in the use of call record data and the specification of models based on such data. The model presented reflects the hierarchical structure of the data, allowing for household and interviewer random effects. The models may inform efficient and effective calling behaviours and may be used in responsive or adaptive survey designs to predict the likelihood of cooperation at the next call (for an early example see Groves and Heeringa, 2006). A particular application may be for longitudinal surveys where call record data and a wide range of information on the sample member is available from previous waves.

The analysis highlights important advantages of collecting call history data, information about the household and the outcome of the survey request at each call via interviewer observations. As argued in Bates et al. (2008) such variables greatly improve models for predicting nonresponse and offer a unique opportunity to provide information on *both* respondents and nonrespondents. Survey agencies may consider routinely collecting and analysing such data to inform the processes leading to cooperation. There is a great need to consider which types of paradata are useful and how they should be collected. Paradata can be subject to measurement error and missing items and careful consideration needs to be given on how to improve the quality of such data. Further work is needed to investigate how best to use such data in practice.

6. Acknowledgement

The research is funded by the Economic and Social Research Council (ESRC), UK: 'Hierarchical analysis of unit nonresponse in sample surveys', grant number: RES-062-23-0458. and 'The Use of Paradata in Cross-Sectional and Longitudinal Research', grant number:

RES-062-23-2997. This work contains statistical data from ONS which is Crown copyright and reproduced with the permission of the controller of HMSO and Queen's Printer for Scotland. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

7. References

Bates, N., Dahlhamer, J. and Singer, E. (2008) Privacy concerns, too busy, or just not interested: Using doorstep concerns to predict survey nonresponse, *J. Off. Statist.,* 24, 4, 591-612.

Bethlehem, J., Cobben, F. and Schouten, B. (2011) *Handbook of nonresponse in household surveys*. John Wiley and Sons, Hoboken, New Jersey.

Biemer, P., Wang, K. and Chen, P. (2010) Using level of effort paradata in nonresponse adjustments with application to field surveys. Invited paper presented at the One-Day International Research Symposium on '*Recent Advances in the Use of Paradata in Social Survey Research*', Royal Statistical Society, London. (available from: http://www.southampton.ac.uk/s3ri/events/workshops/Files/Callback%20Weighting%20Paper%20for%20Paradata%20Workshop%20-%20Final%20Draft.pdf)

Blom, A., Jäckle, A. and Peter, L. (2010) The use of contact data in understanding cross-national differences in unit non-response. Harkness, J. et al. (eds.) *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. Wiley & Sons, New York.

Campanelli, P., Sturgis, P. and Purdon, S. (1997) *Can you hear me knocking*: *An investigation into the impact of interviewers on survey response rates*, Report of the Survey Methods Centre at SCPR, London.

Cobben, F. and Schouten, B. (2007) A follow-up with basic questions of nonrespondents to the Dutch Labour Force Survey. *Discussion paper 07011, Statistics Netherlands*, available from:http://www.cbs.nl/NR/rdonlyres/42B0DC4A-D071-46F8-A249-408D7786E2F3/0/200711x10pub.pdf

Couper, M. and Lyberg, L. (2005) The use of paradata in survey research. *Proceedings of the 54th Session of the International Statistical Institute*, Sydney, Australia.

Couper, M.P. (1998) Measuring survey quality in a CASIC environment. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 48, 743-772.

Durrant, G.B. and Steele, F (2009) Multilevel modelling of refusal and noncontact nonresponse in household surveys: Evidence from six UK government surveys. *J. R. Statist. Soc. A,* 172, 2, 361-381.

Durrant, G.B., D'Arrigo, J. and Steele, F. (2011) Using field process data to predict best times of contact conditioning on household and interviewer influences, *J. R. Statist. Soc. A*, forthcoming.

Goyder, J. (1987) *The silent minority: Nonrespondents on sample surveys*. Boulder: Westview.

Greenberg, B.S. and Stokes, S.L. (1990) Developing an optimal call scheduling strategy for a telephone survey. *J. Off. Statist.,* 6, 4, 421-435.

Groves, R. M., Cialdini, R. B. and Couper, M. (1992) Understanding the decision to participate in a survey. *Publ.Opin. Q.*, 56, 475–495.

