

**Reducing non-response in longitudinal surveys
by improving survey practice**

Lisa Calderwood

**Thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy in
Survey Methodology at UCL**

2016

Acknowledgements

I would like to thank my supervisor, Professor John Micklewright, for his advice and guidance, and Professor Jane Elliott for her support and encouragement.

I would also like to thank my parents for their support and guidance throughout my education and their encouragement and patience throughout my PhD study.

Statement of originality

I confirm that this thesis is not substantially the same as any previously submitted or currently being submitted for a degree, diploma, or similar qualification at any other university or similar institution.

Lisa Calderwood

12 January 2015

List of Publications

Paper 1:

Calderwood, L., Plewis, I., Ketende S.C. and Taylor, R. (2010). Experimental testing of refusal conversion strategies in a large-scale longitudinal study. CLS Working Paper 2010/9. London: Centre for Longitudinal Studies.

This paper was submitted for publication to a journal in May 2010, and revised in response to reviewers comments three times (May 2011, November 2011 and April 2012). The journal finally rejected the paper in October 2012, despite giving it a conditional acceptance in September 2011.

The paper has subsequently been revised to incorporate recently available data from a later wave of the survey. The title and authorship of the paper has also changed to reflect content and contributions of this more recent draft, and is now as follows:

Calderwood, L., Plewis, I., Ketende, S. and Mostafa, T. Evaluating the immediate and longer- term impact of refusal conversion strategies in a large-scale longitudinal study.

This latest version is the version which is being submitted as part of my PhD. However, please note that this paper is referenced through the integrative summary as 'Calderwood et al., 2010', reflecting the original publication.

Paper 2:

Calderwood, L. (2014). Improving between-wave mailings on longitudinal surveys: a randomised experiment on the UK Millennium Cohort Study. *Survey Research Methods*. 8 (2), pp. 99-108.

*This paper is also published as a working paper:
Calderwood, L. (2012). Improving between-wave mailings on longitudinal surveys: A randomised experiment on the UK Millennium Cohort Study. CLS Working Paper 2012/4. London: Centre for Longitudinal Studies.*

Paper 3:

Calderwood, L. (2013). The role of respondent characteristics in tracking on longitudinal surveys: evidence from the UK Millennium Cohort Study. *Lifecourse and Longitudinal Studies*, 4(2), 105-118. DOI: <http://dx.DOI.org/10.14301/lcs.v4i2.231>

*This paper is also published as a working paper:
Calderwood, L. (2010). Keeping in touch with mobile families in the UK Millennium Cohort Study. CLS Working Paper, 2010/7. London: Centre for Longitudinal Studies.*

Paper 4:

Brown, M., and Calderwood, L. (2014). Can encouraging respondents to contact interviewers to make appointments reduce fieldwork effort? Evidence from a randomized experiment in the UK. *Journal of Survey Statistics and Methodology*. 2 (4), pp. 484-498. DOI: 10.1093/jssam/smu017.

This paper is also published as a working paper:

Brown, M. and Calderwood, L. (2013). Can encouraging respondents to contact interviewers to make appointments boost co-operation and save costs? Evidence from a randomised experiment in the UK. CLS Working Paper 2013/9. London: Centre for Longitudinal Studies.

Statement of joint work – Paper 1

Statement of Lisa Calderwood’s contribution to the following paper:

Calderwood, L., Plewis, I., Ketende S.C. and Taylor, R. (2010). Experimental testing of refusal conversion strategies in a large-scale longitudinal study. CLS Working Paper 2010/9. London: Centre for Longitudinal Studies.

Note that this reference refers to the original publication of this paper. A later draft of this paper has been submitted with a different title and authorship as follows:

Calderwood, L., Plewis, I., Ketende, S. and Mostafa, T. Evaluating the immediate and longer term impact of refusal conversion strategies in a large-scale longitudinal study.

The research reported in this paper was carried out as part of a research project entitled ‘Predicting and Preventing Non-Response in Cohort Studies’ (Ref: RES-175-25-0010) funded by the Economic and Social Research Council (ESRC) as part of their Survey Design and Measurement Initiative (SDMI). The project ran from January 2008 to December 2009, and the experiment reported in the paper was carried out on the fourth wave of the Millennium Cohort Study (MCS) in 2008. Professor Ian Plewis was the Principal Investigator of the project. Lisa Calderwood and Rebecca Taylor (who worked at NatCen, the survey delivery partner for the MCS) were co-investigators, and Sosthenes Ketende (then at CLS) was the research officer on the project.

The scientific design of the experiment was led by Plewis, with input from Calderwood and Taylor. The implementation of the experiment was carried out by NatCen field staff and overseen primarily by Taylor, with regular reporting to Plewis and Calderwood.

The data preparation and analysis, including the preparation of almost all of the tables in the paper, was done mainly by Calderwood, with significant input on analysis strategy and presentation of results from Plewis and some support on data preparation and analysis from Ketende. The exception is Table 6, which uses wave five data and has been recently added to the paper, and which was prepared by Mostafa with significant input on analysis strategy from Plewis and Calderwood.

The writing and interpretation of results, including responding to reviewers’ comments, was done mainly by Calderwood, with significant input from Plewis.

Lisa Calderwood

12 January 2015

I confirm that this is an accurate account of how this paper was written.

Professor Ian Plewis

13 January 2015

Statement of joint work – Paper 4

Statement of Lisa Calderwood's contribution to the following paper:

Brown M., and Calderwood, L. (2014). Can encouraging respondents to contact interviewers to make appointments reduce fieldwork effort? Evidence from a randomized experiment in the UK. Journal of Survey Statistics and Methodology. 2 (4), pp. 484-498.
DOI: 10.1093/jssam/smu017.

The research reported in this paper was motivated by a presentation (Kochanek et al., 2010) at the Panel Survey Methods Workshop 2010 in Mannheim, Germany, attended by both Brown and Calderwood,

The paper reports the results of a randomised experiment carried out on the Innovation Panel of *Understanding Society*, the UK Household Longitudinal Survey (UKHLS) in 2011. The application for the inclusion of the experiment onto the panel was jointly written by both Brown and Calderwood.

The scientific design of the experiment was carried out by Brown and Calderwood, with input into implementation from researchers at ISER and NatCen, who were the fieldwork delivery partner for the UKHLS.

The data preparation and analysis, including the preparation of all tables in the paper, was done mainly by Brown, with input on analysis strategy and presentation of results from Calderwood.

The writing of the paper was done jointly by Brown and Calderwood, with Calderwood leading on the background and literature review and Brown leading on the description of the experiment and results. In later iterations, including in response to reviewers comments, Brown took the lead on revising the paper.

Lisa Calderwood

12 January 2015

I confirm that this is an accurate account of how this paper was written.

Matt Brown

13 January 2015

Abstract

My thesis is about the prevention of unit non-response in longitudinal surveys. Non-response can lead to error in survey estimates, meaning they give a biased picture of the true value in the population. The main sources of non-response in surveys are non-contact and refusal. For longitudinal surveys, which follow the same people over time, non-location is an additional source of non-response. Non-response over time in a longitudinal survey is often referred to as attrition.

There are two broad approaches to dealing with non-response in surveys. The first is to make statistical corrections to the survey estimates to take account of non-response error. The second is to try to prevent non-response by improving how surveys are designed and conducted. My research takes a prevention approach, and addresses all three of the major components of non-response in longitudinal surveys. I use randomised experiments to evaluate fieldwork interventions designed to reduce non-response and prior wave data to evaluate the impact of fieldwork interventions on non-response bias. I use data from large-scale, high quality datasets: the UK Millennium Cohort Study and the Innovation Panel of *Understanding Society*, the UK Household Longitudinal Study.

My main findings are: refusal conversion can be an effective way of reducing non-response both in the immediate and longer term and can lead to some modest reductions in non-response bias; improving the design of the covering letter used on between-wave mailings can improve return rates from sample members with lower levels of education and those who speak languages other than English at home; respondent characteristics are related to the success of both office and field tracking; sample members who respond to an invitation to make their own interview appointment require less fieldwork effort overall and they are more likely to do this when a financial incentive is offered.

Table of Contents

Integrative summary	11
Section 1: Introduction and background	12
Section 2: Co-operation	26
Section 3: Location	41
Section 4: Contact	60
References	69
Original Publications	81
Paper 1: Calderwood, Plewis, Ketende and Taylor (2010)	82
Paper 2: Calderwood (2014)	109
Paper 3: Calderwood (2013)	133
Paper 4: Brown and Calderwood (2014)	157
Bibliography	175

Integrative summary

Section 1: Introduction and background

Section 2: Co-operation

Section 3: Location

Section 4: Contact

References

Section 1: Introduction and background

In this section I explain what my PhD thesis is about, why this is an important topic of study and how my research fits with existing knowledge. The topic of my thesis is the prevention of unit non-response in longitudinal surveys. Non-response is failure to obtain measurements for the units, usually people in social surveys, which have been sampled for a particular survey. This can occur either because the sampled person does not take part at all, referred to as unit non-response, or because they take part but fail to provide specific measurements, referred to as item non-response. My thesis is about unit non-response. Throughout my thesis the term non-response is used to refer to unit, rather than item, non-response.

There are two broad approaches to dealing with non-response in surveys. The first is to make statistical adjustments to the survey estimates to take account of non-response error. The second is to try to prevent non-response by improving how surveys are designed and conducted. My thesis is about the prevention of unit non-response. Specifically, it is about how changes or improvements to the way in which longitudinal surveys are conducted can help to reduce, or prevent, non-response. There are also several different types of surveys. Longitudinal surveys collect information from the same people at multiple time points, referred to as ‘data collection waves’, whereas cross-sectional surveys collect information at only one point in time or in the case of repeated cross-sectional surveys from different people at multiple time points. My thesis is about the prevention of unit non-response in longitudinal surveys.

Why is survey non-response important?

Survey research relies on statistical inference from respondents - the group of people who take part in any given survey - to a much bigger group of people, the population of interest or target population. This way results or findings based on information collected from respondents, referred to as survey estimates, can be generalised to the target population. This is a core premise of the survey method. Non-response can lead to error in the survey estimates, meaning they give a biased picture of the true value in the population. This is why non-response is important.

For longitudinal surveys, which follow the same people over time, non-response over time, often referred to as attrition, can threaten the accuracy of the survey findings if those who continue to take part are different from those who do not. Differential drop-

out can also pose a threat to accurate measurement of change over time, which is one of the key purposes of longitudinal research.

As most social surveys are voluntary, non-response is inevitable. For this reason, it is only possible to reduce non-response rather than prevent it completely. The level of non-response on a particular survey is reported via the survey response rate, which is a key indicator of survey quality. The response rate refers to the percentage of eligible sampled units who take part in the survey i.e. for whom at least some measurements are obtained. The non-response rate is the complement of the response rate. Generally speaking, these rates are relatively straightforward to calculate and interpret, though complexities can arise relating to determining or estimating the eligibility of all sampled units, clustering of sampled units and unequal probabilities of selection for different sampled units (for a more detailed discussion of these issues see e.g. Groves et al., 2009).

However, non-response does not *necessarily* lead to bias in the survey estimates. Non-response bias occurs when respondents have different values than non-respondents on the survey variables. Bias is a property of a survey estimate, and can therefore vary across different estimates within the same survey. The extent to which non-response leads to bias in survey estimates depends on whether the factors affecting the propensity to respond among sampled units are also related to the variables of interest in the survey; the more highly they are correlated with each other, the larger the non-response bias will be. In statistical terms, it is the *covariance* of the variable of interest and the response propensity that determines the level of non-response bias. Groves (2006) discusses three alternative causal models underlying the relationship between response propensity and non-response bias. In the 'separate cause' model, the factors affecting propensity to respond are completely separate from the factors affecting the survey variables. In this situation, the covariance is zero and therefore non-response does not lead to bias in the survey estimates. In the 'common cause' model, the same factors predict both propensity to respond and the survey variables. In this situation, the covariance is zero once these factors are controlled for and therefore non-response does not lead to bias in the survey estimates. In the 'survey variable cause' model, the survey variables themselves are related to the propensity to respond. In this situation, the covariance is non-zero and non-response leads to bias in the survey estimates.

One of the main problems in measuring the extent to which survey estimates are biased is that, for most surveys, there is often not very much information available about *all* of the people who are sampled for the survey. Hence, it is often very difficult to measure how similar or different survey respondents and non-respondents are, and therefore to estimate whether the differences between them are likely to lead to bias in the survey estimates. This is one of the main reasons why most survey practitioners tend to focus their efforts on minimising the non-response rate. This is a sensible and reasonable approach as, all other things being equal, minimising the non-response rate should also minimise the *risk* of bias. Even if non-respondents are very different to respondents in ways which are related to the survey variables, if non-response is very low then the impact of this difference on the bias of the survey estimates will be smaller than if non-response is higher. However, high response rates do not necessarily mean that survey results are unbiased, and low response rates do not necessarily mean that survey results are biased (Groves and Peytcheva, 2008).

Although longitudinal surveys have the same difficulty as cross-sectional surveys in assessing bias at their first or baseline wave, for non-response at subsequent waves of data collection information about both respondents and non-respondents is usually available from the baseline or previous wave. This makes it easier to assess the risk of non-response bias due to attrition in longitudinal surveys.

Another reason non-response is important is because it is a key determinant of sample size, and sample size is the main determinant of the precision of survey estimates. Non-response leads to increased variance in survey estimates, which is another type of non-response error. All other things being equal, a larger sample size leads to more precise survey estimates, and more statistically reliable survey findings. For longitudinal surveys, reductions in sample size over time caused by attrition can undermine the statistical reliability of survey findings.

What can be done about survey non-response?

Overall, there are two broad approaches to dealing with non-response in surveys. The first is to make statistical corrections to survey estimates to take account of non-response error. The second is to try to reduce non-response by improving how surveys are designed and conducted.

Weighting is the most commonly used method of statistical adjustment for non-response error. It is also used to correct survey estimates to take account of differential selection probabilities at the sampling stage. There are various different types of non-response weighting, but all depend on the availability of auxiliary information about both the survey population and the survey respondents. The basic idea of non-response weighting is to make statistical corrections to the survey estimates for any known differences between the respondents and the population. Generally speaking, the more auxiliary information is available the more effective the weighting. However, effective weighting also requires the auxiliary information to be related to both the response propensity among the sampled units of the survey population and to the survey estimates (see e.g. Bethlehem, 2002 for a more detailed discussion). As noted above, longitudinal surveys have the benefit of being able to use information collected at the baseline wave in non-response adjustment at later waves. However, longitudinal surveys also have the challenge of producing both cross-sectional and longitudinal weights.

My research takes a prevention approach to dealing with survey non-response. The aim of prevention approaches is to reduce non-response through improvements to survey design and implementation. My professional training is as a survey practitioner and this has strongly motivated my research interests in this area. In order to try to prevent non-response, it is first critical to understand the main sources of sources of non-response, and their causes.

What are the main sources of non-response, and their causes?

On cross-sectional surveys, there are two major sources of non-response: non-contact and refusal. A further residual category of other non-interview is also often used and includes reasons such as incapacity and language barriers. On most surveys, refusals are the biggest reason for non-response, followed by non-contact, with other non-interview being less common (DeMaio, 1980; Durrant and Steele, 2009; Foster, 1998; Lynn and Clarke, 2002). For example, Foster (1998) shows that for five large-scale government cross-sectional surveys in the UK, non-contact rates, as a percentage of all cases, were between 2 per cent and 7 per cent, whereas refusal rates, as a percentage of contacted cases, were between 10 per cent and 29 per cent. On longitudinal surveys, which follow the same people over time, non-location of sample members who move between waves is another major reason for non-response. Many longitudinal surveys combine contact and location when reporting on survey response, though overall non-contact rates are

generally lower than non-location rates, which in turn are lower than non-cooperation rates. Lepkowski and Couper (2002) show that for two longitudinal studies in the US non-location accounts for 4 per cent and 6 per cent of all cases, compared to non cooperation rates of 14 per cent and 19 per cent respectively, as a percentage of located cases. As the determinants of each type of non-response are different, they can therefore have different impacts on non-response bias. Hence, it is important to study each type of non-response separately.

In their classic book “Nonresponse in Household Interview Surveys”, Groves and Couper (1998) propose a conceptual framework for understanding survey participation in cross-sectional surveys. Their framework comprises two separate conceptual models for contact and co-operation. They argue that the determinants of whether or not a sampled person is contacted are the calling patterns of the interviewer, the at-home times of the sampled person and presence of physical impediments such as entry phones. The decision to co-operate with a survey request, after the sampled person has been contacted, is influenced by the social environment, survey design, the characteristics of the householder and the interviewer and, in particular, the doorstep interaction between householder and interviewer.

This model of survey participation is also applicable to non-response at the first or baseline wave of a longitudinal survey. Lepkowski and Couper (2002) extend this theoretical framework to include non-response in the second or subsequent waves of a longitudinal survey. They argue that for contact and co-operation the same broad determinants of non-response apply, but that the availability of additional information at subsequent waves to both the interviewer and the sample member alter the process somewhat. For contact, prior wave call information and information about the likely at-home time of the sample member based on their characteristics can be used to inform calling patterns. For co-operation, the sample member’s decision to take part will be influenced by their experience at the previous wave, whether the same or different interviewer approaches them as well as any feedback they may have received from the survey between waves. Lepkowski and Couper also hypothesise that the importance of the doorstep interaction between the sample member and interviewer may be diminished in second and subsequent waves of longitudinal surveys. In relation to location, they argue that survey design features and sample member’s characteristics are important determinants.

