Applications of Paradata in Survey Design and Analysis

Carlos Lagorio

A thesis submitted for the degree of PhD in Survey Methodology

Institute for Social and Economic Research
University of Essex
Colchester, United Kingdom

Date of revised submission March 31, 2019

Contents

Non-technical Summary

This dissertation discusses three separate but related applications of paradata for survey design and analysis.

Chapter 1 expands on the use of call record data for nonresponse investigation in face-to-face surveys. By focusing on the relatively underexplored analysis of longitudinal call record data in household surveys, it considers association between Wave 1 call record sequences and response outcomes in proceeding Waves (2, 3 and 4) of the UK Household Longitudinal Study (or *Understanding Society*). This chapter addresses the predictive power of predefined call record sequences observed in the baseline wave of this survey by comparing model estimates that employ not just this type of paradata but also more conventional predictors of nonresponse (like sociodemographic characteristics of issued households as well as auxiliary geographic information). Beyond finding associations and comparing model specifications, this analysis is primarily interested in informing response retention strategies for panel surveys based on the calling patterns of earlier waves.

Chapter 2 is similarly concerned with field effort optimization. However, while Chapter 1 uses call record data as predictors of a given outcome of interest (namely future wave contact and cooperation), Chapter 2 proposes models to predict the calling effort inherent in the processes of contact and cooperation conditional on household and aggregate individual-level data (as well as lagged contact record data). Given the onerous fieldwork demands of household longitudinal surveys, the analysis of this chapter aims to inform data collection optimization by identifying predictors (especially those derived from paradata) of differential contactability, cooperation and overall field effort requirements in longitudinal context.

The third and final chapter analyses different types of CAWI generated paradata to assess progress indicator (PI) effects on survey response quality. The data used for this analysis comes from an experiment designed by the author. This chapter seeks to further examine and develop standing theories of progress indicator effects on surveys by focusing on their impact on response quality while also expanding on the uses of web survey paradata and their applications for response quality assessment and respondent behaviour (including satisficing, as well as time and effort management).

Acknowledgements

This dissertation wouldn't have been possible without the support and inspiration of family, friends and colleagues.

Firstly, I would like to thank my mother for teaching me how to read and write, and about the importance of asking questions. To my brothers, Juan Carlos and Esteban, and sister, Cristina, thank you for your continued love, support and encouragement. I would like to dedicate this dissertation to my father who knew well before I did that I would marry my wife, move to the UK and enrol in a PhD programme. Thank you for teaching me about discipline, perseverance and the value of a job well done. I love you and miss you.

To my colleagues at ISER: thank you for the shared experiences, the opportunity to learn from your journeys towards your PhDs and all your valuable input into my own. I would like to specially thank my fellow survey methodology colleagues including: Alex Cernat, Alex Wenz, Jennifer Kelley, Vilma Agalioti, Valerija Kolbas and Yamil Nares. I would also like to thank my office mates: Duygu Ozdemir, Maria Elena Komodromou and Yvonni Markaki for their camaraderie and support.

To the ISER staff: Firstly, I would like to thank Professor Peter Lynn. Since 2012, I've been fortunate to be his student, first as a MSc student and then as a PhD candidate. I am very grateful for his dedication, his patience and his willingness to share his vast experience and knowledge of survey methodology. I am equally grateful to Professor Paul Clarke for his mentorship, guidance, and ability to communicate principles of social statistics and relate them to my research. I would also like to thank the rest of the survey methods team at ISER, including: Tarek Al Baghal, Annette Jackle, Jorre Vannieuwenhuyze, Peter Lugtig, Olena Kaminska, Jon Burton, Laura Fumagalli and Alessandra Gaia. I am also thankful to Emilia del Bono, Adeline Delavande and Angus

Holford for allowing me to collaborate in the BOOST2018 project and letting me design an experiment for the third chapter of this dissertation. Gundi Knies and Alita Nandi: thank you for helping me navigate and make sense of Understanding Society's datasets and technical documentation. Cara Booker, thank you for your counselling and support to me and the rest of the students. Janice Webb, Emily Smith, Kim Bones, Jane Ismail, Tom Butler, Matt Siddall and Paul Siddall; thank you for all your help and for making ISER feel like home.

For the past year and a half I have been employed as Associate Director of the Methods team at Kantar Public UK. I would like to thank Joel Williams for inviting me to join the company and letting me work alongside a team of very smart, dedicated, challenging and inspiring colleagues including George Kyriakopoulos, Jamie Burnett, Luke Taylor, Maria Holcekova and Peter Matthews.

To my family. Olivia, you are my proudest achievement. You were born during my time as a PhD student and helped me find meaning and inspiration as I was completing this dissertation. You also reminded me that there is life outside of work and academia. I love you and I am blessed to be your father. Finally, to my wife: I owe everything to Susana. She is the smartest, kindest, funniest, and most beautiful person I know. Since we met, she has been my companion in this and every journey, and I am grateful every day to have her in my life.

This work was supported by a +3 PhD grant awarded by the UK Economic and Social Research Council.

List of Tables

List of Figures

Introduction

The book "Improving Surveys with Paradata" (Kreuter, 2013a, p. 3–4) defines paradata as

> additional data that can be captured during the process of producing a survey statistic. Those data can be captured at all stages of the survey process and with very different granularities [...] Paradata, as we define them here, are not available prior to data collection but generated within, and they can change over the course of the data collection [...] Paradata can help researchers understand and improve survey data.

This dissertation is guided by this definition. In particular, four elements of the previous citation serve to set the scene for this dissertation and further elaborate on the concept and its applications for survey methodology.

Paradata as additional data

Conventionally, whenever one thinks of survey data, one imagines the responses gathered by a survey questionnaire. Survey data is typically understood as the information that is successfully collected from a group of issued respondents. Anything else, while of potential relevance and use for the survey (including its design, management and/or analysis) is thought of as supplementary information. For example, information about eligible cases or groups as recorded in the sample frame from which eligible respondents are selected is termed "auxiliary data" and is often used not only to design a sample but also for nonresponse bias adjustment (Smith, 2011). Similarly, paradata are information collected by the survey through means other than a questionnaire.

Because they are additional information, paradata not only have the potential to enhance our understanding of the responses gathered by the questionnaire or recorded in the sample frame but also to observe phenomena which would otherwise be unobservable. For example, paradata are often collected not just for respondents but also for nonrespondents to better understand noncontact and noncooperation (Durrant et al., 2011, 2017). They may also provide insight into the response process, and include information about respondents' speed when answering, contactability and propensity to cooperate with a survey, or respondent effort (Couper and Kreuter, 2013; Heerwegh, 2003; Mercer, 2012).

Paradata as process information

Therefore, unlike the information gathered by the survey questionnaire, or even that of the sampling frame, paradata provide insight into the mechanics (and the different stages) of running a survey. In other words, they are data about the information gathering process, and not about the substantive research objectives of a survey. Nor do paradata capture information about the demographic characteristics of potential respondents or their aggregate traits.

Because of the different processes inherent in running a survey there exist different opportunities for collecting process information and therefore different types of paradata. For example, information about interviewer's attempts to contact a respondent (usually referred to as "call records") has been investigated extensively in recent years (Durrant et al., 2013a; Henly and Bates, 2006; Kreuter and Kohler, 2009; Kreuter and Olson, 2013; Wagner, 2013). These data often record time and date of each individual interviewer visit (or call), as well an individual call status (i.e. no contact, partial interview, refusal, ineligible, etc.) and feature in analyses of fieldwork performance, interviewer evaluation, nonresponse assessment and nonresponse bias. Similarly, in Computer Assisted Self-Interview (CASI) surveys, digitally produced paradata can capture a respondent's keystrokes when typing an answer, time spent on survey items, click counts and patterns used when navigating the questionnaire, and information about the device (including screen size, operating system, internet browser, etc.) used by the respondent. These paradata are often used in analyses of measurement error, item / questionnaire evaluation and respondent behaviour (Callegaro, 2013; Couper and Kreuter, 2013; Heerwegh, 2003).

Paradata as differently granular

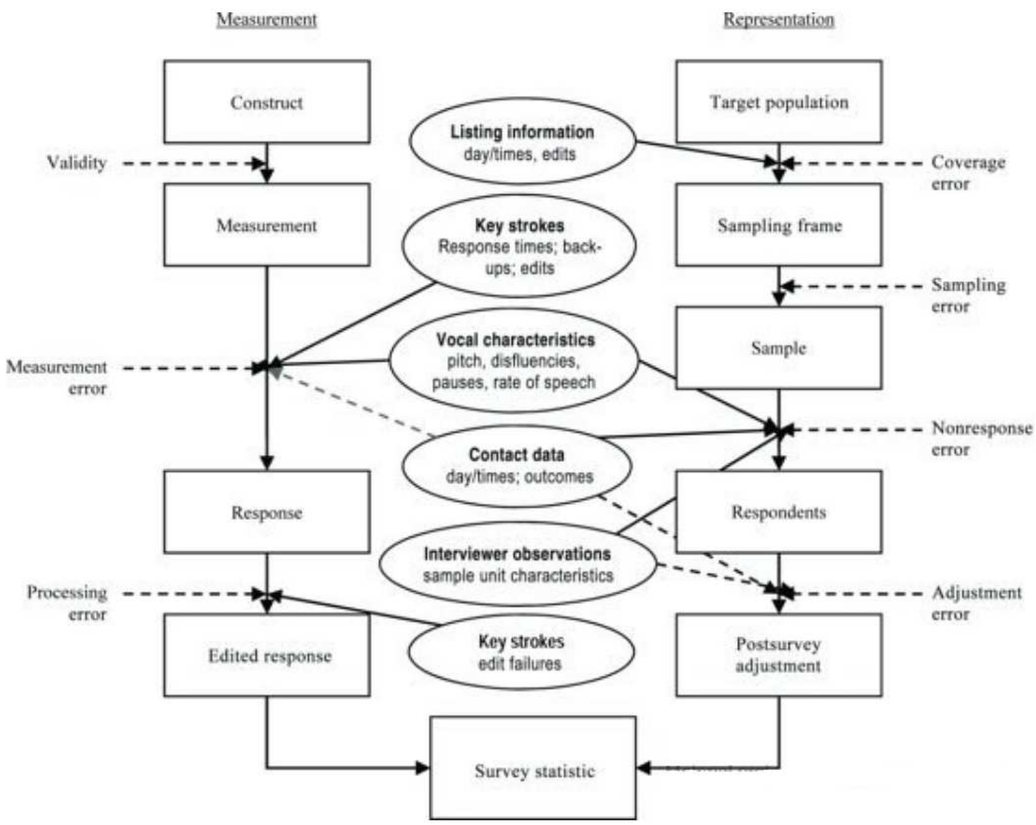
In comparison to other survey-related datasets, paradata can be very messy. They can be considerably larger in size, and often non-rectangular in shape (Yan and Olson, 2013). This is mostly a function of the different levels of granularity at which they can be observed and / or stored. For example, a single household may require anywhere between 1 and 20 different and separate call records to accommodate all of the interviewer's contact attempts (Kreuter, 2013a, p. 7). Similarly, in CASI surveys, timings data can capture not just time spent completing an entire questionnaire, but also, completing a section, an individual question and even durations between individual clicks within a screen. Keystroke data can grow exponentially in size, and vary in nonrectangular shapes, when one considers that keystrokes can be recorded for each individual question by storing not just number of keys typed, but also deletions made, punctuation used, typos made, corrections, etc. (Callegaro, 2013; Couper and Kreuter, 2013; Heerwegh, 2003). Therefore, in contrast with conventional questionnaire data and even auxiliary information, paradata are observed at more levels (often more granular) than the respondent or subsample group levels. As such, paradata pose the additional challenge of complex data management and differing levels of aggregation.

Paradata to understand and improve survey data

Because they provide additional information about the data collection process unobserved by response and auxiliary data with complementary levels of observation and richness, paradata can enhance our understanding of survey data and be used to improve its quality. In this regard, the Total Survey Error (TSE) framework helps inform the analysis and application of paradata in investigations of survey data quality (Biemer, 2010; Groves and Lyberg, 2010; Kreuter and Casas-Cordero, 2010) as well as the analyses of this dissertation.

As Figure 1 shows, paradata could be used to investigate most of the sources of error identified in the TSE framework. Nonetheless, some sources of error have been investigated more extensively than others in the dedicated literature: namely, measurement and nonresponse error. With regards to measurement, paradata capturing response behaviour (such as keystroke data, clicking patterns, response times and even vocal input) have been used to analyse response quality, particularly in web surveys (Callegaro, 2013; Couper and Kreuter, 2013; Heerwegh, 2003; Olson and Parkhurst, 2013; Yan and Olson, 2013). Regarding nonresponse, and especially nonresponse in interviewer-administered household surveys, paradata (including call records, interviewer observations and interviewer-interaction metrics) have been used to formulate response retention strategies, as well as nonresponse bias assessment and adjustment, and fieldwork assessment (Durrant et al., 2013a, 2011, 2017; Henly and Bates, 2006; Kreuter and Kohler, 2009; Kreuter and Olson, 2013; Lynn and Clarke, 2002; Lynn et al., 2002; Olson, 2013; Wagner, 2013).

Figure 1: Paradata in the Total Survey Error Framework.



Graph replicated from Kreuter (2013a, p. 5).

Dissertation Summary

This dissertation discusses three separate but related applications of paradata for survey design and analysis.

Chapter 1 expands on the use of call record data (i.e. structured information about personal visits made by the interviewer to sample addresses) for nonresponse investigation in face-to-face surveys. By focusing on the relatively underexplored analysis of longitudinal call record data in household surveys, it considers association between Wave 1 call record sequences and response outcomes in proceeding Waves (2, 3 and 4) of the UK Household Longitudinal Study (or *Understanding Society*). This chapter addresses the correlations of predefined call record sequences observed in the baseline wave of this survey by comparing model estimates that employ not just this type of paradata but also more conventional predictors of nonresponse (like sociodemographic characteristics of issued households as well as auxiliary geographic information). Beyond finding associations and comparing model specifications, this analysis is primarily interested in informing response retention strategies for panel surveys based on the calling patterns of earlier waves.

Chapter 2 is similarly concerned with field effort optimization. However, while Chapter 1 uses call record data as predictors of a given outcome of interest (namely future wave contact and cooperation), Chapter 2 proposes models to analyse the calling effort inherent in the processes of contact and cooperation conditional on household and aggregate individual-level data (as well as lagged contact record data). Given the onerous fieldwork demands of household longitudinal surveys, the analysis of this chapter aims to inform data collection optimization by identifying predictors (especially those derived from paradata) of differential contactability, cooperation and overall field effort requirements in longitudinal context.

The third and final chapter analyses different types of CAWI (Computer Assisted Web Interviewing) generated paradata to assess progress indicator (PI) effects on survey response quality. PIs are graphical and/or textual elements embedded into the display of a computer-based questionnaire meant to inform the respondent of the relative progress made at any given point of the survey (Couper et al., 2001; Crawford et al., 2001). The

data used for this analysis comes from an experiment designed by the author. This chapter seeks to further examine and develop standing theories of progress indicator effects on surveys by focusing on their impact on response quality while also expanding on the uses of web survey paradata and their applications for response quality assessment and respondent behaviour (including satisficing, as well as time and effort management).

Each chapter is self-contained and can be read independently of each other. Nevertheless, the dissertation as a whole is guided by common themes. Most importantly, this work aims to contribute to the literature on paradata by providing practical examples of their use and analysis for survey methodologists as well as survey practitioners. Beyond the substance of the applications of paradata for nonresponse analysis, fieldwork assessment and measurement error investigation, each chapter hopes to contribute to general discussions on the challenges of generating, managing, and interpreting these complex datasets and deriving meaning (and useful analytical variables) from them. Relatedly, this dissertation is also concerned with identifying appropriate methodological and statistical solutions given not only a particular set of research objectives but also the differential features of paradata. In addition, this dissertation hopes to contribute to ongoing discussions on survey design optimization within the contexts of rising costs, diminishing response rates and increased survey fatigue among potential respondent populations.

Chapter 1

Call and response: Modelling longitudinal contact and cooperation using Wave 1 call records data

Abstract: For longitudinal surveys, there is little discussion on how call record data are able to account for household nonresponse. This chapter uses call records as well as observed data from *Understanding Society's* Wave 1 to model Wave 2, Wave 3 and Wave 4 household contact and cooperation propensities. Multi-level logistic models are used to account for the nested structure of the data (households within interviewers). Results indicate that households which had repeated unproductive contacts, broke appointments, registered above median proportion of "no replies", or began the call sequence with an unproductive contact in Wave 1 are at risk of future nonresponse.

Keywords: Call Records, Contact Propensity, Cooperation Propensity, Household Nonresponse, Household Panel Survey

1.1 Introduction

Survey nonresponse is a prime concern for survey methodologists and practitioners alike. Besides negatively impacting on survey costs and fieldwork efficiency, nonresponse results in diminished statistical power and potentially biased survey estimates. Furthermore, longitudinal surveys carry the compounded problem of attrition (Lynn, 2009a). The sample gathered at the first wave suffers from progressive nonresponse after each wave resulting in additional imprecision and potential bias. Therefore, efforts to reduce nonresponse are motivated not only by considerations of cost but also by concerns with data quality.

Broadly speaking, procedures that deal with nonresponse (including attrition in longitudinal surveys) can be grouped as pre- or post-fieldwork or, said differently, methods to *prevent* or *adjust* for nonresponse. For example, nonresponse prevention may be incorporated as a design feature and could include: cash incentives for respondents, advance letters or increased number of total interview attempts per potential respondent (Laurie and Lynn, 2009; Laurie et al., 1999). While generally effective, these only serve to attenuate (but never eliminate) nonresponse. Additionally, they may also compromise data quality if they exert any other (potentially biasing) effect besides increasing response rates. Lastly, these design features may also involve additional costs and field effort.

Adjustment of nonresponse usually occurs after data collection has finished and entails using auxiliary or associated survey data that can identify the mechanism of nonresponse and therefore minimize its effect via statistical controls such as weighting or imputation (more rudimentary procedures may involve simple case deletion). There is ample work to suggest which variables (whether auxiliary or those found in the questionnaire itself) may be associated with the nonresponse and noncontact mechanisms (Groves and Couper, 1998; Lepkowski and Couper, 2002; Uhrig, 2008). However, the correlations tend to be weak or endogenous. Moreover, while specific survey items may correlate with nonresponse mechanisms in some surveys they may not necessarily do so in others. In other words, the relationship between these survey items and the process of nonresponse may be topic-specific. Auxiliary variables are often collected as demographic aggregates (Callegaro, 2013; Kreuter, 2013b), prone to error (West, 2013), or unable to identify what is most likely (if at least partially) a function of individual traits and decisions. Lastly, these auxiliary variables are not always readily available and, in some cases, no adjustment is possible.

Given all these limitations, a promising source of information to model the mechanisms of contact and cooperation are call records; data about the field process, also referred to as: process data, contact history data or call history data (Durrant et al., 2013a; Henly and Bates, 2006; Kreuter and Olson, 2013; Wagner, 2013). Besides collecting the data associated with a given questionnaire, recording information about each call the interviewer makes with a potential respondent is a common practice among many survey organizations. For example, the time and duration of each call can be recorded along with a codified outcome ("no reply", "completed interview", "some interviewing done", "refusal", "appointment", or "other") (Blom et al., 2010; Kreuter and Casas-Cordero, 2010). Along with call records, the interviewer may also include field observations (Kreuter and Olson, 2013), such as: condition of the household; type of dwelling; presence of alarm systems, gates, guard dogs; evidence of car ownership or infants in the house; etc. Respondent and interviewer identifiers may also be included in addition to derived aggregates from this type of data (e.g. total number of calls, time of first call, average field duration, etc.)

In comparison to conventional survey measurements and other forms of auxiliary data call records generally show stronger associations with the mechanisms of nonresponse. This is largely because of the conceptual proximity to the contact and cooperation processes. Put simply, call records measure the response processes. They register the sequence of events that lead to a given survey target being classified as "completed interview", "noncontact", "refusal", etc. In fact, more detailed classification is possible given the call records, as sample members may be deemed "difficult to contact", "reluctant", "hard-to-get" or "easy-to-get" (Hall et al., 2013; Lynn and Clarke, 2002; Lynn et al., 2002). These data also provide information about the performance of the interviewer and trends of the fieldwork (Kreuter et al., 2010). For example, estimates of average call duration or proportion of call outcomes per interviewer allow for an evaluation of interviewer effort and efficiency as well as the level of difficulty in contacting and eliciting cooperation from sample subjects.

While not completely disregarded, call records remain fairly unexplored when dealing with nonresponse in longitudinal surveys. To be fair, there is a considerable amount of work on using these data to analyse and predict survey nonresponse in cross-sectional surveys or within waves of a panel study. These data have also been used to evaluate nonresponse error (Kreuter and Kohler, 2009; Kreuter and Olson, 2013; Lynn et al., 2002) optimize contact strategies (Wagner, 2013) and assess fieldwork effort (Durrant et al., 2011, 2013b; Kreuter et al., 2010; Mercer, 2012). More specifically, call histories have been used to identify survey bias between early and late respondents (Lynn and Clarke, 2002; Lynn et al., 2002); model response outcomes based on interaction of first contact (Durrant et al., 2015); and determining best times of call to elicit higher response rates (Durrant et al., 2011). However, with regards to longitudinal surveys, the literature is considerably smaller. Call records have been used to understand attrition in panel surveys (Bates, 2004; Henly and Bates, 2006) and flag potential dropouts so interviewers can tailor their field strategy. For example, the total number of contacts per respondent has been shown to be significantly associated with increased nonparticipation in future waves (De Keulenaer, 2005).

However, beyond summary statistics of call records and measures of extended interviewer effort, what remains to be analysed is how specific call sequences or events are able to account for household-level contact and cooperation propensities in subsequent waves. In fact, there is little research on determining the existence of underlying household-level contact and cooperation propensities in longitudinal surveys based on call records. Moreover, there is hardly any work on how call records fare when conditioning on other predictors of nonresponse (i.e. demographic traits and attitudinal characteristics).

Should call records reveal new insights into the response mechanisms of subsequent waves, field strategies to curb noncontact and noncooperation could be designed and implemented at little or no additional cost before data collection begins. While other predictors of nonresponse (like respondent attitudes, demographic traits or social context) are entirely dependent on the respondent, events observed in the call records are in part a function of decisions made by the field office and/or the interviewer. As the product of interviewing interactions, these data are especially convenient for fieldwork adjustment between waves. That these adjustments could potentially be made after only the first wave of a longitudinal study adds more value to their application given the considerable costs of developing and maintaining a survey of this type as well as the crucial importance of its first round of fieldwork.

1.1.1 Research Questions and Objectives

Specifically, this analysis is concerned with two questions:

1. Are there specific events and call sequences observed at Wave 1 of a panel study associated with future contact and cooperation propensities?

2. Do these Wave 1 call sequences reveal additional information about future contact and cooperation unobserved by conventional demographic and attitudinal predictors of nonresponse?

To address these questions this chapter explores the effects of aggregated Wave 1 individual interview and household call record data as well as household traits. Model specification considers the conditionally independent processes of contact and cooperation (Campanelli et al., 1997a; Groves and Couper, 1998; Lepkowski and Couper, 2002; Lynn et al., 2002; Nicoletti and Peracchi, 2005) and is informed by established theories of survey nonresponse (Couper and Groves, 1996; Groves and Couper, 1998; Lepkowski and Couper, 2002). Additional controls include geographical markers[1] and stable household flags[2] to account for interviewer reallocation[3] as well as household alterations between waves. To clarify, the objective of this chapter is to better understand

[1]As geographic markers this chapter uses to the Lower Layer Super Output Area classification as defined by UK's Office for National Statistics. LSOA's are geographical zones with a minimum of 1000 and a maximum of 3000 inhabitants residing in 400 to 1200 households (Office For National Statistics (ONS), 2011).

[2]This analysis defines a household as stable when at least one individual member remains in a household between waves and its physical address stays fixed within a LSOA.

[3]Interviewer reallocation refers to the process (deliberate or otherwise) where households are approached by different fieldworkers across waves. It is important to note that interviewer reallocation is rarely (if ever) a random process. Indeed, interviewer continuity and household response may share com-

the processes of longitudinal contact and cooperation and to determine what role Wave 1 call record sequences play in these processes.

1.2 Theory and concepts

Household surveys may involve (at least) two types of questionnaires: one for the residents and one for the household. While the first measures individual-level items, the latter measures household-level traits such as family composition or general characteristics of the home. Therefore, cooperation with the corresponding survey instrument is the most obvious way of determining whether a household or individual responds or refuses. However, a further and more detailed classification is possible for households given the patterns of response of its corresponding individual residents. A household may be deemed "fully respondent" if all eligible individuals within it participate, "partially respondent" if only some do or "only household questionnaire completed". Because this chapter deals with call interactions with sampled households in the first wave of a longitudinal household survey and given that all individuals from Wave 1 chosen households are tracked in future waves of this study, for the purposes of this analysis a household is considered responding when it completes at the very least the household questionnaire [4].

Cooperation only occurs conditional on contact – it is impossible for a noncontacted individual to cooperate with a survey. However, while they may be procedurally dependent, contact and cooperation are not necessarily statistically correlated. In fact, the literature suggests that these two are driven by different (and independent) trends and traits of the respondent, interviewer protocols, social context, respondent accessibility, and survey design (Groves and Couper, 1998; Lepkowski and Couper, 2002; Lynn et al., 2002; Nicoletti and Peracchi, 2005).

Indeed, Groves and Couper (1998) provide a theoretical framework for contacting sample households and eliciting survey cooperation. Likelihood of contact is a

mon causes (Lynn et al., 2014). Therefore, it is important to take this non-random process into account when interpreting some of the findings of the analysis proposed in this chapter.

[4]It should be noted that this definition for household response status is not the same as the official UKHLS definition.

function of at-home patterns of the sample member and the call strategy employed by the interviewer. In turn, the latter is influenced by attributes of the interviewer while the former is determined further by physical impediments to the respondent (or their household), as well as social-environmental and socio-demographic attributes of the sample member. Once contacted, the decision to cooperate with a survey occurs during householder-interviewer interaction. This interaction can be understood as a function of factors classified as either "out of" or "under" researcher control. Out of researcher control are the social environment (e.g. economic conditions, survey-taking climate, and/or neighbourhood characteristics) in addition to household traits (e.g. household structure, socio-demographic characteristics, and/or psychological predisposition). Controllable by the researcher are survey design (e.g. topic, mode of administration and respondent selection) as well as interviewer traits (e.g. socio-demographic traits, experience and expectations) (p. 25-46).

Thus, understanding and accounting for nonresponse demands careful consideration of these two sequentially linked but ultimately separate processes of contact and cooperation. One approach that takes this into consideration (and which is adopted in this analysis) involves separate (but conditional) modelling of the processes (contact conditional on eligibility; cooperation conditional on contact).

1.2.1 Using call records to model the nonresponse process

Identifying households as likely noncontacts and/or refusals in longitudinal surveys has immediate applications for fieldwork effort allocation and optimizing interviewing strategies. If potential nonrespondents can be flagged before data collection starts, tailored or adaptive approaches (Calinescu et al., 2012; Groves et al., 1992; Groves and Heeringa, 2006; Lynn, 2014; Morton-Williams, 1993; Oksenberg et al., 1986; Schouten et al., 2013; Wagner, 2008) can be implemented to minimize the effect of nonresponse on data quality and cost. Essentially, preventing nonresponse is made easier when it can be anticipated. Indeed, the focus on prevention is based on the assumption that retaining collaborating survey respondents demands less effort than converting dropouts. Moreover, because the processes observed in call records fall (at least partially) under the control of the interviewer, identifying call events or sequences that lead to future nonresponse could inform/modify interviewer behaviour protocols with the aim of reducing nonresponse.

Evaluating household – as opposed to individual – nonresponse is also rooted on a concern for fieldwork optimization. Preventing an entire household from dropping out automatically impacts on the future contactability and cooperation of the individual residents within it. The opposite is not necessarily true – individuals likely to drop out may live in an otherwise highly cooperative and/or contactable household. Thus, limiting analysis solely on the individual may hide higher-level dynamics that determine household (and therefore individual) nonresponse.

Of course, the nature of the data (especially call record data) also places the analysis at the household level. Call records are nested within the home and not the resident. Before interviewing commences, the interviewer's first concern is with finding and eliciting the cooperation of the household to then turn to the individuals within it. Even the "gatekeeper", or resident that first interacts with the interviewer at the doorstep, is best understood as a household artefact as it is likely to 1) not be individually identified 2) be a different individual between calls.

1.3 Data and methods

This analysis attempts to model response observed at Waves 2, 3 and 4 of a household longitudinal survey using covariates observed at Wave 1 (as well as two cross-wave controls). The dependent variables, contact and cooperation, are assumed to be a function of Wave 1 call record sequences, interviewer observations, survey data from the household and individual questionnaires as well as indicators for 1) interviewer continuity (Campanelli and O'Muircheartaigh, 1999; Laurie et al., 1999; Lynn et al., 2014) and 2) geographically-stable households across waves.

1.3.1 Household contact and cooperation in *Understanding Society*

The data used for this analysis come from *Understanding Society*, the United Kingdom's Household Longitudinal Study comprised of approximately 40,000 households and close to 100,000 individuals interviewed in yearly waves. The sample is representative of all UK households with an additional "boost sample" of the five main ethnic minorities in Great Britain. All household members over the age of 16 are eligible to be interviewed. Fieldwork for Wave 1 began in 2009. Survey questions cover multiple topics, including: employment status and history, personal and household finances, health and general well-being, social attitudes, family composition and community attachment (Knies, 2015).

In total 26,200 households issued at Wave 2, 24,425 at Wave 3 and 21,320 at Wave 4 of the survey are analysed. They belong to the General Population Sample Component of Great Britain[5] (Lynn, 2009b). In Wave 2 a total of 382 households are dropped from the analysis as they report: 1) field periods longer than the 3 month limit[6] stipulated by the field protocols of *Understanding Society* or 2) empty call record data. In Wave 3, 336 households are removed for the same reasons as well as 277 in Wave 4.

Table 1.1: Wave 2, Wave 3 and Wave 4 Household Outcomes

Final Outcome	Wave 2	Wave 3	Wave 4
Noncontact	1,467	977	780
Contact: Response	19,928	17,987	16,870
Contact: Refusal	2,991	3,088	2,131
Contact: Other Nonresponse	31	269	180
Non Eligible	279	578	518
Unknown Eligibility	1,504	1,526	841
Total	26,200	24,425	21,320

As previously stated, for this analysis a household is said to cooperate when it answers (at a minimum) the household questionnaire (regardless of the residents' cooperation with the individual questionnaire). Given the sample design and following rules of the UKHLS[7], the households represented in Table 1.1 (those from Waves 2, 3 and

[5]The General Population Sample is a component of *Understanding Society*, representing the UK household population over time (except for those consisting solely of post-2009 immigrants). Other sample components, like the the Ethnic Minority Boost, are excluded from the analysis as they required an additional screening procedure at the doorstep and created a different call record data structure. Because the British Household Panel Survey component was not measured in Wave 1 it is also disregarded from this analysis. Finally, the field management in Northern Ireland did not register call records and thus is also not included.

[6]For 99% of UKHLS households, the field is completed within three months of the date of the first call. Therefore, cases with extended field durations (over three months) are exceptionally rare and do not follow usual protocol.

[7]As Lynn (2009b, p.11-12) explains "All persons identified at Wave 1 as sample members [...] will remain sample members indefinitely regardless of their location or household circumstances. Subsequent

4) contain at least one individual who was sampled and cooperated with an individual and/or household interview at Wave 1. In other words, households that did not register any response at Wave 1 are not included in the final data to be analysed as they were not issued in future waves of the survey.

1.3.2 Variable selection

Besides the call records, information about geographical markers, the characteristics of the dwelling, household size, demographics, levels of political & community attachment, and previous interview experience was used to construct covariates of contact and cooperation. These were selected according to established theories of household nonresponse and comparable empirical studies (Couper and Groves, 1996; Groves and Couper, 1998; Kalsbeek et al., 2002; Lepkowski and Couper, 2002; Uhrig, 2008). In total, 26 variables were considered for the models. While some were originally collected at the household level, others needed to be aggregated from the individual respondent files or call records to fit the structure of the dataset. Further data reduction resulted from the construction of index variables. As previously stated, all these independent variables were gathered in Wave 1 (except for interviewer and LSOA continuity) and are used to model outcomes observed in Wave 2, Wave 3 or Wave 4. For a summarized list of all variables used in these models refer to Table 1.2.

---

to Wave 1, all new births whose mother is a sample member will themselves become a sample member. When sample members move, attempts will be made to follow them to their new location and interview them there. Even when a sample member moves out of the UK they will remain in the sample, though no attempt will be made to carry out face-to-face interviews. [...] At every wave the intention will be to interview all members of each household containing at least one sample member. [...] But at each wave subsequent to Wave 1, there will be many cases where household composition has changed since the previous wave, resulting in sample members being co-resident with non sample members. All such non sample members will be interviewed at any wave when they are co-resident with a sample member, but they do not themselves become sample members and are not therefore followed if they leave the household of sample member(s)."

Table 1.2: Variable Selection

Categories	Variable name	Level of data collection
Call Records	Status of first call	Call
	At least one broken appointment	Call
	Repeat unproductive contacts	Call
	Above median proportion of noncontacts	Call
Geographical markers	Geographical Region	Household
	Urban indicator	Household
Dwelling	Dwelling type	Household
	Groundfloor indicator	Household
	House in worse condition than neighbours	Household
Household size	Number of residents	Household
Demographics	Baby present	Household
	Residents in poor health	Individual
	National origin of household	Individual
	Working status	Individual
	Pensioners in household	Individual
	Deprivation indicator	Household
	Household tenure	Household
Social attachment	No political interest among residents	Individual
	Community attachment	Individual
Previous Interview Experience	Consent to linkage	Individual
	Someone else present during interview	Individual
	Suspicious during interview	Individual
	Understood interview questions	Individual
	Item nonresponse	Individual
Cross-Wave Controls	Same interviewers across waves	Household
	Same LSOA across waves	Household

Call records

With regards to our covariates of interest, four variables are generated from the call records. These derived variables identify problematic call sequences assumed to be associated with reduced household contactability and/or propensity to cooperate.

1. Status of first call are any of the following: "no reply", "appointment set", "unproductive contact", "some interviewing done", "any other status" or "completed interview"[8].

[8]Call statuses have been relabelled from the original dataset to aid interpretation of the data. The original variable includes only five possible values. For this analysis, a sixth value was derived - "completed interview". It distinguishes calls where partial interviewing was done from those cases where the entire interviewing was completed and no more calls were necessary to collect additional information. Moreover, the term "unproductive" was added to the "contact" status as it reflects those calls where the interviewer was able to contact the interviewee but no interviewing or appointment resulted.

2. Houses where an appointment is set but are followed by any call status besides "some interviewing done" or "completed interview" are deemed "at least one broken appointment".

3. A similar event pattern is recorded in houses where "unproductive contact" occurs more than once in a row to single out cases where the interviewer repeatedly engages the respondent but no appointment is made nor any interviewing achieved.

4. Finally, based on the distribution of noncontact calls, households are divided in two groups: those with a proportion of total "no replies" below the corresponding wave sample median and those above it.

Other nonresponse covariates

1. Geographical markers: Usually, population density is associated with decreased contact and cooperation. As proposed by Watson and Wooden (2009), two underlying mechanisms could account for the effects of urbanicity on differential nonresponse. Firstly, social isolation is more prevalent in large urban centers which is linked with increased reluctance to cooperate with surveys. Alternatively, nondomestic routines are more typical in large cities. Therefore, for household surveys, it is harder to establish contact with populations that are less often at home. In this analysis, geographical markers like the UK Government Office Regions and an urbanicity indicator aim to capture and control for these effects. In the UKHLS these two variables are originally derived at the household level.

2. Characteristics of the dwelling: Depending on its physical characteristics, some dwellings may be easier to contact than others (Groves and Couper, 1998). As an example, buildings with locked entrances are most often harder to access than a detached or a semi detached home. Similarly, a ground-floor property will demand less effort from an interviewer than a flat located several stories above; even more so when no elevator is available. Lastly, the overall condition of the dwelling may affect if and when the interviewer decides to visit a particu-

lar household. In this analysis, with regards to the characteristics of the dwelling, a house can be: "detached", "semi-detached", "terraced / end", "flat / maisonette / purpose / converted", "bedsit / with business / sheltered / institution / other"[9]. In addition, there is an indicator to determine whether the household is on a ground floor or elsewhere. Finally, according to the interviewer's observation a house can be deemed to be in a "better / same" or "worse" condition than its neighbours.

3. Household size: Household size refers to the total number of residents (including responding and nonresponding residents). Because the UKHLS is a household survey, probability of contact is likely increased in households with a larger number of eligible respondents.

4. Demographics: The social composition of the household can be related to the at-home patterns associated with increased contact as well as social and psychological dispositions towards survey cooperation (Groves and Couper, 1998). For example, families with small children are more likely to be at home (Nicoletti and Peracchi, 2005). Retired people may be more accessible and have more time and disposition to entertain an interviewer than a younger professional working full hours (Nicoletti and Peracchi, 2005). People whose health status prevents them from leaving the house may be easier to contact but also harder to recruit for a survey interview. Low socioeconomic status can also correlate with reluctance to cooperate with a survey. Similarly, ethnic and national minority groups may feel less socially attached and thus less inclined to respond[10]. Lastly, household tenure is another variable frequently included in analyses of longitudinal nonresponse as it is another indicator of social attachment and/or increased at-home routines (Lepkowski and Couper, 2002; Watson and Wooden, 2009). In this analysis the following variables have been included to account for the social makeup of the household: presence of a baby in the household; aggregated health status; national origin and employment status of the residents; deprivation indicator and

[9] For ease of analysis and to allow for larger cell sizes, these categories were aggregated from an original list of 13 different dwelling types.