Groves, R.M. and Couper, M.P. (1996). Contact-level influences on cooperation in face-to-face Surveys. *J. Off. Statist.*, 12, 1, 63-83.

Groves, R.M. and Couper, M.P. (1998). *Nonresponse in household interview surveys*, New York: Wiley.

Groves, R.M. and Heeringa, S.G. (2006) Responsive design for household surveys: tools for actively controlling survey errors and costs, *J. R. Statist. Soc. A,* 169, 3, 439-459.

Kreuter, F. and Casas-Cordero, C. (2010) Paradata. *Working paper*, *Working paper series of the German Council for Social and Economic Data (RatSWD)*, 136, available from: www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_136.pdf

Kreuter, F., Mueller, G. and Trappman, M. (2011) Nonresponse and measurement error in employment research: Making use of administrative data, *Public Opinion Quarterly*, 74, 5, 880-906.

Kreuter, F., Olson, K., Wagner, J., Yan, T., Ezzati-Rice, T.M., Casas-Cordero, C., Lemay, M, Peytchev, A., Groves, R.M. and Raghunathan, T. E. (2010): Using proxy measures and

other correlates of survey outcomes to adjust for non-response: examples from multiple surveys. *J.R. Statist. Soc. A*, 173, 2, 389-407.

Kreuter, F., and Kohler, U. (2009) Analyzing contact sequences in call record data: Potential and limitations of sequence indicators for nonresponse adjustment in the European Social Survey. *J. Off. Statist.*, 25, 2, 203-226.

Kulka, R.A. and Weeks, M.F. (1988) Towards the development of optimal calling protocolls for telephone surveys: A conditional probabilities approach. *J. Off. Statist.*, 4, 4, 319-358.

Laflamme, F., Maydan, M. and Miller, A. (2008). Using paradata to actively manage data collection survey process, *Proceedings of the Survey Research Methods Section, American Statistical Association*, 630-637.

Lillard, L.A. and Panis, W.A. (2003) aML multilevel multiprocess statistical software, version 2.0. EconWare, Los Angeles, California.

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

O'Muircheartaigh, C. and Campanelli, P. (1999) A multilevel exploration of the role of interviewers in survey nonresponse. *J.R. Statist. Soc. A*, 162, 3, 437-46.

Peytchev, A. and Olson, K. (2007) Using interviewer observations to improve nonresponse adjustments: NES 2004, *Proceedings of the Survey Research Methods Section, American Statistical Association*, 3364-3371.

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

Purdon, S., Campanelli, P. and Sturgis, P. (1999). Interviewer's Calling Strategies on Face-to-Face Interview Surveys. *J. Off. Statist.,* 15, 2, 199-216.

Sangster, R.L. and Meekins, B.J. (2004) Modeling the likelihood of interviews and refusals: Using call history data to improve efficiency of effort in a national RDD survey. *Proceedings of the Survey Research Methods Section, American Statistical Association*, available from: http://www.bls.gov/osmr/abstract/st/st040090.htm

Steeh, C., Kirgis, N., Cannon, B. and DeWitt, J. (2001) Are they really as bad as they seem?: Nonresponse rates at the end of the twentieth century. *J. Off. Statist.*, 17, 227–247.

Steele F., Diamond I. and Wang D.L. (1996) The determinants of the duration of contraceptive use in China: A multilevel multinomial discrete-hazards modeling approach. *Demography*, 33, 12-23.

Stoop, I.A.L. (2005) *The hunt for the last respondent, nonresponse in sample surveys*. Social and Cultural Planning Office of the Netherlands, The Hague.

Sturgis, P. and Campanelli, P. (1998) The scope for reducing refusals in household surveys: An investigation based on transcripts of tape-recorded doorstep interactions. *Journal of the Market Research Society*, 40, 2, 121-139.

Swires-Hennessy, E. and Drake, M. (1992) The optimum time at which to conduct interviews, *International Journal of Market Research (Journal of the Market Research Society)* 34, 1, 61-72.