Couper and Ofstedal (2009) extend this conceptual model for location on longitudinal surveys. They argue that while propensity to move is related to sample member characteristics and societal level factors, the propensity to be located is influenced by survey design factors such as the timing between waves and the methods used to locate sample members who move (referred to as ‘tracking’), and structural factors such as the availability of population registers.

In all of these conceptual models, it should be noted that survey participation consists of multiple processes, some of which are conditional on each other. Specifically cooperation is conditional on contact, and contact is conditional on location. It is important to study each of these processes separately as different types of non-response are caused by different factors. However, it is also important to bear in mind that the conditional nature of these processes means that the empirical literature on non-response is subject to potential selection effects which can vary across different surveys making it difficult to interpret and compare findings. Additionally, much of the literature fails to distinguish between these different types of non-response.

It is also worth noting that these frameworks all take a ‘stochastic’ approach to survey participation. This approach assumes that each sampled person has a probability of being a respondent and a probability of being a non-respondent, and that these probabilities are not fixed and can be influenced by a range of factors, including survey design. So, it is not the case that sampled persons are ‘pre-determined’ to be either respondents or non-respondents.

The conceptual frameworks proposed by these authors are widely used in recent survey methodological literature on non-response, and I use them in my research.

How is non-response studied?

It is by definition difficult to study non-response, as non-respondents do not respond so we often know little about them. Commonly used approaches in the literature on cross-sectional surveys include:

- using information from sampling frames or other supplementary sources including census, register or administrative data to study sub-group variation in response patterns (e.g. Durrant and Steele, 2009),
- comparing survey estimates to other ‘gold standard’ sources such as censuses or larger surveys,

- comparing respondents who are easy to interview with those which require more time and/or effort, including those who are only interviewed following a second attempt (e.g. Curtin, Presser and Singer, 2000; Hall et al., 2013; Heerwegh, Abts and Loosveldt, 2007; Lin and Schaeffer, 1995; Peytchev, Baxter and Carley-Baxter, 2009; Stoop, 2004),
- collecting a sub-set of key information from non-respondents (e.g. Lynn, 2003),
- using interviewer observation variables to compare respondents and non-respondents (e.g. Groves and Couper, 1998; DeMaio, 1980; Durrant and Steele, 2009)
- using information collected during the survey process, often referred to as para-data, to compare respondents and non-respondents (e.g. Kreuter and Olson, 2013)
- evaluating the impact of post-survey adjustments (e.g. Vandecasteele and Debels, 2007).

Many of these approaches have been used for several decades. For example, Smith (1983) uses several of them to examine non-response to the 1980 General Social Survey in the USA. Some of these approaches are more applicable to the study of certain types of non-response than others. For example, collecting a sub-set of information from non-respondents can only be done when contact has been made whereas interviewers in face-to-face surveys are able to record observations for all sampled units regardless of whether contact is made.

Groves (2006) evaluates a wide range of these approaches and concludes that although each of the techniques has some value, all of them also have limitations. For this reason, he advises researchers to study non-response bias using multiple methods simultaneously where possible.

As mentioned earlier, studying non-response bias on the second or subsequent waves of a longitudinal survey is much easier than on a cross-sectional survey or on the baseline wave of a longitudinal survey as information on both respondents and non-respondents is typically available from the previous wave. For this reason, much of the research on non-response in longitudinal surveys focuses on the impact of non-response on bias relative to either the baseline wave or most recent wave (e.g. Aughinbaugh, 2004; Gray et al., 1996; Hawkes and Plewis, 2006).

In relation to survey interventions to reduce non-response, rather than to evaluate non-response bias, much of the research in this area is based on observational studies. These kinds of studies often examine response rates and patterns under an existing survey protocol, or look at the impact on non-response of a change to a survey protocol. Although observational studies can provide interesting insights, they are hampered by the absence of robust information about the counterfactual i.e. what the response rate would have been if the protocol change had not been implemented or if a different survey protocol had been followed. This problem can be addressed, for example by making comparisons with response rates at previous rounds of the survey, but this approach can never completely rule out alternative explanations.

The most methodologically robust approach to evaluating interventions in survey practice involves using experimental approaches which randomise sampled persons to different interventions or treatments (Shadish and Cook, 2009). If the experiment is randomised, well-designed and implemented correctly, these approaches enable the impact of the intervention, or ‘treatment’, on non-response to be robustly evaluated with reference to a counter-factual or ‘control’ condition. However, in survey practice, often protocol changes are implemented non-experimentally, perhaps because survey practitioners feel there is sufficient evidence from the research literature to support the adoption of a particular approach. This is often a reasonable thing to do, but the impact of survey protocols can vary depending on the context, so it is not necessarily the case that if something has been effective on a particular survey it will work on all surveys. Another reason for not implementing interventions experimentally is because experimentation may be viewed as not possible or not practicable. Very few surveys have ‘test-beds’ on which they can try out different interventions before rolling them out, and although pilots can be used to test different survey protocols, they tend not to be large enough to produce statistically robust conclusions, and this is often not their primary purpose. Experiments can be embedded in the main stage fieldwork of surveys, but survey practitioners can reasonably be reluctant to incorporate untested interventions in this way in case they have a detrimental impact on response rates. For this reason, where experiments are embedded in main stage fieldwork on surveys, they often tend to involve relatively modest or low risk interventions.

Other considerations to bear in mind when carrying out randomised experiments on surveys include partial implementation, selection, effect sizes and generalizability. Partial implementation occurs when the experimental treatment is implemented for

some but not all of the cases which should have received the treatment or when the treatment is implemented incorrectly or incompletely for some or all cases. In particular, problems can arise where the implementation cannot be controlled centrally or in an automated way, for example where survey, field staff or interviewers are asked to follow different protocols for different sampled persons. Selection into the experimental treatment can occur when sample members are able to decide whether or not to take-up or opt-in to a particular treatment. In this situation, it is not possible to ensure that all sample members receive the treatment they have been randomly allocated to. These are some of the well-known hazards of carrying out randomised experiments on surveys. There are, however, statistical approaches such as intention-to-treat which can help address this problem. This involves estimating the effect of the intervention for all cases who were allocated to the treatment group, rather than only those who actually received the treatment. However, given that survey practitioners are often not able to guarantee that sample members actually receive a particular intervention, either because they do not take it up or because it is not implemented correctly, and it is often not possible to know in advance which sample members will receive it, there is still value in estimating the overall impact of the treatment even if only a relatively small sub-set of those assigned actually receive the treatment.

Another issue which can arise is lack of statistical power, particularly to detect small effect sizes. Often the types of interventions which are tested in experiments on surveys involve relatively small changes to existing protocols, which are not expected to have big effects, rather than major revisions. Generalisability of results to other surveys, or external validity, is also an inherent problem. Many studies in the literature are based on single survey case studies, and it is often not clear if and under what conditions findings can be generalised to other surveys with different populations and protocols.

What are the recent trends in non-response research?

Recent trends in non-response research have largely been driven by downward trends in response rates and technological developments.

There is lots of evidence that response rates to social surveys in most developed countries declined over the last 30 years of the 20th Century, and this has been a concern since at least the 1970s (e.g. Steeh, 1981). In one of the most comprehensive of the studies in this area, De Leeuw and De Heer (2002) examine response rate trends for 16 countries in Western Europe, North America and Australia across 10 different

government surveys. They conclude that both non-contact and refusal are increasing over time, but that while there are no country differences in the rate of increase in non-contact rates, there are striking country differences in the rate of increase in refusal rates. The response of survey practitioners to this trend was initially to devote additional fieldwork resources to trying to slow the decline in response rates, thereby increasing the costs of survey research. However, more recently, there has been an increasing focus in the survey methodological literature on minimising non-response bias, rather than non-response rates or levels.

Influential in this shift is a study by Groves (2006) which examines the theoretical basis and reviews the empirical literature about the relationship between non-response rates and non-response bias, including several studies about the impact on bias of survey design interventions intended to boost response. He demonstrates, theoretically and empirically, that higher response rates do not necessarily imply lower levels of non-response bias, as this depends on the correlation between the predictors of survey participation and the substantive survey variables. In a subsequent meta-analysis (Groves and Peytcheva, 2008), he further shows that the level of bias can differ between surveys with similar response rates and between variables of interest within the same survey. One of his main recommendations is that ‘blind pursuit of high response rates in probability samples is unwise; informed pursuit of high response rates is wise’. He also recommends the collection of auxiliary variables on respondents and non-respondents which may be useful for statistical adjustment purposes.

These recommendations relate directly to two other major recent trends in non-response research and practice: the use of para-data and responsive or adaptive designs. There is increasing interest in using para-data, which is data generated during the survey process e.g. interviewer call record data, to improve surveys. In relation to non-response, para-data is a major source of auxiliary information which can be captured for both respondents and non-respondents and can therefore be used for non-response error investigation and adjustment (Kreuter and Olson, 2013). Para-data, particularly interviewer call record data, is also a good source of information about fieldwork effort and costs. As such, it is often used in responsive or adaptive designs which involve making adjustments to survey design protocols during data collection in order to improve cost efficiency and to improve the accuracy of survey estimates (Groves and Heeringa, 2006; Couper and Wagner, 2011) and for case prioritisation (e.g. Peytchev et

al., 2010). Related to this, there is increasing interest in monitoring non-response error, and in particular bias, during data collection, and in the use of indicators for evaluating representivity in real-time, such as the 'R' indicator proposed by Schouten et al. (2011). Both of these trends have been facilitated by technological developments making electronic, real-time collection of para-data possible, which in turn has meant that real-time monitoring of non-response bias can be carried out more easily. Technological innovation has also led to a growing interest in the use of the web for data collection, and in the potential for mixed-mode surveys to achieve higher response rates than uni-mode surveys, and to collect data more cost-effectively (e.g. Dillman et al., 2009).

Recent non-response research on longitudinal surveys has also been influenced by these broader trends towards a focus on non-response bias, responsive or adaptive designs, para-data and the use of mixed mode data collection. There has also been a recent growth in research around locating movers in longitudinal surveys (e.g. Fumagalli, Laurie and Lynn, 2013; McGonagle et al, 2011, 2013, which has hitherto been relatively under-researched, including using new technologies such as social media for tracking sample members who move (e.g. Masson et al., 2011). However, of course, in the context of a longitudinal survey, both information on respondent characteristics and para-data is usually available from previous waves. For these reasons, longitudinal surveys are better able to estimate in advance of data collection which cases will require more field effort, and which cases are most important to retain in the sample in order to maintain representivity. This information can therefore be used to target fieldwork interventions to maximise response, and to use fieldwork resources, more effectively. However, there are relatively few examples in the survey literature of longitudinal surveys which have used prior wave information to target response inducement strategies (see Peytchev et al. 2010 for an exception). Lynn (2015) suggests that longitudinal surveys should make greater use of prior wave data to target fieldwork interventions aimed at reducing non-response. He also argues that non-response research should examine the differential impact of fieldwork interventions on sample sub-groups as well as looking at the average impact on the total sample, and that this presents an opportunity to achieve stronger positive effects for these sub-groups of interest as well as helping to explain why interventions that have worked on some surveys are not always effective on others. This approach is supported by the leverage-salience theory of survey participation which argues that the same intervention can have different impacts on different people (Groves, Singer and Corning, 2000).

How does my research fit in?

I am submitting four papers for my PhD by publication. Collectively they address, from a prevention perspective, all three of the major components of non-response in longitudinal surveys. Paper 1 (Calderwood et al., 2010) is about minimising non-response due to refusals, Papers 2 and 3 (Calderwood, 2014; Calderwood, 2013) are about non-response due to failure to locate and Paper 4 (Brown and Calderwood, 2014) is about non-contact. Papers 1, 2 and 4 use randomised experiments to evaluate fieldwork interventions designed to reduce the non-response rate, and Papers 1, 2 and 3 use prior wave data to evaluate the impact of fieldwork protocols designed to minimise the non-response rate on bias for key survey estimates. As discussed earlier, randomised experiments are widely considered to be the most robust source of evidence about the efficacy of fieldwork interventions, and using these randomised experiments to evaluate the impact of interventions on non-response bias as well as on non-response rates is in line with recent trends in the literature.

My papers are based on two major longitudinal surveys in the UK. Papers 1, 2 and 3 are based on the Millennium Cohort Study (MCS), a large-scale birth cohort study in the UK, and Paper 4 uses data from the Innovation Panel of *Understanding Society*, a major household panel study in the UK. Both studies are large-scale, multi-disciplinary, high-quality surveys designed as research resources for the user community and core funded by the Economic and Social Research Council. They are each led by academic teams, at the Centre for Longitudinal Studies at the UCL Institute of Education and the Institute for Social and Economic Research at the University of Essex respectively, with data collection competitively tendered, on a periodic basis, to external survey research agencies. Both use primarily face-to-face data collection. The MCS is a birth cohort study and *Understanding Society* is a household panel survey. These are each particular types of longitudinal survey. Birth cohort studies typically follow a sample of people born at a similar time, such as in the same week or over the same year. The timing of data collection waves is not usually fixed, rather it corresponds to particular scientifically interesting ages or developmental stages, which may be several years apart, and the nature of the data collected at each wave also changes to reflect this. Their intention is therefore to provide rich data about a particular generation at specific ages rather than representing the whole of the population. By contrast, household panel surveys aim to represent the whole of the population at each wave by interviewing all household members in a sample of all households in the country, and by including new

entrants and following those who leave the original household they are designed to maintain population representation over time. The interval between waves tends to be fixed and relatively frequent, typically one year, and the questionnaire content tends to remain broadly the same over time. Their intention is therefore to provide consistent data about the whole population at frequent intervals. A full discussion of the types, strengths and weaknesses and design features of longitudinal surveys can be found in Lynn (2009a).

The MCS is following over 19,000 children born in 2000/1 and is one of four national birth cohort studies in the UK. The sample was recruited through Child Benefit records and is disproportionately stratified and clustered at the level of electoral ward. The data collection for the study takes place in the home and involves face-to-face interviews with up to two co-resident parents and the cohort member. There have been five waves of the study so far: at 9 months (2001-2), age 3 (2003-4), age 5 (2006), age 7 (2008) and age 11 (2012). Cohort members are followed as long as they are alive and living in the UK. The nature and extent of data collection from the cohort member themselves has changed over time as they have got older, and has included cognitive assessments and physical measurements since age 3 and questionnaires since age 7. The length of the home visit has therefore increased over time, and at the age 11 wave it was around 2-2½ hours. Further information on the study design and sample can be found in Plewis (2007).

Understanding Society, also known as the UK Household Longitudinal Survey, began in 2009 and involves annual face-to-face interviews with all individuals, aged 16 and over, in around 40,000 households in the UK. Sample members are followed if they move within the UK including when households split. New household entrants are included in the study. Data is also collected using paper self-completion questionnaires from 10-15 year olds. It incorporates the sample from its predecessor study, the British Household Panel Survey (BHPS), which began in 1991. The Innovation Panel involves an additional sample of around 1,500 households and each wave takes place a year ahead of data collection for the main study. It is a methodological research resource which can be used by both internal and external researchers to develop and evaluate methodologies for longitudinal survey data collection. Interviews, for both the main study and the Innovation Panel, typically take around 30 minutes per person. As all adult household members are interviewed the length of the home visits depends primarily on the household size. Further information on the main study design and

sample design can be found in Buck (2008) and Lynn (2009b), and for the innovation panel in Burton et al. (2008).

The particular nature of the surveys on which my research is based should be borne in mind when considering the generalisability of my results to other longitudinal surveys. In particular, both of these are face-to-face surveys. Data collection mode has a major influence on non-response rates, with face-to-face generally achieving higher response rates than other data collection modes (e.g. Hox and De Leuw, 1994). I discuss the issue of generalisability in relation to each of my papers in later sections.

Overall, these four papers represent a coherent programme of research in the important area of the prevention or reduction of non-response in longitudinal surveys. In the following sections, I provide a summary of each of these papers within the broader context of the survey methodological literature on co-operation, location and contact respectively. I also explain how my research findings make an original contribution to the literature, add value to the existing knowledge in this area and have potential impact on survey practice. The literature reviews in each of the sections focus on the most relevant literature to my papers, and are intended to give a relatively broad summary of the existing evidence in order to position the research reported in my papers in context. As mentioned earlier, much of the literature does not distinguish clearly between different sources of non-response, and some fieldwork interventions can have an impact on more than one source of non-response. There is therefore inevitably some overlap and cross-referencing between sections. This issue is also discussed directly as part of the literature reviews where applicable. Additionally, as my papers are all based on face-to-face surveys, the reviews generally focus on the literature about face-to-face surveys rather than attempting to cover all data collection modes. Finally, as most of the literature on fieldwork interventions to minimise non-response tends to focus on non-response rates rather than on non-response bias, the reviews also focus primarily on the evidence relating to minimising non-response rates rather than non-response bias.