[10] Language difficulties may be more prevalent in these types of households and thus result in less likelihood to respond. Nevertheless, UKHLS included materials in multiple languages to address this possibility.

household tenure. Specifically, presence of a baby is determined by a binary indicator for homes with at least one child under the age of 2 as observed in Wave 1. The health status of a household is determined by the proportion of residents who declared being in "poor" health. The two possible values of the variable are: "all in poor health" or "at least one not in poor health". Additionally, according to the individual national origin, a household is classified as either: "all British", "mixed", or "all non-British". Similar aggregation is used for employment ("no one works", "at least one works, but not long hours"[11], "at least one works long hours" or "all work long hours") and presence of pensioners in the household ("no pensioners", "at least one pensioner", "all pensioners"). A deprivation index is constructed from a battery of questions that ask responding households whether they "have", "can't afford" or "do not need" any of the following: annual holiday, monthly drink/meal out with friends, two pairs of all-weather shoes for all adults in the house, enough money for household repairs, household contents insurance, money for savings/retirement plan, money to replace worn out furniture or money to replace broken appliance. A household is said to be materially deprived if it answers that it cannot afford at least two of these items. Lastly, household tenure is classified as either: "owner / mortgager" or "all others".

5. Social attachment: As mentioned above, measures of social isolation / social attachment are often associated with differential response propensities (Groves and Couper, 1998; Watson and Wooden, 2009). The underlying mechanism argues that participating in surveys (particularly, large, government sponsored surveys) is associated with a civic duty that helps gain a better understanding of society and thus contributes to its general advancement. To the extent that socially excluded groups feel unidentified or isolated from the dominant society, they may not be driven by the same social obligations when asked to participate in a survey. Besides the sociodemographic proxies for social inclusion / attachment previously mentioned (i.e. household tenure, socioeconomic status, national origin, and urbanicity), two additional measures are included in this analysis: politi-

[11] "Long hours" entails working more than 35 hours a week.

cal and community attachment. Firstly, political attachment is derived from the question "How interested would you say you are in politics?". Respondents had a scale of four options to chose from: "very", "fairly", "not very", or "not at all". If no individual residents express any political interest (i.e. all residents answer "not at all"), the household is labelled "no political interest among residents". Otherwise, it becomes "some political interest among residents". Community attachment is determined by an index that totals the responses from eight self-completion questions related to an individual's willingness or habit of interacting with neighbours and/or community organizations. The household average of the aggregated individual responses is placed into quartiles based on the aggregated distributions for all households, where Q1 indicates the lowest possible attachment among the households and Q4 the highest. Because of the high frequency of item missingness (likely due the self-administered component of this questionnaire section) a further category of "missing" is added to the scale for a total of five possible values for this variable.

6. Previous interview experience: Finally, given the longitudinal nature of the study, previous survey experience is also considered to control for possible differential contact and cooperation propensities in future waves. In short, evidence of prior respondent cooperation with the survey is assumed to correlate positively with continued response (Laurie et al., 1999; Lepkowski and Couper, 2002). In total, six variables account for the respondents' previous interview experience (as observed in Wave 1). Firstly, depending on the willingness of individuals to consent to linking their survey data with administrative health records held by the NHS and associated agencies, a household is classified as either "no one consents" or "at least one consents". Secondly, based on the presence of someone else (besides the interviewer and respondent) during any of the household's corresponding individual interviews, a household is classified as having "someone else present" or "no one else present" during the Wave 1 interview. Because multiple household members may be eligible to answer the individual questionnaire, individual interviews should be conducted without the presence of another person when-

ever possible. Thus, this analysis assumes this slight deviation from the proper protocols of a survey interview is possibly indicative of a suboptimal survey experience. Thirdly, based on interviewer assessment, individuals' suspicions with the survey as well as their understanding of the questions are also aggregated to the household level so that these become "some suspicion"/"no suspicion" or "excellent understanding"/"less than excellent understanding". Finally, item nonresponse, is the last variable considered to analyse previous survey experience. As Loosveldt and Billiet (2002) and Laurie et al. (1999) argue, item nonresponse is often associated with negative survey experience and future wave nonresponse. In this analysis, item nonresponse is aggregated from the individual interviews of all Wave 1 responding residents within the household. Given the skewness of the resulting distribution, the results are converted to a logarithmic scale and complemented by an indicator with values "no item nonresponse" or "some item nonresponse".

Cross-wave controls

Finally, two cross-wave controls are derived. The first identifies households as having the same vs. different interviewer across waves[12] while the second identifies households that remain in the same Lower Layer Super Output Area[13] across waves.

Several studies have shown keeping the same interviewer assignments per households across waves is associated with increased contact and cooperation (Campanelli and O'Muircheartaigh, 1999; Laurie et al., 1999; Lynn et al., 2014; Nicoletti and Peracchi, 2005). However, the direction of the association is not always clear and in fact is often confounded by additional mechanisms like area effects (Campanelli and O'Muircheartaigh, 1999; Durrant et al., 2010), household and/or interviewer moving,

[12]In other words, households assigned to the same interviewer for Wave 1 and Wave 2, 3 or 4 depending on the analyses discussed further below.

[13]While not immediately equivalent to a fixed home address, the LSOA variable allows for the analysis of household stability across waves. While it is possible (and even probable) that some movers remain within the vicinity of a previous residence, separating households based on their LSOA across waves correctly identifies those households that move a considerable distance. Further, the analysis of fixed geographical location can be relaxed if it is assumed that LSOAs are homogeneous regarding contact and cooperation propensities. In other words, households that move within a neighbourhood may be assumed to keep certain traits about their structure and other variables possibly correlated with nonresponse vs. households that move outside of the original neighbourhood.

interviewer employment dynamics and other operational decisions made by the field agency between waves. Indeed, and as mentioned previously, in the UKHLS (and likely in most face-to-face household surveys) interviewer allocation between waves is a nonrandom process which in turn is determined by other dynamics associated with differential longitudinal response propensity.

Thus, these cross-wave indicators are included in the models to account for processes likely and substantially confounded with nonresponse, but which are not the main focus of this analysis. To the extent that these covariates are discussed it will not be done to ascertain specific interviewer or area effects but to address (and control for) a particularity of the data as collected by the UKHLS. Indeed, the main focus of this analysis remains Wave 1 call record sequences and their association with Wave 2, 3 and 4 household contact and cooperation.

1.3.3 Modelling strategy

In total, 18 different models are evaluated for this analysis (see Table 1.3). The specifications result from a combination of three different analytical considerations, namely:

1. Outcome of interest (contact vs. cooperation).

2. Wave of observed outcome (Wave 2, Wave 3 or Wave 4).

3. Model specification:

   (a) Reduced form: Other predictors of nonresponse (i.e. demographic and attitudinal variables) + cross-wave controls (indicators of geographical and interviewer continuity across waves per household).

   (b) Expanded form: Call sequences + other predictors of nonresponse (i.e. demographic and attitudinal variables) + cross-wave controls (indicators of geographical and interviewer continuity across waves per household).

   (c) Expanded form: First call status + other predictors of nonresponse (i.e. demographic and attitudinal variables) + cross-wave controls (indicators of geographical and interviewer continuity across waves per household).

Because contact and cooperation produce dichotomous outcomes, logistic models are employed in this analysis. Also, random intercept fixed effect models (Rasbash et al., 2015) account for the hierarchical nature of the data (households nested within interviewers). Within wave and for each outcome of interest, the three corresponding model specifications are analysed comparatively to determine the marginal effects of call record data on the mechanisms of nonresponse after conditioning on demographic and attitudinal predictors of nonresponse (i.e. geographical markers, characteristics of the dwelling, household size, demographics, levels of political & community attachment, and previous interview experience) and accounting for any possible differential effects on contact and/or cooperation attributable to the dynamics of cross-wave household composition, mobility and interviewer allocation as collected in the UKHLS data. Coefficients are reported as odds ratios.

Table 1.3: Model Specifications

<table>
<tr><th></th><th>Outcome</th><th>Specification</th></tr>
<tr><td rowspan="2">Wave 2</td><td>Contact</td><td>1. Other NR predictors + x-wave controls
2. Sequences + Other NR predictors + x-wave controls
3. First call + Other NR predictors + x-wave controls</td></tr>
<tr><td>Cooperation</td><td>4. Other NR predictors + x-wave controls
5. Sequences + Other NR predictors + x-wave controls
6. First call + Other NR predictors + x-wave controls</td></tr>
<tr><td rowspan="2">Wave 3</td><td>Contact</td><td>7. Other NR predictors + x-wave controls
8. Sequences + Other NR predictors + x-wave controls
9. First call + Other NR predictors + x-wave controls</td></tr>
<tr><td>Cooperation</td><td>10. Other NR predictors + x-wave controls
11. Sequences + Other NR predictors + x-wave controls
12. First call + Other NR predictors + x-wave controls</td></tr>
<tr><td rowspan="2">Wave 4</td><td>Contact</td><td>13. Other NR predictors + x-wave controls
14. Sequences + Other NR predictors + x-wave controls
15. First call + Other NR predictors + x-wave controls</td></tr>
<tr><td>Cooperation</td><td>16. Other NR predictors + x-wave controls
17. Sequences + Other NR predictors + x-wave controls
18. First call + Other NR predictors + x-wave controls</td></tr>
</table>

1.3.4 Equations

The following equations summarize the model specifications adopted for this analysis. Equations 1.1 and 1.2 refer to the contact models, with 1.1 representing all components of the random-intercept model (including the outcome of interest, the random intercept, the different vectors corresponding to the variable categories mentioned in Table 1.2 and a composite error term corresponding to the interviewer- and household-level residuals). Equation 1.2 represents the empty (or unspecified) random intercept, which is composed of an interviewer-level grand mean and a corresponding variance term. Equations 1.3 and 1.4 represent the cooperation models. The right-hand side composition of the cooperation models are identical to the contact models. The only difference is found in the outcome of interest which is notated as a conditional mechanism to reflect the conditional relationship between contact and cooperation (Campanelli et al., 1997a; Groves and Couper, 1998; Lepkowski and Couper, 2002; Lynn et al., 2002; Nicoletti and Peracchi, 2005). Equation 1.5 refers to the variance components and the assumptions underlying the models. Further clarification on the notation is found below.

$$\text{logit}\left\{P(Contact_{ij}=1)\right\} = \beta_{0j} + \beta_1 C_{ij} + \beta_2 R_{ij} + \beta_3 W_{ij} + \varepsilon_{ij} \tag{1.1}$$

$$\beta_{0j} = \beta_0 + \upsilon_j \tag{1.2}$$

$$\text{logit}\left\{P(Cooperation_{ij}=1|\ Contact_{ij}=1)\right\} = \beta_{0j} + \beta_1 C_{ij} + \beta_2 R_{ij} + \beta_3 W_{ij} + \varepsilon_{ij} \tag{1.3}$$

$$\beta_{0j} = \beta_0 + \upsilon_j \tag{1.4}$$

$$\left[\ \upsilon_{0j}\ \right] \sim N(0,\ \Omega_u)\ :\ \Omega_u\ =\ \left[\ \sigma^2_{u0}\ \right] \tag{1.5}$$

Where: $\mathbf{P}(\mathbf{Contact}=\mathbf{1})$ is the probability of contact at waves 2, 3 or 4. Similarly, $\mathbf{P}(\mathbf{Cooperation}=\mathbf{1}|\mathbf{Contact}=\mathbf{1})$ is the probability of cooperation conditional on contact at waves 2, 3 or 4. **C** represents a vector of call record predictors of response (first call status or problematic call sequences) observed at Wave 1; vector **R** includes all other predictors of response (geographical markers, characteristics of dwelling, household size, demographic traits, social attachment and previous interview experience) also observed at Wave 1; and finally **W** all cross-wave indicators (interviewer continuity, LSOA continuity and their interaction) observed at Wave 2, 3, or 4 depending on the corresponding wave of the independent variable. i is the household-level identifier and j the interviewer-level identifier[14]. ϵ and υ denote the unobserved error terms for the households and interviewers respectively. The grand mean of the outcome variable (contact or cooperation) is represented by β_0, while β_{0j} is the corresponding mean for any given interviewer. The random intercept residuals are assumed to be normally distributed, centred around a mean of 0 and independent. Their variance-covariance matrix is represented by $[\sigma^2_{u0}]$.

1.4 Results

1.4.1 Are there specific events and call sequences observed at Wave 1 of a panel study associated with future contact and cooperation propensities?

UKHLS data indicate that specific events and call sequences are indeed associated with future contact and cooperation propensities. There are significant and sizeable effects observed in the call records of Wave 1 that are associated with Wave 2, Wave 3 and Wave 4 contact and cooperation propensities.

[14] As previously noted, while these models allow for interviewer-level variation, this analysis will not directly discuss interviewer effects. Instead, these random-intercepts models are meant to simply control for any potential bias introduced by the cross-wave interviewer allocation protocols of the UKHLS and other potentially confounding dynamics, like area effects.

Contact

Table 1.4: Contact Models for Waves 2, 3 and 4. Call Sequences.

	Wave 2	Wave 3	Wave 4
Model Specification	2	8	14
Appointments			
Broke Appointments	1	1	1
Kept Appointments	1.370***	1.449***	1.353**
Did Not Make Appointments	1.399***	1.446***	1.482***
Repeat unproductive contacts	0.918	0.765*	0.811
Above median % of no replies	0.792***	0.764***	0.761***
... remaining coefficients suppressed to ease presentation ...			
Constant	4.354***	9.094***	21.58***
Random Intercept	1.163***	1.100*	1.108
Observations	24104	22038	19699
Log Likelihood	-4817.2	-3555.6	-2889.0
Degrees Of Freedom	48	48	48
AIC	9734.4	7211.2	5878.0
*Estimates as odds ratios. * p<0.05, ** p<0.01, *** p<0.001*			

Estimated coefficients for Wave 2, 3 and 4 contact propensities conditioned on call sequences (specifications 2, 8 and 14). Random-intercept, logistic regression models were specified to account for unobserved interviewer effects. For presentation purposes, only call record variables have been included in this summarized table. For the complete tables with all estimated coefficients specified in the models, refer to Appendix A.

As shown in Table 1.4, an increased proportion of Wave 1 no replies is associated with reduced contactability in Waves 2, 3 and 4. While it is to be expected that non-responding households will likely report increased number of "no reply" calls within a given wave, these data show increased no reply calls continue to have an effect in future waves. It bears repeating that all households considered for this analysis are successfully interviewed in Wave 1. In other words, among a certain segment of Wave 1 households, increased no replies does not immediately result in Wave 1 nonresponse; instead their effect is delayed and observed as Wave 2, Wave 3 or Wave 4 noncontact. While initially contactable (and cooperating) there seems to be an underlying difficulty of contact for these households which eventually results in nonresponse.

Similarly, the occurrence of broken appointments is indicative of future noncontact. Here too, the interviewer is able to engage with the respondent and eventually secure a completed interview in Wave 1. Nevertheless, before securing said interview, the data collection for that household is interrupted: a pre-arranged interview does not occur on the agreed upon date and has to be rescheduled. Even if these households later cooperate, their tendency to break appointments is seemingly associated with a difficulty of contact in subsequent waves.

Table 1.5: Contact Models for Waves 2, 3 and 4. First Call Status.

	Wave 2	Wave 3	Wave 4
Model Specification	3	9	15
First Call Status			
Completed interview	1	1	1
No Reply	0.887	0.656*	0.981
Unproductive Contact	1.001	0.654*	0.791
Appointment made	1.060	0.795	1.046
Some interviewing done	1.272	0.923	1.880
Any other status	0.989	0.612	0.923
... remaining coefficients suppressed to ease presentation ...			
Constant	5.237***	14.18***	24.92***
Random Intercept	1.161***	1.110*	1.098
Observations	24104	22038	19699
Log Likelihood	-4831.3	-3568.7	-2896.6
Degrees Of Freedom	49	49	49
AIC	9764.6	7239.4	5895.8
*Estimates as odds ratios. * p<0.05, ** p<0.01, *** p<0.001*			

Estimated coefficients for Wave 2, 3 and 4 contact propensities conditioned on first call status (specifications 3, 9 and 15). Random-intercept, logistic regression models were specified to account for unobserved interviewer effects. For presentation purposes, only call record variables have been included in this summarized table. For the complete tables with all estimated coefficients specified in the models, refer to Appendix A.

As reported in Table 1.5, the status of the first call at Wave 1 is considerably (yet barely significantly) associated with reduced contactability in Wave 3 alone (first time no replies and unproductive contacts are less likely to be contacted). Similarly, repeated unproductive calls only correlates with reduced contact in Wave 3, but not for Wave 2 or 4 (see Table 1.4).

One could speculate that broken appointments and/or increased proportion of "no replies" signal limited at-home-routines. Even if the interviewer is able to complete an interview with a hard-to-reach household in the first wave, later its underlying difficulty results in future nonresponse. Additionally, broken appointments may also be an indication of less stable (or unpredictable) lifestyle routines of those households less capable of keeping arranged engagements. Otherwise, one could theorize broken appointments and "no replies" as early signs of soft refusals. While unable to say no in Wave 1, these otherwise latent nonrespondents are perhaps more comfortable with not answering the door and showing lack of interest to the survey by Waves 2, 3 or 4.

Obviously, these theories remain to be validated. Nevertheless, UKHLS data show that contact propensities are associated with previous wave call dynamics. The effects reported here (broken appointments and above median proportion of "no replies") behave in expected ways - it is sensible to assume that people who cannot be contacted are likely to have busy schedules or purposely avoid interviewer calls (whether cold or by appointment). What is more remarkable, however, is that these effects replicate not just after one wave but even after two and three waves from the initial round of data collection. This suggests that hard-to-contact patterns observed in the call records hold constant across time. Said differently, a household that is hard to reach once will likely be hard to reach again (and again, and again).

Cooperation

Table 1.6: Cooperation Models for Waves 2, 3 and 4. Call Sequences.

	Wave 2	Wave 3	Wave 4
Model Specification	5	11	17
Appointments			
Broke Appointments	1	1	1
Kept Appointments	1.210**	1.170**	1.254***
Did Not Make Appointments	1.234**	1.171*	1.284***
Repeat unproductive contacts	0.766***	0.725***	0.746***
Above median % of no replies	1.034	1.079	0.961
... remaining coefficients suppressed to ease presentation ...			
Constant	2.837***	6.443***	6.085***
Random Intercept	1.123***	1.083***	1.040*
Observations	22684	21083	18947
Log Likelihood	-7967.1	-8756.5	-6762.9
Degrees Of Freedom	48	48	48
AIC	16034.1	17612.9	13625.8
*Estimates as odds ratios. * p<0.05, ** p<0.01, *** p<0.001*			

Estimated coefficients for Wave 2, 3 and 4 cooperation propensities conditioned on call sequences (specifications 5, 11 and 17). Random-intercept, logistic regression models were specified to account for unobserved interviewer effects. For presentation purposes, only call record variables have been included in this summarized table. For the complete tables with all estimated coefficients specified in the models, refer to Appendix A.

Call data are also significantly and sizeably associated with future household cooperation (see Table 1.6). Broken appointments at Wave 1 are indicative of reduced likelihood of future cooperation. Repeated unproductive contacts are also associated with decreased cooperation in subsequent waves. However, a household's proportion of "no replies" does not show any significant or sizeable association. Here, at-home-routines are no longer tenable assumptions since cooperation is conditional on initial contact. Instead, prior respondent-interviewer interactions likely determine the propensity to successfully complete the interview. Therefore, one could propose that these Wave 1 broken appointments and/or repeated unproductive calls are associated with the household's unwillingness to cooperate and not necessarily with its contactability.

Table 1.7: Cooperation Models for Waves 2, 3 and 4. First Call Status.

	Wave 2	Wave 3	Wave 4
Model Specification	6	12	18
First Call Status			
Completed interview	1	1	1
No Reply	0.888	1.013	1.002
Unproductive Contact	0.676***	0.764**	0.802*
Appointment made	0.842	0.960	0.978
Some interviewing done	1.168	1.830**	1.276
Any other status	0.716*	0.817	1.098
... remaining coefficients suppressed to ease presentation ...			
Constant	4.049***	8.089***	7.494***
Random Intercept	1.122***	1.083***	1.039
Observations	22684	21083	18947
Log Likelihood	-7965.3	-8753.7	-6769.6
Degrees Of Freedom	49	49	49
AIC	16032.6	17609.4	13641.1
*Estimates as odds ratios. * p<0.05, ** p<0.01, *** p<0.001*			

Estimated coefficients for Wave 2, 3 and 4 cooperation propensities conditioned on first call status (specifications 6, 12 and 18). Random-intercept, logistic regression models were specified to account for unobserved interviewer effects. For presentation purposes, only call record variables have been included in this summarized table. For the complete tables with all estimated coefficients specified in the models, refer to Appendix A.

Furthermore, if the status of the first call in Wave 1 is an unproductive contact (i.e. the interviewer engages the respondent but is not able to agree on a future appointment or begin an interview), it is more likely that by Waves 2, 3 or 4 that household will not cooperate with the survey. Similarly, if the first call in Wave 1 is a partial interview, in Wave 3 (and only Wave 3) that household is considerably more likely to cooperate once contacted. Thus, not only is the first wave very important for continued household cooperation, but indeed the very first call with the sampled household.

Here too, the significance and direction of the covariates are consistent with expectations. Moreover, Wave 1 call dynamics are associated not just with the cooperation propensities of the wave immediately after but also for those of Wave 3 and 4.

1.4.2 Do these Wave 1 call sequences reveal additional information about future contact and cooperation unobserved by conventional demographic and attitudinal predictors of nonresponse?

If call records are able to account for future contact and cooperation, can their effects also account for the unexplained variance of response models based on demographic and attitudinal covariates? In other words, do call sequences reveal information about the processes of nonresponse that is not captured by factors such as household's composition, geographical location, family composition, employment status, social attitudes of its residents, or the characteristics of the dwelling?

Table 1.8: Likelihood Ratio Tests

Wave	Outcome	Call	χ^2	d.f.	Prob. > χ^2
2	Contact	Sequence vs. Reduced	36.35	4	0
		First call vs. Reduced	8.12	5	0.1500
	Cooperation	Sequence vs. Reduced	28.86	4	0
		First call vs. Reduced	32.33	5	0
3	Contact	Sequence vs. Reduced	38.68	4	0
		First call vs. Reduced	12.50	5	0.0285
	Cooperation	Sequence vs. Reduced	38.44	4	0
		First call vs. Reduced	43.97	5	0
4	Contact	Sequence vs. Reduced	27.24	4	0
		First call vs. Reduced	11.42	5	0.0437
	Cooperation	Sequence vs. Reduced	29.52	4	0
		First call vs. Reduced	16.24	5	0.0062

To determine whether call records reveal additional information about the contact and cooperation propensities, the 18 models considered for this analysis are grouped in trios of expanded and reduced forms per specification. For every expanded model (i.e. call records + other nonresponse covariates + cross-wave covariates) there exists a corresponding reduced form (i.e. other nonresponse covariates + cross-wave covariates). The comparison allows to test the marginal effects of the call records after conditioning on 1) other covariates of nonresponse as well as 2) any possible cross-wave interviewer allocation dynamics or geographical continuity effects particular to UKHLS that might also account for future nonresponse. As Table 1.8 shows all expanded models (except for the Wave 2 contact model conditioned on first call) are significantly improved when incorporating call record data for Waves 2, 3 and 4. In other words, the call data do account for additional variation that is left unexplained by the other covariates of nonresponse as well as the cross-wave controls.

Control variables: other nonresponse covariates & cross-wave controls

Nevertheless, while call data report sizeable and significant covariates, the same occurs for some of the control variables included in the analysis. Most of these effects are partially consistent with expectations and comparable research.

The strongest associations with contact are found among those households where the same interviewer is kept and/or stayed in the same LSOA between waves. However, it should be stressed that interviewer allocation between waves is not a random process. Instead, this likely reflects the deliberate decision of field offices which in turn might be confounded with the response mechanisms. Indeed, interviewers are likely to be switched between waves in those cases where 1) the household moves, 2) the household is particularly difficult to contact and/or to cooperate with the survey, or 3) the interviewer ceases to work for the survey or the particular area where the household is found. Therefore, when interviewers are changed, this likely reflects a decision by the field office to address a potentially nonresponding household. In other words, interpreting interviewer effects on nonresponse is complicated when one considers the potential circularity of association between interviewer allocation and household response mechanisms.

With regards to the remaining control covariates of nonresponse, few significant associations are found. Certain dwelling types are associated with reduced contact. Pensioner households are also more contactable but less inclined to cooperate. Home ownership is likely to increase contactability, while lack of political interest decreases it. Lack of political interest and suspicion of the previous survey wave are associated with reduced cooperation, while consent to linkage indicates increased cooperation.

In short, while the call events display consistent effects for both Waves 2, 3 and 4 the same cannot be said for the control covariates of nonresponse. In fact, the magnitude and significance (and to a lesser extent the direction) of some of the coefficients (including: material deprivation, national origin, community attachment, self-reported health status, understanding of the questionnaire, and presence of a baby) are altered between waves. Still, for most of the remaining controls no apparent effect is ever observed. Thus, call data account not just for additional significant explanatory power in the models of nonresponse, but indeed report more stable effects across the three waves when compared to the control covariates. Perhaps, this cross-wave consistency is due to the conceptual proximity of the call dynamics to the contact and cooperation processes.

1.5 Discussion

Given the relative magnitude of the estimated coefficients of some of these call covariates, the significant contribution of these data to future nonresponse is potentially applicable for fieldwork design, especially in longitudinal surveys.

More specifically, these findings suggest that households which had repeated unproductive contacts, broke appointments, registered above median proportion of "no replies", or began the call sequence with an unproductive contact are at risk of future nonresponse. The risk is consistent for Waves 2, 3 and 4. This risk is not trivial if one considers the frequency of occurrence of these types of call sequences. Indeed, repeated unproductive contacts occur in 9.5% of the responding households of Wave 1. Similarly, in 13% of these same households broken appointments are observed and for 15.5% the first call is an unproductive contact. Obviously, the risk is magnified when one considers the multiplicative effect of these events (for example, in 2% of Wave 1 responding cases, households whose first call was an unproductive contact would later break an appointment).

If the risks are apparent, so too are the potential applications for nonresponse prevention. Based on their call behaviour on the first wave of *Understanding Society*, households could be grouped by propensity for noncontact and/or noncooperation and exposed to tailored treatments which address theoretical causes of the events observed in the contact data. For example, in households where appointments were broken, a between-wave mailing that acknowledges this occurrence could be drafted and include suggestions for more flexible calling times in the future. Of course, this last approach may be more effective for those cases where a broken appointment was not a proxy for a soft refusal but rather an unexpected change of scheduling plans. Alternatively, the interviewers who registered these broken appointments could be consulted to explore possible drivers of noncontact or noncooperation in these types of households that are not immediately obvious in the data. For future longitudinal household surveys, interviewer training should address the importance of first call interactions and the risk of unproductive contacts while also considering the risk of overly "pushing" reluctant Wave 1 respondents (who may have a different threshold for refusal than those of proceeding waves).

Interestingly, broken appointments are the only call sequence that account for both nonresponse processes. Further research could explore possible drivers of these by addressing the likely different mechanisms of refusing vs. not being contactable for an interview on a date previously agreed upon.

Future research could also address some of the previously mentioned limitations of this analysis. Indeed, the nonrandom allocation of interviewers between waves conditions any discussion of potential application of these findings. Efforts to disentangle the unobserved correlates of nonresponse inherent in the cross-wave interviewer allocation procedures could include randomized experiments where some interviewers are encouraged to incorporate response inducement strategies in their households based on the call record data from previous waves. Replication of this analysis in comparable household longitudinal studies could also serve to validate its findings and potentially resolve this limitation. Should similar findings be found in these comparable studies, new insights on the confounding effect of allocating interviewers nonrandomly between waves could be found.

Additionally, based on the structure of the data, future analysis could explore cross-level interactions between the interviewer clusters and some of the call record covariates. For example, are some interviewers more likely to incur in broken appointments or repeated unproductive contacts? Can new all-purpose fieldwork protocols be designed from the findings of these analysis or do interviewers demand tailored training based on the performance of the call data? These could be explored by including not just random intercepts, but also random slopes to the models already discussed here.

Furthermore, that these data are non-experimental also qualifies the findings discussed previously. The analysis presented here is understood within the context of UKHLS data. Surely, some of the findings suggest effects in expected and reasonable directions. Nevertheless, it bears repeating that since these are observational data, the analysis and findings are not (yet) generalizable.

The analysis presented here focused on the household level given: 1) the structure of the data (call records nested in the household) and 2) an interest in exploring household-level dynamics of contact as well as cooperation with a view towards optimizing possible nonresponse prevention strategies. Nevertheless, individual-level dynamics were ignored (or at best aggregated to the household level) in these analysis. Future work could address this limitation. In particular, the possible effect of individual respondents' attitudinal, behavioural and contextual predispositions on cooperation propensity and how they may correlate with patterns observed in the corresponding call record.

Similarly, while this analysis focused on nonresponse prevention, there is no discussion of call records' impact on nonresponse bias. A first step towards addressing this limitation would be exploring what other differences (if any) exist between the groups identified by the call records. In other words, are those households that break appointments systematically different from those that do not when it comes to basic demographic composition or answers given to particular questions of the survey? Furthermore, do these call record patterns remain constant across the waves within households? In other words, are there "repeat offenders" of problematic call sequences? If so, what possible impact may this have on nonresponse bias and data quality of their responses in general?

Because the primary objective of the models employed in this analysis was to further understand the processes of longitudinal contact and cooperation (and not to predict these outcomes), this chapter did not discuss their predictive ability or other forms of model assessment. Nonetheless, future iterations of this study could benefit from a discussion of model assessment including the use of measures of model fitness as well as classification tables. For further discussion on different types of model assessment in the analysis of call record data see Durrant et al. (2015).

Finally, given the reduction of explanatory power of the models between Waves 2, 3 and 4 the relevance of these findings should be explored further by incorporating data from Waves 5 and onwards as they become available.

Chapter 2

What's taking so long? Field effort dynamics in a longitudinal survey

Abstract: For longitudinal face-to-face surveys one of the components of overall survey costs is the number of call attempts made by the interviewer wishing to locate, make contact and gain the cooperation of the issued sample member. In terms of overall costs, lengthy call sequences necessarily require more resources than shorter ones. Given the need to reduce costs and economize survey resources, understanding what drives number of call attempts is an essential first step towards any attempt to optimize call resources and thus drive down overall costs. Surprisingly, there is very little in the literature of longitudinal, face-to-face surveys concerning drivers of field effort. This chapter will analyse field data (including call records, geographical markers and interviewer observations) as well as household and aggregated individual data from the first four waves of a household longitudinal survey to evaluate their influence on field effort. The analysis will be followed by a discussion of possible design strategies to be employed for curbing this component of survey costs as well as a more general discussion regarding changing field effort dynamics in longitudinal context.

Keywords: Field effort, face-to-face longitudinal surveys, call records, paradata, fieldwork management

2.1 Introduction

A defining feature of face-to-face, longitudinal surveys is the repeated interviewing of sample members by dedicated field staff. Typically, once a respondent cooperates with a baseline wave of a longitudinal study, the survey agency will make the effort to re-interview them in the following waves of the study. To achieve this, interviewers and field staff are commissioned to locate, contact and gain the cooperation of all eligible respondents from a previous wave. When this effort is not successful survey attrition occurs. Because attrition is often systematic (i.e. attriting respondents would have responded differently than those that remain in the survey) and given the loss of statistical power, careful consideration is placed in the design and data collection procedures of longitudinal surveys (Lynn, 2009a). Thus, the success of such types of surveys is a function of the extent to which respondents from a previous wave remain locatable, contactable and cooperative in future waves.

Given this necessity, these surveys are costly, time-consuming and expected to withstand several waves of data collection (often measured in years). One of the components of overall costs in longitudinal, face-to-face surveys is the number of call attempts made by the interviewer wishing to make contact and/or gain the cooperation of the issued sample member. Each call to a potential respondent contributes to the survey's total budget, primarily in terms of interviewer compensation (including hourly/daily rates and travel expenses). Furthermore, each call is associated with additional administrative costs (like continued field office support, supervision, quality control and communication with the respondent). In terms of overall costs, lengthy call sequences necessarily require more resources than shorter ones.

Surprisingly, there is very little in the literature of longitudinal, face-to-face surveys concerning drivers of field effort. The sole exception found in the literature is Durrant et al. (2017). Most studies concerned with longitudinal call records (or longitudinal survey nonresponse) focus instead on the ultimate response outcomes at a given wave, or the prevalence of different forms of attrition patterns, or the biasing effects of these differential response propensities (Lugtig, 2014; Watson and Wooden, 2009). The studies that do analyse field effort tend to rely on cross-sectional data (Hall et al., 2013; Lynn and Clarke, 2002; Lynn et al., 2002).

Because of the need to reduce costs and economize survey resources, understanding what drives number of call attempts is an essential first step towards any attempt to optimize call resources and thus drive down overall costs. Moreover, estimating number of call attempts has immediate applications for data collection design in face-to-face, longitudinal surveys. For instance, prior to securing funding for a given survey, it allows the survey designer to plan and budget according to expected fieldwork performance. Once costs have been approved, it also facilitates the allocation of field resources and allows for a more optimal distribution of efforts conditional on locatability, contactability and cooperation propensities of targeted sample members. Furthermore, to the extent that field effort is attributed to particular factors (for example: demographic composition of the household, attitudes of its residents, survey design, interviewer characteristics and training protocols, previous interview experience, and/or the social environment of the interview), specific design strategies can be drafted and implemented to economize overall data collection resources. Data from longitudinal surveys are uniquely poised for evaluating the role of respondent characteristics, survey design, survey-taking context and interviewer effects in differential field effort given the ready availability of questionnaire and auxiliary information collected in previous waves. Finally, when evaluated in longitudinal context, the possibility that these effects change given their interaction with time and the continued relationship between the respondent and the survey can also be analysed and used to inform data collection policies. At the very least, it allows for an examination of theories regarding determinants of extended calling patterns in cross-sectional surveys and their extension into longitudinal studies.

2.1.1 Research Questions and Objectives

This chapter will analyse field data - including call records (CR), geographical markers and interviewer observations - as well as household and aggregated individual data from the first four waves of the UK's Household Longitudinal Study (also known as *Understanding Society*), to evaluate their influence on field effort and derive associated estimates inherent in the processes of contact and cooperation. More specifically, the analysis presented in this chapter will quantify the number of calls required for sampled households conditional on aggregated individual respondent characteristics, household-level traits, interviewer continuity, geographical controls, previous-wave call record sequences and household response outcomes. This analysis is primarily interested in answering the following questions: what household and/or individual respondent characteristics are associated with increased/diminished calling effort? Do lagged call record patterns correlate with a household's number of calls towards contact and/or cooperation in future waves? Finally, and perhaps most importantly, do these hypothetical correlations change over waves? If so, are these changes due to self-selection of non-attriters or can they be explained as a longitudinal learning effect? The analysis will be followed by a discussion of possible design strategies to be employed for curbing this component of survey costs as well as a more general discussion regarding changing field effort dynamics in longitudinal context.

The rest of this chapter will be divided into four sections. Section 2.2 will discuss the determinants of field effort within the theoretical framework of survey participation (Couper and Groves, 1996; Groves and Couper, 1998) and nonresponse in subsequent waves of longitudinal household surveys (Lepkowski and Couper, 2002). Section 2.3 will describe the data used in this analysis. It will discuss the variables of interest, the structuring of four household-level datasets with information across the first four waves of *Understanding Society* and the modelling strategy adopted for the analysis. Section 2.4 will follow with a discussion of the results structured around the research questions mentioned previously. Finally, section 2.5 will expand on the implications of the findings for longitudinal surveys and fieldwork effort, as well as limitations of the analysis and opportunities for future research.

2.2 Theory and concepts

2.2.1 Deriving indicators of field effort from call record data

In face-to-face household surveys interviewers visit potential respondents in their place of residence. Travelling to the household is done with the ultimate objective of securing the full cooperation of the targeted sample member in the form of a completed survey interview. Of course, before a request for interview can be made to a potential respondent the interviewer must successfully contact the correct eligible sample member. Additionally, in longitudinal surveys, survey staff may also have to spend time reaching out to households not only to secure contact but also to correctly locate or track respondents from previous waves who may have moved address. Thus, interviewer field effort, as measured in number of interviewer visits (or calls) to sample addresses is a function of the processes inherent in securing a response: location, contact and cooperation.

CR data, also called "call history data", "contact data", and "field process data", are defined as information recorded at each contact attempt (or "call") with a potential survey respondent about the call outcome (i.e. "refusal", "successful interview", "non-contact", etc.) as well as time, date and duration of call (Kreuter and Olson, 2013). Thus, CR data constitute a detailed history of all the attempts to secure an interview from all target sample members of a given survey. Recently there has been a rise in the practice of collecting and systematizing this type of data by survey practitioners. Similarly, in survey methods research there has been a growing interest in the analysis of CR data (Bates et al., 2010; Blom, 2009; Blom et al., 2010; Calinescu et al., 2011; De Keulenaer, 2005; Durrant et al., 2013a, 2011; Henly and Bates, 2006; Kreuter and Kohler, 2009).

Generally, CR literature can be grouped in either of two concerns: 1) using CR data to analyse and/or adjust for biases in survey estimates, primarily those resulting from unit nonresponse (Kreuter and Kohler, 2009; Kreuter and Olson, 2013; Lynn et al., 2002) or 2) analysing CR data for fieldwork / data collection design, especially in face-to-face surveys (Durrant et al., 2011, 2013b; Kreuter et al., 2010; Mercer, 2012; Wagner, 2013). This chapter hopes to expand on the latter of these concerns.