Trappman, M. and Mueller, G. (2010) Paradata and linked administrative data of the PASS panel. Poster presented at the One-Day International Research Symposium on '*Recent Advances in the Use of Paradata in Social Survey Research*', Royal Statistical Society, London. (available from: http://www.southampton.ac.uk/s3ri/events/workshops/paradata_posters.html)

Weeks, M.F., Jones, B.L., Folsom, R.E. and Benrud, C.H. (1980). Optimal times to contact sample households. *Public Opinion Quarterly*, 44, 1, 101-114.

Weeks, M.F., Kulka, R.A. and Pierson, S.A. (1987). Optimal call scheduling for a telephone survey. *Public Opinion Quarterly,* 51, 4, 540-549.

West, B.T. and Olson, K. (2010) How much of interviewer variance is really nonresponse error variance?. Special issue on total survey error, *Public Opinion Quarterly*, 74, 5, 1004-1026.

Wood, A.M. and White, I.R. (2006). Using number of failed contact attempts to adjust for non-ignorable non-response. *J.R. Statist. Soc. A,* 169, 3, 525-542.

Table 1: Observed probability of each outcome at first contact, by day and time of call.

		Cooperation	Refusal	Appointment made	Other postponement	Total number of first contact made	% of all first contacts
Monday	am	0.37	0.09	0.34	0.21	381	2.41
	pm	0.37	0.07	0.32	0.24	2162	13.70
	eve	0.25	0.08	0.48	0.20	1648	10.44
Tuesday	am	0.31	0.09	0.34	0.26	279	1.77
	pm	0.31	0.06	0.37	0.26	1919	12.16
	eve	0.23	0.08	0.49	0.21	1649	10.45
Wednesday	am	0.29	0.12	0.40	0.20	214	1.36
	pm	0.26	0.07	0.43	0.24	1544	9.78
	eve	0.20	0.08	0.48	0.24	1472	9.33
Thursday	am	0.28	0.09	0.39	0.25	212	1.34
	pm	0.22	0.08	0.42	0.28	1253	7.94
	eve	0.19	0.08	0.46	0.27	1138	7.21
Friday	am	0.23	0.12	0.39	0.27	151	<1.0
	pm	0.20	0.07	0.46	0.27	735	4.66
	eve	0.18	0.10	0.51	0.22	580	3.68
Saturday	am	0.26	0.05	0.43	0.27	109	<1.0
	pm	0.14	0.08	0.54	0.24	239	1.51
	eve	0.12	0.04	0.52	0.33	52	<1.0
Sunday	am	0.20	0.20	0.30	0.30	10†	<1.0
	pm	0.11	0.05	0.68	0.16	19†	<1.0
	eve	0.06	0.00	0.69	0.25	16†	<1.0
Total		0.26	0.08	0.43	0.24	15782	100

Morning (am): 0.00-12.00, Afternoon (pm): 12.00-17.00, Evening (eve): 17.00-0.00
† indicates cells with a sample size of less than 30

Figure 1: Observed probabilities of each outcome by contact call.

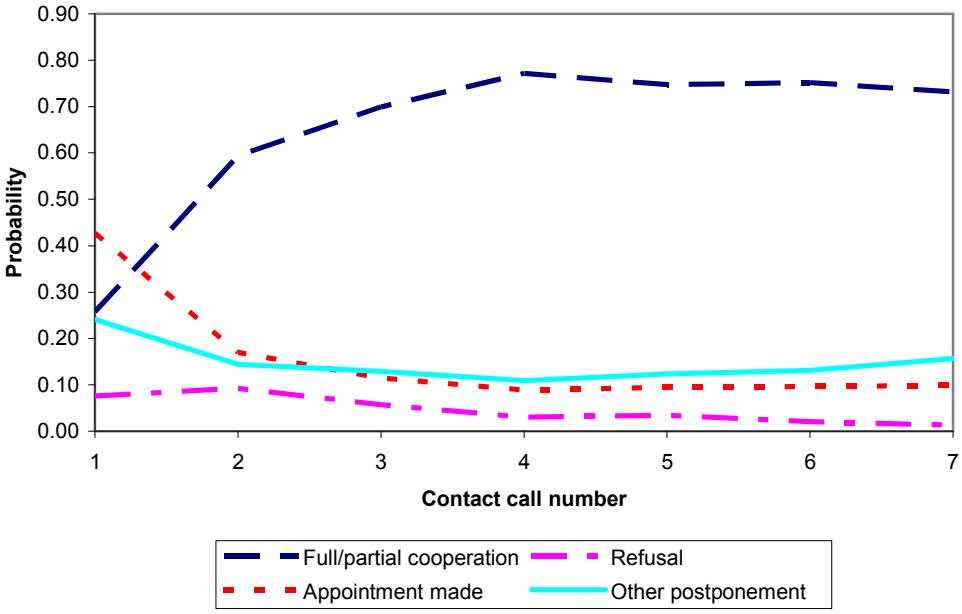

Table 2: Estimated coefficients (and standard errors in parentheses) of multilevel multinomial logistic model controlling for household characteristics.