Section 2: Co-operation

This section is about non-response due to failure to co-operate with the request to participate in a survey. Co-operation is conditional on both successful location and successful contact, which are discussed in subsequent sections. On most surveys, both cross-sectional and longitudinal, refusals are the biggest reason for non-response (DeMaio, 1980; Durrant and Steele, 2009; Lepkowski and Couper, 2002) and therefore preventing refusal has received perhaps the most attention in the survey research literature. Additionally, much of the literature on overall survey non-response is implicitly focused on co-operation as the major source of non-response, although this is often not explicit and it is often not examined separately from other sources of non-response. Moreover, the term ‘participation’ is often used interchangeably to refer to both overall response and co-operation. For these reasons, this section covers some of the more general literature on survey non-response as well as the literature on survey co-operation specifically. For clarity, I have endeavoured to use the term ‘participation’ to refer to overall response and ‘co-operation’ to refer specifically to co-operation.

This section gives a brief review of the main factors affecting co-operation in surveys and examines the survey design features related to co-operation. It then provides a review of the literature in relation to refusal conversion, which involves re-issuing refusals to a different interviewer, and a summary of Paper 1 (Calderwood et al., 2010), which is about the immediate and longer-term impact of refusal conversion on longitudinal surveys.

What are the main factors affecting co-operation in surveys?

Groves and Couper (1998) argue that the decision to co-operate with a survey request is influenced by the social environment, survey design, the characteristics of the householder and the interviewer and, in particular, the interaction between householder and interviewer. This conceptual framework is applicable to cross-sectional surveys and to non-response at the first or baseline wave of a longitudinal survey. Lepkowski and Couper (2002) argue that the availability of additional information at subsequent waves of a longitudinal survey mean that the householder’s decision to take part will be influenced by their experience at the previous wave, whether the same or different interviewer approaches them as well as any feedback they may have received from the survey between waves. They also hypothesise that the importance of the initial interaction may be diminished in subsequent waves of longitudinal surveys.

Social-environmental factors can influence survey participation at both a societal-level and at a local-area level. As neighbourhood characteristics can be relatively easily obtained for both survey respondents and non-respondents through interviewer observations, and as aggregate geographical data can be relatively easily linked to sample frame information, much of the literature in this area has focused on the impact of neighbourhood and area-effects on participation. Groves and Couper (1998) had a particular focus on urbanicity, and found that, controlling for householder characteristics, higher population density was associated with lower response rates. More recent studies have found that, controlling for householder and interviewer characteristics, there is little or no impact of local-area characteristics on survey response (e.g. Durrant and Steele, 2009; O’Muircheartaigh and Campanelli, 1999). Societal-level factors, such as the perceived legitimacy of surveys, trust in important survey sponsors such as government and the scarcity or otherwise of surveys, can also influence survey co-operation. These factors can vary cross-nationally and over-time, and have been used to explain cross-national differences and changes over time in survey response rates. One of the hypotheses surrounding the much-publicised downward trends in response rates over time (De Leeuw and De Heer, 2002) is that societal changes, such as increasing atomisation and societal fragmentation plus reducing amount of discretionary time available, have negatively affected the survey-taking climate in many countries. These societal changes are likely to affect both contact and co-operation. However, the survey-taking climate is a difficult concept to measure, and as a result there has been relatively little empirical literature in this area.

The impact of *householder characteristics* on survey participation is difficult to study in the context of a cross-sectional survey, as by definition the non-respondents are not interviewed so usually very little information is available about them. Several cross-sectional surveys have asked interviewers to make judgements about the likely demographic characteristics of non-responding households based on observations they have made, and these para-data have been shown to be relatively useful for non-response adjustment (Kreuter and Olson, 2013). There are also examples of studies which have been able to obtain information on individual and householder characteristics for all sampled units in cross-sectional surveys through linking to census data (e.g. Groves and Couper, 1998; Durrant and Steele, 2009) and those which have obtained key information on a sub-set of non-respondents (e.g. Lynn, 2003).

In contrast, for second or subsequent waves of longitudinal surveys, extensive information about householder characteristics is available from the baseline wave. As a result, the literature on non-response in longitudinal studies is very dominated by studies which use householder characteristics, typically measured at baseline, to predict non-response at second and subsequent waves of the survey (e.g. Aughinbaugh, 2004; Dex and Rosenberg, 2008; Gray et al., 1996; Hawkes and Plewis, 2006; Watson, 2003). Statistical modelling of this kind is used for non-response adjustment, and it also provides survey practitioners with valuable information about which groups have experienced differential drop-out which they can potentially use to target fieldwork interventions to prevent non-response. Watson and Wooden (2009) provide a review of the empirical literature on the role of respondent characteristics in predicting attrition in longitudinal surveys. They cover all causes of attrition, not just co-operation. Overall they find a reasonable degree of consensus that men, single persons, lower educated people and those in rented accommodation are more likely to drop out of longitudinal surveys than women, married people, higher educated people and those who own their own homes. However, for age, ethnicity, household size and composition, income and employment status, they find no consistent pattern in the literature. They also note that findings often differ between surveys, and these associations can be difficult to interpret, particularly as many papers do not distinguish between contact, location and co-operation. More generally, even if we had clear and consistent evidence about which householder characteristics are associated with failure to co-operate this will not usually tell us *why* particular groups are less likely to take part. For this reason, evidence of this type is of limited utility for survey practitioners wishing to design fieldwork interventions to prevent these groups from dropping out at future waves. As Groves and Couper (1998) argue, householder characteristics are unlikely to be causally related to failure to co-operate but rather they are likely to be associated with particular psychological predispositions which affect the decision to take part, and that understanding these and how to influence them is crucial for encouraging participation. They are also likely to be related to different patterns of residential mobility and time spent at home, which are in turn related to both location and contact as well as co-operation.

Groves, Cialdini and Couper (1992) review a number of *psychological factors* which are likely to be related to survey co-operation; drivers of compliance with requests (reciprocity, authority, consistency, scarcity, social validation and liking), helping

tendencies and opinion change, and their applicability to survey research. Another influential and important attempt at theorising the psychology of survey co-operation is leverage-salience theory (Groves, Singer and Corning, 2000). This theory states that there are a range of different survey attributes which are likely to influence an individual's decision to take part in a survey, and that the level of importance – or leverage - given to any given survey attribute can vary between individuals, and by implication within individuals over time. However, the extent to which these attributes influence the decision to co-operate with the survey request depends not only on the importance attached to them by the sample member but also the salience – or weight – given to the attribute in the survey request. So, there is no 'one size fits all' approach to encouraging co-operation. Instead, the trick is to find out which aspects of the survey request are likely to have most leverage for a particular individual at a given time and to try to increase the influence of these attributes on the decision to co-operate by increasing the salience of these attributes in the survey request. In the longitudinal context, this theory implies that where prior wave information is available about the importance given by sample members to different survey attributes, the survey request can be tailored to increase the salience of the attributes that particular sample members, or groups of sample members, attach most importance to. In many surveys, this will often be down to the interviewer's skills on the doorstep or over the phone, in tailoring their responses in order to increase the salience of survey attributes that the sampled person views as most important. There are also a number of broader survey design implications which follow from this theory about which aspects of the survey to emphasise or de-emphasise in the survey request.

Moving on to *interviewers*, Groves and Couper (1998) argue that it is difficult to interpret the impact of interviewer characteristics on survey participation as these are likely to interact with householder characteristics and other aspects of the survey, such as its topic, and they are also likely to be related to other factors such as their attitudes and behaviours. A further difficulty of studying the impact of interviewers, particularly on face-to-face surveys, is that it can be difficult to disentangle genuine interviewer effects from selection effects, which can occur e.g. if more experienced interviewers are allocated to more difficult cases and area effects, i.e. the fact that the difficulty of the cases varies by area and interviewers are usually allocated to cases based on geographical proximity to their home address in order to minimise travel costs. However, there are a number of studies of interviewers which have succeeded in

disentangling these effects. Overall, they show that more experienced interviewers tend to be able to secure higher co-operation rates, interviewers who have positive attitudes and expectations secure higher co-operation rates and certain personality types are also positively associated with gaining co-operation (Durrant et al., 2010; Jackle et al., 2013).

In relation to longitudinal surveys, there is a strand of research about interviewer continuity, specifically whether or not having the same interviewer at second or subsequent waves is beneficial for participation. A number of longitudinal surveys find that response rates are higher when the same interviewer goes back to the sample member, though there is also evidence that where the sample member did not participate in the previous wave it can be beneficial for a different interviewer to approach them at the next wave (Watson and Wooden, 2014). However, this association between interviewer continuity and higher participation rates is not necessarily causal, and it is important to take account of non-random drop-out among interviewers e.g. weaker interviewers tend not to be asked to work on the survey again and/or may leave the company. There are relatively few methodologically robust studies in this area, and those that are show that interviewer continuity is either not associated at all with higher response rates (Campanelli and O’Muircheartaigh, 1999, 2002) or only associated with higher response rates in very specific circumstances (Lynn, Kaminska and Goldstein, 2014).

In relation to *householder-interviewer interaction*, Groves and Couper (1998) argue that keeping the initial conversation going, trying to ascertain the respondent’s specific concerns and tailoring responses to them are key interviewer behaviours related to securing participation. In addition to Groves and Couper (1998), there are a few other major studies which have recorded and examined respondent-interviewer doorstep interaction on face-to-face surveys (Morton-Williams, 1993; Campanelli, Sturgis and Purdon, 1997). There are also some recent studies using call-record data about the interaction between interviewers and respondents to examine the impact of this on non-response (e.g. Durrant and D’Arrigo, 2014). In the context of the second or subsequent waves of longitudinal studies, Lepkowski and Couper (2002) hypothesise that this interaction is likely to be less important in securing participation as respondents know what to expect from their prior participation. This seems intuitively plausible, though, to my knowledge, this has received little attention in the survey research literature, and I

am unaware of any studies which have recorded interviewer-respondent initial interactions on major longitudinal surveys.

How can survey design minimise non-response due to failure to co-operate?

There are a number of survey design features which are related to non-response due to refusal. The initial core design choices i.e. topic, interview length and data collection mode, all have implications for survey co-operation. The survey sponsor, and the data collection agency, may also influence co-operation. In addition, there are a number of survey processes which will influence co-operation such as whether to send sample members advance notification of the survey request and the form and content of this request, whether to use incentives and the type and value of the incentive and the nature of any follow-up procedures used such as re-issuing cases to different interviewers or different modes, and the use of persuasion letters. Paper 1 (Calderwood et al., 2010) is about refusal conversion, which is a particular type of follow-up procedure involving re-issuing cases, usually to different interviewers. In longitudinal surveys, additional factors such as the feedback given between waves may also influence survey co-operation at subsequent waves.

Going back to the initial design choices, the topic of the survey and how salient it is to sample members is strongly related to survey co-operation. In general, this can be observed by looking at response rate to different kinds of surveys; for example, health surveys tend to have higher response rates than financial surveys (Mindell et al., 2012; ONS, 2013; ONS, 2015). Saliency is not just about topic interest though, it is also about how important and relevant sample members feel the topic is to them. There is clear empirical evidence in the literature that sample members with interest in the survey topic are more likely to co-operate than those who are not (e.g. Groves, Presser and Dipko, 2004) and that more sensitive topics are associated with lower willingness to participate (e.g. Couper et al., 2010). Clearly there is a concern that the survey topic may lead to differential co-operation related to the characteristics associated with the survey measures e.g. if healthy people are more likely to take part in health surveys than non-healthy people this will mean that the survey results are likely to be biased. There are various approaches survey designers can take to mitigate the potentially biasing effects of topic interest, including using other methods e.g. financial incentives to encourage response from sample members who are not interested in the topic (Groves, Singer and Corning, 2000) and de-emphasising the topic in the survey request (Couper et al., 2010; Groves, Presser and Dipko, 2004; Groves et al., 2006).

Interview length is another key design choice which is likely to influence survey co-operation. The number of questions in the data collection instrument is one of the main determinants of interview length. However, the speed of both the interviewer and respondent are also important factors influencing interview length. Additionally, for heavily routed data collection instruments, there can be considerable between-person variation in the number of questions and hence the interview length. Overall, there is evidence that longer interviews are associated with lower response rates on many different types of survey, including face-to-face surveys (e.g. Groves et al., 1999). However, there is relatively little research about the interaction between interview length, respondent burden and topic interest, and how they relate to participation. It is likely that the impact of interview length interacts with both the survey topic and the difficulty of the questions, as well as respondent characteristics. In relation to longitudinal surveys, there is little evidence about the impact of interview length on attrition. An exception is a recent paper by Lynn (2014) which demonstrates that in an on-going household panel survey in the UK, sample members with longer interviews at the previous wave are not more likely to drop out of the panel at the subsequent wave than those with shorter interviews.

Interview mode is also associated with survey co-operation rates. There are a number of different factors associated with this, but in general, interviewer administered modes (face-to-face and telephone) have better response rates than remote modes (web and postal) (Hox and De Leuw, 1994; Dilman et al., 2009). This is largely because, if interviewers are able to contact sample members, they can use their skills to persuade sample members to take part. In general, this is more effective in person rather than over the phone, which is why face-to-face surveys tend to have higher response rates than telephone surveys. However, mode differences in survey response rates are also driven by mode differences in the effectiveness of locating and contacting methods. The survey sponsor and/or data collection agency can also influence survey co-operation, in part as this can help to lend authority, legitimacy and convey the importance of the research. For this reason, this information is almost always included in participant materials, such as advance letters. Groves and Couper (1998) found that in general surveys sponsored by government departments or agencies tended to have higher co-operation rates than those sponsored by academic organisations, and that academic surveys have higher co-operation than commercial surveys, though they note that different types of sponsor are likely to impact differentially on different types of people.

There is lots of evidence that advance notification of the survey request is positively associated with co-operation (e.g. De Leeuw et al., 2007), though there is less evidence about the content of advance materials, and how this relates to survey participation. There is some discussion about the content of advance letters in the De Leeuw et al. (2007) review, and they note that letters which appeal to reciprocity are associated with higher participation rates.

There is a recent trend in many surveys to use info-graphics and modern, attractive designs, and some emerging evidence that this can improve survey participation rates (Taylor, 2013). In the context of longitudinal studies, the information sent to participants between waves may also influence their co-operation at subsequent waves. It is well-established that between-wave feedback mailings are positively related to survey participation at future waves (e.g. Couper and Ofstedal, 2009), though this is in part due to the fact that these mailings help with contact and location at the next wave as well as promoting co-operation. Paper 2 (Calderwood, 2014) is about between-wave mailings and I discuss the literature in this area further in section 3 on location.

Many surveys offer incentives to sample members to encourage them to take part. There is an extensive literature on respondent incentives, which broadly shows that they are effective at boosting response, unconditional incentives are more effective than conditional incentives, cash incentives are more effective than vouchers/payments in kind and that higher value incentives are more effective than lower value incentives (Church, 1993; Singer et al., 1999; Singer, 2002). Similar results are found for longitudinal surveys. Jackle and Lynn (2008) find that unconditional incentives had a larger long-term effect on attrition than conditional incentives, and that different levels of incentives at previous waves did not seem to have an impact on their effectiveness at later waves. Martin, Abreu and Winters (2001) find that paying higher incentives to prior wave non-respondents boosts their participation at later waves and Rodgers (2002) finds that higher value incentives can reduce attrition at subsequent waves. In their review of the literature in this area, Laurie and Lynn (2009), discuss the particular issues regarding the use of incentives on longitudinal surveys including the cumulative effect of incentives on long-term participation and the impact of changing incentive amount or type over time.

Incentivising interviewers may also lead to higher response rates, though there is very little evidence in this area, for either cross-sectional or longitudinal surveys. In one of

the few studies about this, Peytchev et al. (2010) find that this is not an effective method for boosting response.

Even after participants refuse to take part in a survey, this is not necessarily the end of the story. Various follow-up procedures can be used, including re-allocation to different modes or different interviewers. Re-allocating non-respondents in one mode to a different mode can be an effective way of securing participation. This is the approach taken on sequential mixed mode surveys which initially invite participation in a non-interviewer administered mode e.g. web or postal, and then re-allocate non-respondents to an interviewer administered mode e.g. telephone and/or face-to-face. Persuasion letters are also sometimes used (e.g. Olson, Lepkowski and Garabrant, 2011). Re-allocating sample members who do not co-operate to different interviewers, referred to as refusal conversion, is another commonly used survey process to minimise non-response due to refusals. Paper 1 (Calderwood et al., 2010) contains a more extensive review of the literature on refusal conversion and a summary of the literature in this area and this paper follows in the next part of this section.

How effective is refusal conversion at minimising non-response due to failure to co-operate?

Refusal conversion is a well-established fieldwork strategy for reducing refusal rates (see e.g. Stoop, 2004). It involves re-approaching a sampled person who has initially refused to take part and attempting to get them to reconsider their decision, i.e. to 'convert' them from a refusal to an interview. A successful outcome of a refusal conversion attempt is achieving an interview. The proportion of re-issued refusals successfully converted to an interview is usually referred to as the conditional conversion rate and the proportion of all refusals successfully converted to an interview is usually referred to as the unconditional conversion rate. Refusal conversion is usually, but does not necessarily have to be, attempted by a different and often a more experienced interviewer. For this reason it is often called refusal re-issuing, as the case is re-issued to another interviewer.