There exists a considerable literature on the use of CRs for the evaluation and design of survey field strategies. A common thread in this literature is how to use CR data to inform the optimal distribution of available field and data collection resources given a finite budget. A related discussion is how these data serve to evaluate interviewer performance and doorstep interactions. Similarly, these data feature prominently in recent discussions regarding "Responsive" and "Adaptive Designs" (Groves and Heeringa, 2006). For example, there are works on: identifying best times of contact (Durrant et al., 2011, 2013b, 2017), fieldwork "resource allocation" (Calinescu et al., 2011), "level of effort" analysis & general process control (Kreuter et al., 2010). The objective of this analysis is to estimate number of calls observed at the household level when attempting to locate, contact and gain the cooperation for their continued participation in a longitudinal survey.

2.2.2 Field effort and location propensity

Given the possibility for individuals and/or households to move location between waves, most longitudinal surveys spend additional field resources to track and locate target sample members that may have changed physical address. Some studies, including *Understanding Society*, have features built into their design meant to identify potential movers including asking respondents their probability of moving in the future and intra-wave mailer requests for address confirmation / change of address notification (Knies, 2015; Lynn, 2009a). In such cases, the new address is recorded and used as the new point of contact in the subsequent waves. Nevertheless, some sample members may change address between waves in ways not anticipated by the field office and therefore require additional effort to trace and locate once fieldwork for a particular wave has started (Couper and Ofstedal, 2009).

Propensity to move has been found to be associated with life events (like births, deaths, marital splits, employment and/or crime victimization episodes), as well as age, socioeconomic status, employment status, education level, and neighbourhood satisfaction (Clark, 2013; Coulter and Scott, 2015; Rabe and Taylor, 2010). Furthermore, survey design features may also increase the proportion of sample members that move between waves. The longer the survey lasts will automatically increase the opportunities for its targeted respondents to move. Similarly, a larger intra-wave interval is likely to be associated with increased moving propensities at the sample level (Couper and Ofstedal, 2009; Groves and Couper, 1998).

2.2.3 Field effort and contact propensity

Contact propensity has been associated with at-home routines. More specifically, employment and working long hours has been linked to a reduced probability of contact, as well as single-occupancy residences, and age of residents. Similarly, those households with active and/or independent lifestyles which offer many opportunities for out-of-residence activities are naturally going to be less contactable while those households with set domestic routines (i.e. families with children, pensioners) will be easier to contact. Independent of their lifestyle or routines, the respondents' willingness to be found is another relevant factor in estimating propensity for contact. Naturally, the higher the number of people in a sample address, the more likely it is to find someone at home (and thus make contact with) who is eligible to be interviewed in a household survey. Additionally, accessibility to the household is also related to its contactability. Gated residences, buzzers, locked common entrances and other such physical barriers make the task of contact more difficult. Finally, survey design features and interviewer protocols also affect contact propensity: by increasing the number of call attempts at a targeted member the likelihood of contact should also increase. Similarly, varying the time of day and day of the week of call attempts also increases contact propensity given the varied patterns of at-home-routines of sample households (Groves and Couper, 1998; Kalsbeek et al., 2002; Stoop, 2005; Uhrig, 2008). Most surveys, including *Understanding Society* (Knies, 2015), are aware of the role of survey design in ensuring contact and therefore institute minimum number of calls at different times of day and days of the week before deeming a household as "noncontact"[1] (Morton-

[1] Additionally, interviewers may vary in their ability / willingness to work evenings and weekends. Thus, when specifying contact / cooperation models it is common to include interviewer-level random effects (as is done in the analyses for this chapter).

Williams, 1993, p. 54–56).

Thus, conditional on survey design features, some households are going to be easier to contact than others. This relative contactability should be reflected in the call record data. In fact, several authors have proposed "number of calls until first contact" as a proxy for measuring how easy/hard (read: cheap/expensive) it is to contact a given sample member (Hall et al., 2013; Lynn et al., 2002; Mercer, 2012; Stoop, 2004). Once a wave starts, the higher the number of calls until contact the higher the effort required for contacting/locating a target sample member and thus the less contactable they are.

2.2.4 Field effort and cooperation propensity

If contact is largely attributed to socio-demographic and household traits of the sample respondent related to their stay-at-home patterns and accessibility as well as particular survey design features, cooperation is a function of latent social-psychological dispositions that express themselves at the time of the interviewer-interviewee interaction and which are in turn conditioned by individual- as well as household- and ecological-level factors. Once contact has been established, the decision to cooperate or refuse lies primarily with the respondent and can only be influenced to the extent that the interviewer is able to tailor or adapt his/her approach with the sample member and allay any possible drivers or concerns that might push them towards noncooperation. There exists considerable literature on strategies and interviewing training protocols aimed at properly identifying respondent's concerns and adapting / tailoring the interviewing approach to guide the interaction towards a successful completion of the survey interview (Dijkstra and Smit, 2002; Groves and Couper, 1998; Morton-Williams, 1993).

Household and/or individual respondent characteristics have been linked with an underlying propensity for differential cooperation. Often these characteristics and their effects are explained within the frameworks of one of several socio-psychological theories, including: social exchange theory, rational choice, social isolation, and legitimacy / authority (of survey agency) (Groves and Couper, 1998). Briefly, cooperating with a survey can be partially explained as a function of the respondent's: 1) perceived utility in answering the questionnaire, including: being offered a material incentive, or simply the (scarce) opportunity of having "their voice heard"; 2) relative opportunity costs, especially with regard to time and cognitive resources implicit in cooperating with the request (i.e. perception of how busy one may be vs. willingness to cooperate, survey burden); 3) the obligation felt towards a general social good achieved by collaborating with the survey and/or towards the institution behind the request; 4) interest in being consistent with peers who might choose to respond; 5) general interest in the topic (how salient the question items are to the respondent); or 6) affinity or rapport with the interviewer.

It is important to note that while conceptually separate, cooperation and contactability are also difficult to disentangle (Nicoletti and Peracchi, 2005). Cooperation can only occur once a sample target member has been contacted, and it is often stated that the underlying propensity for cooperation / reluctance might very well exist prior to and independent of the event of contact. However, cooperation and contactability may also be driven by similar factors. For example, the socially recluse or the very busy professionals are likely to be less reachable and simultaneously less prone to cooperate. Similarly, while a target sample member uninclined to be found is likely to be recorded as non-contacted, their nonresponse may in fact correspond to reluctance and not necessarily accessibility or stay-at-home routines. Further complicating matters, when achieving full cooperation from a target sample member (especially multi-occupant households), an interviewer may require multiple visits or calls. Thus, even after an initially positive contact, the household's likelihood for successfully completing the questionnaire remains a function not just of its willingness to cooperate once requested but of its continued contactability across the number of calls required by the interviewer. In other words, estimating the level of effort required before achieving a complete interview by exploring the corresponding call records may confound contact as well as cooperation.

2.2.5 Field effort in longitudinal context

Regarding cross-sectional surveys, several authors agree on the importance of initial interviewer-interviewee interactions for future contact and cooperation propensities (Campanelli et al., 1997b; Couper, 1997; Groves and Couper, 1996; Nederhof, 1987;

Pickery et al., 2001). The importance is understood given that following initial contact the survey agency has a lot of information about the respondent. This allows for a better estimation of future response propensities, including ease of contact and proneness for cooperation and thus expected number of calls required. This knowledge comes from the direct experience gained on contact and the possible opinions the respondent might have expressed towards the survey, on the information collected (including that from the questionnaire or interviewer observations such as known socio-demographic and attitudinal correlates of nonresponse), the social context where contact takes place (i.e. the sample member's neighbourhood), the stay-at-home patterns revealed as well as the overall effort inherent in visiting and interacting with the respondent.

Likewise, upon first contact, sample members gain information about the survey agency, the nature of the survey request, the profile of the interviewer, and likely the topic and length of the questionnaire. Indeed, beyond conventional covariates of response (like demographic traits, given social attitudes and/or household composition characteristics), the respondent's evaluation of the first interactions with the interviewer is going to largely determine (or at least be strongly associated with) whether the interviewee continues to collaborate with the survey and the relative ease, in terms of number of calls, that collaboration will require. For example, if a potential respondent expresses negative statements to the interviewer they are much more likely to non-respond at a later call or at least display some reluctance to cooperate or be found. Similarly, while a reclusive respondent living alone may be less likely to answer a door to a stranger for the first time; once contacted, the respondent's traits may become less associated with response propensity in subsequent visits and indeed become a cooperative respondent. In other words, a positive initial interaction is going to be indicative of increased willingness to respond while a negative one is going to be indicative of reduced response likelihood.

The phenomenon extends (and perhaps is magnified) in longitudinal surveys. Lepkowski and Couper (2002) explain that upon completion of the first wave of a longitudinal survey, the interviewer-interviewee interaction occurs in a context where both parties know considerably more about each other. Furthermore, the authors speculate that with each passing wave respondents develop more trust with the survey as well as familiarity with the questionnaire. Thus, correlates of contact and cooperation observed in the initial wave of a longitudinal survey may be moderated in later waves by the relationship developed between the interviewer and interviewee as well as the increased amount of information each has about the other. Additionally, the dynamics of field effort may change in longitudinal context. For example, later waves usually require a lower number of attempts to contact (Lepkowski and Couper, 2002, p. 262). This reduction in calls needed may answer to two distinct processes: 1) longitudinal learning (i.e. interviewers and interviewees become more efficient by continued experience with the survey throughout its waves) or 2) self-selection (i.e. respondents, and possibly interviewers, that require extended calling efforts are more likely to drop out of the survey).

Lastly, certain household- and individual-level traits; survey design and interviewer features; as well as contextual characteristics remain associated to differential field effort and/or response propensity across waves. Indeed, Bates (2004); De Keulenaer (2005); Henly and Bates (2006); Watson and Wooden (2009) find number of calls made at a previous wave to be negatively associated with contact and cooperation propensities in subsequent waves. Thus, it is possible to assume that indicators of field effort can be estimated based on lagged information collected not just from the survey questionnaire, but also from corresponding field data (including interviewer observations and CRs).

2.3 Data and methods

For this analysis, data come from the first four waves of *Understanding Society*, the UK's Longitudinal Household Study. A detailed description of the survey, its design and sample components can be found in Knies (2015).

Four separate datasets will be considered: 1) Wave 1 households 2) Wave 2 households, 3) Wave 3 households and 4) Wave 4 households. Because this analysis is primarily interested in estimating and comparing number of calls based on data collected across all available waves, the first dataset, of Wave 1 households, only includes cases which responded in Wave 1 because nonresponding Wave 1 households were not issued in future waves. The remaining three datasets will progressively incorporate information from the previous waves. In other words, the second dataset, corresponding to households issued for Wave 2, will incorporate information from Wave 1 call records (as well as other household traits from Wave 1), to estimate Wave 2 field effort. The third dataset, of households issued at Wave 3, will include Wave 1 and 2 questionnaire information as well as Wave 1 and Wave 2 CR data to estimate Wave 3 outcomes. The fourth dataset, of households issued at Wave 4, will be the richest analytical sample as it will attempt to model Wave 4 outcomes with information from their corresponding Waves 1, 2, and 3 questionnaires as well as CRs from Waves 1 to 3.

The selection of households to be analysed is based on the sampling protocols of UKHLS. Only those households which contain at least one fully participating respondent from Wave 1 are reissued in future waves. Thus, the Wave 1 dataset only includes households which at the very least include one individual respondent. Because of this condition there is Wave 1 household-level information available for all households in all four analytical samples sourced from the household grid and corresponding individual respondent questionnaire. For waves 2 through 4, all issued households (including responding and non-responding) are considered. Additional questionnaire information is used to feed and/or update information to that collected at Wave 1. Data from the individual- and CR-levels are aggregated at the household level using summary statistics. In case of item nonresponse, most recent previous wave information is used. Further information on data reduction and variable derivation is provided in section 2.3.1.

In total, there are 24638 responding households at Wave 1, 25260 households issued at Wave 2, 22858 households issued at Wave 3 and 20127 households issued at Wave 4. Only the General Population component of Understanding Society (corresponding to England, Scotland and Wales) is used in this analysis as the other subsamples (namely: the Ethnic Minority Boost, the BHPS continuation sample and Northern Ireland) do not have comparable doorstep protocols, are not conducted in the same mode and/or have no available call record data. Minor data management limitations (primarily concerned with linkable information across the waves from eligible households and/or individuals as well as missing data from the original source files) results in ignorable exclusions of a small set of households from the final datasets (accounting for less than 0.65% of original households from waves 1 to Wave 4).

Like many other household longitudinal surveys, *Understanding Society* does not have a constant household identifier across waves. Because of between-wave household alterations (largely due to life events like marital splits, deaths, moves, and new births), individual respondents (and not their corresponding households) are the units tracked by the survey. Thus, at each wave every household is given a new and wave-specific identification number. Unlike households, individual respondents have unique identifiers that remain consistent across waves. For the purposes of this analysis, households are linked across waves through corresponding individuals[2].

2.3.1 Variable selection

Dependent variables

This analysis aims to identify household traits, aggregated individual information, contextual characteristics, and survey design features associated with differences in number of calls inherent in the location, contact and cooperation processes of household response. More specifically, *Total number of calls* is equivalent to the number of calls required by a target sample member at any given wave, while *calls to make contact* indicate ease of contact. Similarly, number of *post-contact calls to completion* of an interview likely captures the level of effort required from a potential respondent household for its cooperation. Of course, this last indicator likely captures continued contactability as the interviewer needs to be able to engage with the respondent at the household, before securing an interview. However, this analysis assumes that following initial contact, the interviewer and interviewee are likely to share information about the respondent's availability and at-home-routines so that future visits have a higher chance of resulting in a successful contact.

[2]In the rare event of a household containing individuals living in different sample households in the preceding wave, the link is formed with that prior-wave household with a larger number of corresponding residents at the current wave. For example, if Homer, Marge, Bart, Lisa and Maggie live at 742 Evergreen Terrace in Waves 2 and 3 of the survey except for Homer who resided at the Bachelor Arms during Wave 2, then 742 Evergreen Terrace at Wave 3 will be linked with 742 Evergreen Terrace at Wave 2 and not the Bachelor Arms

Therefore, calls to make contact will be analysed from a base of all households successfully contacted at a given wave (except for Wave 1 households, where only responding cases are considered since these are the ones that are tracked in future waves of *Understanding Society*). Further, post-contact calls to cooperation includes only those cases that effectively respond at a given wave (in other words, post-contact calls of ultimately nonresponding households are not included in the estimates of this metric). Lastly, total number of calls is estimated from a base of all issued households from waves 2 to 4 and includes noncontacts, nonresponding and responding households (again, Wave 1 is an exception as only responding households are considered for this wave). In other words, the means of these dependent variables (and the coefficients later reported in the analysis) are not additive and correspond to different analytical bases.

Table 2.1: Dependent Variables

		Calls to make contact	Post-Contact Calls to Completion	Total number of calls
Wave 1	Mean (S.E.)	2.266 (0.013)	1.980 (0.014)	4.435 (0.019)
	Cases	*24638*	*24638*	*24638*
Wave 2	Mean (S.E.)	2.094 (0.012)	1.577 (0.013)	4.129 (0.019)
	Cases	*21929*	*19790*	*25260*
Wave 3	Mean (S.E.)	2.030 (0.012)	1.566 (0.013)	3.986 (0.020)
	Cases	*20045*	*17570*	*22858*
Wave 4	Mean (S.E.)	1.954 (0.012)	1.485 (0.014)	3.765 (0.021)
	Cases	*18228*	*16572*	*20127*

As Table 2.1 indicates, with each passing wave total number of calls as well as calls need for contact and cooperation progressively decrease. What remains to be explored is if these reductions are uniform across household types (characteristics of dwelling, household composition, attitudinal dynamics of residents and/or previous interview experience) and/or survey design features (i.e. interviewers). The following independent variables aim to capture potential differences in these differential field effort measures.

Independent variables

Because calling estimates are understood in the context of the conditionally independent processes inherent in securing a response in longitudinal surveys (i.e. location, contact and cooperation), oft-mentioned covariates of response (Couper and Groves, 1996; Groves and Couper, 1998; Lepkowski and Couper, 2002) can also be employed when identifying associated factors of field effort. Thus, the independent variables selected for the analysis are derived from the call records, geographical markers, characteristics of the dwelling, household size, demographics, levels of political & community engagement, previous interview experience, and cross-wave controls for interviewer and geographical continuity and aim to capture the drivers behind the mechanisms of response.

More specifically, the independent variables considered in these analyses can be separated into eight different groups, corresponding to eight different dimensions of association with the mechanisms of location, contact and cooperation:

1. **Geographical controls (Region and Urbanicity):** Geographical region and urbanicity are also included in this analysis to allow for response propensity differences across different socio-geographical contexts. Indeed, population density is commonly found to be negatively associated with contact and/or cooperation. This effect is usually explained as a function of social isolation or increased non-domestic routines typical of large cities (Watson and Wooden, 2009). Assuming these differential propensities entail differential interviewer efforts, geography and urbanicity should also partially explain differences in calling patterns. (See Table B.1 in the appendices).

2. **Dwelling characteristics and accessibility:** Based on dwelling characteristics, some households are easier to contact than others, requiring fewer calls. For example, buildings with physical entry barriers typically result in reduced contact. Furthermore, flats and above ground-floor properties are often more difficult to reach than (semi)detached homes. Lastly, dwelling traits are often related to the immediate geographical context in which they are located. Thus, interviewers may conduct their work differently based on their relative ease with the neighbourhood or local area where the target sample member is located. (See Table B.2 in the appendices).

3. **Sociodemographic composition of household:** The type of residents found

within a household is strongly associated with at-home-routines as well as socio-psychological profiles correlated with disposition towards survey cooperation. For example, families with babies are more likely to be found at home but perhaps be less willing to answer the survey. On the other hand, pensioner households may have the time and inclination to be found and cooperate with the interviewer. Lastly, working long hours means less opportunity to be found at home. (See Table B.3 in the appendices).

4. **Indicators of social inclusion and SES:** Socio-economic status and community attachment are often proxies for social inclusion and could therefore partially explain variability in the cooperation propensity of its residents. (See Table B.4 in the appendices).

5. **Interview experience:** As mentioned in section 2.2.5, the first contact between respondent and interviewer (or survey agency in general) is likely to inform all future call and wave interactions. Thus, interviewer assessment of previous interview experience is included as a dependent variable. Furthermore, item non-response as well as consent to data linkage are likely to correlate to cooperation propensity and thus could also be associated with differential number of calls towards survey cooperation. To account for previous wave effects, the covariates for the models for Waves 2, 3 and 4 correspond to interview experience observed in the previous wave (i.e. Waves 1, 2 and 3 correspondingly). For the Wave 1 models, interview experience correlates are those observed in the same baseline wave (See Table B.5 in the appendices).

6. **Previous wave(s) household response outcome:** Response propensity in longitudinal surveys is correlated within households across waves. In large part this is due to fixed/time-invariant household-level traits associated with ease of contact and stable at-home routines as well as socio-psychological dispositions towards survey cooperation. Said differently, households that are easy to contact once are likely to be easy to contact in the future. Similarly, one-time cooperative respondents are more likely to become loyal (and even efficient) respondents as waves

progress. In turn, this correlation between contact and cooperation propensities across waves could also be associated with differential number of calls in future waves (See Table B.6 in the appendices).

7. **Previous wave(s) household call record sequence:** Because sample households often require more than one contact attempt and given the multiple outcomes each of these attempts may have, summarizing CR data poses an initial challenge when considering possible applications for field management and measuring field performance. In fact, beyond categorizing call sequences as "long" or "short" by tallying the number of contact attempts, households may be further classified given the occurrence of particular call patterns, for example: by their propensity for broken appointments, proportion of calls that result in noncontacts and the occurrence of repeated soft refusals (See Chapter 1). Indeed, Stoop (2005) proposes a categorization of different types of reluctant respondents based, in part, on their propensity for different call and final response outcomes, including: "immediate interviews", "situationally unable", "soft / hard temporary refusals" and "broken appointments". Pollien and Joye (2014) similarly resort to an analysis of broken appointments and other such call sequences when analysis contact data in the ESS and MOSAiCH. While dealing with telephone survey data, Lipps (2012) and Barnes et al. (2008) also investigate the role of broken appointments in explaining reasons for refusal. Similar to lagged response outcomes, particular call record sequences indicative of limited contact and/or cooperation propensity in future waves (See Chapter 1) could also explain variability in subsequent calling patterns. More specifically, this analysis will consider lagged broken appointments, repeated soft-refusals and/or above median proportion of no replies as possible factors associated with differential calling effort (See Table B.7 in the appendices).

8. **Cross-wave continuity controls (Interviewer and geographical location):** Finally, interviewer continuity in longitudinal studies has often been found to be positively associated with continued successful contact and cooperation (Campanelli and O'Muircheartaigh, 1999; Laurie et al., 1999; Lynn et al., 2014; Nico-

letti and Peracchi, 2005). If the repeated visits to a household across waves affords the interviewer with additional contextual information about contact and cooperation retention strategies, changing field staff might result in additional effort (and thus cost) required by the new interviewer. Of course, interviewer reallocation is often the product of interviewer attrition or other similar mechanisms likely confounded with reduced response propensities and other forms of suboptimal field processes. In such cases, interviewer discontinuity might serve to optimize interviewer strategies. Lower Layer Super Output Area (LSOA) are geographical regions comprising between 1000 and 3000 inhabitants residing in 400 to 1200 households (Office For National Statistics (ONS), 2011). They are included as dependent variables to account for households that move physical location between waves. (See Table B.8 in the appendices).

2.3.2 Modelling strategy

In total, this analysis discusses the results of 12 distinct but similarly specified random-intercept linear regression models[3]. There are two analytical dimensions of comparability to consider: Waves (1, 2, 3 & 4) and Dependent Variable (total number of calls, calls to make contact & post-contact calls to completion). To reiterate: the analytical bases are different for each of the three dependent variables. All models contain the same independent variables, with Waves 2, 3 and 4 models containing additional previous wave information which by definition is not available for the Wave 1 equivalents (See Tables 2.2, 2.3 and 2.4). Given the hierarchical structure of the data (households nested within interviewer clusters), random-intercept fixed-effect models are employed to allow for unobserved variability in interviewer field performance. Coefficients are reported as raw number of calls. Reading the estimates across waves will address the

[3]In addition to linear regression, negative binomial and Poisson models were also considered but ultimately not employed for this analysis. While UKHLS call records share some of the properties of typical count datasets, employing continuous regression models yielded very similar results to those reported in count models. Briefly, the magnitudes of virtually all compared coefficients were proportionally similar. The direction of all effects were identical, and most of the significant terms coincided across the three types of models (linear, negative binomial and Poisson). Ultimately, linear regression was chosen to ease model computation and convergence, aid interpretation of the results as raw number of calls, and facilitate the application of Stata's commands for model post-estimation, marginal analyses and graphical outputs (StataCorp, 2017).

longitudinal dynamics of field effort, while the three different dependent variables aim to explore factors associated with number of calls associated with overall interviewer effort, contactability and cooperation propensity.

Table 2.2: Model Specifications: 1 to 3 (Wave 1)

Dependent Variables	Total Number of Calls Calls to make Contact Post-Contact Calls to Completion
Independent Variables	Geographical controls Dwelling characteristics Household composition Social inclusion and SES Interview experience

Table 2.3: Model Specifications: 4 to 6 (Wave 2)

Dependent Variables	Total Number of Calls Calls to make Contact Post-Contact Calls to Completion
Independent Variables	Previous wave(s) Household call sequence Geographical controls Dwelling characteristics Household composition Social inclusion and SES Previous wave interview experience Cross-wave continuity

Table 2.4: Model Specifications: 7 to 12 (Waves 3 and 4)

Dependent Variables	Total Number of Calls Calls to make Contact Post-Contact Calls to Completion
Independent Variables	Previous wave(s) Household response outcome Previous wave(s) Household call sequence Geographical controls Dwelling characteristics Household composition Social inclusion and SES Previous wave interview experience Cross-wave continuity

Equations

The equations notated below summarize the 12 different models employed for this analysis. Equation 2.1 refers to the Wave 1 models for all three outcomes: 1) total number of calls 2) number of calls to make contact (conditional on contacted households) and 3) number of post-contact calls to completion (conditional on responding households), where: β_0 represents the (empty) interviewer-level random intercept; $\boldsymbol{\beta_1 G}_{ij}$ is a vector corresponding to the geographical control variables discussed previously; $\boldsymbol{\beta_2 D}_{ij}$ includes all dwelling type covariates; $\boldsymbol{\beta_3 H}_{ij}$ corresponds to all household composition variables; $\boldsymbol{\beta_4 S}_{ij}$ groups all social inclusion and socioeconomic status covariates; $\boldsymbol{\beta_5 I}_{ij}$ includes all interviewer experience variables; finally, $\boldsymbol{u}_j$ and $\boldsymbol{e}_{ij}$ represent the interviewer- and household-level error terms. Equation 2.2 refers to Wave 2 models for all three outcomes as well. The specification is very similar to the models corresponding to equation 2.1, except for the inclusion of previous wave(s) call sequence variables ($\boldsymbol{\beta_1 C}_{ij}$) and a cross-wave interviewer and LSOA continuity indicator ($\boldsymbol{\beta_7 X}_{ij}$). Furthermore, interview experience is derived from Wave 1 data ($\boldsymbol{\beta_6 I}_{ij}$). Models for Waves 3 and 4 are similarly specified (see equation 2.3). In addition to all covariates described for equation 2.2, the models for the last two waves of this analysis include a $\boldsymbol{\beta_1 R}_{ij}$ term, corresponding to the previous wave(s) household response outcome. For all 12 models, the interviewer- and household-level residuals are assumed to be normally distributed, independent, and with a mean of 0 (see equations 2.4 and 2.5).

Wave 1

$$y_{ij} = \beta_0 + \beta_1 G_{ij} + \beta_2 D_{ij} + \beta_3 H_{ij} + \beta_4 S_{ij} + \beta_5 I_{ij} + u_j + e_{ij} \tag{2.1}$$

Wave 2

$$y_{ij} = \beta_0 + \beta_1 C_{ij} + \beta_2 G_{ij} + \beta_3 D_{ij} + \beta_4 H_{ij} + \beta_5 S_{ij} + \beta_6 I_{ij} + \beta_7 X_{ij} + u_j + e_{ij} \tag{2.2}$$

Waves 3 and 4

$$y_{ij} = \beta_0 + \beta_1 R_{ij} + \beta_2 C_{ij} + \beta_3 G_{ij} + \beta_4 D_{ij} + \beta_5 H_{ij} + \beta_6 S_{ij} + \beta_7 I_{ij} + \beta_8 X_{ij} + u_j + e_{ij} \tag{2.3}$$

Variance components: Waves 1 to 4

$$e_{ij} \sim N(0,\ \sigma_e^2) \tag{2.4}$$

$$u_j \sim N(0,\ \sigma_u^2) \tag{2.5}$$

2.4 Results

In all four waves, *Understanding Society* households display patterns of differential field effort associated with some of the covariates discussed previously. There are sizeable and significant effects associated with overall fieldwork effort as well as the sub-processes of achieving contact and gaining the cooperation of contacted households. Some of these effects are more prevalent in earlier waves, while others remain significant throughout the lifespan of the survey. Nevertheless, while some of the effects lose significance and/or magnitude in later waves, the overall direction of effects does not change with the passing of time (even after conditioning on additional controls like previous household response and previous call record sequence).

Furthermore, even if by Wave 4 geography/urbanicity; dwelling characteristics; household composition; SES and social inclusion indicators; interview experience; and cross-wave continuity indicators continue to account for some of the variability in field effort, previous household response outcomes and call sequences report the strongest and most significant effects. In fact, with regards to previous household response and CR sequences, their magnitude and significance seem to be related to the temporal proximity with the wave of the observed dependent variable. In other words, associations between these covariates and number of calls are more robust and stronger the more recently the wave in which the effects are observed.

2.4.1 Contactability

Geographical controls (Region and Urbanicity):

Table 2.5: Geographical Controls and Contactability

	Wave 1	Wave 2	Wave 3	Wave 4
Geographical Region				
London	0	0	0	0
North East	0.194	0.068	0.223*	0.043
North West	-0.063	0.078	0.077	-0.026
Yorkshire and the Humber	-0.019	0.150	-0.009	0.000
East Midlands	-0.016	0.111	-0.095	-0.036
West Midlands	-0.174*	0.088	-0.017	-0.028
East of England	-0.160*	-0.004	-0.126	-0.094
South East	-0.067	-0.063	-0.087	-0.049
South West	-0.094	-0.007	-0.194*	-0.109
Wales	-0.142	0.142	-0.057	0.159
Scotland	-0.173*	-0.048	-0.159	-0.066
Urban indicator	-0.060	-0.062*	0.002	-0.055
... remaining coefficients suppressed to ease presentation ...				
Constant	2.534***	2.464***	2.674***	2.540***
Random Intercept	0.628***	0.503***	0.466***	0.405***
Observations	24638	21925	20028	18218
** p<0.05, ** p<0.01, *** p<0.001*				

Estimated coefficients for Waves 1, 2, 3 and 4 Calls to Make Contact. Random-intercept, linear regression models were specified to account for unobserved interviewer effects. For presentation purposes, only geographical control variables have been included in this summarized table. For the complete tables with all estimated coefficients specified in the models, refer to Appendix B.

With regards to geographical region or urbanicity, there are no obvious effects on number of calls to make contact (Table 2.5). With respect to London, no one region in Great Britain is consistently harder or easier to contact. Neither is population density/urbanicity associated with differential contactability (at most, there is a slight significant decrease in contactability observed in Wave 2 in urban areas). The absence of effects is observed for most of the four waves considered for this study. While counter to some of the works previously discussed (See sections 2.2.2, 2.2.3, and 2.3.1), it should be noted that other covariates in the specification may explain this absence of effects. Indeed, as previously discussed, social inclusion and community attachment have often been associated with rurality while some dwelling types (i.e. flats and terraced properties which are typically harder to access) are more prevalent in cities than in rural communities.

Dwelling characteristics and accessibility:

Table 2.6: Dwelling Type and Contactability

	Wave 1	Wave 2	Wave 3	Wave 4
Dwelling Type				
Detached	0	0	0	0
Semi	0.022	-0.012	0.004	0.003
Terraced + end	0.143***	0.069*	0.083*	0.036
Flat/Msnette. + Purpose + Conv.	0.277***	0.231***	0.108*	0.156**
Other (+ Missing)	0.371***	0.104	-0.026	-0.174
Groundfloor property	-0.087	-0.014	-0.080	0.039
Property with respect to nbours.				
Better condition	0	0	0	0
Same or missing	0.064	-0.007	0.024	0.035
Worse	0.299***	-0.085	0.173**	0.130*
No Barriers to Dwelling	-0.093	-0.024	-0.058	-0.006
... remaining coefficients suppressed to ease presentation ...				
Constant	2.534***	2.464***	2.674***	2.540***
Random Intercept	0.628***	0.503***	0.466***	0.405***
Observations	24638	21925	20028	18218
				* $p<0.05$, ** $p<0.01$, *** $p<0.001$

Estimated coefficients for Waves 1, 2, 3 and 4 Calls to Make Contact. Random-intercept, linear regression models were specified to account for unobserved interviewer effects. For presentation purposes, only dwelling characteristics variables have been included in this summarized table. For the complete tables with all estimated coefficients specified in the models, refer to Appendix B.

Flats, maisonettes, purpose and converted properties are harder to contact than detached properties for the first four waves of *Understanding Society* (Table 2.6). *Terraced and end properties* are also harder to contact than detached houses, but in lesser magnitude and only for the first three waves. Ground-floor properties are neither harder or easier to reach; however, those in worse condition than those of its neighbours are less contactable (except for Wave 2).

Sociodemographic composition of household:

Table 2.7: Sociodemographics and Contactability

	Wave 1	Wave 2	Wave 3	Wave 4
Number of people in household	-0.205***	-0.139***	-0.112***	-0.097***
At least one baby in household	0.034	0.108*	0.063	0.021
All residents in poor health	-0.170***	-0.140***	-0.138***	-0.122**
Household National Origin				
All British	0	0	0	0
Mixed + Missing	0.039	0.047	0.036	-0.087
All Non-British	0.109*	0.119**	0.083	0.143**
Working Status				
No one works	0	0	0	0
At least 1 works but not long hrs	0.152***	0.135***	0.105**	0.087*
At least 1 (not all) works long hrs	0.242***	0.134***	0.084*	0.142***
All work long hrs	0.865***	0.481***	0.365***	0.339***
Presence of pensioner				
No pensioner	0	0	0	0
At least 1 pensioner	-0.421***	-0.369***	-0.263***	-0.131**
All pensioners	-0.438***	-0.454***	-0.308***	-0.214***
... remaining coefficients suppressed to ease presentation ...				
Constant	2.534***	2.464***	2.674***	2.540***
Random Intercept	0.628***	0.503***	0.466***	0.405***
Observations	24638	21925	20028	18218
				** p<0.05, ** p<0.01, *** p<0.001*

Estimated coefficients for Waves 1, 2, 3 and 4 Calls to Make Contact. Random-intercept, linear regression models were specified to account for unobserved interviewer effects. For presentation purposes, only household composition variables have been included in this summarized table. For the complete tables with all estimated coefficients specified in the models, refer to Appendix B.

As theorized, stay-at-home routines based on the composition and lifestyles of household residents are strongly associated with ease of contact (Table 2.7). For all four waves, working households (and especially those where everyone works long hours) require more calls than those households where no one works. While the effect remains significant and sizeable, the magnitude decreases progressively with each passing wave. Perhaps this reduction in effect size is due to the mutual learning process mentioned in section 2.2.5, where increased awareness of the interviewee's routines afford the interviewer additional information to tailor or adapt his/her field strategies and the respondent's familiarity with the survey request and questionnaire content result in a more expedient interviewing process. Alternatively, self-selection could explain this reduction as respondents who remain in later waves of a longitudinal survey are by definition more loyal and committed while those that will opt out perhaps require more calls to contact.

Presence of pensioners in the household is also strongly and significantly associated with contactability, although comparatively less than household work status. In this case, one could assume that pensioners' increased likelihood of being at home and having more free time than their younger counterparts means they are easier (i.e. requiring less calls) to contact. As is to be expected, an increased number of household residents results in higher contactabilty since at every attempt there is automatically an increased likelihood of someone being at home to answer. Exclusively non-British households are always slightly more difficult to contact than all British households (except in Wave 3). Finally, households where all report being in poor health are associated with increased contactability in all four waves as these are likely to include residents that spend little time outside of the home.

Indicators of social inclusion and SES:

Table 2.8: Social Inclusion Indicators, SES and Contactability

	Wave 1	Wave 2	Wave 3	Wave 4
Material deprivation	-0.011	0.001	0.041	-0.003
Owner/Mortgager	0.097**	0.011	0.025	-0.042
No political interest	0.026	0.044	0.033	0.054
Community Attachment				
Q4	0	0	0	0
Q3	0.118**	0.057	-0.007	0.074*
Q2	0.109**	-0.001	0.023	0.052
Q1	0.190***	0.115**	0.039	0.082*
Missing	0.342***	0.134**	0.014	0.175*
... remaining coefficients suppressed to ease presentation ...				
Constant	2.534***	2.464***	2.674***	2.540***
Random Intercept	0.628***	0.503***	0.466***	0.405***
Observations	24638	21925	20028	18218
** p<0.05, ** p<0.01, *** p<0.001*				

Estimated coefficients for Waves 1, 2, 3 and 4 Calls to Make Contact. Random-intercept, linear regression models were specified to account for unobserved interviewer effects. For presentation purposes, only social inclusion and SES variables have been included in this summarized table. For the complete tables with all estimated coefficients specified in the models, refer to Appendix B.

Social inclusion indicators are significantly associated with ease of contact in the initial wave of *Understanding Society* (Table 2.8). Indeed, those with the strongest levels of community attachment are comparatively easier to contact than every other sample member. Since these effects decrease in magnitude and significance in Waves 2 through 4, one could posit that the initial baseline contact affords each party of the interviewing interaction with enough information about the other so that a minimum of trust is established. Thus, any possible predispositions towards reclusion or social exclusion which might negatively impact on contactability are tempered once the initial wave is completed and a mutual rapport is developed. Material deprivation and political interest are never associated with differential contactability. Lastly, household ownership is slightly associated with increased contactability, but only for Wave 1.

Interview experience:

Table 2.9: Interview Experience and Contactability

	Wave 1	Wave 2	Wave 3	Wave 4
Consent to Data Linkage				
All consent	0	0	0	0
Some consent + missing	-0.168***	-0.087*	-0.027	-0.037
No one consent	0.036	-0.021	0.001	0.020
No one present during interview	0.096***	0.099***	0.104***	0.052*
No suspicion during interview	-0.053	-0.027	-0.107*	-0.082
Excellent understanding of quest.	0.063	0.050	0.009	0.010
Cooperative respondent household	0.058	-0.049	-0.003	0.023
Item nonresponse (log)	0.005	0.030	0.031*	-0.003
Dummy Item nonresponse (log)	-0.006	0.038	0.058	-0.023
... remaining coefficients suppressed to ease presentation ...				
Constant	2.534***	2.464***	2.674***	2.540***
Random Intercept	0.628***	0.503***	0.466***	0.405***
Observations	24638	21925	20028	18218
* *p<0.05*, ** *p<0.01*, *** *p<0.001*				

Estimated coefficients for Waves 1, 2, 3 and 4 Calls to Make Contact. Random-intercept, linear regression models were specified to account for unobserved interviewer effects. Except for Wave 1, all interview experience variables are derived from the wave previous to the wave where the dependent variable is observed. In Wave 1, the interview experience variables are derived from Wave 1. For presentation purposes, only interview experience variables have been included in this summarized table. For the complete tables with all estimated coefficients specified in the models, refer to Appendix B.