Variable (ref = Reference category)	**Categories**	$\hat{\beta}$ $(ste(\hat{\beta}))$ ***Refusal***	$\hat{\beta}$ $(ste(\hat{\beta}))$ ***appointment made***	$\hat{\beta}$ $(ste(\hat{\beta}))$ ***other postponement***
Constant		-2.687 (0.194)***	0.151 (0.126)	-0.636 (0.160)***
Call record variables (time variant)				
Previous contact indicator (ref =First contact)	Contact previously made	-0.251 (0.108)***	-1.606 (0.076)***	-1.849 (0.089)***
Number of contact calls previously made	-	-1.403 (0.051)***	-1.191 (0.036)***	-1.177 (0.038)***
Number of non-contact calls made until first contact	-	-0.051 (0.021)**	-0.162 (0.015)***	-0.261 (0.020)***
Number of intermediate non-contact calls after first contact was made	-	0.532 (0.034)***	0.449 (0.026)***	0.387 (0.032)***
Day and time of contact † (ref =Sun-Mon-Tue eve)	Sun-Mon-Tue am Sun-Mon-Tue pm Wed-Thurs-Fri-Sat am Wed-Thurs-Fri-Sat pm Wed-Thurs-Fri-Sat eve	-0.243 (0.180) -0.634 (0.105)*** -0.570 (0.154)*** -0.477 (0.102)*** -0.231 (0.097)**	-0.425 (0.120)*** -0.737 (0.071)*** -0.763 (0.107)*** -0.511 (0.066)*** -0.122 (0.066)*	-0.246 (0.150) -0.403 (0.087)*** -0.696 (0.131)*** -0.244 (0.084)*** 0.018 (0.080)
Previous Appointment Indicator † (ref =No prior appointment made)	Prior appointment made	-2.770 (0.201)***	-2.100 (0.116)***	-2.100 (0.143)***
How contact was made at doorstep (ref =Face-to-face)	Not face-to-face	2.114 (0.110)***	2.585 (0.077)***	2.319 (0.090)***
Question made by householder during the interviewer introductory conversation (ref =No question made)	At least one question made	-1.483 (0.075)***	-0.430 (0.049)***	-1.278 (0.064)***
Comment made by householder during the interviewer introductory conversation (ref =No comment made)	Positive/neutral comment made At least one negative comment made	-0.668 (0.139)*** 5.704 (0.119)***	0.547 (0.051)*** 2.128 (0.082)***	-0.784 (0.065)*** 3.266 (0.091)***
Age of main person the interviewer talked to (ref =60 and over)	Less than 16 16-34 35-59	3.109 (0.490)*** 0.794 (0.120)*** 0.627 (0.099)***	2.753 (0.305)*** 1.080 (0.082)*** 0.764 (0.071)***	6.144 (0.282)*** 1.660 (0.103)*** 0.870 (0.090)***
Gender of main person the interviewer talked to (ref =Male)	Female	-0.023 (0.066)	0.244 (0.045)***	0.138 (0.056)**
Interviewer Observations (time invariant)				
Type of accommodation (ref =Not house)	House	0.691 (0.109)***	0.800 (0.078)***	0.810 (0.100)***
House in a better or worse condition than others in area (ref =Better/ About the same)	Worse	0.444 (0.131)***	0.336 (0.095)***	0.368 (0.122)***
Household-level variables (time invariant)				
Preschool children present (ref =No)	Preschool children	-0.256 (0.117)**	0.170 (0.076)**	-0.055 (0.099)
Household type (ref =Single household)	Couple household Multiple household	0.566 (0.081)*** 0.284 (0.234)	0.432 (0.057)*** 0.104 (0.164)***	0.540 (0.075)*** 0.263 (0.209)
London indicator (ref =Not London)	London	0.618 (0.108)***	0.513 (0.079)***	0.908 (0.100)***
Urban/rural indicator (ref =Urban)	Rural	-0.294 (0.118)**	-0.232 (0.081)***	-0.363 (0.108)***