The rationale for refusal conversion is that a sample member's decision to co-operate is influenced by their interaction with the interviewer and sending a different interviewer will lead to a different interaction and, hopefully, a different (positive) decision about participating. The approach is also based on the evidence that refusals often result from the particular circumstances of the respondent at the time the interviewer called and that

(the same or a different interviewer) calling back on a different occasion may mean that the circumstances which led to the refusal are no longer pertinent. More broadly, this fits with the 'stochastic' view of survey response which implies that respondents are not generally fixed in their attitude towards participation. As discussed earlier, experienced interviewers tend to have lower refusal rates, and it is for this reason that more experienced interviewers are often used for refusal conversion. Allowing a longer elapsed time between the initial refusal and the conversion attempt can also lead to improved conversion rates (Beullens, Billiet and Loosveldt, 2010). In the context of telephone surveys, there is evidence that interviewers who are able to sustain longer interaction with the sample members are more effective at refusal conversion (Fuse and Xie, 2007). In the context of declining response rates, refusal conversion is becoming more common (Lynn et al., 2002; Curtin et al., 2000). However, although this technique may be effective at boosting response rates, there is also evidence that it may be doing little to reduce non-response bias (Curtin et al., 2000; Keeter et. al. 2000; Lynn et al. 2002).

In longitudinal surveys, refusal conversion can take place both within and between waves of data collection i.e. refusals can be re-approached at subsequent waves as well as (or instead of) during the current wave of data collection. In addition, although maximising response on any given wave of a longitudinal survey is important, it is also crucial to consider the longer-term impact of refusal conversion on both attrition and response bias. Burton, Laurie and Lynn (2006) evaluate the long-term effectiveness of within-wave refusal conversion procedures on the British Household Panel Survey (BHPS). They show that the majority of converted refusals went on to participate in future waves. The paper also shows that refusal conversion on the BHPS improved the representation in the sample of certain groups (such as the geographically mobile, self-employed and local authority renters) and, to the extent that these variables are correlated with other variables of interest, may be expected to reduce non-response bias. Other than this paper, there appears to be no published evidence about the effectiveness of refusal conversion procedures on longitudinal surveys.

Moreover, almost all of the literature on refusal conversion, including Burton et al. (2006), is based on observational studies rather than randomised experimental interventions. This is problematic because estimates of the effectiveness of refusal conversion attempts, including their impact on non-response bias, will be influenced by selection effects caused by non-random choice of cases to allocate for re-issuing by

field staff. There are a few studies which experimentally evaluate the effectiveness of different refusal conversion techniques, though these are all based on cross-sectional telephone surveys (Basson and Chronister, 2006; Keeter et al., 2000).

Against this background, Paper 1 (Calderwood et al., 2010) provides evidence on the immediate and longer-term results of an experimental fieldwork intervention on refusal re-issuing carried out on the fourth wave of the UK Millennium Cohort Study in 2008. The randomised experiment was designed to increase the proportion of refusals successfully converted to an interview at the re-issue stage, and consisted of two interventions which were tested in a crossed design to enable the interventions to be evaluated independently and in combination with each other. The first intervention was to re-issue all refusals rather than a non-random sub-set chosen by field staff, which is the standard re-issuing strategy of the fieldwork agency, and the second intervention involved sending a leaflet addressing common reasons for refusals as part of the re-issue process.

The main findings were that in the intensive re-issuing experimental groups almost a quarter of re-issued refusals were successfully converted to a productive interview at wave four and that around 6 in 10 of the converted refusals in these groups went on to be interviewed again at wave five in 2012 (compared with around 1 in 4 unconverted refusals). In relation to the leaflet treatment, we found that the leaflet did not make a difference to the conditional or unconditional conversion rate in the intensive re-issuing groups. However, we found some indicative evidence that converted refusals who received the leaflet at wave four were more likely to take part at wave five than converted refusals who did not receive the leaflet at wave four, though the difference was not statistically significant.

We also examined the impact of refusal conversion at wave four on response bias, using wave three data as our comparison. Bias reduction was assessed using a series of binary, unordered and ordered logistic regression models with each of the wave three characteristics as the sole dependent variable in separate models and the three interview outcomes – initially interviewed, converted refusal and unconverted refusal – as dummy explanatory variables with initially interviewed as the reference category. Adjusted Wald tests were used to test for significant differences between initially interviewed and converted refusals and initially interviewed and unconverted refusals on each of the dependent variables. We used wave three data as a comparison in order to restrict the

analysis sample to cases which were interviewed at wave three and thereby reduce a major source of variation between cases, and to ensure a common set of reference data for all cases and avoid problems of missing data due to different prior participation histories. Additionally, most of characteristics chosen are changeable over time and this allowed us to use the most recent measurement. The wave three variables were chosen as they are related to both attrition (Plewis et al., 2008) and key substantive outcomes on MCS, and are comparable to those used by Burton et al. (2006). Additionally, they all have low levels of item missing data, which meant that this issue could be addressed by carrying out complete case analysis. We found that refusal conversion led to a statistically significant reduction in bias in the survey estimates for just two of the seven characteristics considered i.e. voting and housing tenure.

Unfortunately, due to problems with how the experiment was implemented, we were unable to compare the effectiveness of intensive re-issuing with standard re-issuing, or to compare the effectiveness of the leaflet in conjunction with both intensive and standard re-issuing.

Despite this, Paper 1 makes three important contributions to the methodological evidence on refusal conversion. Firstly, by providing evidence from an experiment involving intensive re-issuing of all refusals in a randomised sub-sample of cases, we eliminated selection effects into the refusal conversion treatment and thereby addressed a limitation in the existing research in this area, which was largely based on observational studies involving the selection of a non-random sub-sample of refusals into refusal conversion treatments. These findings show that high conversion rates can be obtained, including from groups with high initial refusal rates, and that there is therefore potential for using refusal conversion attempts in a targeted way to reduce non-response bias. Secondly, we also made a significant contribution to knowledge about the effectiveness of refusal conversion strategies in longitudinal surveys, particularly about the longer-term impact of refusal conversion, about which there is relatively little published work. These findings are likely to re-assure longitudinal survey practitioners that within-wave refusal conversion does not appear to have a detrimental effect on long-term participation, supporting the earlier findings of Burton et al. (2006) on the BHPS. Thirdly, we provided evidence about the effectiveness of using leaflets in refusal conversion attempts. The absence of an effect of the leaflet treatment on conversion success rates at wave four suggests that future research should attempt to find ways of making this treatment more effective, such as the use of content

tailored to the specific reasons for refusal given by particular cases rather than generic materials. However, the indicative evidence that this treatment may have had an effect on longer-term outcomes at wave five suggested that such leaflets may be effective at promoting on-going participation in longitudinal surveys. These findings are likely to be generalisable to other major face-to-face longitudinal surveys.

However, it is also important to consider the cost-effectiveness of refusal conversion. Reissuing is an expensive way to obtain an interview. Data from the fieldwork agency showed that the cost of achieving an interview was over three times as high for converted refusals as it was for families who did not refuse initially, though this cost was incurred for a small proportion of the sample. Paper 1 indicates that if the intensive re-issuing had been carried out on the whole sample, it would have reduced the overall refusal rate by around two percentage points and led to an increase in achieved sample size by around 278 cases. Even on MCS, a large-scale study, this is a notable impact. Given that refusal is usually the biggest reason for non-response, interventions targeting refusals are often the more effective at boosting response rates and sample sizes than interventions targeting other reasons for non-response. Having said that, the evidence indicates that refusal conversion does not have a major positive impact on non-response bias. Both Paper 1 (Calderwood et al., 2010) and Burton et al. (2006) find only modest improvements on non-response bias, on a limited number of characteristics. In general, it appears that converted refusals are more similar to sample members who are initially interviewed than unconverted refusals. So, whether or not refusal conversion is cost-effective, may depend on whether we are primarily concerned about sample size or non-response bias.

This clearly illustrates one of the tensions in survey practice and methodology between maximising response rates and minimising non-response bias. There is no right or wrong answer about which of these objectives to prioritise. The important thing is that decisions regarding fieldwork interventions are taken from a position of knowledge about their impact in both of these areas. Paper 1 illustrates how longitudinal surveys can use prior wave data to ensure that, as advocated by Groves (2006), their pursuit of high response rates is informed.

Paper 1 also illustrates some of the difficulties of implementing a randomised experiment in a survey setting. Although over three-quarters of refusals were re-issued in the intensive re-issuing groups, a significant minority of cases were still excluded,

and more generally the exclusions were non-random and varied by experimental group. Although this had the potential to jeopardise the validity of the conclusions based on the initial randomisation, we were able to implement the standard solution to this problem, which is to estimate treatment effects for all cases for which there was an intention to treat as well as to estimate the effects of treatment on the treated (Shadish and Cook, 2009) i.e. compare conversion rates for all refusals as well as treated refusals. However, as noted earlier, a further difficulty in the implementation of the experiment was that the proportion of cases re-issued in ‘non-intensive’ control groups (B and D) was extremely low; much lower than we had anticipated. This appeared to be an unintended consequence of increasing the proportion of cases re-issued in the ‘intensive’ treatment groups (A and C). Unfortunately, this necessitated the exclusion of groups B and D from the evaluation of the experiment, and meant that we were unable to evaluate the impact of the leaflet treatment in combination with the standard approach to re-issuing or to compare the effectiveness of standard re-issuing with intensive re-issuing. Although we were aware of the potential for implementation problems to arise and had taken steps to try to prevent this, it appears that they were not completely effective. This highlights the need for the implementation of experiments in surveys to be carefully monitored, particularly when they are being carried out by sub-contracted survey research agencies, and for the design of experiments to carefully consider and take steps to prevent potential unintended consequences or spill-overs.

In this section, I have reviewed the main factors affecting co-operation in surveys, provided a brief summary of the research literature about how a range of survey design features are related to co-operation and demonstrated how the findings from Paper 1 (Calderwood et al., 2010) have contributed to knowledge in this area.

In the light of my research in this area, survey practitioners should consider carefully the utility of refusal conversion, in particular whether relatively modest improvements in response rates and response bias are a worthwhile return for the considerable investment of resources that this requires. It may be that, particularly for longitudinal surveys, resources would be better spent elsewhere, for example on tracking sample members who move in order to reduce non-location.

More generally, methodological research on the prevention of non-response due to failure to co-operate in longitudinal surveys should focus on the doorstep behaviours of

interviewers and how they differ from cross-sectional surveys, and the content and design of between-wave literature in promoting co-operation at the next wave. Broader research on co-operation in surveys should focus on interviewer incentives and non-monetary interventions to reduce non-cooperation. Greater targeting of fieldwork interventions should also be considered, particularly on longitudinal surveys.

Section 3: Location

This section is about survey non-response due to failure to locate. Although this can occur rarely on cross-sectional surveys, it is primarily an issue for longitudinal surveys which follow people over time. This requirement means that it is necessary to successfully locate sample members at each wave in order to contact them and invite them to take part. Contact and co-operation are therefore both conditional on successful location. Location is generally straightforward for those who have not moved since the last wave of the survey. Sample members who move between waves *may* be difficult to locate, though not necessarily. For example, some sample members who move may provide their new address. Location, in the context of face-to-face longitudinal surveys, refers to establishing the geographical location of sample members i.e. the address at which they are living at the time of each data collection wave. Although location tends to be the second biggest reason for non-response in longitudinal surveys, it has received limited attention in the research literature, perhaps as it is primarily an issue of concern to longitudinal survey methodologists and practitioners.

In practice, it can be quite difficult to disentangle contact and location. For example, unless contact is made with someone at an address there can be uncertainty about whether a particular sample member is living there or not. Often movers are identified using other sources of information, for example neighbours, post-office returns and administrative records. Additionally, some of the interventions used to minimise non-response due to non-location may also impact on other sources of non-response. For example, between-wave mailings which are intended to keep address details up-to-date may also impact on co-operation and contact at future waves. For these reasons, and also because non-contact often accounts for a small proportion of non-response, some of the literature in this area conflates non-location with non-contact and/or non co-operation. This section provides an overview of the problem of locating sample members who move in longitudinal surveys and reviews the existing evidence about how to minimise sample loss due to failure to locate. It also summarises Paper 2 (Calderwood, 2014) and Paper 3 (Calderwood, 2013) which both provide evidence on the effectiveness of the locating methods used on the UK Millennium Cohort Study. Paper 2 presents the results from a randomised experiment carried out on a between-wave mailing and Paper 3 examines the role of respondent characteristics in different types of locating procedures.

What is the location problem in longitudinal surveys?

Groves and Couper's (1998) theoretical framework for survey non-response, discussed in section 2, was expanded by Lepkowski and Couper (2002) to include participation in the second or later waves of a longitudinal survey. They highlighted that as longitudinal surveys follow the same sample members over time, locating them is a crucial first step in securing participation at second or later waves. The other processes involved in survey response, that is making contact and securing co-operation, are both contingent on successful location. These ideas were developed further by Couper and Ofstedal (2009) who explicitly decomposed the location problem into two distinct processes i.e. moving and being located. The overall level of non-response due to failure to locate in a longitudinal survey is the product of the proportion of sample members who move and the proportion of those who move which are not located. For this reason, location is by definition a bigger problem for those longitudinal surveys which have high mobility rates among their study population, regardless of how effective their procedures for locating sample members are.

Couper and Ofstedal (2009) argue that there are a range of factors which influence the mobility of sample members, at both an individual level and societal level. At an individual level, these include a person's age, family, housing and employment situation. At a societal level, these include general patterns of mobility and internal migration within the country, including geographical variations in labour markets, as well as broader societal issues, for example relating to the permanence of people's housing situations. They also argue that there are a range of structural features which usually operate at a societal level and influence the likelihood of locating sample members who move. These include things like the existence of population registers, the extent to which postal services facilitate the forwarding of mail and the degree to which telephone numbers are tied to specific properties. Technological changes, and in particular declining use of landline phone numbers and the increasing use of easily portable mobile phone numbers and email addresses, presents both challenges and opportunities for location. All of these things are outside the control of survey practitioners to influence. However, there are also a number of survey design factors which can influence the likelihood of locating sample members who move. These include general design features of the survey overall, as well as the nature and scope of participant follow-up and how this is done.

Overall mobility rates vary across the life-course, for different sub-populations, between and within countries and over time. For this reason, where and when the survey is taking place as well as the definition of the population can be a major influence on the level of non-response due to failure to locate for any given longitudinal survey. For example, mobility rates tend to be higher during childhood and early adulthood than during late adulthood (ONS, 2004), so we would expect non-response due to non-location to be higher for longitudinal surveys which follow families with children or young adults than those following older people. Similarly particular sub-populations, such as recent immigrants or students, are likely to have higher than average mobility rates, and because of this, locating sample members who move will be a greater challenge for longitudinal surveys following these sub-populations exclusively than those following a general population sample. There may also be some sources of information, for example administrative records, which are applicable only for particular sub-populations or geographical areas and which therefore make location easier for longitudinal surveys following those particular sub-populations or in those areas. In terms of overall survey design, the particular type of longitudinal survey may also be associated with how easy or difficult it is to locate sample members. For example, we could hypothesise that locating movers will be more difficult in longitudinal surveys of households than surveys of individuals, as locating multiple people is likely to be more difficult than locating one individual, particularly if household members have moved to different locations. Conversely, we could hypothesise that locating movers in household surveys would be easier as the existence of multiple household members may provide multiple address leads, and finding one household member may lead to finding the other household members. In both household and individual surveys, locating sample members who have moved is likely to be easier if some of the household members remain at the original address.

Another key survey design feature likely to be associated with how easy or difficult it is to locate sample members is the interval between data collection waves. By definition, longer intervals between waves means that the risk of moving for any given sample member increases. It also increases the risk that the contact information, such as phone number and email addresses, used to locate mobile sample members becomes out-of-date. We would therefore expect that, all other things being equal, longitudinal surveys with longer intervals between waves to have higher levels of non-response due to failure to locate than those with shorter intervals between waves.

Another overall survey design factor that is likely to affect levels of non-response due to failure to locate is the geographical definition of the longitudinal population.

Longitudinal surveys which have more restrictive definitions in this regard are likely to have higher location rates. For example, if a local area longitudinal survey is only interested in following up participants who remain in the same local area, then sample members known to have moved outside this local area can be treated as ineligible rather than as not located, even if a new address for them is not obtained.

Returning to the overall determinants of mobility, moving is often associated with particular changes in personal circumstances. These include family changes such as having children, relationship changes such as partnership formation or dissolution and employment changes such as changes of employer or salary changes associated with changes of job within or between employers. Given that one of the main purposes of longitudinal surveys is to measure change over time, and that many major longitudinal surveys aim to measure change in these specific domains, the problem of non-response due to failure to locate is potentially a serious one. As non-response due to failure to locate is systematically related to these family, relationship and employment changes, this is likely to mean that the survey measures of change over time in these domains, and potentially any associations with change found in the data, are biased. Attrition due to failure to locate also implies that survey measures directly associated with moving, such as mobility rates or housing change may be biased and partial. It is difficult to evaluate whether or not attrition due to failure to locate has a biasing impact on measures of change. Using prior wave data is insufficient for this purpose as it is by its nature fixed at that time. The optimal way of doing this would be to use information from another source, such as administrative data, which provides information about whether all sample members, both those who continue to respond and those who do not, have experienced particular changes of circumstance.