Interestingly, interview experience (either current for Wave 1 or lagged from Waves 2 onwards) is mostly unassociated with differential ease of contact (Table 2.9). For the most part, households require a comparable number of calls to contact regardless of suspicion, tendency to cooperate or understand the survey request. Furthermore, household averages of item nonresponse (represented in a logarithmic scale to account for the skewed distributions of raw means) are also not associated with contactability (except for a slight increase in Wave 3 alone). Attitude towards consent does not show any consistent effect on calls to make contact (a slight effect is observed in Wave 1, and an even smaller one in Wave 2). The only covariate that consistently displays significant (yet slight) effects is "no one present during interview" which displays a positive relationship. Because household size is included as one of the household composition covariates, one could speculate that this suggests that multi-person households where not everyone is at home at the same time require more calls than in those where similar at home patterns are shared by all.

Cross-wave continuity controls (Interviewer and geographical location):

Table 2.10: Cross-wave controls and Contactability

	Wave 2	Wave 3	Wave 4
Same LSOA (up to observed wave)	-0.135	-0.075	-0.060
Same Intervr. (up to observed wave)	0.228*	-0.040	-0.071
Same LSOA and Intervr. (Interaction)	-0.256**	-0.034	0.0029
... remaining coefficients suppressed to ease presentation ...			
Constant	2.464***	2.674***	2.540***
Random Intercept	0.503***	0.466***	0.405***
Observations	21925	20028	18218
** p<0.05, ** p<0.01, *** p<0.001*			

Estimated coefficients for Waves 1, 2, 3 and 4 Calls to Make Contact. Random-intercept, linear regression models were specified to account for unobserved interviewer effects. For presentation purposes, only cross-wave continuity variables have been included in this summarized table. For the complete tables with all estimated coefficients specified in the models, refer to Appendix B.

Contrary to expectations based on the literature, neither geographical nor interviewer continuity impact on the number of calls necessary to make contact (except for Wave 2 where households which kept the same interviewer were harder to reach than those that switched; nevertheless, within households that stayed in the same LSOA interviewer continuity is associated with decreased contactability) (Table 2.10). In any case, that moving households or switching interviewers is not linked to cross-wave differences is good news for survey practitioners as these are processes that lie largely (if not entirely) outside of the survey's control.

Previous wave(s) household response outcome:

Table 2.11: Previous wave(s) household response outcome and Contactability

	Wave 3	Wave 4
Wave 3 HH Response		
Nonresponse		0
Response		-0.410***
Ineligible		-0.141
Wave 2 HH Response		
Nonresponse	0	0
Response	-0.453***	-0.218***
Ineligible	-0.198	0.443
... remaining coefficients suppressed to ease presentation ...		
Constant	2.674***	2.540***
Random Intercept	0.466***	0.405***
Observations	20028	18218
		** p<0.05, ** p<0.01, *** p<0.001*

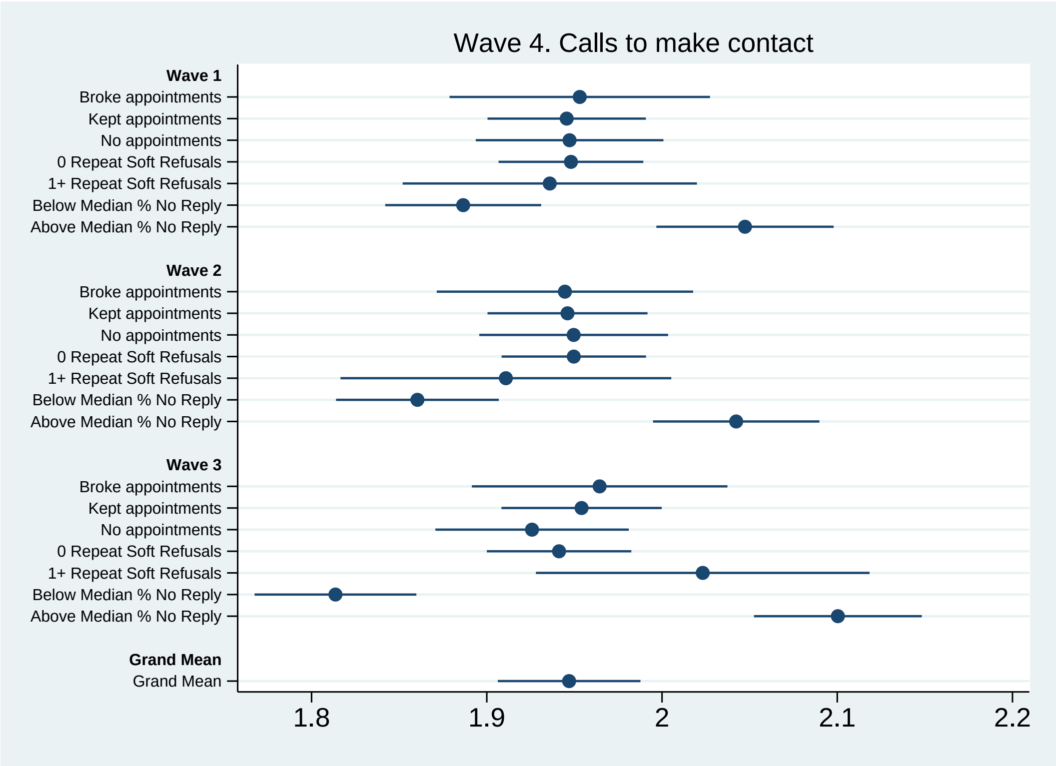
Estimated coefficients for Waves 1, 2, 3 and 4 Calls to Make Contact. Random-intercept, linear regression models were specified to account for unobserved interviewer effects. For presentation purposes, only previous wave household response variables have been included in this summarized table. For the complete tables with all estimated coefficients specified in the models, refer to Appendix B.

Waves 3 and 4 show very strong and significant effects of previous wave household response outcome on ease of contact (Table 2.11). Unsurprisingly, previous wave respondents are considerably less call-intensive to contact than nonrespondents. The effect is apparent not just from the outcome of the immediately previous wave but indeed from response outcomes observed two waves prior.

Previous wave(s) household call record sequence:

Lastly, CR sequences account for additional variability even after conditioning on all the covariates previously mentioned (Figure 2.1). Households which reported an above median proportion of no replies at a previous wave require more calls to contact in the future. One can assume that contactability is correlated across waves; households that are once hard to reach continue to be in future. In fact, while the average Wave 4 household requires 1.95 calls to make contact, those that are above the median proportion of no replies at Wave 3 require approximately 2.1 while those below the median only 1.8. Taking into account *Understanding Society's* absolute costs, these fractional call differences represent substantial field resources once aggregated at overall sample levels.

Figure 2.1: Wave 4. Calls to Make Contact



Model-estimated means for Wave 4 contacted households (n = 18218)

2.4.2 Cooperation

Geographical controls (Region and Urbanicity):

Table 2.12: Geographical Controls and Cooperation

	Wave 1	Wave 2	Wave 3	Wave 4
Geographical Region				
London	0	0	0	0
North East	-0.211	-0.150	-0.215	-0.027
North West	-0.438***	-0.131	-0.278**	-0.150
Yorkshire and the Humber	-0.515***	-0.315**	-0.310**	-0.121
East Midlands	-0.327**	-0.129	-0.187	-0.076
West Midlands	-0.287**	-0.218*	-0.306**	-0.179
East of England	-0.170	-0.154	-0.244**	-0.086
South East	-0.217*	-0.197**	-0.231**	-0.069
South West	-0.609***	-0.375***	-0.506***	-0.271*
Wales	-0.406**	-0.272*	-0.257*	-0.075
Scotland	-0.355***	-0.155	-0.276**	-0.162
Urban indicator	-0.179***	-0.094**	-0.151***	-0.107**
... remaining coefficients suppressed to ease presentation ...				
Constant	1.656***	1.485***	2.450***	2.450***
Random Intercept	0.754***	0.504***	0.471***	0.453***
Observations	24638	19786	17560	16562
** p<0.05, ** p<0.01, *** p<0.001*				

Estimated coefficients for Waves 1, 2, 3 and 4 Post-Contact Calls to Completion. Random-intercept, linear regression models were specified to account for unobserved interviewer effects. For presentation purposes, only geographical control variables have been included in this summarized table. For the complete tables with all estimated coefficients specified in the models, refer to Appendix B.

Contrary to what is observed regarding contact, geography and urbanicity are significantly associated with differential ease of cooperation (Table 2.12). London, as the reference category, is more call-intensive than most other regions in Waves 1, 2 and 3 (but not anymore by Wave 4). However, urbanicity is always associated with decreased number of calls to cooperation, even if the effect is slight. A possible reason, which remains to be confirmed, is the faster and busier lifestyles of urban settings which demand interviewers to allocate their call resources in more optimal ways than their rural counterparts.

Dwelling characteristics and accessibility:

Table 2.13: Dwelling Type and Cooperation

	Wave 1	Wave 2	Wave 3	Wave 4
Dwelling Type				
Detached	0	0	0	0
Semi	0.019	0.042	0.012	-0.036
Terraced + end	0.089*	0.062	0.023	0.014
Flat/Msnette. + Purpose + Conv.	0.008	-0.029	-0.113*	-0.031
Other (+ Missing)	0.633***	0.242*	-0.166	0.107
Groundfloor property	-0.065	-0.001	-0.017	-0.048
Property with respect to nbours.				
Better condition	0	0	0	0
Same or missing	0.069	-0.080	-0.040	-0.001
Worse	0.079	-0.097	-0.009	0.123
No Barriers to Dwelling	0.184**	0.184***	0.011	0.079
... remaining coefficients suppressed to ease presentation ...				
Constant	1.656***	1.485***	2.450***	2.450***
Random Intercept	0.754***	0.504***	0.471***	0.453***
Observations	24638	19786	17560	16562
** p<0.05, ** p<0.01, *** p<0.001*				

Estimated coefficients for Waves 1, 2, 3 and 4 Post-Contact Calls to Completion. Random-intercept, linear regression models were specified to account for unobserved interviewer effects. For presentation purposes, only dwelling characteristics variables have been included in this summarized table. For the complete tables with all estimated coefficients specified in the models, refer to Appendix B.

For the most part dwelling characteristics, previously reporting significant contact effects, do not account for any significant differences in calls to cooperation (Table 2.13). As discussed in the literature, dwelling types are associated with differences in accessibility and ease of reach, but not necessarily with the socio-psychological dispositions towards responding manifested by the respondent once contact is made.

Socio-demographic composition of household:

Table 2.14: Sociodemographics and Cooperation

	Wave 1	Wave 2	Wave 3	Wave 4
Number of people in household	0.197***	0.175***	0.195***	0.176***
At least one baby in household	-0.111*	-0.062	-0.097*	-0.064
All residents in poor health	-0.155***	-0.124**	-0.064	-0.085*
Household National Origin				
All British	0	0	0	0
Mixed + Missing	0.058	-0.039	0.008	-0.052
All Non-British	0.140**	0.147**	0.100*	0.076
Working Status				
No one works	0	0	0	0
At least 1 works but not long hrs	0.357***	0.116**	0.121**	0.135***
At least 1 (not all) works long hrs	0.416***	0.040	0.187***	0.116**
All work long hrs	0.340***	0.008	0.127*	0.022
Presence of pensioner				
No pensioner	0	0	0	0
At least 1 pensioner	-0.117*	-0.190***	-0.087	-0.049
All pensioners	-0.212***	-0.307***	-0.199***	-0.163***
... remaining coefficients suppressed to ease presentation ...				
Constant	1.656***	1.485***	2.450***	2.450***
Random Intercept	0.754***	0.504***	0.471***	0.453***
Observations	24638	19786	17560	16562
** p<0.05, ** p<0.01, *** p<0.001*				

Estimated coefficients for Waves 1, 2, 3 and 4 Post-Contact Calls to Completion. Random-intercept, linear regression models were specified to account for unobserved interviewer effects. For presentation purposes, only household composition variables have been included in this summarized table. For the complete tables with all estimated coefficients specified in the models, refer to Appendix B.

Household composition is linked with differential number of calls to cooperation, but for the most part effects disappear or decrease by Waves 3 and 4 (Table 2.14). Perhaps this attenuation is a function of the self-selection effect previously discussed (loyal and efficient respondents remain through later waves of a longitudinal survey, while the rest drop out). Alternatively, a learning effect may also explain the disappearance of most demographic effects on cooperation in the last two waves analysed here. As discussed in section 2.2.5 once rapport between interviewer and interviewee is established, both sides learn more about each other and developed trust and/or familiarity conducive to more efficient response processes.

On the other hand, a demographic composition variable that is associated with a robust cooperation effect across all waves is number of people in household. Obviously, the more potential respondents in the household the more effort required by the interviewer to complete all individual questionnaires and the increased probability of not finding all residents at home in one single occasion. Lastly, all pensioner households not only demand fewer calls to contact but also fewer to gain their cooperation. Here too, the effect remains strongly significant across all first four waves of *Understanding Society*.

Indicators of social inclusion and SES:

Table 2.15: Social Inclusion Indicators, SES and Cooperation

	Wave 1	Wave 2	Wave 3	Wave 4
Material deprivation	0.1000**	0.136***	0.094**	0.127***
Owner/Mortgager	-0.060	-0.125***	-0.165***	-0.160***
No political interest	0.072	0.081*	0.047	0.078*
Community Attachment				
Q4	0	0	0	0
Q3	0.029	0.019	0.055	-0.044
Q2	0.065	-0.003	-0.004	-0.018
Q1	0.142**	0.099**	0.067	0.026
Missing	0.312***	0.137**	0.225*	0.030
... remaining coefficients suppressed to ease presentation ...				
Constant	1.656***	1.485***	2.450***	2.450***
Random Intercept	0.754***	0.504***	0.471***	0.453***
Observations	24638	19786	17560	16562
** p<0.05, ** p<0.01, *** p<0.001*				

Estimated coefficients for Waves 1, 2, 3 and 4 Post-Contact Calls to Completion. Random-intercept, linear regression models were specified to account for unobserved interviewer effects. For presentation purposes, only social inclusion and SES variables have been included in this summarized table. For the complete tables with all estimated coefficients specified in the models, refer to Appendix B.

Confirming findings from literature on nonresponse (See section 2.2.4), low SES is associated with decreased cooperation ease (Table 2.15). Indeed, households that are materially deprived and those not owned/mortgaged by the residents require additional call effort. Nevertheless, indicators of community attachment or political interest are not consistently associated with cross-wave patterns of increased calls to cooperation. To the extent that these indicators are proxies for psychological inclinations towards efficient responding patterns or conversely illustrate the preferences/prejudices of the interviewer who has already made contact with a household remains to be determined.

Interview experience:

Table 2.16: Interview Experience and Cooperation

	Wave 1	Wave 2	Wave 3	Wave 4
Consent to Data Linkage				
All consent	0	0	0	0
Some consent + missing	0.094	0.000	0.011	0.053
No one consent	0.147***	0.010	0.061*	0.014
No one present during interview	0.132***	0.033	0.079**	0.030
No suspicion during interview	-0.114**	-0.141***	-0.134*	-0.207**
Excellent understanding of quest.	-0.099**	-0.048	0.016	-0.015
Cooperative respondent household	-0.126**	-0.003	-0.271***	-0.105**
Item nonresponse (log)	-0.005	0.004	0.024	0.001
Dummy Item nonresponse (log)	0.039	0.037	0.091*	0.042
... remaining coefficients suppressed to ease presentation ...				
Constant	1.656***	1.485***	2.450***	2.450***
Random Intercept	0.754***	0.504***	0.471***	0.453***
Observations	24638	19786	17560	16562
				* $p<0.05$, ** $p<0.01$, *** $p<0.001$

Estimated coefficients for Waves 1, 2, 3 and 4 Post-Contact Calls to Completion. Random-intercept, linear regression models were specified to account for unobserved interviewer effects. For presentation purposes, only interview experience variables have been included in this summarized table. For the complete tables with all estimated coefficients specified in the models, refer to Appendix B.

Interview experience is highly correlated with ease of cooperation in the baseline wave of the survey (Table 2.16). Those that do not consent to data linkage, answer without the presence of a third party, are suspicious of the survey, do not fully understand the questionnaire and/or are not fully cooperative require more calls towards cooperation in Wave 1. However, these effects mostly disappear from Waves 2 onwards except for: suspicion and cooperative attitude. These two psychological traits are associated with differential cooperation ease in subsequent waves.

Cross-wave continuity controls (Interviewer and geographical location):

Table 2.17: Cross-wave controls and Contactability

	Wave 2	Wave 3	Wave 4
Same LSOA (up to observed wave)	0.392***	0.153**	0.166***
Same Intervr. (up to observed wave)	-0.056	-0.053	-0.121
Same LSOA and Intervr. (Interaction)	-0.060	0.026	0.061
... remaining coefficients suppressed to ease presentation ...			
Constant	1.485***	2.450***	2.450***
Random Intercept	0.504***	0.471***	0.453***
Observations	19786	17560	16562
** p<0.05, ** p<0.01, *** p<0.001*			

Estimated coefficients for Waves 1, 2, 3 and 4 Post-Contact Calls to Completion. Random-intercept, linear regression models were specified to account for unobserved interviewer effects. For presentation purposes, only cross-wave continuity variables have been included in this summarized table. For the complete tables with all estimated coefficients specified in the models, refer to Appendix B.

A seemingly counter-intuitive finding is that geographical continuity is associated with increased number of calls to cooperation (Table 2.17). However, this can be explained as a function of a common field practice given the finite calendar and call resources available for households that move (but do not notify the survey agency) between waves. To clarify: once an interviewer realizes that a target household has moved they have to attempt to locate the new address and/or wait for further instructions from the field office. Simultaneously, they continue contacting other households that did not move. By the time the movers have been found, so much calendar time has passed that only a certain number of calls (usually less than those afforded to non-movers) remains available.

Contrary to expectations, interviewer continuity does not result in more efficient cooperation field strategies. The absence of effect is observed for all waves: 2, 3 and 4. Here too, this might be a beneficial finding for survey practitioners as interviewer attrition / switching is often a process that is partially (if not entirely) uncontrolled by the field agency.

Previous wave(s) household response outcome:

Table 2.18: Previous wave(s) household response outcome and Cooperation

	Wave 3	Wave 4
Wave 3 HH Response		
Nonresponse		0
Response		-0.313***
Ineligible		-0.165
Wave 2 HH Response		
Nonresponse	0	0
Response	-0.219***	-0.100
Ineligible	-0.263	-0.515
... remaining coefficients suppressed to ease presentation ...		
Constant	2.450***	2.450***
Random Intercept	0.471***	0.453***
Observations	17560	16562
		** p<0.05, ** p<0.01, *** p<0.001*

Estimated coefficients for Waves 1, 2, 3 and 4 Post-Contact Calls to Completion. Random-intercept, linear regression models were specified to account for unobserved interviewer effects. For presentation purposes, only previous wave household response variables have been included in this summarized table. For the complete tables with all estimated coefficients specified in the models, refer to Appendix B.

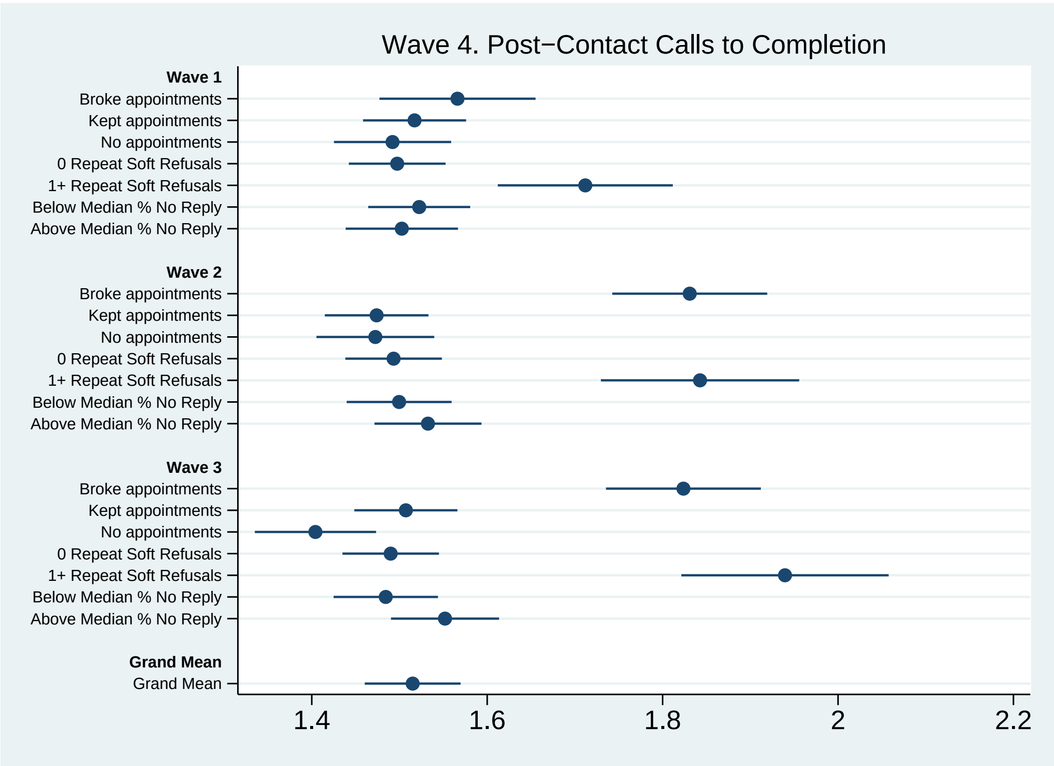
Previous wave response is correlated with increased cooperation ease, suggesting a pattern of cross-wave response propensity similar to the one observed for contact (Table 2.18). Nevertheless, while also highly significant, this effect is not as large as the corresponding contactability correlation. Furthermore, by Wave 4 the effect is only significant for Wave 3 (i.e. the immediately previous wave) but not for Wave 2 response outcomes.

Previous wave(s) household call record sequence:

Finally, CR patterns again account for the most significant and strongest effects on cooperation for all four waves of the survey. However, if proportion of no reply call outcomes explains contactability, for cooperation it is primarily broken appointments and repeated soft refusals. In fact, Figure 2.2 shows that by Wave 4, the average household requires roughly 1.5 post-contact calls to completion; however, for those households that in Wave 3 broke appointments or refused to answer in two or more successive calls, 1.82 and 1.95 calls are required respectively. The effects are comparable in magnitude for Wave 2 CR sequences. And Wave 1 repeated soft refusals are also associated with significantly higher number of calls to cooperation by Wave 4.

If for all other covariates, socio-psychological attributes explain most of the variation in cooperation, the same stands for lagged CR sequences. Breaking an appointment and/or "soft refusing" necessitates a face-to-face interaction between interviewer and respondent (the same is not necessarily true for "above median proportion of no replies"). Therefore, one can assume that respondent's dispositions towards the survey are at play when deciding to cancel an appointment or engage with an interviewer in a nonproductive way. In other words, contactability or access do not seem to drive these two CR sequences.

Figure 2.2: Wave 4. Post-Contact Calls to Completion



Model-estimated means for Wave 4 responding households (n = 16562)

2.4.3 Field effort in longitudinal context: Learning vs. self-selection

Thus far, this analysis has highlighted the possibility that differential calling effort is not only explained by the covariates selected but in fact answers to two additional factors: 1) a *learning* process whereby interviewers and interviewees develop more efficient interactions as survey waves progress or 2) a *self-selection* bias where only respondents in households that are inherently efficient continue to participate with the survey while the remaining drop out eventually. The following sections discuss learning and self-selection effects in ease of contact, cooperation and overall field effort.

Learning effects

To address the possibility that the effects thus far discussed are a function of learning effects, Table 2.19 shows global estimates for a balanced panel of all households issued at all Waves (1, 2, 3 and 4). As reported in the figures for households in the balanced panel, there is evidence of learning effects for ease of contact, cooperation and overall field effort between waves 1 and 4.

While the typical household of the balanced panel requires an average of 2.21 calls to make contact in Wave 1, by Wave 4 the number has been reduced to 1.91 (representing a statistically significant 14% reduction in field effort). Additionally, the reduction is monotonic, with each wave being significantly easier to contact than the previous one. However, the overall trend is not linear; instead the rate of reduction diminishes with each wave. Similarly, effort towards cooperation following initial contact also reduces considerably and significantly between waves 1 and 4 (from 1.85 to 1.44 calls, or 22% percent). Nevertheless, while there is a considerable reduction between Waves 1 and 2, field effort remains comparable for Waves 2 and 3. A further drop, slight but significant, is again observed at Wave 4. Lastly, overall effort reduces significantly between Waves 1 and 4: from 4.38 to 3.77 (or 14%). Here too, as with contactability, the reduction is monotonic and significant across all waves.

Table 2.19: Learning Effects

	Wave	Mean	95% C.I.		Total Households
Calls to Make Contact	1	2.21	2.18	2.24	16,844
	2	2.04	2.02	2.07	
	3	1.95	1.93	1.98	
	4	1.91	1.89	1.94	
Post-Contact Calls to Cooperation	1	1.85	1.82	1.89	14,724
	2	1.50	1.47	1.52	
	3	1.51	1.49	1.54	
	4	1.44	1.42	1.47	
Total Number of Calls	1	4.38	4.34	4.42	19,751
	2	4.00	3.96	4.04	
	3	3.95	3.91	3.99	
	4	3.77	3.73	3.81	

To create the balanced panels within each of the three metrics reported in the table, households were only included if: 1) they were issued in all of the four waves and 2) did not contain missing values for the corresponding field effort variable in any of the waves. Unlike the analytical bases reported in Tables B.9, B.10, B.11 and B.12, the balanced panels do not include ineligible cases for any of the waves.

Self-selection effects

In addition to the learning effects previously discussed, evidence suggests that self-selection also accounts for some changes in field effort across the first four waves of the survey. Table 2.20 tests differences between households issued in Wave 4 and households not issued in Wave 4. Given following and sample issue rules of *Understanding Society*, a household not issued at Wave 4 is effectively a household that has dropped out of the survey not to be reissued in any future wave. Therefore, in this analysis, self-selection effects are a function of attrition and are captured in those cases were a significant difference is observed.

Households that were not issued in Wave 4 were not significantly harder to contact at Wave 1 than Wave 4 households (2.29 calls to contact vs. 2.24). However, with regards to cooperation and overall field effort, there are significant ($p < .001$) effects observed between drop-outs and remainers. There is an observable difference in post-contact calls required to cooperation between attriters and Wave 4 issued households (from 2.11 to 1.97, or 6.64%). For total number of calls the difference is also significant (4.65 vs. 4.38, or 5.81%).

In short, self-selection accounts for changes in field-effort dynamics. Cooperation and overall field effort diminishes by Wave 4 as a function of attriters' propensity to demand higher number of calls.

Table 2.20: Self-Selection Effects. Wave 1 Field Effort.

	Not issued in Wave 4		Issued in Wave 4	
	Mean	Households	Mean	Households
Calls to Make Contact	2.29	6,098	2.24	19,991
Post-Contact Calls to Cooperation	2.11***	6,062	1.97	19,899
Total Number of Calls	4.65***	6,109	4.38	20,020
** p<0.05, ** p<0.01, *** p<0.001*				

In the table above, Wave 1 field effort metrics are compared between households issued at Wave 4 and households not issued at Wave 4. For each of the three dependent variables, independent sample t-tests (assuming equal variance) determine if the difference in the means between households is significantly different from 0.

2.5 Discussion

Literature on longitudinal survey field management often focuses on response retention strategies. Common concerns include: the importance of locating and tracking intra-wave movers, improving interviewer calling strategies to ensure contact, and tailoring the request in a way that encourages continued cooperation and loyalty from respondents. While these remain crucial concerns, it is important not to overlook another important aspect of field management in longitudinal surveys: namely, calling effort. This last point gains credence once one considers that nonresponse (in the forms of nonlocation, noncontact and noncooperation) progressively decreases with the passing of each wave. Thus, the potential for improving field efficiency also decreases if the focus lies exclusively with response rate retention. This analysis proposes a look at the field sub-processes inherent in achieving location/contact and cooperation. By focusing on relative calling effort (measured in total number of calls, calls to contact and, calls to cooperation) a more detailed analysis of factors associated with differential contactability and cooperation propensities is possible. In turn, recommendations can be made to inform not just *effective* response retention strategies but also *efficient* ones.

This analysis suggests that household traits are associated with differential ease of contact and/or cooperation. For the most part, these associations can be understood within the context of established literature on drivers of nonresponse. Furthermore, these associations are also impacted by the temporal nature of longitudinal surveys: some effects remain constant while others disappear from one wave to the next.

In sum, only a few of the covariates selected for this analysis remain significant and substantial by Wave 4. Thus, possible application of these findings could focus on 1) work status, 2) presence of pensioners, 3) national origin, 4) material deprivation and property ownership, 5) levels of suspicion and cooperation towards interview and, 6) urbanicity. Since these household traits consistently report differential contact and/or cooperation effects, strategies should be explored that address possible mechanisms behind these differences. For example, households with busy professionals could be identified before fieldwork starts and offered the opportunity to schedule an interview based on their availability. Alternatively, previous wave CR information should be made available and further analysed to identify in which day of the week and at what time of the day successful contact attempts were achieved. Lastly, material incentives could curb the number of calls required to contact and/or gain the cooperation of working households. When designing these incentives, careful consideration should be placed on the relative costs of the estimated extra calls per interviewer vs. monetary amount to be offered to selected households.

Beyond demographic composition, attitudinal correlates may also be used to design targeted call-reduction strategies. If households that are suspicious and/or noncooperative of the interviewer demand additional resources in the future, perhaps a tailored intra-wave mailing can be drafted to address possible motivations behind these attitudes. Additionally, interviewers could be notified and/or reminded of these households before a new wave commences and given additional training or scripting materials aimed at dissuading these negative attitudes.

Previous CR patterns and household response outcomes report the strongest, most significant and most consistent effects. Broken appointments, above median proportion of no replies and repeated soft refusals are not only commonplace and associated with differential response propensities (See Chapter 1), but account for substantial differences in subsequent number of calls required. Similarly, nonresponding and noneligible households will require considerably more field effort in future. Here too, approaches can be developed to identify and tailor targeted strategies to households which report either problematic call sequences or nonproductive response outcomes in previous waves. At the very least, this new finding suggests that there is practical value in collecting and archiving this type of paradata.

Although requiring further exploration, this analysis also indicates that temporal effects of field effort covariates are a function of longitudinal learning as well as self-selection. Briefly, sample members that remain in the survey after four waves seemingly become easier to contact, easier to cooperate and require less overall effort as a function of time. Self-selection is also associated with longitudinal field effort reduction. In fact, post-Wave 1 drop outs are significantly more onerous than Wave 4 issued households with regards to cooperation and overall field effort.

Perhaps the most obvious application of these findings is to use the estimated calls towards contact, cooperation and overall field effort as planning inputs, targets for live monitoring of data collection and post-fieldwork performance assessment. For example, call scheduling could be informed by expected field effort: more onerous households should be called on as soon as field starts to allow enough calendar time for all the calls required. Conversely, less demanding cases could be deprioritised and called on after more difficult households have been approached. Furthermore, if a given household is expected to require an overall effort of 4 calls, with 2 towards contact and another 2 towards cooperation, field supervisors can keep and update their balance sheet of allotted field effort for that household (and all households in the aggregate) throughout data collection. So long as calls do not exceed their expected targets for a given household or group of households, one can assume that associated costs and other non-monetary resources will remain in line with planned performance. Instead, if the effort exceeds the expected number of calls, the field agency should alert the interviewer and supervisor and consider adjusting strategies. Immediate actions could entail a revision of the original targets, tailored approaches, a change in the interviewing staff, or ceasing to call on the case altogether. Lastly, following data collection, individual interviewers and field offices can be evaluated based on their performance against the set targets.

The findings of this analysis may also have implications on budgeting and costs.

Nevertheless, it should be noted that field effort, as measured in this chapter, is not proportionally or even directly associated with field costs at the household or even interviewer assignment level. For one, interviewers are commonly paid based on their hourly rate (with additional bonuses based on performance metrics like strike and response rates). In other words, an interviewer who calls on 6 houses within an hour will receive the same base compensation as another one who calls on only 3. Additionally, field effort is necessarily conditional on sample design choices (like sample size, following rules, and distance between issued cases within an interviewer assignment, etc.). Therefore, estimating how expensive an individual household will be in a future wave is beyond the objectives (and capabilities) of this analysis. Nevertheless, aggregate measures of field effort could inform global estimates of expected field costs and therefore be used as another input in the budgeting of future waves. For example, mean unit costs of call effort towards contact, cooperation and overall interviewing activity can be derived by dividing total calls for each of the three dependent variables of this analysis by the total field budget. From these global unit costs, yearly comparisons would give the survey planner an idea of expected costs in future and provide him/her with an additional metric of budgetary evaluation.

However, certain limitations from this analysis should also be noted. Firstly, these data were not experimentally manipulated. Thus, the findings are not necessarily generalizable nor do they confirm causal relationships. A future area of research could entail the replication of these findings in future waves of *Understanding Society* and other comparable household longitudinal surveys.

Secondly, because of limited interviewer-level information, intra-wave field staff reallocation and the confounding of area- and interviewer-effects, this analysis focused primarily on household-level characteristics and correlates of calling effort. Thus, some of these findings are potentially sensitive to unobserved interviewer traits (like ability, experience, psychological profile and workload, among others). To test these possible associations, future work could incorporate additional data from interviewers to the extent that they become available. Additionally, cross-classified models could be implemented in future replications of this study to control for interviewer allocation across waves and area effects. For further discussion on cross-classified models to analyse interviewer and area effects see: Brunton-Smith et al. (2017); Durrant et al. (2011, 2010); Vassallo et al. (2017, 2014).

Lastly, this analysis has not considered measurement error or the possible effect of field effort on response bias. Should there be a relationship between field effort and response, the findings and recommendations could be revisited. For example, reducing number of calls to contact might make field work more efficient but could also result in negative externalities like decreased data quality. Indeed, it is possible to assume that too fast or too cheap field processes might result in hurried interviewing and potentially create poorer quality data. A final area of possible research could entail defining call reduction thresholds conditional on minimum data quality parameters.

Chapter 3

Measuring progress indicator effects on respondent effort, time management, and response quality

Abstract: Literature on progress indicator (PI) effects is primarily concerned with nonresponse. The bulk of this research has focused on how the inclusion and design of PIs may impact the probability of a survey respondent to abandon the survey once started. To the extent that nonresponse can be curbed by manipulating the design features of a web questionnaire, continued research on this topic is warranted. However, because data quality may also be compromised in those cases where a response is recorded what remains to be explored is the effect of PIs on measurement error of observed data. Building on previous research on the mediating effects of PIs on respondent motivation and perception of survey burden, and using an experiment from a survey of 1221 university students, this study aims to determine whether the presence of a PI has an effect on 1) respondent's management of time and effort and 2) response quality of observed responses. The findings will be accompanied by a discussion on measurement error and implications for questionnaire design.

Keywords: Progress indicators, response quality, survey burden, CAWI paradata

3.1 Introduction

A common feature of web surveys is progress indicators (PIs). Typically, PIs are displayed as horizontal, rectangular bars steadily increasing in length with each passing question, and provide the respondent with a visual estimate of how much of the survey has already transpired and how much more there is to go. Thus, they also provide sense of rate of progression, or speed, of a questionnaire as well as the burden / effort required to complete it. Despite their prevalence, there is no consensus on the virtues of including (or excluding) them in web surveys. This is not to say that the question of PI effects in web surveys has not been addressed. Indeed, since the early 2000s, a growing body of work has been concerned with identifying possible mechanisms behind these effects and measuring their possible impact on the quality of survey data.

Literature on PI effects is primarily concerned with nonresponse. The bulk of this research has focused on how the inclusion and design of PIs may impact the probability of a survey respondent to abandon the survey once started. On this first general question results remain mixed: some have found a negative impact on response rates, while others have found a positive or null relationship (Couper et al., 2001; Crawford et al., 2001). In an attempt to qualify these initial findings, some have explored the relationship of PIs and nonresponse as a function of: the respondent's expectation of survey duration (Heerwegh and Loosveldt, 2006); the rate of visual progression of the indicator (Conrad et al., 2005, 2003a,b, 2010; Matzat et al., 2009; Peytchev, 2009); the objective length/duration of the survey (Conrad et al., 2010; Heerwegh and Loosveldt, 2006; Matzat et al., 2009; Yan et al., 2011); and/or survey enjoyment / satisfaction, topic saliency, task difficulty (Conrad et al., 2010; Crawford et al., 2001; Matzat et al., 2009). However, none of these were able to conclude that PIs by themselves lower (or increase) nonresponse. Rather, the relationship is often mediated by other effects.

Perhaps the sole incontrovertible result is that so long as the information from PIs is encouraging (i.e. progression is happening faster than expected by the respondent), their inclusion will be an effective way to curb nonresponse. Several studies have experimented with slow-to-fast, fast-to-slow and constant-speed PIs to manipulate the respondent's 1) perception of survey burden and 2) encouragement towards task completion (Bohme, 2011; Conrad et al., 2010; Heerwegh and Loosveldt, 2006; Kaczmirek, 2008; Matzat et al., 2009). Briefly, where PIs were purposely altered to show faster progression during the earlier stages of a survey followed by a slowing down in later stages of the survey (fast-to-slow PIs), respondents were less likely to drop off than in surveys with constant PI speed or slow-to-fast progression. But because survey duration expectations are 1) respondent-specific 2) difficult to identify before fieldwork starts, and given that encouraging PI information is often the result of artificially manipulating the visual rate of progress of the PI, these findings are hard to generalize and their application to survey design remains limited and perhaps ethically questionable. Most surveys cannot measure duration expectations from all its respondents prior to starting the questionnaire without inadvertently risking response rates, data quality and/or affecting respondent motivation. While altering the rate of a PI may have positive effects for respondent retention, manipulating its speed does not constitute best practice. For a thorough review of these discussions see: Tourangeau et al. (2013); Villar et al. (2013).