Indicator if adults in employment (ref =No adults)	One or more adults	0.233 (0.095)**	0.302 (0.067)***	0.578 (0.087)***
Educational attainment of Household Reference Person (ref =No educational attainment/ A levels, GCSEs)	First/Higher/College degree/Other attainment	-0.433 (0.088)***	-0.188 (0.060)***	-0.323 (0.078)***
Survey indicator (ref =EFS)	FRS	-0.019 (0.123)	-0.079 (0.088)	-0.151 (0.114)
	GHS	-0.371 (0.108)***	-0.143 (0.077)*	-0.204 (0.100)**
	OMN	-0.401 (0.115)***	-0.878 (0.084)***	-0.611 (0.106)***
	NTS	-0.887 (0.108)***	-0.418 (0.076)***	-0.491 (0.099)***
	LFS	-4.136 (0.169)***	-3.752 (0.118)***	-4.735 (0.153)***
Interaction between interviewer observation and previous outcome				
Day and time of call * Previous Appointment Indicator (ref = Sun-Mon-Tue eve and No prior appointment made)	Sun-Mon-Tue am * Appointment	0.585 (0.373)	0.170 (0.243)	0.254 (0.288)
	Sun-Mon-Tue pm * Appointment	0.481 (0.287)*	0.701 (0.164)***	0.592 (0.198)***
	Wed-Thurs-Fri-Sat am * Appointment	0.562 (0.326)*	0.571 (0.199)***	0.515 (0.244)**
	Wed-Thurs-Fri-Sat pm * Appointment	0.606 (0.265)**	0.278 (0.154)*	0.349 (0.185)*
	Wed-Thurs-Fri-Sat eve * Appointment	0.101 (0.257)	0.006 (0.146)	0.067 (0.178)

The model is estimated using full information maximum likelihood. Where a closed form solution to the maximum likelihood function does not exist the residuals at each level are 'integrated out' numerically using Gauss-Hermite quadrature. The number of quadrature points used is 16. Approximate standard errors are computed based on an approximation to the Hessian matrix. The missing value categories have been suppressed to save space.

* significant at the 10% level
** significant at the 5% level
*** significant at the 1% level
† variable included in an interaction

Table 3: Predicted probabilities of each outcome (in %) for two-way interaction.(row-percentages shown).

Interaction between day and time of call and previous appointment made						
		Prior appointment made	**Outcome at current call**			
			Cooperation	Refusal	Appointment made	Other postponement
Day and time of call	Sun-Tue am	Yes	71.8	2.4	14.2	11.6
		No	21.0	11.9	39.1	28.1
	Sun-Tue pm	Yes	69.3	1.6	16.3	12.8
		No	30.2	9.3	34.3	26.3
	Sun-Tue eve	Yes	70.0	1.8	16.7	11.5
		No	9.3	13.8	46.1	30.9
	Wed- Sat am	Yes	73.0	1.8	14.8	10.4
		No	33.3	9.6	33.9	23.2
	Wed- Sat pm	Yes	71.3	2.1	14.4	12.2
		No	23.9	10.3	37.8	28.1
	Wed- Sat eve	Yes	70.7	1.6	15.5	12.1
		No	12.9	12.0	44.1	31.1

Table 4: Estimated household and interviewer random effect parameters from the multilevel multinomial logistic regression model (standard errors in parentheses).

Parameter	**Estimate (Standard Error)**
Household common standard deviation σ_u	1.900 (0.096)***
Household random effect loadings $\lambda^{(s)}$	
$\lambda^{(1)}$ Refusal	1[a]
$\lambda^{(2)}$ Appointment made	0.873 (0.041)***
$\lambda^{(3)}$ Other postponement	1.280 (0.055)***
Interviewer common standard deviation σ_v	0.755 (0.043)***

a Constrained to equal 1
*** Significantly different from zero at the 1% level