As well as the overall design of the survey, the specific processes or procedures adopted for preventing non-response due to failure to locate on any given longitudinal survey are likely to impact on the proportion of movers which are located. Given that longitudinal surveys are not able to influence the overall mobility rate in their population, survey methodology and practice has tended to focus on how best to locate sample members who do move. In the survey research literature, the process of attempting to locate sample members who move is usually called ‘tracking’ (as well as ‘locating’ and ‘tracing’), the ways in which this is done are referred to as tracking methods or

procedures (or ‘locating’ or ‘tracing’) and the proportion of movers who are successfully located, that is a new, correct address is found, is usually referred to as the tracking rate or tracking success rate (again the term ‘location’ or ‘tracing’ can also be used).

What do we know about the effectiveness of tracking procedures in longitudinal surveys?

Many of the examples in the survey research literature about tracking are descriptive accounts of how specific surveys have attempted to locate sample members who move and how successful these tracking methods have been overall (e.g. Eckland, 1968; McAllister, Goe and Butler, 1973; Crider, Willits, and Bealer, 1971; Burgess, 1989). Often these examples are about surveys of specific difficult-to-track sub-populations (e.g. Cotter et al., 2005 who considers a cohort of disruptive boys recruited to a clinical study) and/or surveys which have attempted to re-establish contact with a group of people many years after they were first surveyed for a follow-up which was not part of the original study protocol (e.g. Clarridge, Sheehy and Hauser, 1978; Hampson et al., 2001).

Various typologies or categorisations of tracking procedures have been proposed in the literature. The main distinction made is between prospective (or forward) tracking, which involves attempts to prevent loss of contact with sample members due to mobility or mitigate the impact of this, for example by collecting alternative contact details and retrospective tracking, which involves trying to find sample members when it is discovered that they have moved and where they have not provided their new address (Burgess, 1989; Laurie, Smith and Scott, 1999). Prospective tracking involves things like encouraging sample members to update their address if they move and facilitating this through the provision of multiple, cheap and easy ways to contact the study, collecting a full-range of contact information from sample members, including information about a ‘contact person’ or ‘stable contact’ who can be approached if the sample member moves and maintaining contact with sample members between waves, through between-wave mailings. Retrospective tracking includes things like attempting to get in touch with sample members through personal visits to last known address, letters, phone calls, emails and text messages and asking neighbours and new occupants for forwarding addresses and/or asking them to forward information on behalf of the survey. The increasing use of email addresses and mobile phone numbers, and their

portable nature, has provided additional channels through which to attempt to contact people who have moved from their residential address.

Couper and Ofstedal (2009) extend this typology to further distinguish between field tracking, which is usually done by interviewers and centralised tracking, done in the office by a central team of specialists. The key distinction here seems to be between tracking which is done remotely such as phone calls, which may be made either by interviewers or office staff, and ground tracking, which involves in-person visits to the local area, usually by interviewers. Generally field (or ground) tracking is only done on face-to-face surveys, though in principle it can also be done as a bespoke exercise on other types of survey. Centralised, or office tracking, may also involve using specialist databases, internet or social media searches and administrative records such as electoral rolls and records held by government departments or agencies. A useful distinction can also be made between tracking which must be done on an individual basis for each sample member (case-level tracking) and tracking which can be done in an automated manner for large numbers of sample members at the same time (batch-level tracking). The use of administrative records for tracking is one of the main types of batch-level tracking used, and differences in the availability of these types of records, particularly population registers, is one of the main structural factors influencing differences in tracking rates between longitudinal surveys in different countries. Field tracking and case-level office tracking are usually only done retrospectively i.e. when it has been established that a sample member has moved. However, batch-level tracking can be used either prospectively for all cases, including those who are not known movers, or retrospectively only for known movers.

These distinctions can sometimes be confusing. For example, the term ‘tracking’ is sometimes used to only refer to the situation where specific additional effort is needed to attempt to find a sample member i.e. retrospective tracking. However, prospective methods can be effective without any specific additional effort being required, for example where a sample member who moves proactively gets in contact with the study to supply a new address. So, not all movers will require retrospective tracking, and both the proportion of movers who require retrospective tracking and the effectiveness of retrospective tracking is likely to be influenced by the effectiveness of prospective tracking. Additionally, some prospective tracking methods identify sample members for whom retrospective tracking is needed, for example, through post-office returns to between-wave mailings.

Many large-scale longitudinal surveys devote considerable resources to tracking and report high levels of tracking success. For example, the German Socio-Economic Panel, the British Household Panel Survey (BHPS), the Health and Retirement Study and the Panel Study of Income Dynamics in the USA all report tracking success rates of between 94-98% (Couper and Ofstedal, 2009). However, very few studies report the proportion of cases found through different tracking methods (see Hampson et al., 2001 for an exception). In general, there is also a lack of robust experimental evidence about the relative success of different types of tracking procedures and relatively little discussion in the literature about the cost-effectiveness of tracking overall, and of specific tracking methods. Laurie, Smith and Scott (1999) discuss the cost-effectiveness of the tracking procedures used on BHPS, noting that as the proportion of cases which need to be tracked is small (less than 3% in their study), the average costs across all cases in the study is low (around 50 pence per case). We can hypothesise that prospective methods which aim to prevent the loss of contact with sample members are likely to be more cost-effective than retrospective methods used to try to find sample members who move, particularly if this retrospective tracking is done by interviewers in the field rather than remotely in the office. Methods which involve automated or batch processing for large numbers of sample members at the same time are also likely to be more cost-effective than those which involve case-by-case tracking, which can be labour intensive.

As noted earlier, technological development presents both opportunities and challenges for surveys in general, and for tracking in particular. The use of the internet and social media for tracking, in addition to email and mobile phones, is potentially an opportunity to improve the effectiveness of tracking, and as these are remote methods they are likely to be cost-effective techniques. There has been a growing literature in this area recently, though mainly based on small-scale surveys. Koo and Rohan (2000) compare the effectiveness two different web-based directories in Canada for tracking and there are also a few studies which report the effectiveness of using social media websites such as Facebook and MySpace for tracking in the US (Nwadiuko et al., 2011) and UK (Masson et al., 2011). The emergence of these new methods of tracking has led to a recent focus on ethical and privacy concerns. Concern about this issue is not new, and has been a common theme in research in this area for several decades. For example, Burgess (1989) includes an extended discussion of these issues. In general, tracking methods aim to safeguard the confidentiality of the sample members. However, this can

be difficult as it is generally necessary to reveal the identity of the person being sought when seeking help to find them. The approach usually taken is to reveal the name of the person but not to reveal the specific name of the study they are involved in. Although such approaches are possible to maintain in internet and social media tracking, arguably they restrict the ability of surveys to fully exploit these technologies for tracking. More broadly, ethical concerns about tracking usually relate to whether explicit consent is needed from participants to try to find them through particular ways and more recently the acceptability to participants of using methods such as internet searches and social media, as well as through administrative data. Privacy laws also differ between countries, and this is another one of the main structural factors influencing between-country differences in tracking rates.

Another area in which there has been a recent growth in evidence is the literature on between-wave mailings, which is the topic of my Paper 2 (Calderwood, 2014). This method is discussed in the next part of this section. The role of respondent characteristics in tracking is a particularly under-researched area. The final part of this section summarises the literature on this topic and summarises my Paper 3 (Calderwood, 2013) which examines the role of respondent characteristics in both office and field tracking.

What do we know about the effectiveness of between-wave mailings for tracking on longitudinal surveys?

Between-wave mailings are one of the most commonly used prospective tracking methods on longitudinal surveys. They involve a mailing, usually by post, to sample members between waves of data collection. As part of this mailing, sample members are asked to inform the study if they have moved address or intend to soon. Sometimes sample members are also asked to confirm that the address to which the mailing was sent is still correct. Post-office returns from these mailings also alert survey managers that retrospective tracking is needed. Between-wave mailings also often include feedback leaflets and also sometimes prompt sample members to update other contact details. As such, they can also influence co-operation and contact at future data collection waves as well as location.

There are several design considerations which apply to these mailings and some variation between surveys in their practice. However, until recently, there has been little or no methodological literature in this area to inform the choice of design. In the last

few years there has been a flurry of research about these mailings involving randomised experiments on several major longitudinal surveys, addressing this gap in the evidence. Randomised experiments have been carried out on the between-wave mailings on the Panel Study of Income Dynamics (PSID) in 2008 and 2010 (McGonagle, Couper and Schoeni, 2011; McGonagle, Schoeni and Couper, 2013), the British Household Panel Survey (BHPS) in 2008 (Fumagalli, Laurie and Lynn, 2013) and on the UK Millennium Cohort Study (MCS) in 2010 (Calderwood, 2014).

A primary design consideration for between-wave mailings is the number and spacing of these mailings and how this relates to the interval between waves of data collection. The residential moves of sample members are likely to be distributed across the time interval between waves and therefore, all other things being equal, the more frequently between-wave mailings take place the more effective they are likely to be at identifying new addresses. The timing of these between-wave mailings is also important and survey designers need to balance the need to allow sufficient time to elapse since the previous contact for moves to occur with the desire to identify moves in a timely way soon after they occur to maximise the effectiveness of retrospective tracking and allow sufficient time for this to be completed before the next wave of data collection begin. On the PSID, data collection waves take place every other year. McGonagle et al. (2011) experimented with carrying out the mailing at the mid-point between waves (control) and roughly three-quarters of the way between waves (treatment). They found no statistically significant difference between the treatment and control groups in the proportion of cards returned or in the proportion providing a new telephone number.

Another design consideration relates to whether or not to include a return card in the mailing for sample members to return and if so, what the design and content of this card should be. Related to this, a further consideration is whether all sample members are asked to return their card (address confirmation) or only those whose address or other contact details have changed (change of address). Fumagalli et al. (2013) found that sending any form of return card i.e. either a change of address card or an address confirmation card elicited a higher proportion of address changes than not sending any card for sample members to use to return their change of details. PSID use an address confirmation card and experimented with a re-design of this card. McGonagle et al. (2011) found, contrary to their expectations that a re-designed, colourful address confirmation card led to lower return rates than the 'traditional' black-and-white card. Their post-hoc explanation for this was because the newer design was more complex to

complete and return than the traditional design. Fumagalli et al. (2013) found that there was very little difference in the proportion of new addresses obtained using a change of address card compared with an address confirmation card and, as a result, concluded that as the mailing costs associated with a change of address card are much lower (as return postage is only paid if the card is returned), this was a more cost-effective approach. This conclusion was further supported by evidence that at the subsequent wave fewer interviewer field visits were needed for the cases in the change of address card group than for those who received address confirmation cards.

Where address confirmation cards are used, a further consideration is whether or not to send follow-up letters to sample members who don't respond to the initial mailing to encourage them to return the card. McGonagle et al. (2011) found that sending a reminder mailing significantly increased the proportion of forms that were returned. This is consistent with the literature on postal surveys. Dillman, Smyth and Christian (2009) recommend the use of reminder mailings for postal surveys as part of a 'tailored design method' and there is much empirical evidence that reminders are effective at boosting response rates on these surveys (see e.g. Fox et al., 1988 for a meta-analysis).

Similarly respondent incentives may be used, with both address confirmation and change of address cards, to encourage their return. The PSID and BHPS both use incentives in their between-wave mailings and McGonagle et al. (2011), McGonagle et al. (2013) and Fumagalli et al. (2013) all tested the effect of incentives in their experiments. They explored the relative effectiveness of conditional incentives, which are provided after the card has been returned, and unconditional incentives, which are sent up front to all sample members. However, their results were not consistent with each other. On PSID, there was no difference in the return rate between conditional and unconditional incentives, though they did find that fewer calls were needed in the field to finalise the case if an unconditional incentive was used. On BHPS, an unconditional incentive led to a much higher return rate than an incentive that was conditional on the return of the card. This finding is consistent with the broader literature on the use of incentives in surveys which recommends the use of unconditional incentives to boost response rates (e.g. Singer, 2002). However, as this higher return rate was not associated with a higher proportion of new addresses, overall unconditional incentives were viewed as less cost effective than conditional incentives. Both studies also experimented with different values of the incentives. On BHPS, they found that a higher value incentive led to a higher return rate only when it was offered conditional on return

of an address confirmation card. On PSID, they found that offering an incentive led to a higher return rate than offering no incentive, and that higher value incentives led to a higher return rates.

A final consideration in the design of between-wave mailings is whether to include a letter and/or a newsletter or findings leaflet and if so, what the content and design of these additional materials should be. The standard practice on PSID is to send a newsletter separately a few months before the mailing in which sample members are asked to confirm their contact details. McGonagle et al. (2011) tested this approach by sending the newsletter to a randomised half of their sample and not sending it to the remainder. They found that receiving a newsletter a few months earlier made no difference to whether or not sample members returned their address confirmation cards. BHPS includes a respondent report with their between-wave mailing. Fumagalli et al. (2013) experimentally tested tailored versions of this report for young people and busy people. They found that the tailored report for young people led to slightly higher face-to-face response rates for this group at the subsequent wave of data collection but had no impact on the overall proportion completing either a face-to-face or telephone interview (offered to those who are not willing or able to respond face-to-face). The tailored report for busy people had no effect on the response rate to the face-to-face visits but did lead to a slightly higher response rate overall due to an increase in take-up of the telephone interviews.

McGonagle et al. (2013) also experimented with not sending a between-wave mailing at all. This was a particularly interesting intervention as although there is lots of evidence that responding to a between-wave mailing is positively associated with both participation and reduced field effort at subsequent waves, this had hitherto not been tested experimentally with the counter-factual of not sending a mailing at all. They demonstrated that receiving the mailing did lead to cost savings, as these cases required less tracking in the field and fewer calls to finalise the case. Finally, McGonagle et al. (2011), McGonagle et al. (2013) and Fumagalli et al. (2013) all include evaluations of the cost-effectiveness of their between-wave mailings, and conclude that they are cost-effective.

Paper 2 (Calderwood, 2014) adds to this literature by providing further evidence about the design of between-wave mailings from a randomised experiment carried out in 2010 on the UK Millennium Cohort Study. The experiment was designed to increase the

return rate to the between-wave mailing overall, and in particular from two groups in particular who have high attrition rates in the survey: sample members from minority ethnic groups and those with lower education. The rationale for this design was that as responding to between-wave mailings is positively associated with response at subsequent waves, an approach which explicitly aimed to increase the return rate among the groups most likely to drop-out could lead to improved response rates from these groups and a reduction in overall bias.

The main objective of the experiment was to increase the proportion of forms that were returned. This is referred to as the 'return rate'. The MCS between-wave mailings take an address confirmation approach, so sample members are asked to return the form they are sent with their contact details regardless of whether there are any changes. We expected that the experimental treatment would have most effect on the proportion of forms returned in response to the initial mailing. This is referred to as the 'early return rate'.

The experimental manipulation involved the design and content of the covering letter included in the mailing. Covering letters in between-wave mailings perform a similar function to advance letters; that is they are designed to promote compliance with a request. As discussed in section 2, advance letters are standard practice on most surveys and there is strong evidence that they are associated with higher response rates (e.g. De Leeuw et al. 2007). Design principles regarding advance letters include using headed paper to demonstrate the authority of the request and explaining how sample members were chosen and the purpose of the study. However, the empirical literature on the design of advance letters is relatively limited, and recent evidence suggests that the content of letters may not make a difference to response rates (Olson et al., 2011). In their meta-analysis exploring the influence of advance letters on response in telephone surveys, De Leeuw et al. (2007) found that letters which appealed to the psychological concept of reciprocity (discussed by Groves, Cialdini and Couper, 1992) were most effective.

Our experimental treatment involved re-designing the content of the covering letters used on the between-wave mailing for both the initial and reminder mailings. There were four components to the re-design. The first two components – simplifying the language and reducing the length – were intended to make the letter easier to read for those with lower levels of education and who speak languages other than English at

home. We hypothesised that this would lead to higher return rates from these groups. Assuming that the return rates from other groups would not be negatively affected by this treatment, higher return rates from these targeted groups would also imply a higher return rate overall. The second two components – changing the style from formal to informal and changing the signatory from the study’s principal investigator to the study’s cohort maintenance officer - were intended to enhance the likelihood of compliance with the request to return the form for all sample members and particularly those in the targeted groups. We hypothesised that this would lead to higher return rates from all sample members and, in particular, higher return rates from those in the targeted groups. Other elements of the design and content of the covering letters - the study branding, the broad content and order - were held constant as far as possible between the control and treatment versions of the letters. No other changes to the standard protocol for the between-wave mailing were made. An address confirmation approach was used, with two reminders and without incentives, but with a leaflet feeding back results.

Our analysis examined whether there were any significant differences by experimental group in the overall return rate and early return rate. We also explored variations in the return rates by education level and languages spoken at home between experimental groups, and the impact this had in each experimental group on the composition of those who returned their forms. Design-based F tests were used to test for significant differences and the analysis was carried out in STATA using *svy* commands to adjust for the clustered and stratified nature of the sample design. Bonferroni corrections for multiple comparisons were made to the critical value for statistical significance for education and language.