Beyond the question of survey drop-off, very little work has been devoted to PI effects on measurement error. While the impact of PIs on nonresponse continues to be explored, no work yet has focused on response quality as a function of PIs. Thus, the discussion of the potential benefits of PIs remains incomplete so long as the quality of the responses observed is not also addressed. Building on previous research on the mediating effects of PIs on respondent motivation and perception of survey burden, and using an experiment from a survey of 1221 university students, this study aims to address this gap by investigating PI effects on measurement error. The experiment follows a split-sample approach with a treatment group representing respondents exposed to a PI and a control group representing respondents not exposed to a PI. Particularly, this analysis is concerned with potential effect of PIs on respondents' effort, time management and the quality of their responses. It hypothesizes that PIs may affect the respondents' perception of burden and in turn result in differential propensities to satisfice and/or engage in other forms of time- and effort-reduction (Krosnick, 1991; Krosnick et al., 1996). Satisficing, broadly understood as taking cognitive shortcuts when answering a survey (or answering suboptimally), reduces the quality of the responses recorded. To test these assumptions, this chapter will analyse and compare estimates from the observed responses between the treatment and control groups as well as paradata measuring question times, survey duration and click patterns.

3.2 Theory and concepts

The theoretical framework used to understand PI effects is informed by the concepts of: respondent burden (Crawford et al., 2001), respondent motivation (Heerwegh and Loosveldt, 2006), and/or survey experience (Conrad et al., 2010; Matzat et al., 2009). Briefly, it is assumed that the respondent's initial and continued motivation towards completing the questionnaire is always set against the burden that participating in the survey entails. Thus, a respondent's motivation to cooperate with a survey is partly conditioned by their perception of survey burden. Because PIs provide the respondent with a graphical (if not also textual) representation of their rate of completion, the respondent is able to make (and periodically update) assumptions about survey length, speed of progression, time spent and time remaining. All these are proxies for how burdensome the questionnaire is (or is perceived to be). When the balance between respondent motivation and survey burden tips to the latter, the respondent might be more likely to drop-off. This last point contextualizes the mixed results relating to nonresponse discussed earlier. PIs do not necessarily make surveys more or less burdensome. Instead, they provide the respondent with additional information about their survey experience and overall progress throughout the questionnaire. Sometimes that information is encouraging, sometimes it is discouraging, and sometimes it is neither: it all depends on the expectations the respondent brings to the survey and the changing dynamics of burden (e.g. interest, saliency, effort and/or difficulty) of the questionnaire with each passing question.

The focus on nonresponse in PI literature thus far is understood given the relatively lower response rates of web surveys when compared against other modes like CATI or face-to-face (Manfreda et al., 2008). To the extent that nonresponse can be curbed by manipulating the design features of a web questionnaire (including PIs), further investigation is warranted. Especially, when nonresponse negatively impacts data quality in the form of nonresponse bias or reduced statistical precision. However, data quality may also be compromised in those cases where a response is recorded. Indeed, one topic that remains to be explored is the effect of PIs on measurement error of observed data. More specifically, what impact (if any) do PIs have on data quality when respondents choose to complete the survey?

In this regard, it should be noted that the extent of item nonresponse has also been examined in PI literature (Conrad et al., 2005, 2010; Couper et al., 2001; Villar et al., 2013; Yan et al., 2011). However, the focus on item nonresponse has been mostly secondary to the mechanisms of drop-off and not investigated within the contexts of response quality or measurement error. This analysis aims to look not just at item nonresponse but also at other forms of suboptimal responding associated with measurement error and response quality.

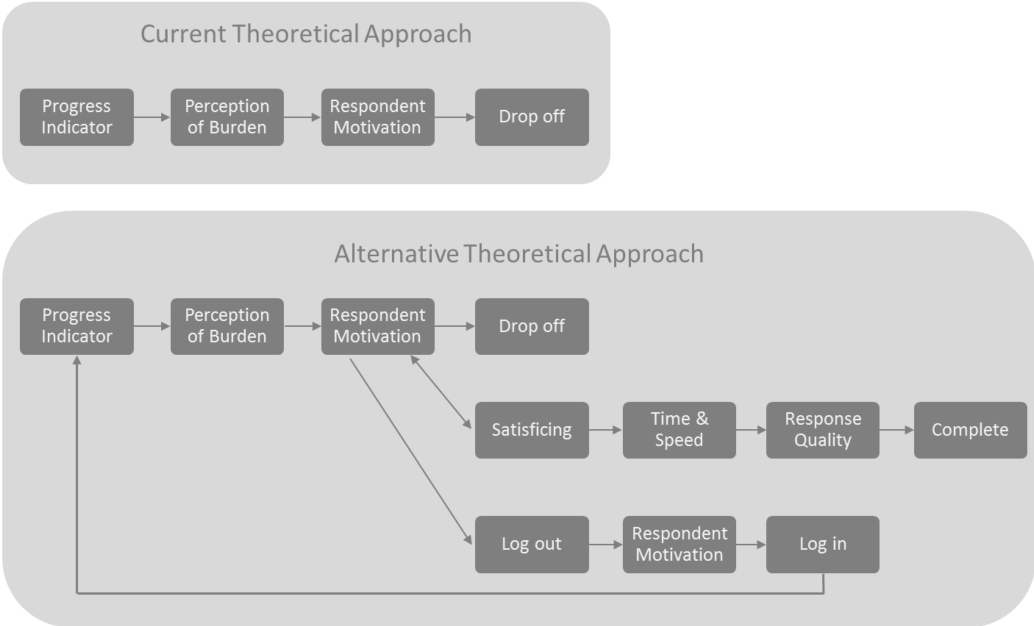
This last aim rests on the following assumption: when respondent motivation is dominated by survey burden, it may not always result in nonresponse but rather in suboptimal responding (Krosnick, 1991; Krosnick et al., 1996). While it is possible that as a survey progresses, and becomes more difficult / tedious / less interesting the respondent becomes more inclined towards abandoning the survey completely (Galesic, 2006), it is also possible that they persist given the *sunk cost* (Yan et al., 2011) of the time and effort they have spent on the survey so far. Further, the respondent may remain motivated to fulfill the *social exchange* (Heerwegh and Loosveldt, 2006) with the survey sponsor and complete the survey while at the same time have no more motivation to fully engage with the remaining questions. In other words, an unmotivated respondent may rush through a questionnaire and/or go through the motions of answering it without devoting all their cognitive resources to the response process (Tourangeau et al., 2004). This becomes more plausible in surveys with monetary incentives that reward full completion of the questionnaire (as is the case with the data analysed here).

In this way, PIs may have an impact on the respondents' propensity to satisfice or engage in suboptimal responding (instead of dropping out or not responding altogether) if the information interpreted from the PI discourages the respondent who otherwise chooses to continue to cooperate with the survey. Under this assumption, PIs should produce data of differential quality within the observed responses of a web survey.

Alternatively, a third option (besides survey drop-off or satisficing) is to control burden by managing one's time with the questionnaire and complete it in multiple sessions instead of in one sitting. Web surveys, as other forms of self-administered surveys, allow the respondent to do that. The question then arises if PIs have an effect on the propensity for multiple sessions. If so, the discussion of their effect on burden and data quality could be revisited. Two hypotheses emerge: PIs may have an effect on the propensity for multiple sessions. If so, conditional on PI presence, some respondents might be more prone to *sticking it out* while others decide to manage their time by logging on and off the survey multiple times. Thus, if a respondent sticks it out despite their burden, response quality may be negatively affected. Alternatively, the respondent that distributes their interactions with the survey across several sessions may diminish the negative effect of perceived accumulated burden and potentially give higher quality responses.

In sum, and as reflected in Figure 3.1, PI effects could correspond not just to one mechanism (i.e. survey drop-off) but instead to three. While some respondents might interpret the information from the PI as too discouraging to continue answering a survey and thus drop out altogether, others might continue to *soldier on* or *stick it out*. If this decision to continue is taken despite the burden, there exists a risk for the respondent to answer questions suboptimally (without devoting all their cognitive resources available). With each passing question, this second type of respondent will continue to negotiate their motivation and thus revisit their decision to cooperate (and likely to keep satisficing) or to finally drop out. A third type of respondent might decide to temporarily leave the survey until a later time when their motivation allows them to continue. Should this momentary disengaging from the survey serve to attenuate the perception of increased burden response quality shouldn't be negatively affected.

Figure 3.1: Theoretical Approaches



3.2.1 Research Questions and Objectives

Specifically, this analysis is concerned with the following questions:

1. Does the presence of a PI have an effect on:

    (a) the respondents' management of time and effort?

    (b) the quality of observed responses?

2. If PIs have an effect on response quality, is it mediated by the respondents' management of time and effort?

The rest of this chapter will be divided into three sections. Section 3.3 will describe the data used in this analysis, with special emphasis on: the recruitment, sample composition and field protocols of the BOOST2018 study; the experimental design; structure of the questionnaire; survey paradata collected (including durations per survey screen, timestamps and click counts); and variables of interest used to measure response quality as well as respondent time management. Section 3.4 will follow with a discussion of the results structured around the research questions mentioned previously. Finally, section 3.5 will expand on the implications of the findings for measurement error and survey design, limitations of the analysis and opportunities for future research.

3.3 Data and methods

3.3.1 The BOOST2018 Survey

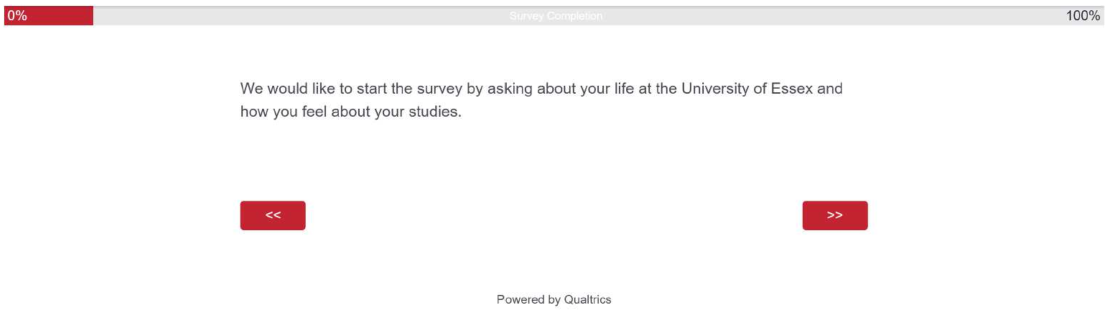
BOOST2018 is an active longitudinal study of first year undergraduate students at the University of Essex. Its main objective is to measure attitudes and behaviours related to study habits and performance as well as academic and employment expectations. Data is collected via online surveys, lab experiments as well as university administrative records (including marks and personal background information given at the time of registration). The first wave was a survey fielded from November 2015 to January 2016, followed by a centrally-located lab experiment in January/February 2016. Two more surveys conducted in March/April and May/June 2016 constitute the third and fourth wave of the study.

The data considered in this analysis comes from the first wave of the study. In total, 1882 participants gave consent to be contacted by BOOST2018 research team out of an eligible 2621 students (72% enrollment rate) by the start of the first wave. The survey was launched on 23 November, 2015 and reminders programmed for 26 November, 2 December, 9 December, 21 December and 9 January, 2016. The survey was closed on 15 January. Because of an unanticipated glitch with the university email server an additional reminder was sent on 27 November. By the end of data collection, a total of 1221 students fully responded to the survey (65% response rate) with an additional 130 partially completing the survey.

A benefit of this sample is that, while not representative of the general population, these students are considerably more homogeneous in terms of age, cognitive ability, computer literacy, email providers/spam filters (all email invites were sent to their corresponding @essex.ac.uk address) and geographical proximity and therefore potentially less prone to selection bias effects. Moreover, as a comparatively more captive audience they demand less fieldwork and data collection resources.

The online survey platform Qualtrics was used for programming of the questionnaire, data collection and fieldwork management (Qualtrics, 2015). The questionnaire was programmed to be compatible with different devices, including: desktop computers, laptops, as well as mobile devices like smart phones and tablets. Qualtrics' default PI was used as the treatment. The PI sits horizontally on top of the page and progresses from left to right, colouring the background from light grey to red. The middle of the bar reads *Survey Completion* while the extreme left and right ends read *0%* and *100%* correspondingly. See Figure 3.2.

Figure 3.2: Progress Indicator



BOOST2018 relies on monetary incentives to promote enrolment into the survey and continued cooperation with each wave. All students who enrolled in the survey panel were given £5 in cash. All eligible enrollees were then invited to participate in Wave 1 and were paid £10 once they completed the survey.

3.3.2 Experimental design

Before fieldwork started, a PI treatment was randomly allocated to 50% of the sample of potential respondents (n=941) after stratifying by: Academic Department, Age group (Young/Mature), Mode of study (Part Time/Full Time), Tuition Fee Classification (EU/Home/Overseas), Sex (Male/Female) and Entered through Clearing (Yes/No). Because of the nature of the experimental treatment and that the survey was conducted online, the stratification aimed to reduce the possibility of sample composition bias with regards to computer literacy, English literacy and computer usage. The other 941 sample participants were administered the same questionnaire and fieldwork protocol without being exposed to a progress indicator.

As evidenced by Table 3.1, sample balance was achieved for the experimental and treatment groups as respondents were distributed proportionally across the strata. Therefore, differences between the treatment and control subsamples are assumed to be independent of respondent selection and instead dependent on exposure to the PI.

Table 3.1: Stratification of experimental groups

	Sampled			Respondents	
	Total	PI	No PI	PI	No PI
Department					
Art History and Theory	0.4%	0.3%	0.4%	0.5%	0.0%
Biological Sciences	9.8%	9.8%	9.7%	10.8%	9.4%
CISH	1.6%	1.6%	1.6%	1.5%	1.8%
Computer Science & Engineering	8.4%	8.3%	8.6%	9.1%	9.9%
Economics	8.8%	8.7%	8.8%	8.0%	8.7%
Essex Business School	14.5%	14.5%	14.4%	11.6%	12.5%
Government	6.6%	6.6%	6.5%	6.5%	6.1%
Health and Human Sciences	2.0%	2.0%	2.0%	2.2%	1.8%
History	5.2%	5.2%	5.1%	5.0%	4.4%
Language and Linguistics	4.3%	4.3%	4.4%	5.0%	5.2%
Law	11.7%	11.7%	11.7%	11.8%	11.5%
Literature, Film, and Theatre Studies	6.6%	6.6%	6.6%	6.6%	7.0%
Mathematical Sciences	2.2%	2.2%	2.1%	2.3%	2.4%
Philosophy	2.6%	2.6%	2.6%	2.8%	2.6%
Psychology	9.3%	9.4%	9.3%	10.1%	10.5%
Sociology	6.1%	6.2%	6.1%	6.1%	6.3%
Age classification					
Mature	8.3%	8.5%	8.2%	9.0%	7.8%
Young	91.7%	91.5%	91.8%	91.0%	92.2%
Study mode					
Full-time	99.7%	99.7%	99.7%	99.5%	99.5%
Part-time	0.3%	0.3%	0.3%	0.5%	0.5%
Fee					
Home & EU	86.1%	85.9%	86.3%	86.7%	87.9%
Overseas	13.9%	14.1%	13.7%	13.3%	12.1%
Sex					
Male	48.0%	47.9%	48.1%	43.2%	44.0%
Female	52.0%	52.1%	51.9%	56.8%	56.0%
Mode of acceptance					
Clearing	21.3%	21.7%	21.0%	20.9%	18.4%
Other	78.7%	78.3%	79.0%	79.1%	81.6%
Base	1882	941	941	602	619

For sampled as well as responding cases, none of the differences between PI and No PI respondents are statistically significant ($p < 0.05$).

3.3.3 Questionnaire

The following features and question types were programmed into the questionnaire in accordance to the research objectives of BOOST2018's principal investigators, but also to allow for the analysis of response quality, speed, burden, satisficing, and time management strategies. What follows is a brief description of these features and question types that were later used to derive the variables of interest of this analysis (See section 3.3.5).

Questionnaire structure

The BOOST2018 Wave 1 questionnaire was comprised of 362 items, nested in 98 questions, divided into 21 separate topical sections, and displayed in a total of 86 separate screens. Six questions contained skip logics which guide the respondent to different sections of the questionnaire depending on the response[1]. As with any Qualtrics survey, there was an introductory landing page preceding the questionnaire as well as an exit page with a short message thanking the responder for participating. In addition, a respondent could go back on a previously answered question by clicking the *back* button, which was present at each of the 85 screens following the introduction page. Going back was only possible between consecutive questionnaire screens; skipping further back to particular sections was not possible and no navigation menu was programmed into the questionnaire. Lastly, a respondent could exit the web browser and come back to the survey at a later time. Upon return, the last visited page would reload.

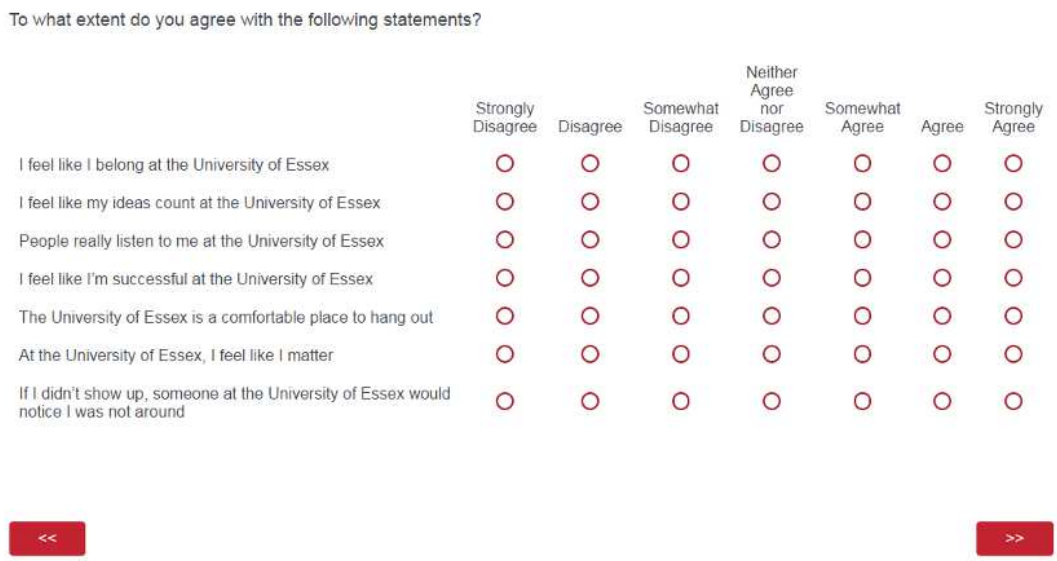
Question matrices and grids

Not all question items had an individually corresponding screen. This is due to the presence of several matrix questions and other question formats where multiple statements were displayed simultaneously due to conceptual proximity between items and/or to ease cognitive burden. Most screens were headed by a single question and then fol-

[1] In other words, not all respondents follow the same exact question flow. However, based on the data collected, exposure to treatment was not associated with any of the skip patterns. That PIs did not have an effect on skip patterns allows us to conclude that any possible response quality or time management effects discussed in the Results section (See 3.4) are not due to different questionnaire flows.

lowed by several corresponding response options. For example, consider Figure 3.3.

Figure 3.3: Example multiple item question



In total, there were 13 grid questions with at least 4 (and up to 7) response statements displayed in a single page and prefaced by a common response scale.

Motivational statements

To curb nonresponse, the principal investigators of BOOST2018 requested that *motivational statements* (Al Baghal and Lynn, 2015; Holland and Christian, 2009; Oudejans and Christian, 2010; Smyth et al., 2009) be programmed into the questionnaire. These statements appeared in a separate pop-up window whenever the respondent left a question blank and clicked on the *next* button. There were two types of motivational statements: *soft* and *hard* response requests.

The first type of statement (soft request) asked the respondent to confirm if they intended to leave the question blank by giving them the option to go back and select a response option or proceed to the next screen without answering. Almost all close-ended questions included a statement with a soft request.

For a smaller number of questions, the motivational statements were accompanied by a hard request prompt. This second type of statement forced the respondent to select a response option before proceeding to the next question. In other words, leaving the question blank was not possible. Depending on the skip pattern, a respondent could be faced with a maximum of seven hard requests.

Lastly, an even smaller number of open-ended questions relating to personal information (i.e. phone number, alternative email address), course assessment, and evaluation of the survey did not have any statements at all. Respondents could leave these open-ended questions blank and not be exposed to any message.

Cumulative sum questions

The questionnaire also included 3 questions asking the respondent to weigh the importance of 4 (and up to 8) related items by typing in numbers that in total must add up to 100. The idea behind the question was to make the respondent think about the relative importance between these related items. Here too a motivational statement popped up whenever a question was left blank or the typed values did not add up to 100 exactly.

Open-ended questions

Excluding demographic questions, there were two open-ended questions in the questionnaire. In section 14 (of 21), students were asked to identify the 4 course modules to which they are currently enrolled by writing out the full name and/or course code (recorded by Qualtrics as a text string). Later in the survey, in the second to last question students were asked to leave a final comment to evaluate their perception of the survey and provide any recommendations to the BOOST2018 team. Here, an essay-style textbox was displayed under the question wording to allow for longer responses.

Item nonresponse options

BOOST2018's investigators also requested that the questionnaire omit explicit item nonresponse / noninformative options. In other words, except for one demographic survey item (asking about parental education), student respondents were not offered *don't know* or *prefer not to say* response options for any question. Rather, item nonresponse was registered only when questions were left blank (which, as previously noted and except for the small number of open-end questions, always prompted a motivational message with a soft request). Thus, item nonresponse was purposely made burdensome (by requiring at least two extra mouse clicks and a motivational statement pop-up) to encourage the respondent to give substantive responses throughout the survey. The investigators' assumption was that the questionnaire's salience with regards to the student experience should yield high rates of informed answers.

3.3.4 Paradata

At each of the 86 pages, clicks and timings data were collected by relying on Qualtrics' *metadata* survey item bank. Additionally, custom JavaScript code was embedded at the beginning of each of the 21 sections to collect full date timestamps. As with the questionnaire features and question types discussed previously, these types of paradata helped derive the variables of interest discussed in section 3.3.5.

Screen durations

Given Qualtrics' default metadata capturing protocol, each page recorded three timings variables: *time to first click*, *time to last click* and *time to page submission* measured in seconds (rounded to the nearest millisecond). The counters for all three started at 0 upon the loading of every new page. *Time to first click* equalled the amount of time between the complete loading of a page and the first click. Any clicking, and not just those that select a given answer option, counted toward this duration. The exception to the rule is the *next* button which recorded the *time to page submission* but did not count as a *first click* or *last click* duration measure. *Time to last click* measured the time between page loading and the last click before a respondent hit the *next* button. Therefore, if a respondent clicked *next* without answering the *time to first click* and *time to last click* both read 0 seconds. Also, if a respondent made just one click (for example on a response option) and then clicked on "next", *time to first click* and *time to last click* would have the same value. *Time to page submission* always recorded the total amount of time spent on a page from the time it finished loading to the time of clicking *next*. When a respondent went back on a question or logged out of the survey, all duration counters were reset to 0.

Timestamps

Timestamp data also captured time and duration information. However, in contrast to screen duration data, timestamps logged the respondents' behaviour through the questionnaire in calendar time (i.e. Friday, January 1st, 2016 09:00:00 GMT+1).

For reasons of data management, timestamps were not recorded at each of the 86 screens but rather at the beginning of the 21 sections of the questionnaire as well as at the last screen of the survey. In other words, there were a total of 22 timestamps recorded per completed interview. As with screen durations, timestamp data were also overwritten when a respondent reloaded (by going back or logging out) any of the 22 screens containing the embedded JavaScript code.

While screen duration data only accounted for screen-level durations as measured by the last interaction with a given survey item, timestamps measured all calendar time between the first time a respondent logged in to the survey until they finished it (as well as all the time between sections).

Click Data

Besides measures of duration, the survey also recorded total number of clicks per page. The tally included all manners of clicking, both inside and outside response markers. In other words, clicking on an empty section of the questionnaire screen counted as an observed click. The only exception was the *next* button which was not added to the total number of clicks per page. Keyboard typing did not count towards click counts or screen duration data but finger tapping on tactile screen devices did and was equivalent to a conventional mouse click.

As mentioned previously regarding motivational statements, when a prompt popped up with a soft request the respondent was given two options: *continue without answering* or *go back and answer the question*. If the respondent chose *continue without answering*, no additional click (or screen duration) data was recorded and the respondent proceeded to the next question. However, if the respondent clicked on *go back and answer the question*, that click counted towards the 1) tally and the 2) first click screen duration. In contrast, a hard request prompt to answer before proceeding to the next question automatically reset all click counts (and screen durations) data to 0. Finally, if the respondent went back or abandoned the survey, all click count (and screen durations) data of the last-visited page were reset to 0.

3.3.5 Variables of Interest

Given the theoretical assumptions and mechanisms proposed in section 3.2, the variables used in this analysis are those that can be assumed to be associated with the related concepts of survey burden and satisficing. More specifically, these variables should identify - at the aggregate level - the extent to which respondents take shortcuts or adopt other burden-mitigating strategies instead of fully engaging with the questionnaire or a survey item.

With regards to satisficing, the specialized literature often relies on similar metrics across a wide variety of survey topics, populations and modes (Barge and Gehlbach, 2012; Holbrook et al., 2003; Kaminska et al., 2010; Krosnick et al., 1996). Authors suggest looking at the frequency of nondifferentiation (or "straightlining"), don't know / refusal answers (or item nonresponse), and response inconsistencies (i.e. when a respondent offers contradictory information in two or more responses within a survey). Additionally, question durations and response speed have also been explored (Callegaro et al., 2009; Malhotra, 2008; Yan and Olson, 2013). All these indicators can be understood as possible attempts at minimizing effort and thus could reflect suboptimal forms of responding. What follows is a description of the variables used in this study.

Response speed

Respondents' strategies for time management can be derived by the speed with which they respond a questionnaire. For example, speedy answers can be the result of deliberate suboptimal responding: very quick durations may be associated with a truncated reading (or complete skipping) of the question/option wording. Conversely, very slow durations may be linked with other forms of suboptimal responding like dithering / hesitation or distracted reading.

This analysis considered question-to-question durations as well as cumulative questionnaire durations measured by screen durations, and specifically by *time to page submission*. Because of the highly skewed durations data and based on similar examples from the specialized literature (Malhotra, 2008; Ratcliff, 1993; Yan and Olson, 2013; Yan and Tourangeau, 2008; Zandt, 2002), median durations were used in this analysis to avoid excessive influence of outliers. Overly long question durations may not necessarily represent the amount of time a respondent is engaging with a question. Instead, these are most likely due to a browser windows left unattended by a respondent who drops off or temporarily abandons the survey.

Statistically, median point estimates were compared between the treatment and control groups to assess PI effects. Additionally, simultaneous quantile regression models were included to account for the possible effects of progress indicators on overly long response durations. Briefly, simultaneous quantile regression relies on conditional quantile (rather than conditional mean) estimation. This type of model allows the exploration of correlations at several sections of a dependent variable's distribution (including those near the median as well as those nearer to the tails of the distribution) by fitting multiple equations at given quantiles. Because of the need to fit separate equations (corresponding to the different quantiles specified in the model), the variance-covariance matrix is estimated via bootstrapping (Gould, 1998; Hao et al., 2007; Koenker and Bassett Jr, 1978; Koenker and Hallock, 2001).

Multiple sessions

A second form of time management is the completion of the questionnaire in multiple sessions. As already noted, BOOST2018 allowed respondents to log out mid-questionnaire and log back in at a later time, picking up from the last screen visited. Unfortunately, BOOST2018 did not record log-ins or log-outs: no variable in the dataset captured if/when a respondent exited the survey or returned to it[2].

[2] Even if these data were recorded, interpreting them is not without its challenges. Initially, one could determine that so long as a respondent does not actively log out of the survey by closing the browser window, the session remains the same. However, given the ability for devices to load multiple browser windows simultaneously, this definition conflates sessions which are inadvertently left open with those which are being actively worked on. Furthermore, closing a browser does not necessarily mean a respondent wishes to leave the survey. Instead it could mean that they are switching browsers or devices midstream. Or perhaps that the respondent mistakenly clicked on the wrong icon while attempting to move, resize or minimize the window.

Instead, for this analysis, multiple sessions were identified based on timestamp data. Because they measured all calendar time transpiring between different section markers, in this analysis timestamp data were used to capture not just the time spent actively engaged with the survey but also instances of *survey downtime*. Therefore, whenever a respondent spent far longer in a given section of the questionnaire than a typical respondent it was assumed that he/she likely had logged out of the survey. As part of the analysis, different thresholds for "far longer" were tested and crosstabulated with the experimental conditions to assess PI effects on the tendency to answer the survey in one or multiple sessions.

Click counts

As proxies for respondent effort and response quality, click counts may be driven by different mechanisms. Stieger and Reips (2010) proposed using click data to investigate different respondent behaviours that hamper data quality. Similarly, Heerwegh and Loosveldt (2002) studied questionnaire trajectories, response option formats (including radio buttons vs. drop-down menus), and backing up to previous questions by examining respondents' click data. This analysis uses click data to investigate respondent effort and response quality.

For example, a more diligent respondent (i.e. one who does not satisfice) is likely to click more times in questions that require additional mechanical and/or cognitive effort than a less diligent one (i.e. one who rushes through the questionnaire and satisfices). For example, in BOOST2018, cumulative-sum and open-ended questions ask the respondent to click on as many open-text and numeric input boxes as apply to them before typing in the answer. The less diligent respondent would likely click on as few as necessary in order to save time and effort, while the more diligent respondent will click on as many as apply to him/her. Moreover, one can assume that cognitively challenging questions (like those requiring mathematical calculations or added recall effort) will cause more diligent respondents to hesitate and change their answers at a higher rate than less diligent respondents who are comparatively more likely to satisfice and stick to the original response option.

Of course, the diligent respondent may alternatively click less frequently than their counterpart if one considers that rushing through a questionnaire is likely to result in more inaccurate clicking (i.e. missing the response marker) and therefore require additional "corrective" clicking. Similarly, the less diligent respondent may pre-emptively and repeatedly click the "next" upon the loading of each screen in an attempt to bypass as many questions as possible.

Nevertheless, given 1) the questionnaire features previously discussed (particularly the motivational statements) and 2) that PI and No PI respondents did not, in the aggregate, route through the questionnaire in systematically different ways, this analysis assumed that decreased clicking is more likely a byproduct of reduced effort from less diligent respondents and not more accurate / efficient clicking from more diligent respondents. In other words, because of the item nonresponse curbs programmed into the questionnaire, this analysis assumes that all respondents are forced to devote a minimum amount of effort to allow them to properly navigate the survey (even if just to satisfice).

This analysis relied on question-to-question click counts as well as cumulative click counts over the entire questionnaire as a proxy for respondent effort.

Item Nonresponse

Item nonresponse has long been considered in investigations of survey response quality. It is often investigated as a function of differential burden or propensity to satisfice (Barge and Gehlbach, 2012; Kaminska et al., 2010; Krosnick, 1991; Leeuw et al., 2003; Sharp and Frankel, 1983).

For any given survey item in BOOST2018, leaving a question blank may legitimately represent the true answer of the respondent. At times, the respondent does not know, has no opinion or can't otherwise answer a given question. However, a respondent may have other reasons for not responding to a survey item. For example, one may choose not to respond in the interest of saving time and/or not fully engaging with a given question. This bypassing of the response process constitutes another form survey satisficing. Thus, if choosing an option demands a certain amount of cognitive effort, overburdened or otherwise unmotivated respondents may bypass the full steps of reading, understanding, judging and/or answering a question and leave a question blank.

This analysis estimated item nonresponse rates as the proportion of questions left unanswered / or where a noninformative option was used over the number of applicable questions presented to the respondent.

Response Length in Open-Ended Questions

As an alternative to not responding, a respondent wishing to mitigate burden may shorten his/her answers to open-ended questions by typing fewer words (Barrios et al., 2011; Denscombe, 2008; Emde and Fuchs, 2012; Ganassali, 2008; Kaczmirek et al., 2017; Reja et al., 2003; Zuell et al., 2015). Thus, the number of characters used in responses to these types of questions is another proxy for respondent effort. For this analysis, response length was calculated as the sum of characters used in typing an answer.

Inconsistencies

Because overburdened respondents devote less cognitive effort to their responses, they are more likely to contradict themselves or otherwise respond inconsistently across different questions of the questionnaire (Kaminska et al., 2010; Krosnick et al., 1996). In the first wave of BOOST2018 there were three opportunities for respondents to report contradicting or inconsistent information.

Early in the questionnaire, the respondent was asked about their expected final marks in sequential questions. In the first question, the respondent was asked to type a number between 0 and 100. Immediately after, the respondent was asked to estimate the probability of earning either a First, 2:1, 2:2, a Third or a Failing grade by plotting a distribution chart against all classifications. For the respondent to be consistent, the expected final marks reported should correspond with the median of the distribution.

Similarly, there was a pair of questions about study habits. In the first question, the respondents were asked how many hours a week they devote to a list of different activities, including *testing oneself*. In the second question, the respondents were asked about the reasons why they may test themselves. Here, one of the options is *I don't test myself*. To answer consistently, one would expect respondents to either state that they spend 0 hours testing themselves and then choose the corresponding option (i.e. *I don't test myself*); or, if they claim to spend at least 1 hour testing themselves per week they should not select the *I don't test myself option* later on.

Finally, section 13 of the questionnaire asked respondents to fill in a time diary of activities during their last weekday in a matrix composed of 24 1-hour slots. The list included 16 different possible activities (such as sleeping, eating, shopping, studying, etc.). Because the University of Essex does not schedule classes between 8pm and 8am, a respondent was inconsistent if he/she selected the options *class* or *lecture* during this time period.

Nondifferentiation

Another common form of satisficing is nondifferentiation (also called straightlining) (Kaminska et al., 2010; Krosnick et al., 1996). Faced with a grid of radio buttons mapping scale positions with the corresponding response option, a respondent may straightline by giving the same answer within a single column in an attempt to save clicking and scrolling effort and likely spend as little time as possible answering the question. Again, it is possible (and probable) that choosing the same scale option for a battery of related items constitutes the true answer of a given individual respondent. However, given the experimental design, differential subsample-level propensities for straightlining could be indicative of PI effects.

Because of routing in the BOOST2018 questionnaire, every respondent had the opportunity to straightline between 0 and 13 times. In this analysis, a respondent was defined as a having straightlined if he/she chose the same response option for all items in at least one grid question (conditional on being presented with at least one of such grids).

Single 100s in cumulative sum questions

Because they contain motivational statements with hard response requests, in cumulative sum questions typing a single '100' in just one of the options requires the least amount of effort. Doing so precludes the respondent from having to type in additional numbers for other response options and eliminates the necessity for arithmetic calculation. Therefore, these types of answers can be the result of survey satisficing. For this analysis, typing '100' in at least one of the three cumulative sum questions was considered satisficing.

3.4 Results

3.4.1 Does the presence of a PI have an effect on the respondents' management of time and effort?

Response speed

As Table 3.2 shows, the median duration as measured by timestamp data, shows that the No-PI median respondent spends 48:49 (48.82) minutes on the survey while the PI respondent spends 54:13 (+5.4). However, this difference is not statistically significant. Instead, to the extent that there are significant differences, they are observed in the longer tails of the distribution: larger overall durations are observed among those exposed to a PI. Indeed, the differences become not only significant but substantial at the 95th and 99th percentile with differences close to 33000 minutes (or almost 23 calendar days) and 18000 minutes (12 days) respectively. While no significant differences are observed at the 50th and 75th percentiles, the direction of the effect remains constant throughout: PI respondents tend to take much longer with the survey among those above the median duration. On the opposite end of the distribution, speeders are not necessarily influenced by presence of a PI. There are no significant differences found in the 1st, 5th or 25th centiles of duration as measured by timestamp data.

While timestamp data evidence a treatment effect on the longer ends of the distribution, screen durations do not seem to be affected by the PI in either ends of the distribution. In fact, the median duration differences between the PI and No PI subsamples are nonsignificant and only amount to 1 minute. Unlike with timestamp data, there does not appear to be a significant effect observed in the longer tails of the question screen duration distributions.