The measure of education used was based on highest qualification, including both vocational and academic qualifications. The education level of the ‘main respondent’, usually the mother of the MCS cohort member, was used. Language spoken at home was a household-level measure reported by the main respondent. Both education and language are measured at the most recent prior wave in 2008 when the cohort member was aged seven. The analysis using these variables is therefore restricted to families who took part in that wave. Missing data rates were very low for these variables and cases with missing data were excluded from the analysis.

Our results showed that the experimental treatment did not succeed in boosting the overall return rate; there was no difference between experimental groups in this overall return rate (55 per cent for the control group and 56 per cent for the treatment group). However, it did lead to big increases in the return rates among sample members with no qualifications (42 per cent compared with 34 per cent) and among those who speak languages other than English at home (55 per cent compared with 51 per cent). These differences are statistically significant at conventional levels. This shows that the treatment improved the effectiveness of the mailing for these important groups and given the positive association between responding to between-wave mailings and taking part in subsequent waves of data collection, it is possible that this may lead to a reduction of bias at future waves.

We also found that the treatment led to a modest increase in the proportion of sample members returning their forms without the need for a reminder (34 per cent compared with 31 per cent). This difference was statistically significant at conventional levels. This meant that the cost-effectiveness of the mailing was improved, as reminders had to be sent to fewer sample members, and this finding is consistent with previous experimental evidence from the postal surveys literature (Taylor and Lynn, 1998).

However, our main finding, that the re-design of the covering letters had no impact on the overall return rate, implies that return rates on these mailings are unlikely to be strongly influenced by the design of the covering letter included in the mailing. It also provides indicative evidence that the willingness of sample members to respond to survey related requests may not be strongly influenced by the content of advance letters, which is consistent with other recent experimental literature (Olson et al., 2011). Having said that, it is important to bear in mind that the control letter used in the experiment was well-designed and followed best practice guidelines, and that the recipients were members of an established longitudinal survey, and hence used to receiving such letters annually.

Our findings about the design of covering letters are likely to be of relevance to survey practitioners in both cross-sectional and longitudinal surveys, including those using other modes of data collection.

Reflecting on the design of the experiment, it is clearly a limitation that the four components of the re-design of the letter were conflated in a single treatment. As

mentioned earlier, this means that it is not possible to disentangle the effects of the different components of the re-design either on their own or in combination with each other. Additionally, the paper is limited in the range of outcomes considered. It would have been illuminating to explore whether the higher return rate from those with lower educational qualifications and/or those whose speak languages other than English at home meant that they were more likely to participate in the subsequent wave than in the control group.

More generally, it is interesting, though perhaps not surprising, that this recent growth in tracking research has been about between-wave mailings, rather than other tracking methods. This is an area in which it is relatively easy to conduct randomised experiments, as it is straightforward to control the allocation of cases to different mailing conditions. It is also a relatively low-risk to intervene in these mailings, particularly on established studies whose sample members are used to responding to them.

The findings from Fumagalli et al. (2013) and Paper 2 (Calderwood, 2014) demonstrating the impact of an intervention on particular sub-groups of the study populations – younger people and busy people in the BHPS and lower educated and those who speak other languages at home in the MCS – chime with the reflections of Lynn (2015) regarding the use targeted response inducement strategies on longitudinal surveys. Together, they provide evidence that different versions of survey materials for specific sub-groups can improve compliance with survey-related requests. In the light of this research, survey practitioners should consider making greater use of targeted content in survey materials, including on between-wave mailings.

What do we know about the role of respondent characteristics in tracking on longitudinal surveys?

There is relatively little attention in the literature to the role of respondent characteristics in tracking on longitudinal surveys. This is surprising as many tracking methods, particularly prospective ones, rely on sample members taking particular actions e.g. informing the study when they move address, or returning a change of address card, and it seems reasonable to expect that some types of people will be more likely to do this than others. Despite this, tracking is often portrayed as something that is ‘done to’ sample members, rather than a process in which they are active agents. As

discussed earlier, this is in part because the term tracking is often, either explicitly or implicitly, used to refer to retrospective tracking.

Call (1990) examines respondent co-operation with requests for contact information on the National Survey of Families and Households in the US, and concludes that respondent demographics can influence the type and amount of contact information collected. This has implications for prospective tracking, particularly if those most likely to move are also the least likely to provide contact information, and suggests that greater emphasis should be given to the collection of this information in longitudinal surveys. The findings also imply that a targeted approach to the collection of contact information may be worth pursuing, with more information collected from sample members most at risk of dropping out.

Paper 3 (Calderwood, 2013) examines the role of respondent characteristics in tracking success rates, and compares how they are related to the relative success of both office tracking and field tracking. It is an observational study using data from the UK Millennium Cohort Study. This paper was motivated by the desire to increase the proportion of cases tracked in the office rather than in the field as a way of improving the cost-effectiveness of tracking procedures. As discussed earlier, office tracking is likely to be more cost-effective than field tracking, which implies that longitudinal surveys should try to track as many movers as possible prior to the start of data collection. In this way, more expensive field tracking can be restricted to cases which cannot be located through office tracking or which are only discovered to be movers during fieldwork. Our aim was to examine the characteristics of those sample members tracked through office and field tracking, with the intention of identifying particular groups for whom office tracking was not effective and which could therefore be targeted for additional work as a way of boosting office tracking rates.

We define office tracking rate as the percentage of all movers who were successfully located using office tracking methods prior to the start of the data collection wave, the field tracking rate as the percentage of movers not located by office tracking who were successfully located by interviewers in the field, and overall tracking rate as the percentage who were successfully located by either office or field tracking. Successful location means that a correct, new address was found. In our definition, office tracking includes proactive notifications of address changes by sample members, for example in response to a prospective tracking procedure such as a change of address. It should also

be noted that, in these definitions, field tracking necessarily happens after office tracking, and that field tracking is conditional on unsuccessful office tracking. For these reasons, these rates are not directly comparable. It should also be noted that office tracking was also used during the fieldwork. Therefore the field tracking rate includes some cases who were found using a combination of these methods. Similarly, some cases were not identified as movers until after the start of fieldwork, and so did not receive office tracking prior to fieldwork.

In our analysis we first use descriptive statistics to compare the rates of office, field and overall tracking for a range of different respondent characteristics. We then used multiple logistic regression analysis to examine the relationship between each of our chosen respondent characteristics and tracking rates, controlling for the other respondent characteristics. Separate logistic regression models were run to examine the relationship between respondent characteristics and each of the three tracking rates. Design-based F tests were used to test for significance differences and the analysis was carried out in STATA using *svy* commands to adjust for the sample design. All respondent characteristics were measures at the prior wave and no adjustments were made to take account for missing data, though in general missing data levels were relatively low for our chosen variables.

We find that a range of respondent characteristics are related to the success of office, field and overall tracking in the MCS. Controlling for other characteristics, ethnic group and age were both associated with successful office tracking; non-white and younger sample members were less likely to be found through this method. Ethnic group, employment status and accommodation type were associated with the success of conditional field tracking; sample members who were non-white, non-employed and living in flats were less likely to be located in the field. These groups of sample members were also less likely to be tracked overall, along with those from 'other' family types i.e. not lone parents or couple parents. Overall, the results show that respondent characteristics are important determinants of tracking success. However, it should also be noted that the overall goodness of fit of our logistic regression models was relatively low, indicating that other factors are also important.

Our finding that relatively few respondent characteristics were associated with office tracking success is reassuring in some ways as it shows that, with the exception of younger and non-white respondents, office tracking procedures are not systematically

failing to locate certain types of sample members. However, as overall tracking success was related, at least in the bivariate analysis, to all of the respondent characteristics, this implies that improving the effectiveness of all types of tracking for hard-to-locate groups should be the primary aim of further research and improvements to survey practice in this area. Further research is needed to examine the mechanisms through which these characteristics are related to tracking success. However, given the difficulty associated with tracking families living in flats and families from non-white ethnic groups, these findings imply that longitudinal surveys should consider implementing additional and/or targeted tracking methods for these groups, including the collection of additional contact information for those living in flats and the use of translated tracking materials and office tracking staff/interviewers who speak minority languages for non-white minority ethnic groups.

Paper 3 (Calderwood, 2013) also demonstrated that it is possible to locate a relatively high proportion of movers (over half) using office tracking prior to the start of fieldwork, and that this is an important component of achieving high overall tracking rates. On the one hand, this is encouraging news for longitudinal surveys which use solely remote data collection methods or which cannot afford to carry out field tracking. On the other hand, it also underlines the importance of field tracking to achieving high tracking rates.

Overall, this paper shows that although high overall tracking rates can be achieved on large-scale longitudinal surveys through a combination of office and field tracking, there is still a significant minority of sample members who are not located through these existing methods, and that these sample members have different characteristics than those who are located. This is another example of the tension in survey practice and methodology between maximising response rates and minimising non-response bias.

Our findings will be of relevance to survey practitioners running longitudinal surveys, particularly those who use both office and field tracking, which are primarily face-to-face surveys. However, managers of longitudinal surveys using only remote data collection, for whom field tracking is unlikely to be possible will also be interested, as for them improving the effectiveness of office tracking is a major concern. The main limitations of this research are that it combines both prospective and retrospective office tracking and that it is not able to explore the causal processes through which respondent characteristics are linked to tracking success.

In the light of this research, longitudinal survey practitioners should consider making greater use of additional targeted tracking methods for sample members who are most at risk of non-response due to failure to locate, for example those living in flats and families from non-white ethnic groups.

In this section, I have reviewed the main factors affecting location in longitudinal surveys, provided a brief summary of the research literature about how a range of survey design features are related to location and demonstrated how the findings from Paper 2 (Calderwood, 2014) and Paper 3 (Calderwood, 2013) have contributed to knowledge in this area. Both of these papers used prior wave data to evaluate the impact of survey processes used to track sample members on non-response bias.

Future methodological research on the prevention of non-response due to failure to locate in longitudinal surveys should focus on the effectiveness of specific different tracking methods, both 'traditional' methods such as using stable contacts and contacting neighbours and 'new' methods such as using administrative data and new technologies and social media for tracking.

Section 4: Contact

This section is about survey non-response due to failure to make contact. On most surveys, non-contact is the least common reason for non-response (Durrant and Steele, 2009; Foster, 1998). As discussed in section 3, for longitudinal surveys contact is contingent on successful location. Although the processes of location and contact are distinct in both cross-sectional and longitudinal surveys, they are often combined in the literature.

This section provides a brief review of the main factors affecting contact in surveys and the literature in this area, and summarises Paper 4 (Brown and Calderwood, 2014). This paper uses the *Understanding Society* Innovation Panel to test, through a randomised experiment, an innovative approach to making contact which involved encouraging sample members to initiate contact with their interviewer and book an interview appointment.

What are the main factors affecting contact?

Groves and Couper (1998) argue that the main factors associated with successful contact on face-to-face surveys are a) the at-home times of the sample members or householders b) any physical barriers to making contact and c) when and how many times the interviewer visits. This conceptual framework can also be applied to telephone surveys using landline numbers, though the nature of the barriers to contact are different in that context e.g. caller ID, answerphones.

At-home times vary with the socio-demographic characteristics of sample members. For example, non-employed or retired persons are generally more likely to be at home during the daytime than employed persons. However, for cross-sectional surveys, interviewers do not know the characteristics of sample members before they contact them, so they are not usually able to vary their approach to take account of this.

Physical barriers to making contact include devices such as entry phones and concierge desks which make it difficult for interviewers to actually alert the sample member directly to their presence. Interviewers usually become aware of such barriers to contact when they first visit the property, and need to try to find strategies to get around them where possible. Both at-home times and physical barriers are generally not under the control of the survey, and so the behaviour of the interviewer, in terms of when, how many times and over what period they attempt to contact sample members is the main way in which survey practitioners can influence successful contact.

The calling pattern of the interviewer clearly interacts with the at-home times of the sample member in influencing the likelihood of contact. The most difficult to contact sample members are generally those who spend little time at home, and vice versa. All other things being equal, the more contact attempts made by the interviewer the greater the likelihood of making contact. In this context, the duration of the fieldwork period can also have a major influence on non-contact rates; the shorter the fieldwork period the higher the rate of non-response due to non-contact.

It is standard practice on large-scale surveys to define minimum call requirements which interviewers must meet prior to recording non-contact outcomes. These call requirements specify the minimum number of calls and the period over which these calls must be made. There is also usually a requirement to vary the pattern of calls, for example, at least one call at weekends and/or weekday evenings. In general, for most large-scale surveys which follow this approach, the level of non-response due to failure to contact sample members tends to be relatively low. For this reason, and because interviewers' attempts to make contact with sample members are a major component of survey costs, research in this area has tended to focus on how to improve the efficiency of interviewers' contact attempts rather than on methods to reduce non-response due to non-contact.

Optimising calling strategies is a concern for both telephone and face-to-face surveys, though the context of interviewer decisions about when to call is very different in the two types of survey. For many telephone surveys, particularly those carried out in centralised telephone interviewing call-centres in large fieldwork agencies, interviewers have limited, if any, discretion about which cases to call and when to call them as this is determined by automated call scheduling procedures. For face-to-face surveys, and some telephone surveys, interviewers have much more discretion about when to attempt contact with their assigned cases. Additionally, interviewer workloads on face-to-face surveys are usually geographically clustered, which means that decisions about when to attempt to contact cases are not always taken in isolation from decisions about other cases. For example, if an interviewer is visiting the area for an interview appointment for a particular case, they may decide to call on other cases which live nearby while they are in the area. The costs associated with contact attempts are higher for face-to-face surveys than for telephone surveys, as they involve payment of interviewers for travel time to the area and the re-imburement of the direct costs associated with this travel. The availability of data on interviewer calling patterns is also different in

telephone and face-to-face surveys. Para-data on interviewer calls has been available for several decades for telephone surveys. For face-to-face surveys, this information has also been collected for many decades, but this has typically been done on paper and the data was rarely captured electronically. Recently major survey organisations in the US and UK are increasingly making use of electronic sample management systems for face-to-face surveys which require this data to be entered by interviewers in real-time. The increased availability of this information has contributed to the recent burgeoning of research in the area of para-data and responsive or adaptive designs, as discussed in section 1.

There are several studies looking at how to optimise calling patterns, for both telephone (e.g. Weeks et al., 1987; Greenberg and Stokes, 1990) and face-to-face (e.g. Weeks et al., 1980; Kalsbeek et al., 1994; Groves and Couper, 1998; Campanelli, Sturgis and Purdon, 1997; Durrant, D'Arrigo and Steele, 2011). In general, these studies show that the best times to make contact with sample members are during weekday evenings and at weekend daytimes, and that while the majority of sample members are contacted after a relatively small number of calls, there is a substantial minority of cases which require many more calls in order to make contact.

Research about the relationship between the characteristics of sample members and interviewer contact attempts is more limited, primarily because most cross-sectional surveys do not have information about the characteristics of sample members who are not contacted. There are, however, a number of studies which have successfully linked cross-sectional surveys to census data, meaning that this information is available for both respondents and non-respondents. These include Groves and Couper (1998) in the USA and Durrant and Steele (2009) and Durrant, D'Arrigo and Steele (2011) in the UK. In general, they find that, as expected, households with children and retired persons are easier to contact than single-person households and those with adults in employment. Durrant, D'Arrigo and Steele (2011) also find that interviewer observations about the household and area, and area-level variables are useful at predicting best times to contact and that higher grade, better educated and older interviewers are more successful at establishing contact.

One specific strand of research in this area has focused on evaluating the impact of extended efforts to make contact with sample members who are difficult to contact, and in particular whether these efforts are worthwhile in terms of reducing non-response

bias as well as improving response rates. Extended call attempts are becoming more common, on both telephone and face-to-face surveys, as making contact with sample members has become increasingly difficult over time. This is in part due to demographic changes such as increasingly female labour force participation and a growth in single-person households, technological changes which introduce new barriers to contact such as caller ID and societal changes which mean that people are less likely to respond to unsolicited calls. A number of studies examining the impact of extended call attempts on telephone surveys in the US have concluded that they have little impact on non-response bias (Curtin, Presser and Singer, 2000; Keeter et al., 2000; Peytchev, Baxter and Carley-Baxter, 2009). However, analysis of the impact of extended call attempts on a number of UK face-to-face cross-sectional surveys shows that they do pay off in terms of a reduction in non-response bias (Lynn and Clarke, 2002; Hall et al., 2013).

More generally, most of the literature in this area, for both telephone and face-to-face surveys, relies on the analysis of observed interviewer calling patterns, which are by definition clustered at the times most conducive to making contact and inter-related. For face-to-face surveys in particular, calls patterns for different cases may also not be independent of each other as assignments are often clustered and interviewers often call on multiple cases in the same visit to the area. There are relatively few studies which attempt to experimentally manipulate calling patterns in order to examine the impact of different patterns on non-contact or fieldwork efficiency. This is perhaps unsurprising as this is difficult and expensive to do, particularly on face-to-face surveys.

What are the main factors affecting contact on longitudinal surveys?