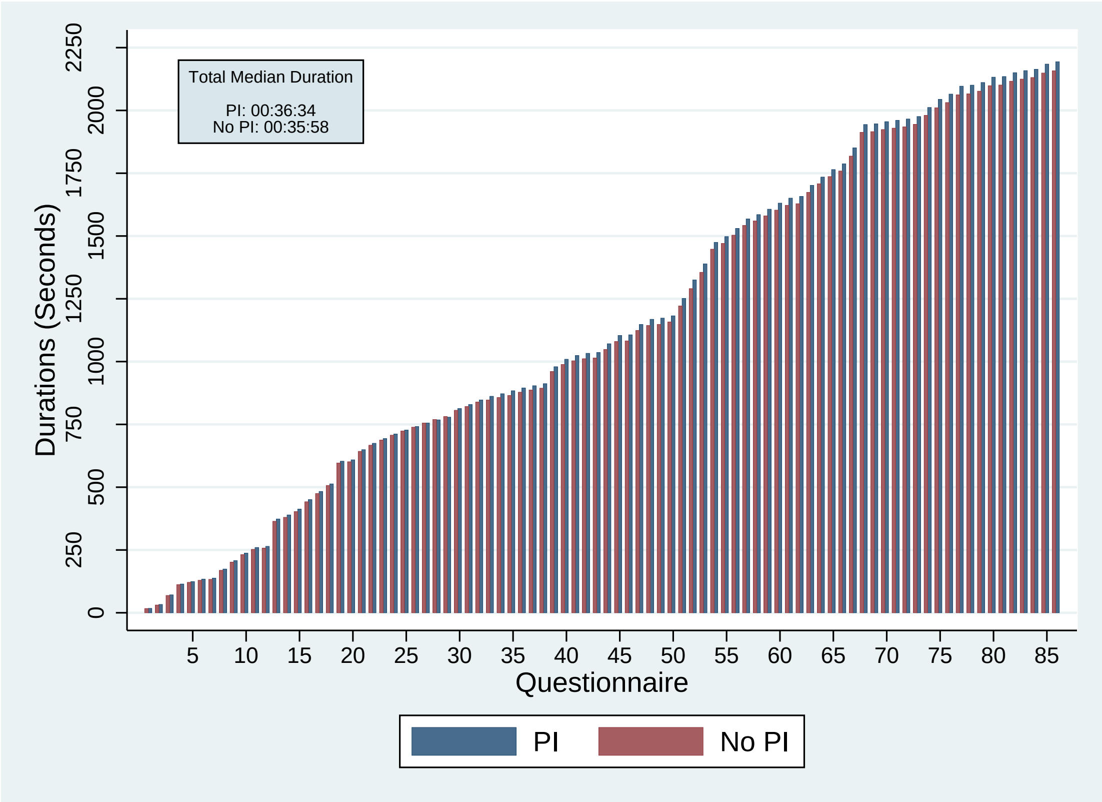
Table 3.2: Quantile Regression. Duration in Minutes

Simultaneous quantile regression*. Bootstrapped SEs. 200 iterations.		Number of observations = 1221	
		Timestamp Data	Screen Duration Data
Quantile (1%)	Treatment (PI)	3.18	1.68
	Constant (No PI)	16.82	12.69
Quantile (5%)	Treatment (PI)	1.68	1.92
	Constant (No PI)	24.32	19.53
Quantile (25%)	Treatment (PI)	0.12	0.19
	Constant (No PI)	36.83	29.44
Quantile (50%)	Treatment (PI)	5.40	1.00
	Constant (No PI)	48.82	38.00
Quantile (75%)	Treatment (PI)	193.18	-0.03
	Constant (No PI)	159.07	55.51
Quantile (95%)	Treatment (PI)	32747***	41.89
	Constant (No PI)	10453.48	266.23
Quantile (99%)	Treatment (PI)	17895.97*	395.48
	Constant (No PI)	43201.68	1352.65
			** p<0.05, ** p<0.01, *** p<0.001*

*Treatment (PI) values represent the estimated additive effects with respect to the constant (No PI) values, their significance is reported whenever the difference between PI and No PI effects is different than 0. R^2 values for each of the quantile regressions (which never add up to more than 0.08) are suppressed from the table to aid interpretation of the treatment effects. All Constant (No PI) terms (representing estimated durations for non-PI quantiles) are significantly different from 0 at p<0.001.

If total durations as measured by screen duration data are not affected by PIs, could the same be said about individual screen durations? In other words, what is the speed of BOOST2018 respondents as they progress throughout the questionnaire? Are these speeds influenced by exposure to the treatment? Consider Figure 3.4.

Figure 3.4: Cumulative Median Question Durations



Within each experimental group, median question durations are estimated for each of the 86 screens and summed cumulatively starting with the first screen.

By plotting the cumulative median durations of the 86 screens of the questionnaire of the treatment and control subsamples the speed patterns of both can be assessed and compared. At first glance, what is most apparent is the very similar shapes of the distributions: both subsamples display a constant climb punctuated by periods of slight increases in slope. There are no apparent tendencies for rushing or slowing down as a function of treatment. This steadiness of the speed becomes more evident once the questionnaire is broken into quarters: by screen 21, approximately 600 seconds have transpired; by screen 43 1000 seconds have gone by; 1750 seconds by screen 65; and finally, 2200 seconds (or 36:36 minutes) by screen number 86. It should be noted that there is a slight difference observed around screen number 33. In fact, the PI sample's cumulative median duration remains higher for the remainder of the questionnaire. However, the difference is very slight and only amounts to 36 seconds by the end of a survey. Considering a total median screen duration of approximately 36 minutes, this represents less than a 2% difference. In sum, screen durations are proportionally distributed throughout the questionnaire and are largely unaffected by PI presence.

PIs are not correlated with screen-level response speed differences but are nonetheless associated with a propensity for longer fieldwork durations within the long tail of the distribution. More specifically, among those respondents who spend considerably longer with a survey, PIs substantially increase the amount of time between survey start and survey completion. Given that these long durations are measured in multiple hours (if not days), it is sensible to assume that these types of respondents are distributing their interactions with the survey into more than one session. This last point explains the large differences observed between overall durations as measured by screen time vs. timestamp data: as already discussed, screen durations measure duration spent during an active session; timestamps log all time (including downtime) spent from the beginning of the survey, between sections, and until completion.

It seems that among those that take a very long time with the survey, respondents exposed to a PI allow for longer periods of downtime than those not presented with a PI. To elaborate on this assumption, the next section discusses possible PI effects on the propensity for multiple sessions and extended periods of survey downtime.

Multiple sessions

As previously discussed, this analysis assumes that a respondent is not actively engaging with the questionnaire whenever he/she spends an overly long amount of time in completing a section the survey. After a certain section duration threshold, it becomes more likely that the respondent is not actively responding to the survey but rather that he/she has taken a break from it.

To illustrate how multiple sessions may be identified from timestamp data consider the following example. In Table 3.3 timestamps from a single respondent are used to measure the time spent completing each of the different sections of the questionnaire. In addition, summary statistics from the entire sample (median, mean, and standard deviation) are included for reference.

Table 3.3: Identifying multiple sessions from timestamps

	Example Respondent		Pooled Sample Statistics		
Section	Timestamp	Duration	Median	Mean	s.d.
1	23 Nov 2015 18:47:02	0:00:26	0:00:19	1:10:43	17:00:07
2	23 Nov 2015 18:47:28	0:02:41	0:02:21	3:11:50	9:49:40
3	23 Nov 2015 18:50:09	0:01:07	0:00:38	0:24:21	7:01:04
4	23 Nov 2015 18:51:16	0:00:27	0:00:34	0:24:32	10:37:50
5	23 Nov 2015 18:51:43	0:00:36	0:00:31	1:11:51	9:10:00
6	23 Nov 2015 18:52:19	0:00:23	0:00:23	1:10:38	9:10:46
7	23 Nov 2015 18:52:42	0:06:31	0:03:57	4:10:59	4:43:13
8	23 Nov 2015 18:59:13	0:01:04	0:01:13	1:50:13	22:17:46
9	23 Nov 2015 19:00:17	16:07:36 (+50 days)	0:05:39	4:31:53	4:22:37
10	13 Jan 2016 11:07:53	0:02:31	0:02:18	0:11:54	2:36:28
11	13 Jan 2016 11:10:24	0:01:28	0:01:18	1:03:47	18:49:04
12	13 Jan 2016 11:11:52	0:06:26	0:06:22	1:39:40	19:44:49
13	13 Jan 2016 11:18:18	0:03:18	0:03:17	0:55:39	21:29:52
14	13 Jan 2016 11:21:36	0:03:30	0:02:51	0:18:57	3:52:43
15	13 Jan 2016 11:25:06	0:00:12	0:00:14	0:01:12	0:19:27
16	13 Jan 2016 11:25:18	0:00:06	0:00:07	0:00:09	0:00:18
17	13 Jan 2016 11:25:24	0:00:14	0:00:18	0:00:38	0:06:24
18	13 Jan 2016 11:25:38	0:02:03	0:02:19	1:17:31	11:20:06
19	13 Jan 2016 11:27:41	0:00:51	0:00:50	0:01:43	0:18:51
20	13 Jan 2016 11:28:32	0:00:41	0:01:07	0:01:38	0:03:36
21	13 Jan 2016 11:29:13	0:00:40	0:00:37	1:11:13	5:07:39
End Page	13 Jan 2016 11:29:53	N/A	N/A	N/A	N/A

Throughout most of the sections of the questionnaire the exemplary respondent tracks the duration of the median respondent. However, in section 9 the respondent strays considerably from the median. Almost 51 days pass between the time the first screen of the section loads to the time the last answer of the section is submitted. Clearly, this respondent did not spend all that time actively engaged with the survey. Instead, one can assume that he/she likely took a break and completed the survey in (at least) two sessions.

While it is sensible to assume that section durations over 51 days likely entail multiple sessions, it is not immediately obvious when an overly long duration is indicative of survey downtime (and not of slow responding). This analysis experimented with multiple thresholds to test PI effects on propensity to complete the survey in more than one session. Briefly, if a respondent took longer than any of the different thresholds in completing at least one section of the questionnaire he/she was defined as having stopped engaging with the survey. In total 7 different thresholds were tested against the experimental conditions (See Table 3.4).

Table 3.4: Sensitivity Analysis: Thresholds for Logging out

Threshold Definition	PI	No PI	Total
30 minutes	27.29	24.88	26.07
1 hour	23.79	20.68	22.21
2 hours	19.63	15.67	17.62
3 hours	17.8	14.7	16.23
6 hours	14.64	11.31	12.95
12 hours	12.98	9.69	11.31
24 hours	10.32	7.43	8.85
1 week	5.32**	2.42	3.85
* *p<0.05*, ** *p<0.01*, *** *p<.001*			

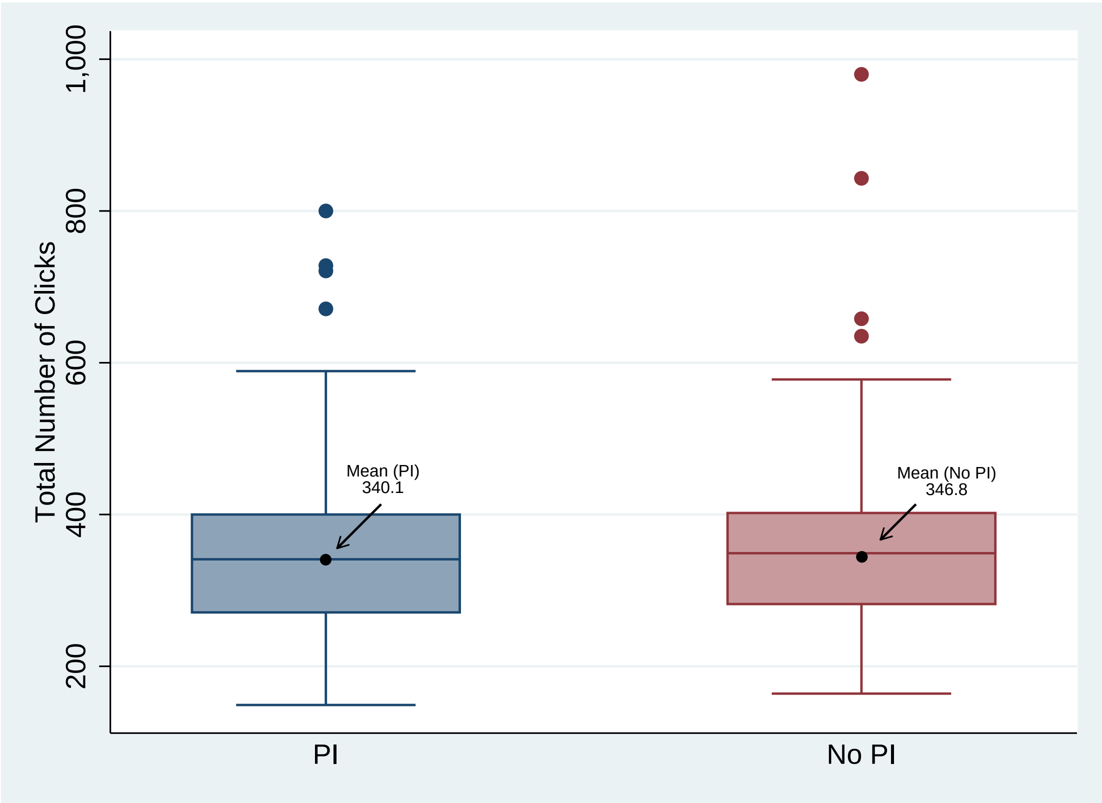
Cell entries represent proportion of respondents who spent at least 30 minutes / 1 hour / 2 hours / 3 hours / 6 hours / 12 hours / 24 hours or 1 week in at least one of the sections of the questionnaire. Significant differences are tested between the experimental groups.

Evidence of PI effects on propensity for logging out of the survey is weak. The only significant difference is observed within the small proportion of the sample that spend 1 week or more completing any section of the questionnaire. However, this finding alone does not suggest that PIs increase the respondent's likelihood to complete a session in multiple sessions. In fact, for the threshold durations between 30 minutes and 1 full day, no PI effects were observed. When one considers the median durations reported in table 3.3 against the marginal totals of table 3.4 it is fair to assume that some BOOST2018 respondents (at least 26.07%) likely completed the survey in more than one session, but their decision to do so was not affected by the presence of a PI.

Respondents' click patterns

As illustrated in figure 3.5, PI and No PI respondents click on average 340.1 and 346.8 times per survey. This difference is not significant (p = 0.2052). What's more the mean clicks-per-survey distributions of the experimental subsamples are very similar. In other words, there do not seem to be any effects of PIs on overall clicking effort.

Figure 3.5: Total Clicks Per Survey



Box-whisker plot of total number of clicks by treatment groups. Middle box represents the distribution between the 25th and 75th percentile, with the middle line plotting the median value. Additional mean scores have been highlighted by a black dot and accompanying text. Length of the whiskers is calculated as 1.5 times the value of the inter-quartile range.

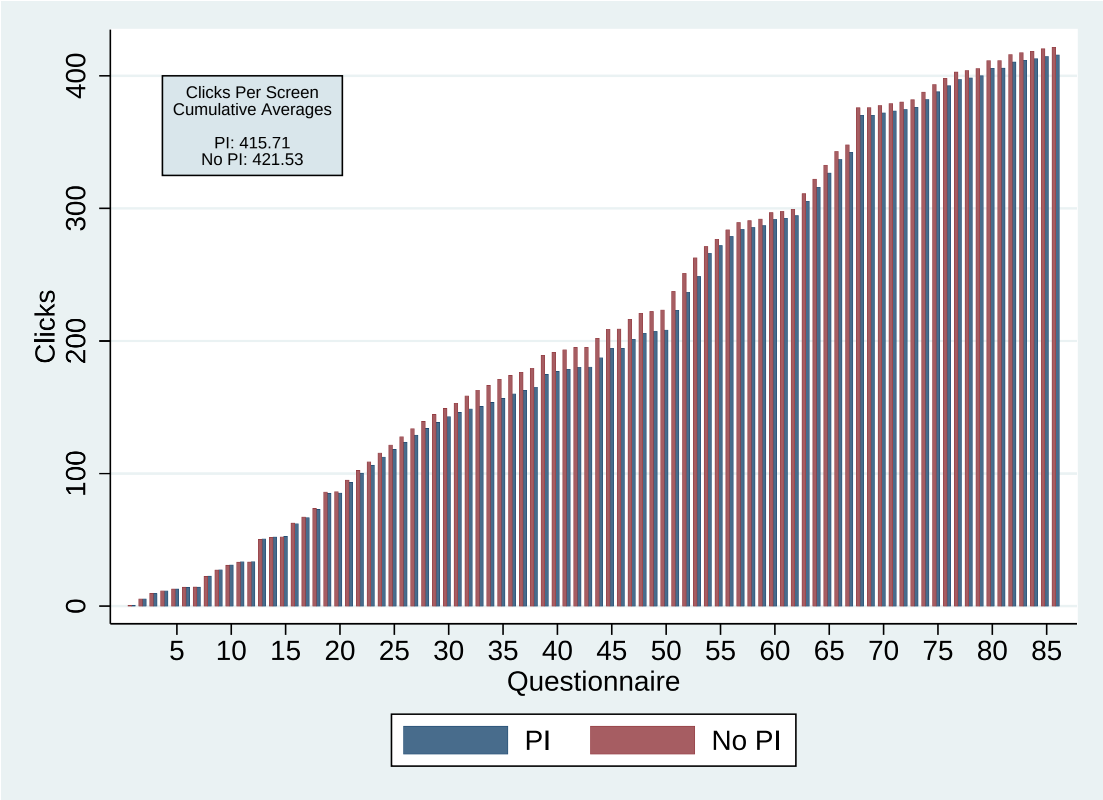
Figure 3.6 plots the cumulative mean number of clicks per screen corresponding to all 86 screens of the BOOST2018 questionnaire. These screen-level clicking patterns do not seem to be affected by PI presence. Overall, the graph's climb is fairly even and is only marked by periods of slight slope adjustments[3]. In addition, the total difference is very small: PI respondents' cumulative mean clicks per screen is 416 while those of No PI respondents is 422[4]. Said differently, there is no evidence to suggest that effort is affected by PIs at the individual screen level. Not only is the overall effort comparable

[3]There is a visible gap between questions 33 and 53. However, this section coincides with a couple of skip questions with very small cell sizes (n = 21) whose means (while not significantly different) create a large differences between the cumulative means of both experimental groups. The gap is reduced almost entirely by question 54 (another skip question, n = 12) where PIs average 9 more clicks than No PIs).

[4]While conceptually similar, the mean of total clicks per survey (See Figure 3.5) and the cumulative mean of total clicks per screen (See Figure 3.6) are not similarly derived. While the mean total of clicks per survey is calculated by dividing the total number of clicks by number of respondents, the cumulative mean of clicks per screen calculates the mean number of clicks per screen and then sums them up across all 86. Because of skip patterns and given that not all respondents go through all 86 screens, this second metric (corresponding to a *questionnaire-level* calculation) is necessarily going to be higher than the first (corresponding to a *respondent-level* calculation).

between the experimental groups, no differential effort as a function of question type or question placement is evident from the data. This null finding suggests that PIs do not affect a respondent's likelihood of engaging in effort-saving strategies.

Figure 3.6: Clicks Per Screen. Cumulative Averages



Within each experimental group, mean number of clicks are estimated for each of the 86 screens and summed cumulatively starting with the first screen.

3.4.2 Does the presence of a PI have an effect on the response quality of observed responses?

Thus far, there is no evidence to suggest that PIs affect response speed, time management or clicking effort. What remains to be tested is the possibility for response quality to be affected by exposure to treatment.

Item nonresponse in open text questions

Exposure to PIs does not have an effect on the respondent's propensity for leaving open-ended questions blank (See Table 3.5). There are no differences observed in the proportion of students who left at least one 1) module name or 2) mark unanswered in the corresponding questions (section 14) nor in the proportion who left the 3) final comment question blank (section 21). These kinds of questions require comparatively more effort than closed ones: in the absence of ready-made options, the respondent is required to formulate an answer and then type it. Furthermore, and as previously stated, BOOST2018 did not force or request completion of open-ended questions left blank upon the first attempt: no motivational statements were programmed for these questions. Thus, the lack of any observable effect suggests that PI does not affect a respondent's effort or his/her perceptions of burden with regards to these types of questions.

Open text character length

Similarly, Table 3.5 reports no effect in the number of characters devoted to the final comment among those respondents that decide to leave a comment. On average, respondents leave comments 86 characters long irrespective of treatment. Presumably, if there were different perceptions of burden conditional on the information interpreted from a PI, the amount of effort devoted to typing a comment should translate into different number of characters. However, none of this is evident from the data.

Nondifferentation

No PI respondents do not seem to straightline more often than PI respondents (See Table 3.5). Furthermore, typing a single '100' (instead of multiple numerical amounts adding up to '100') is as equally likely to occur within PI and No PI respondents. Here too differential effort/burden is not confirmed given the comparable quality of responses between the treatment and control groups.

Response consistency

Lastly, presence of PI has no effect on the rate of inconsistent answers. If burden makes respondents more careless and assuming PIs have an effect on perception of burden, it follows that the likelihood of giving contradictory answers should increase when observed at subsample levels. However, response quality is also unaffected by PI when measured in terms of respondent consistency.

Table 3.5: Response Quality Indicators

	PI	No PI
Left at least 1 module name blank	8.8	7.11
Left at least 1 module mark blank	55.8	56.9
Left final comment blank	40.86	41.84
Character mean in final comment (Conditional on leaving a comment)	87.03	85.61
Straightlined at least once	30.90	34.89
1+ '100' in matrix question	18.60	18.90
Final marks inconsistent	0.83	0.32
Test oneself inconsistent	1.33	1.29
Class / Lecture between 8pm and 8am	3.32	2.42
Number of observations	602	619

*Except for number of observations & character means, quantities expressed in %. * p<0.05, ** p<0.01, *** p<0.001*

3.4.3 If PIs have an effect on response quality, is it mediated by the respondents' management of time and effort?

In section 3.2, this chapter suggested that PI effects are driven by three different mechanisms. The first mechanism posits that PIs cause people to drop off if the perception of burden interpreted from them exceeds the maximum amount of effort the respondent is willing to devote to the survey. This first mechanism informs most of the literature on PI effects.

As a theoretical alternative, this chapter proposed that in addition to propensity to drop off, PIs could still affect those respondents who stick with the survey. As a result, this chapter argued, PI presence / absence would result in differential perceptions of burden among otherwise cooperative respondents. From this point the last two mechanisms emerge: variations in burden perceptions would drive some respondents to satisfice and others to complete the survey in multiple sessions.

These two mechanisms suppose a mediating relationship (Baron and Kenny, 1986): PIs affect burden which in turn affect time and effort management which in turn affect response quality. For these theoretical assumptions to hold, evidence of these two mechanisms would be observed in the respondent's speed, clicking effort, response quality and tendency to log out of the survey. Furthermore, response quality would have to be mediated by the respondents' time and effort management strategies (i.e. response speed, propensity to complete the survey in multiple sessions and click counts). In terms of statistical estimation and given the experimental design of this study, this would entail that 1) PIs would have a significant effect on response quality 2) PIs would have a significant effect on time and effort management 3) in a multivariate analysis, where PI and time / effort management are associated with response quality PI effects would cease to be significant but time / effort management would continue to remain significantly associated with response quality. Sections 3.4.1 and 3.4.2 have shown that neither of the first two conditions are backed by the evidence and therefore no further analysis is warranted.

3.4.4 Additional considerations: PIs and longitudinal nonresponse

Data from BOOST2018 Wave 1 show that response rates, response quality and time- and effort-management strategies are largely unaffected by the inclusion of a PI. However, one additional analytical consideration remains: does PI presence have an impact on future wave nonresponse? In other words, even if data quality is not affected within the same wave of the inclusion of the PI, does PI presence have a delayed effect on response quality and/or time- and effort-management in the form of future wave nonresponse? Table 3.6 shows that for Waves 2, 3 and 4, presence of PI at Wave 1 does not have a differential effect on rates of full response, partial response or nonresponse.

Table 3.6: PIs and Longitudinal Response

		PI at Wave 1 (602 cases)	No PI at Wave 1 (619 cases)
Wave 2	Non response	31.4%	29.9%
	Partial response	0.3%	0.5%
	Complete response	68.3%	69.6%
Wave 3	Non response	16.9%	17.6%
	Partial response	2.7%	2.7%
	Complete response	80.4%	79.6%
Wave 4	Non response	23.8%	26.3%
	Partial response	6.6%	4.5%
	Complete response	69.6%	69.1%
*Quantities expressed in %. * p<0.05, ** p<0.01, *** p<0.001*			

3.5 Discussion

3.5.1 Response quality & measurement error

PIs are very common in surveys today. However, relatively little is known about their effect on data quality. This analysis has found no evidence to suggest PIs affect how fast respondents answers a survey (or specific questions), how they negotiate their time with the overall allotted fieldwork duration, how much effort they devote to the survey or the quality of their responses.

The speed with which respondents answer particular items or the overall duration a respondent spends on the survey is seemingly unaffected by PIs. This is true for two conceptualisations of time: active time (as measured by screen-level duration data) as well as calendar time (as measured by timestamp data which includes not just active time but also downtime). The only exception to this last point remains for respondents who took a break longer than 1 week[5]. However, this finding alone does not support the theoretical mechanisms discussed previously. In fact, no strong evidence of PI effects on the propensity to complete the survey in multiple sessions was found. Evidence of PI effects on respondent effort, measured in number of clicks, was also absent. This is true for overall click means as well as within questionnaire screen clicking. Lastly, when considering conventional metrics for satisficing effects on response quality, PI presence did not result in any significant differences. In other words, there is no evidence of measurement error attributable to PI presence.

Given these largely null findings, the question remains: should we (continue to) use progress indicators? One could be tempted to recommend that they be used at the discretion of the survey practitioner. Nonetheless, future research should be explored

[5] As previously stated a total of 6 reminders were programmed into BOOST2018. These reminders were sent to all respondents (irrespective of experimental condition) who had not completed nor opted out of the survey by the time the reminders were scheduled to be sent. Thus, a possible explanation behind the significant difference in proportion of respondents who took 1 week or more to complete a given questionnaire section is that reminders have a differential effect on survey completion conditional on PI presence. More specifically, no PI respondents could be less likely to come back after longer breaks because of differential (and differentially enduring) perceptions of burden. Because reminders were not experimentally manipulated, and given the wording on field duration on the communications sent to the respondent, this hypothesis cannot be tested with the data as collected by BOOST2018. Future research could experiment not just with PI presence but also reminder presence to further examine this hypothesis.

to further test the theoretical assumptions set forth in this study as well as address some of its limitations.

3.5.2 Limitations

The analysis presented here has some limitations worthy of discussion. Firstly, while the sample's homogeneity reduces the possibility for selection bias, it also produces estimates that cannot be directly extrapolated to the general population. Students are younger, more educated, more computer literate and are likely to spend more time on the computer than the average respondent. Additionally, one can assume that students are considerably more motivated by the subject of BOOST2018 than the average respondent is to the average survey. Therefore, to the extent that the findings discussed here can be generalized, they remain to be confirmed in more heterogeneous populations and for less salient survey topics. As one example, one could assume that burden is perceived by other subpopulations differently (i.e. less educated, less computer literate individuals) and thus PI effects could be more obvious among different types of respondents[6].

Secondly, the response incentive structure of this survey is not typically replicated in other surveys and likely caused higher-than-usual levels of survey cooperation (the 72% enrolment and 65% response rates are especially high for a survey of young respondents). The monetary incentives (£5 for signing up, £10 for participating in the first wave and a total of at least £100 upon completion of all waves) in addition to a very active and visible enrolment campaign (extended over several weeks and relying on heavy institutional support from the university) likely created respondents who are very loyal to the survey. Therefore, there is a considerable possibility that perceptions of burden and careless responding were overpowered by this loyalty. Another opportunity for future research would be to test PI effects on less motivated / loyal respondents.

Thirdly, this survey included motivational statements and request prompts aimed to

[6]Unfortunately, no research on PIs has been conducted on samples representative of the general population. In fact, analyses have relied on university students (Bohme, 2011; Couper et al., 2001; Crawford et al., 2001; Heerwegh and Loosveldt, 2006) or respondents recruited from opt-in access panels, online banner ads, printed fliers or specialist online communities (Conrad et al., 2005, 2003a,b, 2010; Kaczmirek, 2008; Matzat et al., 2009; Peytchev, 2009; Yan et al., 2011).

curb item nonresponse. As previously stated, these were a requirement of BOOST2018's principal investigators who wished to make this kind of nonresponse burdensome and instead incentivise substantive responding. Because the theory behind this analysis is that burden is the mechanism by which response quality and time and effort management are affected, one could assume that the findings of this analysis may differ for surveys designed with less burdensome features. Indeed, these features could have inadvertently resulted in increased perceptions of burden and/or drop-off propensities that in turn could have overpowered any direct or indirect PI effects. If this experiment was to be replicated, one could test these effects in the absence of any of these prompts which likely impact on the respondent's motivation to continue responding or giving quality answers.

Fourthly, the absence of an indicator of "survey downtime" posed a challenge to the analyses discussed here. More specifically, not being able to identify when: 1) a respondent disengaged from the survey or 2) a new survey session was initiated, forced us to consider alternative proxies for "survey downtime" based on timestamp data. In future, additional paradata and survey features could be programmed into the survey to better capture survey downtime and other forms of respondent disengagement. For example, browser minimization / closures could be incorporated into the client-side paradata to be collected. Further, extended periods of mouse, keyboard, or touch screen inactivity could also prompt the appearance of a motivational statement inviting the respondent to continue responding or consider dropping off momentarily and come back at a later time.

In addition, the paradata collected for this analysis was not able to capture when (how, and for how long) respondents backed up in the questionnaire. Instead, as previously explained, the screen durations, timestamp and click data (as well as the responses themselves) only reflected the final response process at each screen / questionnaire section. Thus, the data considered here offers only partial insight into the time- and effort-management strategies as well as the perceptions of burden (cognitive and otherwise) of the respondent. In large part, this is a function of the data structure and volume constraints put on the paradata collection to make the analytical datasets manageable.

Future research could explore new forms of paradata collection that include not just the information gathered at the last visit of a questionnaire screen or section but also all the different pathways considered by the respondent as well as any hesitation reflected in changed response options.

Arguably, PI effects may be stronger in devices with smaller screens (such as smartphones or small tablets) than in devices with larger screens (like desktop or laptop computers). However, this analysis does not consider the potential effects of type of device used to complete the survey. To consider these types of effects paradata collecting device type would have needed to be collected at each of the 86 screens of the questionnaire. Unfortunately, this was not possible as it would have resulted in too large a dataset and likely hampered the loading and navigation of the online questionnaire. Instead, device data (as with timestamp data) was only collected at the beginning of each of the 21 sections of the survey as well as the last screen. At most, overall response rates based on device type and PI presence can be derived from the data collected. For example, Table 3.7 shows that PI does not have an effect on drop out based on device type. Device switching is also unaffected by PI presence. Future analyses should consider PI effects on response quality and respondent time- and effort-management strategies as mediated/moderated by device type.

Table 3.7: Device Type and Response Rates

	Started on a mobile device		Started on a computer	
	PI	No PI	PI	No PI
Ended on a mobile device	74.88	74.87	0.44	0.2
Ended on a computer	11.59	9.23	91.7	92.65
Dropped out	13.53	15.9	7.86	7.14
Total	100	100	100	100

*Quantities expressed in %. * p<0.05, ** p<0.01, *** p<0.001*

Lastly, BOOST2018 is a very long and complicated survey. A survey of reduced length and complexity would be easier to complete and induce different perceptions of burden. Another possible avenue for future research is to test PI effects on time and effort management and response quality in questionnaires of differing difficulties and lengths.

Conclusion

This dissertation discussed three practical applications of paradata in survey design and analysis. Beyond the particular objectives of each chapter, this dissertation contributed to ongoing debates among survey methodologists and practitioners regarding the management and interpretation of complex survey paradata, devising appropriate methodological and statistical tools to analyse them, and deriving practical strategies towards survey design optimization within the contexts of rising costs, diminishing response rates and increased survey fatigue.

Summary

Chapters 1 and 2 were primarily concerned with call record data and their potential for nonresponse assessment and prevention, as well as fieldwork optimisation in household longitudinal surveys. These first two chapters relied on data from the first four waves of the United Kingdom's Household Longitudinal Study, also called *Understanding Society*.

The objective of Chapter 1 was to investigate drivers of longitudinal nonresponse in household face-to-face surveys. More specifically, this chapter aimed at finding Wave 1 call record sequences associated with future wave contact and cooperation propensities and was informed by standing theories of nonresponse in household surveys (Couper and Groves, 1996; Groves and Couper, 1998; Lepkowski and Couper, 2002). The focus on call record data was based on the limitations of conventional nonresponse predictors, the relatively underexplored analysis of this type of data in studies of longitudinal nonresponse, and that call record data capture processes more directly under the control of the interviewer and the survey designer.

In short, this first chapter posed two questions: 1) are there specific call sequences observed at Wave 1 of a household longitudinal survey associated with future contact and cooperation propensities? and 2) do these Wave 1 call sequences reveal additional information about future contact and cooperation unobserved by conventional demographic and attitudinal predictors of nonresponse? To answer these questions, 18 different model specifications were evaluated. These specifications allowed for the comparative analysis of specific call sequences as well as more conventional predictors of nonresponse (like geographic markers, household size and composition, dwelling types, demographic and socioeconomic traits of the household residents, attitudes towards social and community attachment, as well as previous interviewer experience) and their associations with future wave contact and cooperation. To account for unobserved interviewer effects, the nonrandom allocation of interviewers between waves, and the effect of intra-wave household moves, random-intercept models were used in the analysis as well as cross-wave interviewer- and geographical-continuity controls.

The analysis showed significant associations between specific call record sequences observed at Wave 1 and differential contact and cooperation propensities in Waves 2, 3 and 4. More specifically, Wave 1 households which had repeated unproductive contacts, broke appointments, registered above median proportion of "no replies", or began the call sequence with an unproductive contact were at risk of future nonresponse for Waves 2, 3 and 4. Following the results section, this chapter suggested ways in which these findings can help inform the design of future nonresponse prevention strategies, including interviewer training as well as tailored household contact protocols. Finally, it concluded with a discussion of the limitations of the analysis as well as avenues for future research.

Chapter 2 analysed factors associated with differential field effort in household longitudinal surveys. In particular, this chapter analysed different drivers of field effort within the processes of contacting and eliciting cooperation from a household across waves of a longitudinal survey. Because household longitudinal surveys are resource-intensive, time-consuming and expected to withstand several waves of data collection, analysing drivers of field effort can help quantify associated survey costs, optimise the use of fieldwork resources, as well as inform survey design to make data collection more efficient. As with the first chapter, this analysis was informed by standing theories of household nonresponse (Couper and Groves, 1996; Groves and Couper, 1998; Lepkowski and Couper, 2002) but incorporated the use of call sequence variables to expand on them.

This analysis was primarily interested in answering the following questions: what household and/or individual respondent characteristics are associated with differential calling effort? Do lagged call record patterns correlate with a household's number of calls towards contact and/or cooperation in future waves? Finally, and perhaps most importantly, do these hypothetical correlations change over waves? If so, are these changes due to self-selection of nonattriters or can they be explained as a longitudinal learning effect? To answer these questions, 12 random-intercept linear regression models were employed to estimate number of calls towards contact and cooperation for the first four waves of *Understanding Society*. As in Chapter 1, in addition to the conventional covariates of nonresponse, particular call sequences were included in the models to analyse and compare their association with differential field effort towards contact and cooperation.

This analysis found that number of calls towards contact and cooperation was found to be associated with observable household characteristics as well as particular lagged call record sequences. This analysis also revealed that longitudinal learning as well as self-selection effects contribute to field effort reduction across the waves of a longitudinal survey. This chapter also proposed ways in which the findings could help optimise field resources in household longitudinal surveys. Specifically, the estimates of calls needed towards contact and cooperation could be used to plan and budget resources in future waves as well as monitor real-time data collection operations. Tailored contact approaches can also be designed to prioritise those households deemed more costly to engage. Finally, interviewer training could incorporate the lessons learned from this type of analysis. The chapter ended with a discussion of its limitations and opportunities for continued research.

Chapter 3 analysed the use of web survey paradata (including timestamps, response durations and click data) to further examine and expand on standing theories of progress indicator (PI) effects on the quality of survey response. Most analyses of PI effects are primarily concerned with nonresponse: how the presence of a progress indicator may drive some respondents abandon the survey once started (Conrad et al., 2005, 2003a,b, 2010; Couper et al., 2001; Crawford et al., 2001; Heerwegh and Loosveldt, 2006; Matzat et al., 2009; Peytchev, 2009; Yan et al., 2011). There is very little research on how PI affects other components of data quality, like measurement error. Based on the theory of survey burden and survey satisficing (Conrad et al., 2010; Crawford et al., 2001; Heerwegh and Loosveldt, 2006; Krosnick, 1991; Krosnick et al., 1996; Matzat et al., 2009), this final chapter analyses data from an original experiment to determine if PI presence affects perceptions of burden and therefore results in differential response quality. In other words, beyond the propensity to drop out of a survey, does the inclusion of a PI cause a respondent to 1) perceive the required effort to answer differently and therefore 2) devote more or less cognitive resources when answering questions? To answer these questions a split-sample experiment was designed with a treatment group representing respondents exposed to a PI and a control group representing respondents not exposed to a PI. Survey effort and response quality were measured using different variables, including some derived from web paradata like click patterns, response speed and survey duration.

The experiment conducted for this chapter yielded no significant response quality effects from PIs. The presence of a progress indicator did not result in differential perception of burden or in differential response quality. Survey durations, response speeds and survey satisficing proxies were not significantly different between the treatment and control groups. While further research is recommended, at first glance these null finding suggest that PIs can be employed (or not) by practitioners without fear of unduly biasing survey estimates. In addition to these findings, this chapter also devoted some discussion to the practical considerations of managing and interpreting meaning from the paradata generated by web surveys. It concluded with a discussion of some of its limitations as well as suggested areas of further research.

Contributions to the existing literature

Chapters 1 and 2 expanded on existing studies of call record data to investigate and assess nonresponse in household surveys. While recent works have explored the use of call record data in the analysis of fieldwork protocols and nonresponse mechanisms (Durrant et al., 2011, 2013b, 2010; Henly and Bates, 2006; Lugtig, 2014; Wagner, 2013; Watson and Wooden, 2009), there is relatively little research on the effect of call record sequences in future wave nonresponse propensities in longitudinal surveys (Durrant et al., 2017). Furthermore, there is no other analysis that looks directly at how particular call sequences (such as broken appointments, repeated unproductive calls and increased proportion of no replies) are associated with differential contact and cooperation propensities in future waves of a longitudinal survey. In addition, few analyses have comparatively assessed the magnitude and significance between associations of conventional predictors of nonresponse and these types of call record sequences. Thus, the first chapter of this thesis fills a gap in the literature by finding considerable and significant associations between particular call sequences observed in the baseline wave of a longitudinal survey and contact and cooperation propensities in future waves.