For longitudinal surveys, the factors influencing successful contact with sample members are the same as for cross-sectional surveys. However, one of the main differences is that on longitudinal surveys interviewers are, in principle, able to use information from prior waves about sample member characteristics and when contact has successfully been made previously to inform their call patterns. It is hard to judge how widespread this practice actually is on longitudinal surveys as, until recently, there has been relatively little published research in this area. Lipps (2012) has recently shown, using the telephone Swiss Household Panel, that making the first call at the time the household was first contacted at the previous wave increases both the contact probability and the overall likelihood of co-operation at the current wave. However, Kreuter and Muller (2014) were unable to replicate this finding in a randomised

experiment on the German PASS study, also a longitudinal survey conducted by telephone. In a randomised experiment on a face-to-face longitudinal study in the US, Kreuter, Mercer and Hicks (2014) found that modest gains in fieldwork efficiency were possible through using pre-specified appointment times, based on the prior wave interview date. One of the reasons why few longitudinal surveys appear to have taken this approach may be that for longitudinal telephone surveys the costs associated with making contact are relatively low and therefore efficiency gains in this area have a relatively minimal impact on survey costs, whereas for face-to-face longitudinal surveys where the potential cost-saving are greater, the manipulation of interviewer calling patterns in this way for specific cases is more difficult due to geographical clustering of cases and the efficiency gains this clustering allows as well as the difficulty of controlling the behaviour of field interviewers.

It is also worth bearing in mind that on the second or later waves of a longitudinal survey the interviewer's initial task is slightly different, in ways which have implications for how contact can be made. In this context, sample members generally have a good understanding of the survey request through their prior experience of taking part and many will require minimal persuasion to participate again. The interviewer's initial task is often therefore to make contact and set up an appointment for the interview, rather than to persuade sample members to take part. This can be done much more efficiently over the phone than in person, and for this reason several large-scale face-to-face longitudinal surveys allow first contact by phone, particularly for those sample members who are most likely to take part.

Another response to the different nature of making contact in longitudinal surveys is to encourage sample members to initiate contact with the interviewer to set up an interview appointment, rather than the other way around. This approach has been pioneered by the National Opinion Research Centre (NORC) on the 1979 cohort of the National Longitudinal Surveys of Youth (NLSY), a large-scale telephone longitudinal survey in the US. From 2002, they have offered incentives (of between \$40-\$80) to sample members who had been co-operative in the previous wave to contact the study to set up an appointment for their interview, and found that almost half of those eligible for this offer took it up (Kochanek et al., 2010). As interviewers did not need to spend time making calls to set up appointments for these sample members, they found that these cases took much less interviewer time to complete. However, as this innovation was not tested through a randomised experiment, it is not possible to definitively

attribute the reduction in fieldwork costs to this change as alternative explanations such as selection effects, for example that those taking up the offer were the most co-operative and hence would have had lower costs anyway, cannot be ruled out.

Paper 4 (Brown and Calderwood, 2014) provides evidence from a randomised experiment, conducted on the Innovation Panel of *Understanding Society*: the UK Household Longitudinal Study (UKHLS) in 2011, which sought to evaluate whether this ‘early bird’ approach used on the NLSY could be successful in a UK context. To our knowledge, this was the first time that such an approach had been used on a major UK longitudinal survey, and the first time anywhere that this approach has been robustly evaluated using a randomised experimental approach. As UK surveys are typically carried out face-to-face, rather than by phone as is often the case in the US, there is potential for this ‘early-bird’ approach to deliver even greater cost-savings in the UK than in the US.

Another major difference between the US and the UK, and one reason why survey costs in the US are typically higher, is that high-value respondent incentives are standard practice on most US surveys. Although incentives are becoming more widespread in the UK, and are used on the UKHLS, they tend to be much lower value. There are also several major longitudinal surveys in the UK, including the British Birth Cohorts, which do not use incentives at all. For this reason, our experiment also included a treatment group who were not offered an additional incentive and instead encouraged to become an ‘early-bird’ by an appeal to altruism, motivated by Groves, Cialdini and Couper (1992). The incentive treatment involved offering a conditional £5 incentive per person in households where the ‘early bird’ offer was taken up. The altruism treatment involved telling sample members that pre-booking their appointment would ‘make your interviewer’s life easier as they will not need to make repeated calls or visits to your home in order to try and reach you’.

We found that the take-up rate for the ‘early-bird’ offer was higher for those in the incentive treatment group compared with those who received the treatment which appealed to altruism (10 per cent compared with 6 per cent). Another experiment involving the use of higher-value standard incentives (£10 rather than £5) was fielded in parallel to ours, and we found that the take-up rate for the early bird offer was the highest among those who received a higher-value standard incentive (17 per cent). These differences were all statistically significant at conventional levels. This suggests

that take-up rates to the ‘early bird’ offer are maximised when the combined total incentive value is maximised. However, formal testing of the interaction between the offer of the early bird incentive and the standard incentive, using a logistic regression model, found that this was not statistically significant.

In relation to overall fieldwork effort, we were not able to detect any statistically significant differences between experimental groups in the average number of interviewer visits required complete interviewing in productive households. However, when we looked only at productive households in which the early-bird offer was taken up, we did find that fewer calls were required compared with households in which the early-bird was either not offered or offered and not taken up (1.4 compared with 3.4 on average). This difference was statistically significant at conventional levels. In order to attribute this reduction to the take-up of the early-bird offer, we examined the number of calls required at the previous wave. This showed that although those taking up the early-bird offer had required fewer calls than other cases at the previous wave, the early-bird offer had led to a statistically significant reduction in the number of calls required (from 3.2 at the previous wave to 1.4 at the current wave) which was not observed for other groups. In this way, we were able to attribute the reduction in calls found among those taking up the early-bird offer to the offer itself.

In relation to response rates, we had expected that the early-bird offer may have a modest positive impact on response rates for the incentive treatment group due to the increase in the total value of the incentive available. We found that the response rate was actually slightly higher for both the incentive treatment group (77 per cent) and the no incentive treatment group (78 per cent) compared with the control group (73 per cent). It was highest among those offered the early-bird incentive and the higher value standard incentive (79 per cent). Although these differences were not statistically significant due to small sample sizes, they do provide indicative evidence that the early-bird could have a potentially beneficial impact on response rates.

However, overall the take-up of the early-bird offer was relatively low, and much lower than on NLSY. Although there were other differences between the studies, it seems likely that the main reason for this was that the value of the incentives used on the UKHLS experiment was much lower than in the US study. Our findings imply that the early-bird approach has the potential to reduce costs, but that further research is needed to examine ways of encouraging a higher proportion of respondents to set-up

appointments for themselves. In summary, Paper 4 (Brown and Calderwood, 2014) provides interesting comparative evidence from the UK about an innovative fieldwork approach pioneered on a US study, and makes an important contribution to the survey methodological literature on improving the efficiency of fieldwork approaches to make contact and on the use of incentives.

In the light of this research, survey practitioners, particularly on longitudinal surveys should consider how to leverage the loyalty and commitment of highly engaged sample members to improve the cost-effectiveness of the contacting and appointment making process. This may include encouraging sample members to make appointments as in our experiment. Allowing centralised booking of interview appointments and/or allowing sample members to proactively give their availability and preferences for slots or particular times and dates, may be a way of facilitating this which is simpler for respondents than liaising with a specific interviewers on a bi-lateral basis. However, this is likely to be easier to facilitate for telephone surveys, where interviewer availability can be more easily guaranteed for a particular time than for face-to-face surveys where this is usually depending on the availability of a particular interviewer.

It should also be noted that any cost-savings as a result of the reduction of interviewer effort in appointment making need to be set against the direct costs of encouraging sample members to set-up their own appointments such as additional incentives and new technology needed such as the provision of mobile phones to face-to-face interviewers in our experiment.

More generally, this paper illustrates the advantages of using a methodological ‘test bed’, in the shape of the *Understanding Society* Innovation Panel for carrying out experiments with fieldwork practice. As noted in section 1, it is relatively rare that survey practitioners are able to do this. However, it also shows that there can be unanticipated consequences when multiple experiments are fielded in parallel with each other. For our paper, this brought some benefits as we were able to evaluate the impact of experimental intervention in conjunction with a parallel experiment on standard incentive amounts, and this led to some interesting findings. However, this also led to some very small sample sizes in particular groups when the two experiments were combined. This exacerbated our existing difficulty of detecting statistically significant effects for our experimental treatment due to small sample sizes. Although we were able to detect some statistically significant effects of the experimental treatment, some of our

findings failed to reach conventional significance levels due to small sample sizes. This illustrates one of the challenges of carrying out randomised experiments in surveys discussed in section 1.

In this section, I have reviewed the main factors affecting contact in surveys, provided a brief summary of the research literature in this area and demonstrated how the findings from Paper 4 (Brown and Calderwood, 2014) have contributed to knowledge in this area.

Future methodological research on the prevention of non-response due to failure to make contact in longitudinal surveys should focus improving the cost effectiveness of the process, and in particular making greater use of the wider range of contact details which are often available for sample members in longitudinal studies. For example, text messages and emails as an alternative to phone and personal visits.

In my PhD thesis I have explored how to reduce non-response in longitudinal surveys by improving survey practice. Collectively my four papers have addressed all three of the major components of non-response in longitudinal surveys; co-operation, location and contact, and represent a coherent programme of research. I have used randomised experiments and prior wave data to evaluate the effectiveness of fieldwork interventions on both the non-response rate and non-response bias. This summary has shown how my research fits within the broader context of survey methodological literature on non-response, and demonstrates its contribution to knowledge in this area.

References

- Aughinbaugh, A. (2004). The impact of Attrition on the Children of the NLSY79, *The Journal of Human Resources*, Vol.39 (2), pp.536-563. DOI: 10.3368/jhr.XXXIX.2.536
- Basson, D. and Chronister, M. (2006). Recordings of prior refusals: do they improve later conversion attempts? Paper presented at the Methodology of Longitudinal Surveys workshop, Essex, UK, July 2006.
- Bethlehem, J.G. (2002). Weighting Nonresponse Adjustments Based on Auxiliary Information. In Groves, R.M., Dillman, D.A., Eltinge, J., and Little, R. (eds.). *Survey Nonresponse*. Wiley: New York.
- Beullens, K., Billiet, J. and Loosveldt, G. (2010). The effect of the elapsed time between the initial refusal and conversion contact on conversion success: evidence from the 2nd round of the European Social Survey. *Quality and Quantity*. 44, pp.1053-1065. DOI: 10.1007/s11135-009-9257-4
- Brown, M., and Calderwood, L. (2014). Can encouraging respondents to contact interviewers to make appointments reduce fieldwork effort? Evidence from a randomized experiment in the UK. *Journal of Survey Statistics and Methodology*. 2 (4), pp. 484-498. DOI: 10.1093/jssam/smu017.
- Buck, N. (2008). Understanding Society: Design overview. *Understanding Society Working Paper Series*, 2008-01.
- Burgess, R.D. (1989). Major Issues and Implications of Tracing Survey Respondents, In Kasprzyk, D., Duncan, G., Kalton, G., and Singh, M.P., *Panel Surveys*, Wiley: New York pp.52-74
- Burton, J., Laurie, H. and Lynn, P. (2006). The long-term effectiveness of refusal conversion procedures on longitudinal surveys. *Journal of the Royal Statistical Society Series A*, Royal Statistical Society, 169(3), pp.459-478. DOI: 10.1111/j.1467-985X.2006.00415.x
- Burton, J., Laurie, H., and Uhrig, S.C. (2008). Understanding Society: Some preliminary results from Wave 1 of the Innovation Panel. *Understanding Society Working Paper Series*, 2008-03.

Calderwood, L., Plewis, I., Ketende S.C. and Taylor, R. (2010). Experimental testing of refusal conversion strategies in a large-scale longitudinal study. *CLS Working Paper 2010/9*. London: Centre for Longitudinal Studies.

Calderwood, L. (2013). The role of respondent characteristics in tracking on longitudinal surveys: evidence from the UK Millennium Cohort Study. *Lifecourse and Longitudinal Studies*, 4(2), pp.105-118. DOI: [http://dx.Doi.org/10.14301/lles.v4i2.231](http://dx.doi.org/10.14301/lles.v4i2.231)

Calderwood, L. (2014). Improving between-wave mailings on longitudinal surveys: a randomised experiment on the UK Millennium Cohort Study. *Survey Research Methods*. 8 (2), pp. 99-108.

Call, V.R.A. (1990). Respondent Cooperation and Requests for Contacts in Longitudinal Research. *National Survey of Families and Households, Working Paper No.35*. Madison: University of Wisconsin-Madison.

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

Campanelli, P., and O'Muircheartaigh, C. (2002). The importance of experimental control in testing the impact of interviewer continuity on panel survey nonresponse. *Quality & Quantity* 36, pp.129-144.

Campanelli, P., Sturgis, P. and Purdon, S. (1997). Can you hear me knocking? An investigation into the impact of interviewers on survey response rates. London, GB, The Survey Methods Centre at SCPR, 212pp.

Church, A. H. (1993). Estimating the effect of incentives on mail survey response rates: a meta-analysis. *Public Opinion Quarterly*, 57, pp.62-79. DOI: 10.1086/269355

Clarridge, B.R., Sheehy, L.L. and Hauser, T.S. (1978). Tracing members of a panel: a 17-year follow-up. *Sociological Methodology*, 9, pp.185-203.

Cotter, R.B., Burke, J.D., Stouthamer-Loeber, M., and Loeber, R. (2005). Contacting Participants for Follow-up: How Much Effort is Required to Retain Participants in Longitudinal Studies?. *Evaluation and Program Planning*, 28, pp.15-21. DOI: 10.1016/j.evalprogplan.2004.10.002.

- Couper, M.P., Singer, E., Conrad, F.G. and Groves, R.M. (2010). Experimental Studies of Disclosure Risk, Disclosure Harm, Topic Sensitivity, and Survey Participation. *Journal of Official Statistics*, 26, 2, pp. 287-300. DOI: 10.1111/j.1369-7625.2011.00739.x
- Couper, M.P. and Ofstedal, M.B. (2009). Keeping in Contact with Mobile Sample Members. In Lynn, P. (ed.), *Methodology of Longitudinal Surveys*, Chichester: John Wiley & Sons, Inc, pp. 183-203.
- Couper, M.P. and Wagner, J. (2011). Using Paradata and Responsive Design to Manage Survey Nonresponse. *Proceedings of the 58th World Statistical Congress, Dublin*, 542-548. <http://2011.isiproceedings.org/papers/450080.pdf>
- Crider, D.M., Willits, F.K and Bealer, R.B (1971). Tracking Respondents in Longitudinal Surveys. *Public Opinion Quarterly*, 35, pp. 613-620.
- Curtin, R., Presser, S. and Singer, E. (2000). The Effects of Response Rate Changes on the Index of Consumer Sentiment. *Public Opinion Quarterly*, 64, pp.413-428. DOI: 10.1086/318638
- De Leeuw, E., and De Heer, W. (2002). Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. In Groves, R.M., Dillman, D.A., Eltinge, J., and Little, R. (eds.). *Survey Nonresponse*. Wiley: New York.
- De Leuw, E., Callegaro, M., Hox, J. Korendijk, E., and Lensvelt-Mulders, G. (2007). The Influence of Advance Letters in Telephone Surveys. *Public Opinion Quarterly*, 71(3), pp.413-443. DOI: 10.1093/poq/nfm014
- DeMaio, T.J. (1980). Refusals: Who, Where, and Why? *Public Opinion Quarterly*, 44, pp. 223-233. DOI: 10.1086/268586
- Dex, S. and Rosenberg, R. (2008). Ethnic minorities and nonresponse in the Millennium Cohort Study. *CLS Working Paper, 2008/11*. London: Centre for Longitudinal Studies.
- Dillman, D.A., Smyth, J.D., and Christian, L.M. (2009). *Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method*. Wiley: New York

- Durrant, G.B. and Steele, F. (2009). Multilevel Modelling of Refusal and Non-contact in Household Surveys: Evidence from six UK Government Surveys. *Journal of the Royal Statistical Society Series A*. 172(2), pp. 361-381. DOI: 10.1111/j.1467-985X.2008.00565.x
- Durrant, G.B., Groves, R.M., Staetsky, L., and Steele, F. (2010). Effects of Interviewer Attitudes and Behaviours on Refusal in Household Surveys. *Public Opinion Quarterly*, 74(1), pp.1-36. DOI: 10.1093/poq/nfp098
- Durrant, G.B., D'Arrigo, J. and Steele, F. (2011). Using field process data to predict best times of contact conditioning on household and interviewer influences. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174, (4), pp. 1029-1049. DOI:10.1111/j.1467-985X.2011.00715.x.
- Durrant, G.B., and D'Arrigo, J. (2014). Doorstep interactions and interviewer effects on the process leading to cooperation or refusal. *Sociological Methods & Research*, pp.1-29. DOI:10.1177/0049124114521148
- Eckland, B.K (1968). Retrieving Mobile Cases in Longitudinal Surveys. *Public Opinion Quarterly*, 32, pp.51-64. DOI: 10.1086/267578
- Foster, K. (1998). Evaluating non-response on household surveys. *Government Statistical Service Methodology Series, No.8*.
- Fox, R.J., Crask, M.R., and Kim, J. (1988). Mail Survey Response Rate, A Meta Analysis of Selected techniques for Inducing Response. *Public Opinion Quarterly* 52 (4): pp.467-491. DOI: 10.1086/269125
- Fumagalli, L., Laurie, H. and Lynn, P. (2013). Experiments with Methods to Reduce Attrition in Longitudinal Surveys. *Journal of the Royal Statistical Society Series A*, 176(2), pp.499-519. DOI: 10.1111/j.1467-985X.2012.01051.x
- Fuse, K. and Xie, D. (2007). A successful conversion or double refusal: a study of the process of refusal conversion in telephone survey research. *The Social Science Journal*, Vol. 44, pp.434-446. DOI: 10.1016/j.soscij.2007.07.011