Regarding Chapter 2, there is very little in the literature of longitudinal, face-to-face surveys on drivers of field effort associated with the process of contact and cooperation. Similarly, there is little if any discussion on the role that longitudinal learning or self-selection play in determining the effort a household requires towards contact and/or cooperation. The analysis presented in Chapter 2 suggest that differential field effort towards contact and cooperation is associated with particular household traits and previous wave call interactions. Furthermore, while field operations become more efficient as waves progress, the findings of this chapter indicate that efficiencies are a result of both self-selection as well as learning mechanisms.

Lastly, Chapter 3 shifts the discussion of the associated literature on PI effects from nonresponse to response quality and survey satisficing. It does so by analysing web paradata as gathered by an original experiment. While no significant findings were observed in this analysis suggesting the existence of PI effects on perceptions of burden, time- and effort-management strategies, or the propensity to satisfice, the possibility that these effects do exist could be further explored. In particular, the experiment could be tested among less heterogeneous populations and for different types of surveys (of varying topics, lengths and incentive structures).

Limitations and future research

The findings of Chapters 1 and 2 provide insight into associations observed within the context of *Understanding Society* data. Further confirmation of these findings could be achieved by successful replication of these analysis in additional waves of the study, or alternative datasets from comparable household longitudinal surveys and/or experimental approaches. Notwithstanding the replicability of these findings, one should note that given the cost and resource constraints on these types of surveys a recent generalized push towards alternative modes of data collection means that call record data may no longer be available in future waves of *Understanding Society* or similar studies (Burton, 2016). It remains to be seen if the findings discussed in this dissertation are translatable to these other modes of data collection, specially CAWI surveys. Finally, replicability is also hampered by the fact that there exist no uniform call outcome coding system which makes analysis and cross-survey comparisons difficult. An opportunity for future researchers and practitioners could entail the drafting of a shared system of call record codification similar to AAPOR's Response Rate definitions (AAPOR, 2016).

Further research should consider the role call record data may play in not just prevention of nonresponse and field optimization, but also in measurement error and data quality. While these data have already been explored for nonresponse bias assessment in cross-sectional surveys (Kreuter and Kohler, 2009; Lynn et al., 2002), no comparable literature is found for longitudinal household surveys. If particular call record sequences have been found to be associated with differential response propensity as well as field effort requirements, can they also signal differential response quality? Finally, call records should continue to be used in interviewer performance assessment and analysis of interviewer-interviewee interaction. Again, if problematic call sequences are adversely affecting response rates and field effort allocation, what role do interviewers play? Are some interviewers more likely to incur broken appointments or repeated soft refusals?

Regarding Chapter 3, and as previously noted, the null findings do not necessarily mean that PIs have no effect on response quality; rather that these have not (yet?) been found. Given the homogeneous population of the BOOST2018 survey and the saliency of the questionnaire topic, one can hypothesize that a similar experimental design among a general population sample with a less engaging survey topic may result in different perceptions of respondent burden and therefore yield significant PI effects on response quality. In addition, the challenges of identifying survey downtime from the paradata as collected by BOOST2018 not only conditioned the results of this chapter but also highlighted the type of problems researchers and practitioners often face when dealing with paradata.

Future research should continue to explore and develop ways to generate, manage and analyse survey paradata that 1) better capture respondent behaviour (and how this behaviour may affect the quality of their responses) and 2) aid in survey design (for example, questionnaire development and item evaluation). As technology continues to evolve respondents' engagement with online surveys will necessarily change in parallel. In particular, changes in devices used to answer CAWI surveys as well as changes in how people use these devices and incorporate them into their daily routines may afford new opportunities for paradata collection. Handheld devices are becoming more prevalent, more portable and more sophisticated in their ability to collect and manage additional sources of information including, among others, GPS tracking and geolocation, data linkage across a user's collection of devices (smartphone, personal computer, work computer, etc.), as well as augmented data through auxiliary non-survey sources (like social networks and instant messaging apps). All these new sources of data, while demanding careful consideration of data protection, consent and anonymity, have the potential to enhance our understanding of respondents and respondent behaviour.

Appendices

Appendix A

Additional material for Chapter 1

Table A.1: Wave 2. Contact.

		1	2	3
Wave 1 Call Records	First Call Status			
	Completed interview			1
	No Reply			0.887
	Unproductive Contact			1.001
	Appointment made			1.060
	Some interviewing done			1.272
	Any other status			0.989
	Appointments			
	Made & Broke Appointments		1	
	Made & Kept Appointments		1.370***	
	Did Not Make Appointments		1.399***	
	Repeat unproductive contacts		0.918	
	Above median % of no replies		0.792***	
Other Wave 1 NR Covariates	Geographical Region			
	London	1	1	1
	North East	0.729	0.731	0.733
	North West	0.857	0.852	0.855
	Yorkshire And The Humber	0.587***	0.575***	0.586***
	East Midlands	0.904	0.899	0.908
	West Midlands	0.649**	0.648**	0.645**
	East Of England	1.453*	1.438*	1.449*
	South East	1.252	1.229	1.245
	South West	0.897	0.875	0.897
	Wales	0.587**	0.580**	0.585**
	Scotland	0.682**	0.670**	0.677**
	Urban Indicator	1.039	1.031	1.037
	Dwelling type			
	Detached	1	1	1
	Semi	0.988	0.995	0.987
	Terraced + End	0.738**	0.756**	0.742**
	Flat/Maisonette + Purpopse + Converted	0.587***	0.597***	0.590***
	Bedsit + W/Bsiness + Sheltr + Inst + Oth	0.917	0.922	0.913
	Groundfloor property	0.972	0.977	0.972
	Property with respect to neighbours			
	Better Or Same Condition	1	1	1
	Worse	0.740**	0.750**	0.745**
	No Info + Miss	1.115	1.072	1.087
	Number of people in household	0.977	0.966	0.968
	At least one baby in household	1.222	1.227*	1.217
	All residents in poor health (self-reported)	0.754	0.743*	0.748
	National origin of household			
	All Nonbritish	1	1	1
	Mixed	1.470*	1.422*	1.464*
	All British	1.291**	1.283**	1.287**
	Working status			
	No One Works	1	1	1

Continued on next page

Continued from previous page

		1	2	3
	At Least 1 Works But Not Long Hours	1.028	1.062	1.036
	At Least 1 Works Long Hours	0.925	0.950	0.935
	All Work Long Hours	0.652***	0.685***	0.670***
	Presence of pensioner			
	No Pensioner	1	1	1
	At Least 1 Pensioner	0.841*	0.860	0.848
	All Pensioners	1.553***	1.486***	1.530***
	Deprivation indicator	0.908	0.913	0.906
	Owner/Mortgager	1.178*	1.168*	1.181*
	No political interest	0.722***	0.732***	0.723***
	Community Attachment			
	Q4	1	1	1
	Q3	1.101	1.101	1.102
	Q2	0.981	0.981	0.984
	Q1	1.057	1.064	1.062
	Missing	0.784*	0.809*	0.793*
	No one consents to data linkage	0.797***	0.804**	0.801***
	No one present during interview	0.828**	0.831**	0.829**
	No suspicion during interview	1.078	1.076	1.075
	Excellent understanding of questions	1.097	1.090	1.101
	Item Nonresponse (log)	0.891**	0.889**	0.891**
	Dummy Item Nonresponse (log)	1.076	1.067	1.074
x-wave	Same LSOA across Waves (1 & 2)	2.900***	2.962***	2.920***
	Same interviewer across Waves (1 & 2)	2.257***	2.257***	2.262***
	Same LSOA and interviewer (interaction)	1.913***	1.864***	1.902***
	Constant	4.964***	4.354***	5.237***
	Constant (Random Intercept)	1.160***	1.163***	1.161***
	Observations	24104	24104	24104
	Log Likelihood	-4835.4	-4817.2	-4831.3
	Degrees Of Freedom	44	48	49
	aic	9762.7	9734.4	9764.6

*Odds ratios (Exponentiated coefficients). * p<0.05, ** p<0.01, *** p<0.001*

Estimated coefficients for Wave 2 Contact Propensities. Random-intercept, logistic regression models were specified to account for unobserved interviewer effects.

Table A.2: Wave 2. Cooperation.

		4	5	6
Wave 1 Call Records	First Call Status			
	Completed interview			1
	No Reply			0.888
	Unproductive Contact			0.676***
	Appointment made			0.842
	Some interviewing done			1.168
	Any other status			0.716*
	Appointments			
	Made & Broke Appointments		1	
	Made & Kept Appointments		1.210**	
	Did Not Make Appointments		1.234**	
	Repeat unproductive contacts		0.766***	
	Above median % of no replies		1.034	
Other Wave 1 NR Covariates	Geographical Region			
	London	1	1	1
	North East	0.970	0.957	0.946
	North West	0.932	0.923	0.916
	Yorkshire And The Humber	0.814	0.799	0.800
	East Midlands	1.097	1.089	1.083
	West Midlands	0.940	0.940	0.940
	East Of England	1.298*	1.287*	1.295*
	South East	1.135	1.135	1.139
	South West	1.206	1.184	1.195
	Wales	1.154	1.141	1.133
	Scotland	0.899	0.893	0.892
	Urban Indicator	0.970	0.966	0.970
	Dwelling type			
	Detached	1	1	1
	Semi	1.011	1.015	1.008
	Terraced + End	0.987	0.989	0.974
	Flat/Maisonette + Purpopse + Converted	0.875	0.871	0.861
	Bedsit + W/Bsiness + Sheltr + Inst + Oth	0.943	0.947	0.940
	Groundfloor property	0.979	0.986	0.979
	Property with respect to neighbours			
	Better Or Same Condition	1	1	1
	Worse	1.149	1.157	1.149
	No Info + Miss	1.300	1.344	1.418
	Number of people in household	1.018	1.031	1.027
	At least one baby in household	1.224*	1.214*	1.212*
	All residents in poor health (self-reported)	0.744**	0.746**	0.746**
	National origin of household			
	All Nonbritish	1	1	1
	Mixed	1.486**	1.449*	1.476**
	All British	0.902	0.892	0.902
	Working status			
	No One Works	1	1	1

Continued on next page

Continued from previous page

		4	5	6
	At Least 1 Works But Not Long Hours	0.755***	0.764***	0.757***
	At Least 1 Works Long Hours	0.893	0.898	0.892
	All Work Long Hours	0.902	0.901	0.894
	Presence of pensioner			
	No Pensioner	1	1	1
	At Least 1 Pensioner	0.644***	0.647***	0.646***
	All Pensioners	0.609***	0.609***	0.611***
	Deprivation indicator	1.171**	1.182**	1.169**
	Owner/Mortgager	1.022	1.017	1.025
	No political interest	0.794***	0.799***	0.794***
	Community Attachment			
	Q4	1	1	1
	Q3	0.898	0.897	0.895
	Q2	0.855*	0.856*	0.854*
	Q1	0.845*	0.842*	0.846*
	Missing	0.611***	0.615***	0.615***
	No one consents to data linkage	0.579***	0.583***	0.582***
	No one present during interview	1.024	1.026	1.025
	No suspicion during interview	1.580***	1.570***	1.565***
	Excellent understanding of questions	1.342***	1.333***	1.338***
	Item Nonresponse (log)	0.900***	0.901***	0.900***
	Dummy Item Nonresponse (log)	0.910*	0.911*	0.910*
x-wave	Same LSOA across Waves (1 & 2)	1.667***	1.697***	1.679***
x-wave	Same interviewer across Waves (1 & 2)	3.114***	3.140***	3.094***
x-wave	Same LSOA and interviewer (interaction)	1.018	0.990	1.009
	Constant	3.413***	2.837***	4.049***
	Constant (Random Intercept)	1.125***	1.123***	1.122***
	Observations	22684	22684	22684
	Log Likelihood	-7981.5	-7967.1	-7965.3
	Degrees Of Freedom	44	48	49
	aic	16055.0	16034.1	16032.6

*Odds ratios (Exponentiated coefficients). * p<0.05, ** p<0.01, *** p<0.001*

Estimated coefficients for Wave 2 Cooperation Propensities. Random-intercept, logistic regression models were specified to account for unobserved interviewer effects.

Table A.3: Wave 3. Contact.

		7	8	9
Wave 1 Call Records	First Call Status			
	Completed interview			1
	No Reply			0.656*
	Unproductive Contact			0.654*
	Appointment made			0.795
	Some interviewing done			0.923
	Any other status			0.612
	Appointments			
	Made & Broke Appointments		1	
	Made & Kept Appointments		1.449***	
	Did Not Make Appointments		1.446***	
	Repeat unproductive contacts		0.765*	
	Above median % of no replies		0.764***	
Other Wave 1 NR Covariates	Geographical Region			
	London	1	1	1
	North East	0.782	0.787	0.777
	North West	0.854	0.837	0.838
	Yorkshire And The Humber	0.860	0.833	0.852
	East Midlands	1.209	1.201	1.210
	West Midlands	0.788	0.781	0.780
	East Of England	1.344	1.322	1.341
	South East	1.444*	1.413*	1.446*
	South West	0.867	0.834	0.860
	Wales	0.580**	0.563**	0.569**
	Scotland	0.746	0.724*	0.736
	Urban Indicator	1.117	1.102	1.115
	Dwelling type			
	Detached	1	1	1
	Semi	0.813	0.820	0.814
	Terraced + End	0.741**	0.757*	0.742**
	Flat/Maisonette + Purpopse + Converted	0.605***	0.610***	0.605***
	Bedsit + W/Bsiness + Sheltr + Inst + Oth	0.594	0.591	0.594
	Groundfloor property	1.136	1.131	1.134
	Property with respect to neighbours			
	Better Or Same Condition	1	1	1
	Worse	0.815	0.829	0.825
	No Info + Miss	0.664	0.656	0.682
	Number of people in household	0.897***	0.887***	0.895***
	At least one baby in household	1.107	1.099	1.093
	All residents in poor health (self-reported)	1.499	1.475	1.482
	National origin of household			
	All Nonbritish	1	1	1
	Mixed	1.061	1.027	1.060
	All British	1.011	0.999	1.010
	Working status			
	No One Works	1	1	1

Continued on next page

Continued from previous page

		7	8	9
	At Least 1 Works But Not Long Hours	0.995	1.035	1.014
	At Least 1 Works Long Hours	0.976	1.003	1.000
	All Work Long Hours	0.813	0.860	0.847
	Presence of pensioner			
	No Pensioner	1	1	1
	At Least 1 Pensioner	1.003	1.036	1.008
	All Pensioners	2.787***	2.660***	2.716***
	Deprivation indicator	0.793**	0.798**	0.789**
	Owner/Mortgager	1.276**	1.260**	1.285**
	No political interest	0.821*	0.837*	0.820*
	Community Attachment			
	Q4	1	1	1
	Q3	1.017	1.016	1.024
	Q2	1.085	1.089	1.094
	Q1	1.016	1.021	1.030
	Missing	0.782	0.816	0.799
	No one consents to data linkage	0.899	0.917	0.910
	No one present during interview	0.903	0.908	0.910
	No suspicion during interview	1.004	0.998	0.997
	Excellent understanding of questions	1.104	1.100	1.110
	Item Nonresponse (log)	0.907	0.906	0.908
	Dummy Item Nonresponse (log)	0.967	0.959	0.967
x-wave	Same LSOA across Waves (1 & 3)	2.720***	2.774***	2.739***
x-wave	Same interviewer across Waves (1 & 3)	2.365***	2.337***	2.368***
x-wave	Same LSOA and interviewer (interaction)	0.940	0.919	0.931
	Constant	10.16***	9.094***	14.18***
	Constant (Random Intercept)	1.104*	1.100*	1.110*
	Observations	22038	22038	22038
	Log Likelihood	-3574.9	-3555.6	-3568.7
	Degrees Of Freedom	44	48	49
	aic	7241.9	7211.2	7239.4

*Odds ratios (Exponentiated coefficients). * p<0.05, ** p<0.01, *** p<0.001*

Estimated coefficients for Wave 3 Contact Propensities. Random-intercept, logistic regression models were specified to account for unobserved interviewer effects.

Table A.4: Wave 3. Cooperation.

		10	11	12
Wave 1 Call Records	First Call Status			
	Completed interview			1
	No Reply			1.013
	Unproductive Contact			0.764**
	Appointment made			0.960
	Some interviewing done			1.830**
	Any other status			0.817
	Appointments			
	Made & Broke Appointments		1	
	Made & Kept Appointments		1.170**	
	Did Not Make Appointments		1.171*	
	Repeat unproductive contacts		0.725***	
	Above median % of no replies		1.079	
Other Wave 1 NR Covariates	Geographical Region			
	London	1	1	1
	North East	0.918	0.906	0.897
	North West	1.162	1.153	1.145
	Yorkshire And The Humber	1.154	1.129	1.135
	East Midlands	1.197	1.191	1.189
	West Midlands	1.024	1.024	1.027
	East Of England	1.343**	1.337**	1.345**
	South East	1.224*	1.227*	1.232*
	South West	1.188	1.167	1.180
	Wales	1.159	1.143	1.144
	Scotland	0.911	0.904	0.909
	Urban Indicator	0.870**	0.865**	0.868**
	Dwelling type			
	Detached	1	1	1
	Semi	0.961	0.963	0.957
	Terraced + End	0.988	0.988	0.976
	Flat/Maisonette + Purpopse + Converted	1.007	1.000	0.989
	Bedsit + W/Bsiness + Sheltr + Inst + Oth	0.962	0.961	0.953
	Groundfloor property	0.824*	0.827	0.819*
	Property with respect to neighbours			
	Better Or Same Condition	1	1	1
	Worse	1.186	1.190	1.183
	No Info + Miss	2.826	2.946	2.958
	Number of people in household	0.975	0.991	0.981
	At least one baby in household	1.381***	1.365***	1.368***
	All residents in poor health (self-reported)	0.831	0.836	0.839
	National origin of household			
	All Nonbritish	1	1	1
	Mixed	1.258	1.233	1.234
	All British	1.142	1.130	1.135
	Working status			
	No One Works	1	1	1

Continued on next page

Continued from previous page

		10	11	12
	At Least 1 Works But Not Long Hours	0.746***	0.755***	0.746***
	At Least 1 Works Long Hours	0.834**	0.832**	0.825**
	All Work Long Hours	0.762**	0.750***	0.751***
	Presence of pensioner			
	No Pensioner	1	1	1
	At Least 1 Pensioner	0.763***	0.762***	0.769***
	All Pensioners	0.718***	0.720***	0.726***
	Deprivation indicator	1.070	1.076	1.067
	Owner/Mortgager	1.068	1.062	1.066
	No political interest	0.808***	0.812***	0.809***
	Community Attachment			
	Q4	1	1	1
	Q3	0.876*	0.875*	0.871*
	Q2	0.961	0.965	0.959
	Q1	0.832**	0.830**	0.832**
	Missing	0.747***	0.753***	0.751***
	No one consents to data linkage	0.610***	0.614***	0.615***
	No one present during interview	1.047	1.049	1.049
	No suspicion during interview	1.358***	1.354***	1.349***
	Excellent understanding of questions	1.103*	1.095*	1.099*
	Item Nonresponse (log)	0.950	0.952	0.952
	Dummy Item Nonresponse (log)	0.959	0.962	0.958
x-wave	Same LSOA across Waves (1 & 3)	0.861*	0.867	0.863
	Same interviewer across Waves (1 & 3)	1.180	1.175	1.173
	Same LSOA and interviewer (interaction)	1.580***	1.567***	1.584***
	Constant	7.577***	6.443***	8.089***
	Constant (Random Intercept)	1.083***	1.083***	1.083***
	Observations	21083	21083	21083
	Log Likelihood	-8775.7	-8756.5	-8753.7
	Degrees Of Freedom	44	48	49
	aic	17643.4	17612.9	17609.4

*Odds ratios (Exponentiated coefficients). * p<0.05, ** p<0.01, *** p<0.001*

Estimated coefficients for Wave 3 Cooperation Propensities. Random-intercept, logistic regression models were specified to account for unobserved interviewer effects.

Table A.5: Wave 4. Contact.

		13	14	15
Wave 1 Call Records	First Call Status			
	Completed interview			1
	No Reply			0.981
	Unproductive Contact			0.791
	Appointment made			1.046
	Some interviewing done			1.880
	Any other status			0.923
	Appointments			
	Made & Broke Appointments		1	
	Made & Kept Appointments		1.353**	
	Did Not Make Appointments		1.482***	
	Repeat unproductive contacts		0.811	
	Above median % of no replies		0.761***	
Other Wave 1 NR Covariates	Geographical Region			
	London	1	1	1
	North East	0.791	0.790	0.772
	North West	0.887	0.877	0.871
	Yorkshire And The Humber	0.984	0.962	0.973
	East Midlands	1.089	1.079	1.083
	West Midlands	0.812	0.805	0.813
	East Of England	1.523*	1.516*	1.523*
	South East	1.033	1.016	1.038
	South West	1.125	1.095	1.120
	Wales	0.911	0.880	0.890
	Scotland	0.763	0.741	0.766
	Urban Indicator	1.031	1.019	1.029
	Dwelling type			
	Detached	1	1	1
	Semi	0.850	0.854	0.849
	Terraced + End	0.854	0.869	0.851
	Flat/Maisonette + Purpopse + Converted	0.562***	0.565***	0.557***
	Bedsit + W/Bsiness + Sheltr + Inst + Oth	0.715	0.720	0.708
	Groundfloor property	0.795	0.799	0.790
	Property with respect to neighbours			
	Better Or Same Condition	1	1	1
	Worse	0.742*	0.762*	0.742*
	No Info + Miss	0.798	0.774	0.834
	Number of people in household	0.860***	0.850***	0.862***
	At least one baby in household	1.084	1.076	1.065
	All residents in poor health (self-reported)	0.736	0.714	0.741
	National origin of household			
	All Nonbritish	1	1	1
	Mixed	0.955	0.926	0.936
	All British	1.141	1.128	1.129
	Working status			
	No One Works	1	1	1

Continued on next page

Continued from previous page

		13	14	15
	At Least 1 Works But Not Long Hours	1.143	1.205	1.143
	At Least 1 Works Long Hours	0.919	0.951	0.915
	All Work Long Hours	0.778	0.834	0.777
	Presence of pensioner			
	No Pensioner	1	1	1
	At Least 1 Pensioner	0.839	0.863	0.849
	All Pensioners	2.414***	2.319***	2.423***
	Deprivation indicator	0.783**	0.789**	0.777**
	Owner/Mortgager	1.504***	1.488***	1.507***
	No political interest	0.707***	0.716***	0.708***
	Community Attachment			
	Q4	1	1	1
	Q3	1.148	1.150	1.142
	Q2	0.952	0.958	0.943
	Q1	1.030	1.041	1.036
	Missing	0.637**	0.667**	0.644**
	No one consents to data linkage	0.833*	0.846	0.845
	No one present during interview	1.012	1.022	1.018
	No suspicion during interview	0.958	0.945	0.952
	Excellent understanding of questions	1.113	1.111	1.118
	Item Nonresponse (log)	0.966	0.964	0.970
	Dummy Item Nonresponse (log)	0.920	0.911	0.918
x-wave	Same LSOA across Waves (1 & 4)	2.702***	2.748***	2.706***
x-wave	Same interviewer across Waves (1 & 4)	1.247	1.244	1.239
x-wave	Same LSOA and interviewer (interaction)	1.561**	1.527*	1.569**
	Constant	23.87***	21.58***	24.92***
	Constant (Random Intercept)	1.101	1.108	1.098
	Observations	19699	19699	19699
	Log Likelihood	-2902.6	-2889.0	-2896.9
	Degrees Of Freedom	44	48	49
	aic	5897.2	5878.0	5895.8
	*Odds ratios (Exponentiated coefficients). * p<0.05, ** p<0.01, *** p<0.001*			

Estimated coefficients for Wave 4 Contact Propensities. Random-intercept, logistic regression models were specified to account for unobserved interviewer effects.

Table A.6: Wave 4. Cooperation.

		16	17	18
Wave 1 Call Records	First Call Status			
	Completed interview			1
	No Reply			1.002
	Unproductive Contact			0.802*
	Appointment made			0.978
	Some interviewing done			1.276
	Any other status			1.098
	Appointments			
	Made & Broke Appointments		1	
	Made & Kept Appointments		1.254***	
	Did Not Make Appointments		1.284***	
	Repeat unproductive contacts		0.746***	
	Above median % of no replies		0.961	
Other Wave 1 NR Covariates	Geographical Region			
	London	1	1	1
	North East	0.793	0.786	0.779
	North West	0.926	0.918	0.916
	Yorkshire And The Humber	1.026	1.003	1.010
	East Midlands	1.031	1.027	1.023
	West Midlands	0.982	0.976	0.982
	East Of England	1.135	1.127	1.138
	South East	1.034	1.029	1.034
	South West	1.201	1.180	1.193
	Wales	0.872	0.857	0.857
	Scotland	0.765*	0.756*	0.765*
	Urban Indicator	1.057	1.049	1.052
	Dwelling type			
	Detached	1	1	1
	Semi	0.965	0.970	0.964
	Terraced + End	0.946	0.954	0.940
	Flat/Maisonette + Purpopse + Converted	0.999	0.999	0.987
	Bedsit + W/Bsiness + Sheltr + Inst + Oth	0.948	0.951	0.934
	Groundfloor property	0.913	0.918	0.912
	Property with respect to neighbours			
	Better Or Same Condition	1	1	1
	Worse	1.099	1.112	1.096
	No Info + Miss	0.948	0.956	0.984
	Number of people in household	0.946*	0.953	0.951*
	At least one baby in household	1.207*	1.194*	1.198*
	All residents in poor health (self-reported)	0.632***	0.627***	0.637***
	National origin of household			
	All Nonbritish	1	1	1
	Mixed	1.262	1.225	1.249
	All British	1.067	1.048	1.062
	Working status			
	No One Works	1	1	1

Continued on next page

Continued from previous page

		16	17	18
	At Least 1 Works But Not Long Hours	0.874	0.898	0.875
	At Least 1 Works Long Hours	0.961	0.972	0.957
	All Work Long Hours	0.942	0.959	0.935
	Presence of pensioner			
	No Pensioner	1	1	1
	At Least 1 Pensioner	0.819**	0.825**	0.821**
	All Pensioners	0.761***	0.754***	0.766**
	Deprivation indicator	1.002	1.012	1.001
	Owner/Mortgager	0.985	0.979	0.985
	No political interest	0.804***	0.811***	0.803***
	Community Attachment			
	Q4	1	1	1
	Q3	0.893	0.895	0.890
	Q2	0.869	0.871	0.865
	Q1	0.877	0.878	0.877
	Missing	0.778**	0.795**	0.782**
	No one consents to data linkage	0.718***	0.726***	0.723***
	No one present during interview	1.040	1.044	1.040
	No suspicion during interview	1.232***	1.224***	1.226***
	Excellent understanding of questions	1.185***	1.179**	1.185***
	Item Nonresponse (log)	0.984	0.986	0.986
	Dummy Item Nonresponse (log)	0.960	0.960	0.959
x-wave	Same LSOA across Waves (1 & 4)	1.241**	1.256**	1.246**
	Same interviewer across Waves (1 & 4)	1.275*	1.270*	1.280*
	Same LSOA and interviewer (interaction)	1.318*	1.302*	1.313*
	Constant	7.196***	6.085***	7.494***
	Constant (Random Intercept)	1.038	1.040*	1.039
	Observations	18947	18947	18947
	Log Likelihood	-6777.7	-6762.9	-6769.6
	Degrees Of Freedom	44	48	49
	aic	13647.3	13625.8	13641.1
	*Odds ratios (Exponentiated coefficients). * p<0.05, ** p<0.01, *** p<0.001*			

Estimated coefficients for Wave 4 Cooperation Propensities. Random-intercept, logistic regression models were specified to account for unobserved interviewer effects.

Appendix B

Additional material for Chapter 2

Table B.1: Geographical Controls

	Wave 1	Wave 2	Wave 3	Wave 4
Geographical Region				
London	9.96%	10.06%	9.58%	9.01%
North East	4.85%	4.84%	4.80%	4.85%
North West	12.18%	12.12%	12.04%	12.12%
Yorkshire and the Humber	9.05%	9.05%	8.92%	9.15%
East Midlands	8.04%	8.10%	8.27%	8.52%
West Midlands	9.01%	8.94%	8.81%	8.69%
East of England	9.64%	9.65%	9.63%	9.94%
South East	13.68%	13.73%	13.74%	13.97%
South West	9.11%	9.13%	10.05%	9.58%
Wales	5.44%	5.42%	5.45%	5.61%
Scotland	9.04%	8.95%	8.72%	8.56%
Urban indicator				
Urban	77.71%	77.92%	77.52%	77.13%
Rural	22.29%	22.08%	22.48%	22.87%
Total	24796	25260	22857	20127

Table B.2: Dwelling characteristics & Access

	Wave 1	Wave 2	Wave 3	Wave 4
Dwelling Type				
Detached	23.34%	23.38%	23.70%	23.86%
Semi	30.46%	30.41%	30.98%	31.64%
Terraced + end	28.54%	28.85%	29.47%	29.20%
Flat/Maisonette + Purpose + Converted	15.88%	15.68%	14.83%	14.30%
Other (+ missing)	1.78%	1.68%	1.04%	0.98%
Groundfloor property				
Not groundfloor (+ missing)	7.84%	7.75%	7.46%	7.21%
Groundfloor	92.16%	92.25%	92.54%	92.75%
Property with respect to neighbours				
Better condition	9.49%	9.42%	8.33%	8.87%
Same (+ missing)	84.84%	84.87%	86.28%	86.25%
Worse	5.67%	5.70%	5.39%	4.88%
Barriers to Dwelling				
At least 1 barrier	11.32%	11.10%	10.19%	9.62%
No Barriers (+ missing)	88.68%	88.90%	89.50%	89.85%
Total	24796	25260	22857	20127

Table B.3: Household Composition

	Wave 1	Wave 2	Wave 3	Wave 4
Mean number of people in household	2.44	2.50	2.50	2.50
Presence of baby in household				
No baby (+ missing)	90.74%	90.62%	91.21%	91.53%
At least one baby in household	9.26%	9.38%	8.79%	8.47%
HH Resident's health				
Other (+ missing)	84.03%	88.59%	88.36%	88.64%
All residents in poor+fair health	15.97%	11.41%	11.65%	11.36%
Household National Origin				
Mixed (+ missing)	4.26%	4.38%	4.23%	4.36%
All British	84.92%	84.77%	86.00%	86.48%
All Non-British	10.82%	10.84%	9.77%	9.17%
Working Status				
No one works	34.84%	33.75%	33.41%	32.78%
At least 1 works but not long hours	23.72%	23.86%	23.48%	23.39%
At least 1 works long hours	29.86%	30.91%	31.78%	32.02%
All work long hours	11.58%	11.49%	11.33%	11.81%
Presence of pensioner				
No pensioner	74.10%	75.16%	74.80%	73.44%
At least 1 pensioner	6.48%	6.43%	6.71%	7.00%
All pensioners	19.42%	18.40%	18.50%	19.56%
Total	24796	25260	22857	20127

Table B.4: SES and Social Inclusion Indicators

	Wave 1	Wave 2	Wave 3	Wave 4
Deprivation				
Not materially deprived	68.12%	67.89%	68.84%	68.32%
Deprived (at least 2 items)	31.88%	32.10%	31.16%	31.68%
Home ownership				
Non-Owner/Mortgager	34.18%	34.25%	33.55%	32.64%
Owner/Mortgager	65.82%	65.74%	66.45%	67.36%
HH Political Interest				
Some interest	80.82%	81.20%	81.47%	80.26%
No political interest (+ missing)	19.18%	18.79%	18.53%	19.74%
Community Attachment				
Missing	13.79%	13.44%	12.27%	2.93%
Q1	21.90%	22.48%	22.27%	25.63%
Q2	21.72%	21.83%	22.31%	26.19%
Q3	22.63%	22.51%	22.95%	24.69%
Q4	19.96%	19.75%	20.21%	20.55%
Total	24796	25260	22857	20127

Table B.5: Previous Interview Experience

	Wave 1	Wave 2	Wave 3	Wave 4
Consent to Data Linkage				
All consent	63.38%	63.54%	64.80%	65.93%
Some consent (+ missing)	9.07%	9.46%	9.96%	10.57%
No one consents	27.56%	27.01%	25.24%	23.50%
Other person present during interview				
Someone else present (+ missing)	41.68%	42.15%	39.50%	31.95%
No one else present	58.32%	57.85%	60.51%	68.05%
Respondent suspicion during interview				
Some suspicion (+ missing)	19.68%	19.34%	7.36%	4.79%
No suspicion	80.32%	80.66%	92.64%	95.21%
Understanding of interview questions				
Less than excellent (+ missing)	39.49%	39.23%	37.31%	36.95%
Excellent understanding	60.51%	60.77%	62.70%	63.05%
Cooperative respondents				
Not cooperative respondents (+ missing)	28.74%	28.54%	23.14%	21.89%
Very cooperative respondents	71.26%	71.45%	76.87%	78.11%
Mean Item nonresponse (log)	-0.652	-0.664	-0.305	-0.772
Dummy Item nonresponse (log)				
No item nonresponse	36.94%	36.78%	4.81%	21.19%
Some item nonresponse	63.06%	63.22%	95.20%	78.81%
Total	24796	25260	22857	20127

Table B.6: Household Response Outcomes

	Wave 1	Wave 2	Wave 3	Wave 4
HH Response				
Response	100%	78.79%	78.52%	83.62%
Nonresponse	0%	16.55%	17.36%	13.77%
Ineligible	0%	4.66%	4.12%	2.61%
Total	24796	25260	22857	20127

Table B.7: Call Record Sequences

	Wave 1	Wave 2	Wave 3	Wave 4
Appointments in Call Sequence				
Broke appointments	12.95%	13.42%	12.50%	11.78%
Kept appointments	55.35%	46.39%	48.53%	49.22%
No appointments	31.69%	40.19%	38.97%	39.00%
Repeat Soft Refusals in Call Sequence				
0 Repeat Soft Refusals	90.50%	89.15%	90.16%	90.75%
1+ Repeat Soft Refusals	9.50%	10.85%	9.84%	9.25%
Proportion of No replies in Call Sequence				
Below Median % No Replies	60.70%*	47.97%	49.84%	47.96%
Above Median % of No Replies	39.30%	52.03%	50.17%	52.04%
Total	24796	25260	22857	20127

*Median proportion of no replies is estimated from all issued households at a given wave. Because the analytical base of Wave 1 households only contains responding households, the number of Wave 1 households with an above median proportion of no replies is considerably less than 50%.

Table B.8: Cross-wave Continuity Indicators

	Wave 1	Wave 2	Wave 3	Wave 4
Same LSOA*				
Different	N/A	11.46%	16.12%	20.80%
Same	N/A	88.54%	83.89%	79.20%
Same Interviewer				
Different	N/A	32.45%	47.87%	59.49%
Same	N/A	67.55%	52.13%	40.51%
Same LSOA and Interviewer (Interaction)				
Different	N/A	39.36%	54.32%	65.74%
Same	N/A	60.64%	45.68%	34.26%
Total	24796	25260	22857	20127

*"Same Interviewer" means that a given household has kept the same interviewer for all observed waves. Similarly, "Same LSOA"means that a household has stayed in the same LSOA across all observed waves. For example, at Wave 3 a household that has stayed in the same LSOA since Wave 1 but has changed interviewers at some point between Wave 1 and Wave 3 will have a value of 1 for "Same LSOA" and 0 for "Same Interviewer" as well as 0 for the corresponding interaction term.

Table B.9: Wave 1. Field Effort Dynamics

	Total Number of Calls	Calls to Make Contact	Post-Contact Calls to Completion
Geographical Region			
London	0	0	0
North East	-0.091	0.194	-0.211
North West	-0.579***	-0.063	-0.438***
Yorkshire and the Humber	-0.654***	-0.019	-0.515***
East Midlands	-0.446**	-0.016	-0.327**
West Midlands	-0.531***	-0.174*	-0.287**
East of England	-0.355**	-0.160*	-0.170
South East	-0.325**	-0.067	-0.217*
South West	-0.768***	-0.094	-0.609***
Wales	-0.619***	-0.142	-0.406**
Scotland	-0.576***	-0.173*	-0.355***
Urban indicator	-0.260***	-0.060	-0.179***
Dwelling Type			
Detached	0	0	0
Semi	0.055	0.022	0.019
Terraced + end	0.284***	0.143***	0.089*
Flat/Maisonette + Purpose + Converted	0.318***	0.277***	0.008
Other (+ Missing)	0.959***	0.371***	0.633***
Groundfloor property	-0.127	-0.087	-0.065
Property with respect to neighbours			
Better condition condition	0	0	0
Same or missing	0.132*	0.064	0.069
Worse	0.396***	0.299***	0.079
No Barriers to Dwelling	0.084	-0.093	0.184**
Number of people in household	0.065***	-0.205***	0.197***
At least one baby in household	-0.143*	0.034	-0.111*
All residents in poor health (self-reported)	-0.302***	-0.170***	-0.155***
Household National Origin			
All British	0	0	0
Mixed + Missing	0.054	0.039	0.058
All Non-British	0.328***	0.109*	0.140**
Working Status			
No one works	0	0	0
At least 1 works but not long hours	0.608***	0.152***	0.357***
At least 1 (but not all) works long hours	0.697***	0.242***	0.416***
All work long hours	1.165***	0.865***	0.340***
Presence of pensioner			
No pensioner	0	0	0
At least 1 pensioner	-0.523***	-0.421***	-0.117*
All pensioners	-0.648***	-0.438***	-0.212***
Material deprivation	0.089*	-0.011	0.1000**
Owner/Mortgager	-0.010	0.097**	-0.060
No political interest	0.148**	0.026	0.072

Continued on next page

Continued from previous page

Community Attachment			
Q4	0	0	0
Q3	0.152**	0.118**	0.029
Q2	0.170**	0.109**	0.065
Q1	0.335***	0.190***	0.142**
Missing	0.747***	0.342***	0.312***
Consent to Data Linkage			
All consent	0	0	0
Some consent + missing	-0.094	-0.168***	0.094
No one consent	0.174***	0.036	0.147***
No one present during interview	0.323***	0.096***	0.132***
No suspicion during interview	-0.148**	-0.053	-0.114**
Excellent understanding of questions	-0.039	0.063	-0.099**
Cooperative respondent household	-0.064	0.058	-0.126**
Item nonresponse (log)	0.025	0.005	-0.005
Dummy Item nonresponse (log)	0.037	-0.006	0.039
Constant	4.171***	2.534***	1.656***
Constant (Random Intercept)	1.013***	0.628***	0.754***
Observations	24638	24638	24638
Log likelihood	-60346.8	-50717.9	-53907.4
AIC	120787.6	101529.7	107908.8
*Degrees of Freedom (44). * p<0.05, ** p<0.01, *** p<0.001*			

Estimated coefficients for Wave 1 Field Effort Outcomes (Calls to Make Contact, Post-Contact Calls to Cooperation, and total Number of Calls). Random-intercept, linear regression models were specified to account for unobserved interviewer effects.

Table B.10: Wave 2. Field Effort Dynamics

	Total Number of Calls	Calls to Make Contact	Post-Contact Calls to Completion
Wave 1 Appointment			
Broke appointments	0	0	0
Kept appointments	-0.571***	-0.104**	-0.368***
No appointments	-0.756***	-0.116**	-0.500***
W1 1+ Repeat Soft Refusals	0.476***	0.007	0.399***
W1 Above Median % of No Replies	0.582***	0.372***	0.063*
Geographical Region			
London	0	0	0
North East	-0.077	0.068	-0.150
North West	-0.173	0.078	-0.131
Yorkshire and the Humber	-0.381*	0.150	-0.315**
East Midlands	-0.074	0.111	-0.129
West Midlands	-0.386**	0.088	-0.218*
East of England	-0.363**	-0.004	-0.154
South East	-0.407***	-0.063	-0.197**
South West	-0.608***	-0.007	-0.375***
Wales	-0.261	0.142	-0.272*
Scotland	-0.440**	-0.048	-0.155
Urban indicator	-0.238***	-0.062*	-0.094**
Dwelling Type			
Detached	0	0	0
Semi	0.045	-0.012	0.042
Terraced + end	0.255***	0.069*	0.062
Flat/Maisonette + Purpose + Converted	0.421***	0.231***	-0.029
Other (+ Missing)	0.461**	0.104	0.242*
Groundfloor property	0.094	-0.014	-0.001
Property with respect to neighbours			
Better condition	0	0	0
Same or missing	0.005	-0.007	-0.080
Worse	-0.024	-0.085	-0.097
No Barriers to Dwelling	0.143	-0.024	0.184***
Number of people in household	0.038*	-0.139***	0.175***
At least one baby in household	0.097	0.108*	-0.062
All residents in poor health (self-reported)	-0.332***	-0.140***	-0.124**
Household National Origin			
All British	0	0	0
Mixed + Missing	-0.089	0.047	-0.039
All Non-British	0.287***	0.119**	0.147**
Working Status			
No one works	0	0	0
At least 1 works but not long hours	0.251***	0.135***	0.116**
At least 1 (but not all) works long hours	0.191**	0.134***	0.040
All work long hours	0.571***	0.481***	0.008
Presence of pensioner			

Continued on next page

Continued from previous page

No pensioner	0	0	0
At least 1 pensioner	-0.627***	-0.369***	-0.190***
All pensioners	-1.013***	-0.454***	-0.307***
Material deprivation	0.241***	0.001	0.136***
Owner/Mortgager	-0.228***	0.011	-0.125***
No political interest	0.309***	0.044	0.081*
Community Attachment			
Q4	0	0	0
Q3	0.020	0.057	0.019
Q2	0.003	-0.001	-0.003
Q1	0.181**	0.115**	0.099**
Missing	0.309***	0.134**	0.137**
Consent to Data Linkage			
All consent	0	0	0
Some consent + missing	-0.148*	-0.087*	-0.000
No one consent	0.065	-0.021	0.010
No one present during interview	0.192***	0.099***	0.033
No suspicion during interview	-0.165***	-0.027	-0.141***
Excellent understanding of questions	-0.052	0.050	-0.048
Cooperative respondent household	-0.085	-0.049	-0.003
Item nonresponse (log)	0.063*	0.030	0.004
Dummy Item nonresponse (log)	0.112*	0.038	0.037
Same LSOA (W1 & W2)	0.942***	-0.135	0.392***
Same Interviewer (W1 & W2)	0.201	0.228*	-0.056
Same LSOA and Interviewer (Interaction)	-0.383***	-0.256**	-0.060
Constant	3.845***	2.464***	1.485***
Constant (Random Intercept)	1.010***	0.503***	0.504***
Observations	25253	21925	19786
Log likelihood	-61753.5	-42373.2	-38389.5
AIC	123614.9	84854.3	76887.0
*Degrees of Freedom (51). * p<0.05, ** p<0.01, *** p<0.001*			

Estimated coefficients for Wave 2 Field Effort Outcomes (Calls to Make Contact, Post-Contact Calls to Cooperation, and total Number of Calls). Random-intercept, linear regression models were specified to account for unobserved interviewer effects.

Table B.11: Wave 3. Field Effort Dynamics

	Total Number of Calls	Calls to Make Contact	Post-Contact Calls to Completion
Wave 2 HH Response			
Nonresponse	0	0	0
Response	-0.844***	-0.453***	-0.219***
Ineligible	-0.633	-0.198	-0.263
Wave 1 Appointment			
Broke appointments	0	0	0
Kept appointments	-0.398***	0.003	-0.290***
No appointments	-0.454***	-0.013	-0.348***
W1 1+ Repeat Soft Refusals	0.295***	0.008	0.296***
W1 Above Median % of No Replies	0.309***	0.239***	-0.004
Wave 2 Appointment			
Broke appointments	0	0	0
Kept appointments	-0.614***	-0.027	-0.341***
No appointments	-0.714***	-0.030	-0.475***
W2 1+ Repeat Soft Refusals	0.402***	-0.092*	0.414***
W2 Above Median % of No Replies	0.415***	0.311***	0.026
Geographical Region			
London	0	0	0
North East	0.019	0.223*	-0.215
North West	-0.324*	0.077	-0.278**
Yorkshire and the Humber	-0.394*	-0.009	-0.310**
East Midlands	-0.386*	-0.095	-0.187
West Midlands	-0.441**	-0.017	-0.306**
East of England	-0.472***	-0.126	-0.244**
South East	-0.567***	-0.087	-0.231**
South West	-0.912***	-0.194*	-0.506***
Wales	-0.386	-0.057	-0.257*
Scotland	-0.756***	-0.159	-0.276**
Urban indicator	-0.222***	0.002	-0.151***
Dwelling Type			
Detached	0	0	0
Semi	0.007	0.004	0.012
Terraced + end	0.211***	0.083*	0.023
Flat/Maisonette + Purpose + Converted	0.193*	0.108*	-0.113*
Other (+ Missing)	-0.075	-0.026	-0.166
Groundfloor property	-0.117	-0.080	-0.017
Property with respect to neighbours			
Better condition	0	0	0
Same or missing	0.069	0.024	-0.040
Worse	0.431***	0.173**	-0.009
No Barriers to Dwelling	-0.047	-0.058	0.011
Number of people in household	0.097***	-0.112***	0.195***
At least one baby in household	-0.022	0.063	-0.097*
All residents in poor health (self-reported)	-0.306***	-0.138***	-0.064

Continued on next page

Continued from previous page

Household National Origin			
All British	0	0	0
Mixed + Missing	0.095	0.036	0.008
All Non-British	0.183**	0.083	0.100*
Working Status			
No one works	0	0	0
At least 1 works but not long hours	0.230***	0.105**	0.121**
At least 1 (but not all) works long hours	0.216***	0.084*	0.187***
All work long hours	0.472***	0.365***	0.127*
Presence of pensioner			
No pensioner	0	0	0
At least 1 pensioner	-0.540***	-0.263***	-0.087
All pensioners	-0.815***	-0.308***	-0.199***
Material deprivation	0.222***	0.041	0.094**
Owner/Mortgager	-0.180***	0.025	-0.165***
No political interest	0.130**	0.033	0.047
Community Attachment			
Q4	0	0	0
Q3	0.007	-0.007	0.055
Q2	-0.056	0.023	-0.004
Q1	0.067	0.039	0.067
Missing	0.513***	0.014	0.225*
Consent to Data Linkage			
All consent	0	0	0
Some consent + missing	-0.084	-0.027	0.011
No one consent	-0.027	0.001	0.061*
No one present during interview	0.184***	0.104***	0.079**
No suspicion during interview	-0.044	-0.107*	-0.134*
Excellent understanding of questions	-0.078	0.009	0.016
Cooperative respondent household	-0.115*	-0.003	-0.271***
Item nonresponse (log)	0.163***	0.031*	0.024
Dummy Item nonresponse (log)	0.361***	0.058	0.091*
Same LSOA (W1, W2 & W3)	0.615***	-0.075	0.153**
Same Interviewer (W1, W2 & W3)	-0.065	-0.040	-0.053
Same LSOA and Interviewer (Interaction)	-0.039	-0.034	0.026
Constant	5.208***	2.674***	2.450***
Constant (Random Intercept)	0.993***	0.466***	0.471***
Observations	22770	20028	17560
Log likelihood	-55374.4	-38018.9	-33488.3
AIC	110868.7	76157.8	67096.5
*Degrees of Freedom (57). * p<0.05, ** p<0.01, *** p<0.001*			

Estimated coefficients for Wave 3 Field Effort Outcomes (Calls to Make Contact, Post-Contact Calls to Cooperation, and total Number of Calls). Random-intercept, linear regression models were specified to account for unobserved interviewer effects.

Table B.12: Wave 4. Field Effort Dynamics

	Total Number of Calls	Calls to Make Contact	Post-Contact Calls to Completion
Wave 2 HH Response			
Nonresponse	0	0	0
Response	-0.603***	-0.218***	-0.100
Ineligible	-0.527	0.443	-0.515
Wave 3 HH Response			
Nonresponse	0	0	0
Response	-0.788***	-0.410***	-0.313***
Ineligible	0.994**	-0.141	-0.165
Wave 1 Appointment			
Broke appointments	0	0	0
Kept appointments	-0.132*	-0.008	-0.049
No appointments	-0.158*	-0.006	-0.074
W1 1+ Repeat Soft Refusals	0.168*	-0.012	0.214***
W1 Above Median % of No Replies	0.187***	0.161***	-0.020
Wave 2 Appointment			
Broke appointments	0	0	0
Kept appointments	-0.392***	0.002	-0.357***
No appointments	-0.427***	0.005	-0.358***
W2 1+ Repeat Soft Refusals	0.357***	-0.039	0.349***
W2 Above Median % of No Replies	0.251***	0.182***	0.033
Wave 3 Appointment			
Broke appointments	0	0	0
Kept appointments	-0.549***	-0.010	-0.316***
No appointments	-0.661***	-0.039	-0.420***
W3 1+ Repeat Soft Refusals	0.458***	0.082	0.450***
W3 Above Median % of No Replies	0.417***	0.287***	0.068*
Geographical Region			
London	0	0	0
North East	0.196	0.043	-0.027
North West	-0.107	-0.026	-0.150
Yorkshire and the Humber	-0.105	0.000	-0.121
East Midlands	-0.215	-0.036	-0.076
West Midlands	-0.218	-0.028	-0.179
East of England	-0.186	-0.094	-0.086
South East	-0.006	-0.049	-0.069
South West	-0.339	-0.109	-0.271*
Wales	-0.017	0.159	-0.075
Scotland	-0.291	-0.066	-0.162
Urban indicator	-0.223***	-0.055	-0.107**
Dwelling Type			
Detached	0	0	0
Semi	-0.011	0.003	-0.036
Terraced + end	0.103	0.036	0.014
Flat/Maisonette + Purpose + Converted	0.226**	0.156**	-0.031

Continued on next page

Continued from previous page

Other (+ Missing)	-0.284	-0.174	0.107
Groundfloor property	0.069	0.039	-0.048
Property with respect to neighbours			
Better condition	0	0	0
Same or missing	-0.039	0.035	-0.001
Worse	0.352**	0.130*	0.123
No Barriers to Dwelling	-0.035	-0.006	0.079
Number of people in household	0.104***	-0.097***	0.176***
At least one baby in household	0.049	0.021	-0.064
All residents in poor health (self-reported)	-0.270***	-0.122**	-0.085*
Household National Origin			
All British	0	0	0
Mixed + Missing	-0.281**	-0.087	-0.052
All Non-British	0.185*	0.143**	0.076
Working Status			
No one works	0	0	0
At least 1 works but not long hours	0.257***	0.087*	0.135***
At least 1 works (but not all) long hours	0.237***	0.142***	0.116**
All work long hours	0.414***	0.339***	0.022
Presence of pensioner			
No pensioner	0	0	0
At least 1 pensioner	-0.349***	-0.131**	-0.049
All pensioners	-0.501***	-0.214***	-0.163***
Material deprivation	0.234***	-0.003	0.127***
Owner/Mortgager	-0.298***	-0.042	-0.160***
No political interest	0.240***	0.054	0.078*
Community Attachment			
Q4	0	0	0
Q3	0.001	0.074*	-0.044
Q2	0.032	0.052	-0.018
Q1	0.144*	0.082*	0.026
Missing	0.381**	0.175*	0.030
Consent to Data Linkage			
All consent	0	0	0
Some consent + missing	0.023	-0.037	0.053
No one consent	0.003	0.020	0.014
No one present during interview	0.089*	0.052*	0.030
No suspicion during interview	-0.246**	-0.082	-0.207**
Excellent understanding of questions	-0.055	0.010	-0.015
Cooperative respondent household	-0.191	0.023	-0.105**
Item nonresponse (log)	0.002	-0.003	0.001
Dummy Item nonresponse (log)	0.502	-0.023	0.042
Same LSOA (W1, W2, W3 & W4)	0.426***	-0.060	0.166***
Same Interviewer (W1, W2, W3 & W4)	-0.050	-0.071	-0.121
Same LSOA and Interviewer (Interaction)	-0.114	0.0029	0.061
Constant	5.559***	2.540***	2.450***
Constant (Random Intercept)	0.950***	0.405***	0.453***
Observations	20012	18218	16562

Continued on next page

Continued from previous page

Log likelihood	-47853.2	-33500.2	-31375.3
AIC	95838.4	67132.4	62882.7
*Degrees of Freedom (63). * p<0.05, ** p<0.01, *** p<0.001*			

Estimated coefficients for Wave 4 Field Effort Outcomes (Calls to Make Contact, Post-Contact Calls to Cooperation, and total Number of Calls). Random-intercept, linear regression models were specified to account for unobserved interviewer effects.

Bibliography

AAPOR (2016). Standard definitions: Final dispositions of case codes and outcome rates for surveys. *American Association for Public Opinion Research*, 9th edition.

Al Baghal, T. and Lynn, P. (2015). Using motivational statements in web-instrument design to reduce item-missing rates in a mixed-mode context. *Public Opinion Quarterly*, 79(2):568–579.

Barge, S. and Gehlbach, H. (2012). Using the theory of satisficing to evaluate the quality of survey data. *Research in Higher Education*, 53(2):182–200.

Barnes, W., Bright, G., and Hewat, C. (2008). Making sense of labour force survey response rates. *The Labour gazette*, 2(12):32–42.

Baron, R. M. and Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6):1173.

Barrios, M., Villarroya, A., Borrego, Á., and Ollé, C. (2011). Response rates and data quality in web and mail surveys administered to phd holders. *Social Science Computer Review*, 29(2):208–220.

Bates, N. (2004). Contact histories: A tool for understanding attrition in panel surveys. In *Paper presented at the 58th Annual AAPOR Conference. May 13-16 2004. Phoenix, Arizona*.

Bates, N., Dahlhamer, J., Phipps, P., Safir, A., and Tan, L. (2010). Assessing contact history paradata quality across several federal surveys. In *Proceedings of the American Statistical Association 2010 Joint Statistical Meeting*.

Biemer, P. P. (2010). Total survey error: Design, implementation, and evaluation. *Public Opinion Quarterly*, 74(5):817–848.

Blom, A. G. (2009). Nonresponse bias adjustments: what can process data contribute? Technical report, ISER Working Paper Series.

Blom, A. G., Jäckle, A. E., and Lynn, P. (2010). The use of contact data in understanding cross-national differences in unit non-response. In Harkness, J. A., Braun, M., Edwards, B., Johnson, T. P., Lyberg, L. E., Mohler, P. P., Pennell, B.-E., and Smith, T. W., editors, *Survey methods in multinational, multiregional, and multicultural contexts*, Wiley Series in Survey Methodology, pages 335–355. John Wiley & Sons Inc., Hoboken, NJ, US.

Bohme, R. (2011). Research note on progress indicators in online surveys. Technical report, Westfalische Wilhelms-Universitat Munster.

Brunton-Smith, I., Sturgis, P., and Leckie, G. (2017). Detecting and understanding interviewer effects on survey data by using a cross-classified mixed effects location–scale model. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(2):551–568.

Burton, J. (2016). Moving online with Understanding Society. *Institute for Social and Economic Research Blog*, July, 16th.

Calinescu, M., Bhulai, S., and Schouten, B. (2011). Optimal scheduling of contact attempts in mixed-mode surveys. *Statistics Netherlands, Hague, The Netherlands*.

Calinescu, M., Schouten, B., and Bhulai, S. (2012). Adaptive survey designs that minimize nonresponse and measurement risk. *Statistics Netherlands*.

Callegaro, M. (2013). Paradata in web surveys. In Kreuter, F., editor, *Improving Surveys with Paradata: Analytic Uses of Process Information*, Wiley Series in Survey Methodology, pages 261–279. Wiley, Hoboken, N.J.

Callegaro, M., Yang, Y., Bhola, D. S., Dillman, D. A., and Chin, T.-Y. (2009). Response latency as an indicator of optimizing in online questionnaires. *Bulletin de Méthodologie Sociologique*, 103(1):5–25.

Campanelli, P. and O'Muircheartaigh, C. (1999). Interviewers, interviewer continuity, and panel survey nonresponse. *Quality and Quantity*, 33(1):59–76.

Campanelli, P., Sturgis, P., and Purdon, S. (1997a). *Can You Hear Me Knocking: an Investigation into the Impact of Interviewers on Survey Response Rates.* Social and Community Planning Research, London.

Campanelli, P., Sturgis, P., and Purdon, S. (1997b). *Exploring the impact of survey doorstep interactions*. Social and Community Planning Research, London.

Clark, W. A. (2013). Life course events and residential change: Unpacking age effects on the probability of moving. *Journal of Population Research*, 30(4):319–334.

Conrad, F., Couper, M., Tourangeau, R., and Peytchev, A. (2005). Impact of progress feedback on task completion: First impressions matter. In *CHI'05 Extended Abstracts on Human Factors in Computing Systems*, pages 1921–1924. ACM.

Conrad, F. G., Couper, M. P., and Tourangeau, R. (2003a). Interactive features in web surveys. In *joint meetings of the American Statistical Association, San Francisco, CA*.

Conrad, F. G., Couper, M. P., Tourangeau, R., and Peytchev, A. (2003b). Effectiveness of progress indicators in web surveys. it's what's up front that counts. In *Proceedings of the Fifth International ASC Conference*, Chesham, UK.

Conrad, F. G., Couper, M. P., Tourangeau, R., and Peytchev, A. (2010). The impact of progress indicators on task completion. *Interacting with computers*, 22(5):417–427.

Coulter, R. and Scott, J. (2015). What motivates residential mobility? re-examining self-reported reasons for desiring and making residential moves. *Population, Space and Place*, 21(4):354–371.

Couper, M. P. (1997). Survey introductions and data quality. *Public Opinion Quarterly*, pages 317–338.

Couper, M. P. and Groves, R. M. (1996). Household-level determinants of survey nonresponse. *New Directions for Evaluation*, 1996(70):63–79.

Couper, M. P. and Kreuter, F. (2013). Using paradata to explore item level response times in surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(1):271–286.

Couper, M. P. and Ofstedal, M. B. (2009). Keeping in contact with mobile sample members. In Lynn, P., editor, *Methodology of Longitudinal Surveys*, Wiley Series in Survey Methodology, pages 183–203. Wiley, Chichester.

Couper, M. P., Traugott, M. W., and Lamias, M. J. (2001). Web survey design and administration. *Public opinion quarterly*, 65(2):230–253.

Crawford, S. D., Couper, M. P., and Lamias, M. J. (2001). Web surveys perceptions of burden. *Social science computer review*, 19(2):146–162.

De Keulenaer, F. (2005). Using process data to predict attrition from a panel survey: a case study. In *Proceedings of the Survey Research Methods Section, American Statistical Association*.

Denscombe, M. (2008). The length of responses to open-ended questions: A comparison of online and paper questionnaires in terms of a mode effect. *Social Science Computer Review*, 26(3):359–368.

Dijkstra, W. and Smit, J. H. (2002). Persuading reluctant recipients in telephone surveys. *Survey nonresponse*, pages 121–34.

Durrant, G. B., D'Arrigo, J., and Muller, G. (2013a). Modeling call record data: Examples from cross-sectional and longitudinal surveys. In Kreuter, F., editor, *Improving Surveys with Paradata: Analytic Uses of Process Information*, pages 281–308. Wiley, London.

Durrant, G. B., D'Arrigo, J., and Steele, F. (2011). Using paradata to predict best times of contact, conditioning on household and interviewer influences. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(4):1029–1049.

Durrant, G. B., D'Arrigo, J., and Steele, F. (2013b). Analysing interviewer call record data by using a multilevel discrete time event history modelling approach. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(1):251–269.

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

Durrant, G. B., Maslovskaya, O., and Smith, P. W. (2015). Modeling final outcome and length of call sequence to improve efficiency in interviewer call scheduling. *Journal of Survey Statistics and Methodology*, 3(3):397–424.

Durrant, G. B., Maslovskaya, O., and Smith, P. W. (2017). Using prior wave information and paradata: Can they help to predict response outcomes and call sequence length in a longitudinal study? *Journal of Official Statistics*, 33(3):801–833.

Emde, M. and Fuchs, M. (2012). Using adaptive questionnaire design in open-ended questions: A field experiment. In *American Association for Public Opinion Research (AAPOR) 67th Annual Conference, San Diego, USA*.

Galesic, M. (2006). Dropouts on the web: Effects of interest and burden experienced during an online survey. *Journal of Official Statistics*, 22(2):313–328.

Ganassali, S. (2008). The influence of the design of web survey questionnaires on the quality of responses. *Survey Research Methods*, 2(1):21–32.

Gould, W. (1998). Interquartile and simultaneous-quantile regression. *Stata Technical Bulletin*, 7(38).

Groves, R. M., Cialdini, R., and Couper, M. P. (1992). Understanding the decision to participate in a survey.pdf. *Public Opinion Quarterly*, 56(4):475–495.

Groves, R. M. and Couper, M. P. (1996). Contact level influences on cooperation in face-to—face surveys. *Journal of Official Statistics*, 12(1):63–83.

Groves, R. M. and Couper, M. P. (1998). *Nonresponse in Household Interview Surveys*. Wiley series in Probability and Statistics: Survey Methodology Section. Wiley, New York, 1 edition.

Groves, R. M. and 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.

Groves, R. M. and Lyberg, L. (2010). Total survey error: Past, present, and future. *Public Opinion Quarterly*, 74(5):849–879.

Hall, J., Brown, V., Nicolaas, G., and Lynn, P. (2013). Extended field efforts to reduce the risk of non-response bias: Have the effects changed over time? Can weighting achieve the same effects? *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 117(1):5–25.

Hao, L., Naiman, D. Q., and Naiman, D. Q. (2007). *Quantile regression*. Quantitative Applications in the Social Sciences. Sage, Thousand Oaks.

Heerwegh, D. (2003). Explaining response latencies and changing answers using client-side paradata from a web survey. *Social Science Computer Review*, 21(3):360–373.

Heerwegh, D. and Loosveldt, G. (2002). An evaluation of the effect of response formats on data quality in web surveys. *Social science computer review*, 20(4):471–484.

Heerwegh, D. and Loosveldt, G. (2006). An experimental study on the effects of personalization, survey length statements, progress indicators, and survey sponsor logos in web surveys. *Journal of Official Statistics*, 22(2):191.

Henly, M. and Bates, N. (2006). Using call records to understand response in panel surveys. In *American Statistical Association. Survey Research Methods Section*, Alexandria, Virginia.

Holbrook, A. L., Green, M. C., and Krosnick, J. A. (2003). Telephone versus face-to-face interviewing of national probability samples with long questionnaires: Comparisons of respondent satisficing and social desirability response bias. *Public Opinion Quarterly*, 67(1):79–125.

Holland, J. L. and Christian, L. M. (2009). The influence of topic interest and interactive probing on responses to open-ended questions in web surveys. *Social Science Computer Review*, 27(2):196–212.

Kaczmirek, L. (2008). *Human-survey interaction: Usability and nonresponse in online surveys*. PhD thesis, Universität Mannheim.

Kaczmirek, L., Meitinger, K., and Behr, D. (2017). *Higher data quality in web probing with EvalAnswer: a tool for identifying and reducing nonresponse in openended questions*, volume 2017/01. GESIS Papers.

Kalsbeek, W. D., Yang, J., and Agans, R. P. (2002). Predictors of nonresponse in a longitudinal survey of adolescents. In *Proceedings of the American Statistical Association (Survey Research Methods Section)*, pages 1740–1745.

Kaminska, O., McCutcheon, A. L., and Billiet, J. (2010). Satisficing among reluctant respondents in a cross-national context. *Public Opinion Quarterly*, 74(5):956–984.

Knies, G. E. (2015). Understanding society-the uk household longitudinal study: Waves 1–5, 2009-2014, user manual. *Colchester, UK: University of Essex*.

Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50.

Koenker, R. and Hallock, K. F. (2001). Quantile regression. *Journal of economic perspectives*, 15(4):143–156.

Kreuter, F., editor (2013a). *Improving surveys with paradata: analytic uses of process information*. Wiley Series in Survey Methodology. Wiley, Hoboken,NJ.

Kreuter, F. (2013b). Improving surveys with paradata: Introduction. In Kreuter, F., editor, *Improving Surveys with Paradata: Analytic Uses of Process Information*, Wiley Series in Survey Methodology, pages 1–9. Wiley, Hoboken, N.J.

Kreuter, F. and Casas-Cordero, C. (2010). Paradata. *German Council for Social and Economic Data Working Paper*, 2010. Berlin.

Kreuter, F., Couper, M. P., and Lyberg, L. E. (2010). The use of paradata to monitor and manage survey data collection. In *Proceedings of the Joint Statistical Meetings, American Statistical Association*, pages 282–296.

Kreuter, F. and Kohler, U. (2009). Analyzing contact sequences in call record data. potential and limitations of sequence indicators for nonresponse adjustments in the european social survey. *Journal of Official Statistics*, 25(2):203–226.

Kreuter, F. and Olson, K. (2013). Paradata for nonresponse error investigation. In Kreuter, F., editor, *Improving Surveys with Paradata: Analytic Uses of Process Information*, Wiley Series in Survey Methodology, pages 13–42. Wiley, Hoboken, N.J.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied cognitive psychology*, 5(3):213–236.

Krosnick, J. A., Narayan, S., and Smith, W. R. (1996). Satisficing in surveys: Initial evidence. *New directions for evaluation*, 1996(70):29–44.

Laurie, H. and Lynn, P. (2009). The use of respondent incentives on longitudinal surveys. In Lynn, P., editor, *Methodology of Longitudinal Surveys*, Wiley Series in Survey Methodology, pages 205–233. Wiley, Chichester.

Laurie, H., Smith, R. A., and Scott, L. (1999). Strategies for reducing nonresponse in a longitudinal panel survey. *Journal of Official Statistics*, 15(2):269–282.

Leeuw, E. D. d., Hox, J. J., and Huisman, M. (2003). Prevention and treatment of item nonresponse. *Journal of Official Statistics*, 19(2):153–176. Universiteit Utrecht.

Lepkowski, J. M. and Couper, M. P. (2002). Nonresponse in the second wave of longitudinal household surveys. In Groves, R. M., Dillman, D. A., Eltinge, J. L., and Little, R. J. A., editors, *Survey Nonresponse*, Wiley series in Probability and Statistics, pages 259–272. Wiley, New York.

Lipps, O. (2012). Using information from telephone panel surveys to predict reasons for refusal (lipps 2012).

Loosveldt, G. and Billiet, J. (2002). Item nonresponse as a predictor of unit nonresponse in a panel survey. *Journal of Official Statistics*, 18(4):545.

Lugtig, P. (2014). Panel attrition separating stayers, fast attriters, gradual attriters, and lurkers. *Sociological Methods & Research*, 43(4):699–723.

Lynn, P. (2009a). Methods for longitudinal surveys. In Lynn, P., editor, *Methodology of Longitudinal Surveys*, Wiley Series in Survey Methodology, pages 1–19. Wiley, Chichester.

Lynn, P. (2009b). Sample design for understanding society. *ISER Working Paper Series*, 2009(01). Colchester: Institute for Social and Economic Research, University of Essex.

Lynn, P. (2014). Coping with nonresponse. In Engel, U., Jann, B., Lynn, P., Scherpenzeel, A., and Sturgis, P., editors, *Improving Survey Methods: Lessons from Recent Research*, European Association of Methodology Series, pages 319–338. Routledge, New York.

Lynn, P. and Clarke, P. (2002). Separating refusal bias and non-contact bias: Evidence from UK national surveys. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 51(3):319–333.

Lynn, P., Clarke, P., Martin, J., and Sturgis, P. (2002). The effects of extended interviewer efforts on nonresponse bias. In Groves, R. M., Dillman, D. A., Eltinge, J. L., and Little, R. J. A., editors, *Survey Nonresponse*, Wiley series in Probability and Statistics, pages 135–147. Wiley, New York, 1 edition.

Lynn, P., Kaminska, O., and Goldstein, H. (2014). Panel Attrition: How Important is Interviewer Continuity? *Journal of Official Statistics*, 30(3).

Malhotra, N. (2008). Completion time and response order effects in web surveys. *Public Opinion Quarterly*, 72(5):914–934.

Manfreda, K. L., Bosnjak, M., Berzelak, J., Haas, I., Vehovar, V., and Berzelak, N. (2008). Web surveys versus other survey modes: A meta-analysis comparing response rates. *Journal of the Market Research Society*, 50(1):79.

Matzat, U., Snijders, C., and van der Horst, W. (2009). Effects of different types of progress indicators on drop-out rates in web surveys. *Social Psychology*, 40(1):43–52.

Mercer, A. (2012). Using paradata to understand effort and attrition in a panel survey. *Section on Survey Research Methods – JSM*.

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

Nederhof, A. J. (1987). When neutrality is negative. *Quality and Quantity*, 21(4):425–432.

Nicoletti, C. and Peracchi, F. (2005). Survey response and survey characteristics: microlevel evidence from the european community household panel. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(4):763–781.

Office For National Statistics (ONS) (2011). Super Output Areas (SOA). http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census/super-output-areas–soas-/index.html.

Oksenberg, L., Coleman, L., and Cannell, C. F. (1986). Interviewers' voices and refusal rates in telephone surveys. *Public Opinion Quarterly*, 50(1):97–111.

Olson, K. (2013). Paradata for nonresponse adjustment. *The ANNALS of the American Academy of Political and Social Science*, 645(1):142–170.

Olson, K. and Parkhurst, B. (2013). Collecting paradata for measurement error evaluations. In *Improving Surveys with Paradata: Analytic Uses of Process Information*. wiley.

Oudejans, M. and Christian, L. M. (2010). Using interactive features to motivate and probe responses to open-ended questions. *Social and behavioral research and the Internet*, pages 215–244.

Peytchev, A. (2009). Survey breakoff. *Public Opinion Quarterly*, 73(1):74–97.

Pickery, J., Loosveldt, G., and Carton, A. (2001). The effects of interviewer and respondent characteristics on response behavior in panel surveys a multilevel approach. *Sociological Methods & Research*, 29(4):509–523.

Pollien, A. and Joye, D. (2014). Patterns of contact attempts in surveys. In *Advances in Sequence Analysis: Theory, Method, Applications*, pages 285–304. Springer.

Qualtrics (2015). Qualtrics (version 11.2015). http://www.qualtrics.com.

Rabe, B. and Taylor, M. (2010). Residential mobility, quality of neighbourhood and life course events. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 173(3):531–555.

Rasbash, J., Charlton, C., Browne, W., Healy, M., and Cameron, B. (2015). *MLwiN Version 2.34*. Centre for Multilevel Modelling, University of Bristol.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological bulletin*, 114(3):510.

Reja, U., Manfreda, K. L., Hlebec, V., and Vehovar, V. (2003). Open-ended vs. close-ended questions in web questionnaires. *Developments in applied statistics*, 19(1):159–177.

Schouten, B., Calinescu, M., and Luiten, A. (2013). Optimizing quality of response through adaptive survey designs. *Survey Methodology*, 39(1):29–58.

Sharp, L. M. and Frankel, J. (1983). Respondent burden: A test of some common assumptions. *Public Opinion Quarterly*, 47(1):36–53.

Smith, T. W. (2011). The report of the international workshop on using multi-level data from sample frames, auxiliary databases, paradata and related sources to detect and adjust for nonresponse bias in surveys. *International Journal of Public Opinion Research*, 23(3):389–402.

Smyth, J. D., Dillman, D. A., Christian, L. M., and McBride, M. (2009). Open-ended questions in web surveys: Can increasing the size of answer boxes and providing extra verbal instructions improve response quality? *Public Opinion Quarterly*, 73(2):325–337.

StataCorp (2017). *Stata Statistical Software: Release 15*. StataCorp LP, College Station, TX.

Stieger, S. and Reips, U.-D. (2010). What are participants doing while filling in an online questionnaire: A paradata collection tool and an empirical study. *Computers in Human Behavior*, 26(6):1488–1495.

Stoop, I. A. (2004). Surveying nonrespondents. *Field Methods*, 16(1):23–54.

Stoop, I. A. (2005). *The hunt for the last respondent: Nonresponse in sample surveys*. Sociaal en Cultureel Planbu.

Tourangeau, R., Conrad, F. G., and Couper, M. P. (2013). *The science of web surveys*. Oxford University Press.

Tourangeau, R., Rips, L. J., and Rasinski, K. (2004). *The psychology of survey response*. Cambridge University Press, Cambridge.

Uhrig, S. C. (2008). The nature and causes of attrition in the british household panel study. *ISER Working Paper Series*, 2008(05). Colchester: Institute for Social and Economic Research, University of Essex.

Vassallo, R., Durrant, G., and Smith, P. (2017). Separating interviewer and area effects by using a cross-classified multilevel logistic model: simulation findings and implications for survey designs. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(2):531–550.

Vassallo, R., Durrant, G. B., Smith, P. W., and Goldstein, H. (2014). Interviewer effects on non-response propensity in longitudinal surveys: a multilevel modelling approach. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*.

Villar, A., Callegaro, M., and Yang, Y. (2013). Where am I? A meta-analysis of experiments on the effects of progress indicators for web surveys. *Social Science Computer Review*, 31(6):744–762.

Wagner, J. (2013). Using paradata-driven models to improve contact rates in telephone and face-to-face surveys. In Kreuter, F., editor, *Improving Surveys with Paradata: Analytic Uses of Process Information*, Wiley Series in Survey Methodology, pages 145–170. Wiley, Hoboken, N.J., 1 edition.

Wagner, J. R. (2008). *Adaptive survey design to reduce nonresponse bias, PhD Thesis*. University of Michigan, Ann Arbor.

Watson, N. and Wooden, M. (2009). Identifying factors affecting longitudinal survey response. In Lynn, P., editor, *Methodology of Longitudinal Surveys*, Wiley Series in Survey Methodology, pages 157–181. Wiley, Chichester.

West, B. T. (2013). The effects of errors in paradata on weighting class adjustments: A simulation study. In Kreuter, F., editor, *Improving Surveys with Paradata: Analytic Uses of Process Information*, Wiley Series in Survey Methodology, pages 361–388. Wiley, Hoboken, N.J.

Yan, T., Conrad, F. G., Tourangeau, R., and Couper, M. P. (2011). Should I stay or should I go: The effects of progress feedback, promised task duration, and length of questionnaire on completing web surveys. *International Journal of Public Opinion Research*, 23:131—-147.

Yan, T. and Olson, K. (2013). Analyzing paradata to investigate measurement error. In Kreuter, F., editor, *Improving Surveys with Paradata: Analytic Uses of Process Information*, Wiley Series in Survey Methodology, pages 73–95. Wiley, Hoboken, N.J.

Yan, T. and Tourangeau, R. (2008). Fast times and easy questions: the effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology*, 22(1):51–68.

Zandt, T. V. (2002). Analysis of response time distributions. In *Stevens' Handbook of Experimental Psychology*. John Wiley & Sons, Inc.

Zuell, C., Menold, N., and Körber, S. (2015). The influence of the answer box size on item nonresponse to open-ended questions in a web survey. *Social Science Computer Review*, 33(1):115–122.