- Gray, R., Campanelli, P., Deepchand, K. and Prescott-Clarke, P. (1996). Exploring survey non-response: The effect of attrition on a follow-up of the 1984-85 Health and Life Style Survey. *The Statistician*, Vol. 45. No 2. pp.163-183. DOI: [10.2307/2988406](https://doi.org/10.2307/2988406)
- Greenberg, B.S. and Stokes, S.L. (1990). Developing an Optimal Call Scheduling Strategy for a Telephone Survey. *Journal of Official Statistics*, 6, pp.421-435. DOI:10.1111/j.1467-985X.2006.00423.x
- Groves, R.M., Cialdini, R.B and Couper, M.P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly*, 56, 4, pp.475-495. DOI: 10.1086/269338.
- Groves, R.M. and Couper, M.P. (1998). *Nonresponse in Household Interview Surveys*. Wiley: New York. DOI: 10.1002/9781118490082.index
- Groves, R.M., Singer, E., Corning, A., and Bowers, A. (1999). A Laboratory Approach to Measuring the Joint Effects of Interview Length, Incentives, Differential Incentives and Refusal Conversion on Survey Participation. *Journal of Official Statistics*, 15, pp. 251-268. DOI:10.1093
- Groves, R.M., Singer, E., and Corning, A. (2000). A Leverage-Saliency Theory of Survey Participation. *Public Opinion Quarterly*, 64, pp.299-308. DOI: 10.1086/317990
- Groves, R.M., Presser, S. and Dipko, S. (2004). The Role of Topic Interest in Survey Participation Decisions. *Public Opinion Quarterly*, 68, pp.2-31. DOI: 10.1093/poq/nfh002.
- Groves, R.M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly*, 70, pp. 646-675. DOI: 10.1093/poq/nfl033
- Groves, R.M., Couper, M., Presser, S., Singer, E., Tourangeau, R., Piani Acosta, G. and Nelson, L. (2006). Experiments in Producing Nonresponse Bias. *Public Opinion Quarterly*, 70, pp.720-736. DOI: 10.1093/poq/nfl036
- 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*, 169, 3, pp.439-459. DOI: 10.1111/j.1467-985X.2006.00423.x

- Groves, R. M. and Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. *Public Opinion Quarterly*, 72(2), pp.167-189. DOI: 10.1093/poq/nfn011
- Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E., Tourangeau, R. (2009). *Survey Methodology*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- 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*, No. 117: 5-25. doi:10.1177/0759106312465545
- Hampson, S.E., Dubanoski, J.P., Hamada, W., Marsella, A.J., Matsukawa, J., Suarez, E. and Goldberg, L.R. (2001). Where are they now? Locating former elementary-school students after nearly 40 years for a longitudinal study of personality and health. *Journal of Research in Personality*, 35, pp.375-387. DOI:10.1006/jrpe.2001.2317
- Hawkes, D. and Plewis, I. (2006). Modelling non-response in the National Child Development Study. *Journal of the Royal Statistical Society Series A (Statistics in Society)*. 169 (3), pp.479-491. DOI: 10.1111/j.1467-985X.2006.00401.x
- Heerwegh, D., Abts, K., and Loosveldt, G. (2007), Minimising Survey Refusal and Noncontact rates: Do our efforts pay off? *Survey Research Methods*, 1, pp.3-10.
- Hox, J.J. and De Leeuw, E.D. (1994). A Comparison of Nonresponse in mail, telephone, and face-to-face surveys. Applying multilevel modelling to meta-analysis. *Quality and Quantity*, 329-344.
- Jackle, A. and Lynn, P. (2008). Respondent incentives in a multi-mode panel survey: cumulative effects on non-response and bias. *Survey Methodology*, Vol. 34, No. 1, pp105-117.
- Jackle, A., Lynn, P., Sinibaldi, J. and Tipping, S. (2013). The Effect of Interviewer Experience, Attitudes, Personality and Skills on Respondent Co-operation with Face-to-Face Surveys. *Survey Research Methods*. Vol.7, No.1. <http://dx.doi.org/10.18148/srm/2013.v7i1.4736>

- Kalsbeek, W.D., Botman, S.L., Massey, J.T., and Liu, P.W. (1994). Cost-efficiency and the number of allowable call attempts in the National Health Interview Survey. *Journal of Official Statistics*, 10, pp.133-152.
- Keeter, S., Miller, C., Hohut, A., Groves, R.M., and Presser, S. (2000). Consequences of Reducing Nonresponse in a National Telephone Survey. *Public Opinion Quarterly*, 64, pp.125-148. DOI: 10.1086/317759
- Kochanek, K., Krishnamurty, P., and Michael, R. (2010). The 'Early Bird' Field Innovation on a 30-year-old Longitudinal Survey. Presentation at the 2nd International Panel Survey Methods Workshop, July, 2010. Mannheim, Germany.
- Koo, M.M. and Rohan, T.E. (2000). Use of the world wide web-based directories for tracing subjects in epidemiological studies. *American Journal of Epidemiology*, 152, pp.889-894. DOI: 10.1093/aje/152.9.889
- Kreuter, F., and Olson, K. (2013). Paradata for nonresponse error investigation. In Kreuter, F. (ed.) *Improving Surveys with Paradata. Analytic Uses of Process Information*. Wiley: New Jersey. DOI: 10.1002/9781118596869.ch2
- Kreuter, F., Mercer, A., and Hicks, W. (2014). Increasing Fieldwork Efficiency Through Prespecified Appointments. *Journal of Survey Statistics and Methodology*, 2 (2), pp. 210-223. DOI: 10.1093/jssam/smu005
- Kreuter, F., and Muller, G. (2014). A note on improving process efficiency in Panel Surveys with Paradata. *Field Methods*, DOI:10.1177/1525822X14538205.
- Laurie, H., Smith, R. and Scott, L. (1999). Strategies for Reducing Nonresponse in a Longitudinal Panel Survey. *Journal of Official Statistics*, 15(2), pp. 269-282. DOI: 10.1111/j.1467-985X.2012.01051.x
- Laurie, H. and Lynn, P. (2009). The Use of Respondent Incentives on Longitudinal Surveys. In Lynn, P. (ed). *Methodology of Longitudinal Surveys*, Chichester: Wiley.
- 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. (eds.). *Survey Nonresponse*, New York: John Wiley and Sons. pp. 259-272.

Lin, I-F., and Schaeffer, N.C. (1995). Using Survey Participants to Estimate the Impact of Nonparticipation. *Public Opinion Quarterly*, 59(2), pp.236-258.

DOI:10.1086/269471

Lipps, O. (2012). A note on improving contact times in Panel Surveys. *Field Methods*, vol. 24, 1, pp. 95-111. DOI: 10.1177/1525822X14538205

Lynn, P. and Clarke, P. (2002). Separating Refusal Bias and Non-Contact Bias: Evidence from UK National Surveys. *The Statistician*, 51(3), pp. 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. (eds.). *Survey Nonresponse*, New York: John Wiley and Sons.

Lynn, P. (2003). PEDAKSI: Methodology for Collecting Data about Survey Nonrespondents. *Quality and Quantity*, 37, pp. 239-261.

Lynn, P. (2009a). Methods for Longitudinal Surveys. In Lynn, P. (ed.). *Methodology of Longitudinal Surveys*. Chichester: John Wiley & Sons, Inc, pp. 1-20.

Lynn, P. (2009b). Sample design for Understanding Society. *Understanding Society Working Paper Series*, 2009-01.

Lynn, P., Kaminska, O. and Goldstein, H. (2014). Panel attrition: how important is interviewer continuity? *Journal of Official Statistics*. 30(3), pp. 443-457.

DOI: [10.2478/jos-2014-0028](https://doi.org/10.2478/jos-2014-0028).

Lynn, P. (2015). Targeted response inducement strategies on longitudinal surveys. In Engel, U., Jann, B., Lynn, P., Scherpenzeel, A. and Sturgis, P. (eds.). *Improving Survey Methods: Lessons from Recent Research*. Routledge.

Lynn, P. (2014). Longer Interviews May Not Affect Subsequent Survey Participation Propensity. *Public Opinion Quarterly*. 78(2), pp. 500-509.

Martin, E., Abreu, D., and Winters, F. (2001). Money and Motive: Effects of Incentives on Panel Attrition in the Survey of Income and Program Participation, *Journal of Official Statistics*, 17 (2), pp.267-284.

- Masson, H., Balfe, M., Hackett, S. and Phillips, J. (2011), Lost without a trace? Social networking and social research with a hard-to-reach population. *British Journal of Social Work*, 43(1), pp. 24-40. DOI: 10.1093/bjsw/bcr168
- McAllister, R.J., Goe, S.T. and Butler, E.W. (1973). Tracking Respondents in Longitudinal Surveys: Some preliminary considerations. *Public Opinion Quarterly*, 37, pp.413-416. DOI: 10.1086/268103
- McGonagle, K.A., Couper, M.P., and Schoeni, R.F. (2011). Keeping Track of Panel Members: An Experimental Test of a Between-Wave Contact Strategy. *Journal of Official Statistics*, 27, pp. 319-338. DOI: 10.2478/jos-2013-0022
- McGonagle, K.A., Schoeni, R.F., and Couper, M.P. (2013). The Effects of a Between-Wave Incentive Experiment on Contact Update and Production Outcomes in a Panel Study. *Journal of Official Statistics*, 29(2), pp. 261-276.
- Mindell, J., Biddulph, J.P., Hirani, V., Stamatakis, E., Craig, R., Nunn, S. and Shelton, N. (2012). Cohort Profile: The Health Survey for England. *International Journal of Epidemiology*, doi: 10.1093/ije/dyr199
- Morton-Williams, J. (1993). *Interviewers Approaches*, Aldershot: Dartmouth
- Nwadiuko, J., Isbell, P, Zolotor, A.J., Hussey, J., and Kotch, J.B. (2011). Using Social Networking Sites for Subject Tracing, *Field Methods*, 23 (1), pp. 77-85. DOI: 10.1177/1525822X10384088
- Office for National Statistics (2004). Census 2001 National Report for England and Wales, part 2. London: TSO.
- Office for National Statistics (2013). Family Resources Survey, United Kingdom, 2011/2012. London: TSO.
- Office for National Statistics (2015). Family Resources Survey, United Kingdom, 2013/2014. London: TSO.
- Olson, K., Lepkowski, J.M. and Garabrant, D.H. (2011). An Experimental Examination of the Content of Persuasion, Letters on Nonresponse Rates and Survey Estimates in a

Nonresponse Follow-Up Study. *Survey Research Methods*, 5, pp. 21-26. DOI: 10.1111/j.1467-985X.2012.01042.x

O’Muircheartaigh, C., and Campanelli, P. (1999). A Multi-level Exploration of the Role of Interviewers in Survey Nonresponse. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 162(3), pp.437-446.

Peytchev, A., Baxter, R.K., and Carley-Baxter, L.R. (2009). Not All Survey Effort is Equal. Reduction of Nonresponse Bias and Nonresponse Error. *Public Opinion Quarterly*, 73, 4, pp. 785-806. DOI: 10.1093/poq/nfp037

Peytchev, A., Riley, S., Rosen, J., Murphy, J. and Lindblad, M. (2010). Reduction of Nonresponse bias through Case Prioritisation. *Survey Research Methods*, 4, 1, pp.21-29.

Plewis, I. (ed.) (2007). Millennium Cohort Study: Technical Report on Sampling (4th edition). London: Centre for Longitudinal Studies.

Plewis, I., Ketende, S.C., Joshi, H. and Hughes, G. (2008). The Contribution of Residential Mobility to Sample Loss in a British Birth Cohort Study: Evidence from the first two waves of the Millennium Cohort Study. *Journal of Official Statistics*, 24, 364-385.

Rodgers, W. (2002). Size of Incentive Effects in a Longitudinal Study, *American Association for Public Research: Strengthening Our Community – Section on Survey Research Methods*.

Schouten, B., Shlomo, N., and Skinner, C. (2011). Indicators for Monitoring and Improving Representativeness of Response. *Journal of Official Statistics*, 27, 2, pp. 231-253.

Shadish, W. R. and Cook, T.D. (2009). The Renaissance of Field Experimentation in Evaluating Interventions. *Annual Review of Psychology*, 60, pp. 607-629. DOI: 10.1146/annurev.psych.60.110707.163544

Singer, E., Hoewyk, J. V., Gebler, N., Raghunathan, T. and McGonagle, K. (1999). The effects of incentives on response rates in interviewer-mediated surveys. *Journal of Official Statistics*, 15(2), pp. 217-230. DOI: 10.4135/9781446262269

- Singer, E. (2002). The Use of Incentives to Reduce Nonresponse in Household Surveys. In Groves, R.M., Dillman, D.A., Eltinge, J.L., and Little, R.J.A. (eds.). *Survey Nonresponse*, New York: John Wiley and Sons.
- Smith, T.W. (1983). The Hidden 25 Percent: An Analysis of Nonresponse on the 1980 General Social Survey. *Public Opinion Quarterly*, 47 (3), pp. 386-404. DOI: 10.1086/268797
- Steeh, C.G (1981). Trends in Nonresponse Rates, 1952-1979. *Public Opinion Quarterly*, 45 (1), pp. 40-57. DOI: 10.1086/268633
- Stoop, I. (2004). Surveying nonrespondents. *Field Methods*, 16, pp.23-54. DOI:10.1177/1525822X03259479
- Taylor, M. (2013). Better communication with participants. Presentation to International Workshop on Household Survey Non-Response, London.
- Taylor, S. and Lynn, P. (1998). The effect of a preliminary notification letter on response to a postal survey of young people. *Journal of the Market Research Society*, 40, pp. 165-173.
- Vandecasteele, L. and Debels, A. (2007). Attrition in Panel Data: The effectiveness of weighting. *European Sociological Review*, 23, pp. 81-97. DOI: 10.1093/esr/jcl021
- Watson, D. (2003). Sample Attrition between waves 1 and 5 in the European Community Household Panel. *European Sociological Review*, 19, 4, pp. 361-378. DOI: 10.1093/esr/19.4.361
- Watston, N. and Wooden, M. (2009). Identifying factors affecting longitudinal survey response, In Lynn, P. (ed.) *Methodology of Longitudinal Surveys* Chichester: Wiley.
- Watson, N. and Wooden, M. (2014). Re-engaging with survey non-respondents: evidence from three household panels. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 177 (2), pp.499–522. DOI: 10.1111/rssa.12024
- Weeks, M.F., Jones, B.L., Folsom, R.E., and Jr. Benrud, C.H. (1980). Optimal Times to Contact Sample Households. *Public Opinion Quarterly*, 44(1), pp. 101-114. DOI – 10.1086/26856

Weeks, M.F., Kulka, R., and Pierson, S. (1987). Optimal Call Scheduling for a Telephone Survey. *Public Opinion Quarterly*, 51(4), pp. 540-549. DOI:10.1086/269056

Original Publications

Paper 1: Calderwood, Plewis, Ketende and Taylor (2010)

Paper 2: Calderwood (2014)

Paper 3: Calderwood (2013)

Paper 4: Brown and Calderwood (2014)

Paper 1:

Calderwood, L., Plewis, I., Ketende S.C. and Taylor, R. (2010). Experimental testing of refusal conversion strategies in a large-scale longitudinal study. CLS Working Paper 2010/9. London: Centre for Longitudinal Studies.

Note that this paper has subsequently been revised to incorporate recently available data from a later wave of the survey. This latest version is the version which is being submitted as part of my PhD. The title and authorship of the paper has also changed to reflect content and contributions of this more recent draft, and is now as follows:

Calderwood, L., Plewis, I., Ketende, S. and Mostafa, T. Evaluating the immediate and longer- term impact of refusal conversion strategies in a large-scale longitudinal study.

Abstract

Refusal conversion is one of the fieldwork strategies commonly used to minimise non-response in surveys. However, there is relatively little evidence about the effectiveness of this strategy, particularly in face-to-face longitudinal surveys. Moreover, much of the existing evidence is based on observational studies. This paper evaluates the effectiveness of fieldwork strategies to convert refusals using evidence from a randomised experiment implemented in wave four of a large-scale longitudinal study in the UK. We show that intensive re-issuing is an effective way of increasing the proportion of refusals converted to a productive interview and hence reducing the refusal rate. We also show that refusal conversion may have led to a reduction in non-response bias in the survey estimates for several key variables. The longer-term impact of refusal conversion is also a key concern in longitudinal surveys. We use wave five outcomes to demonstrate that the majority of converted refusals go on to participate in the subsequent wave.

Key words: non-response: fieldwork intervention: longitudinal study: treatment effects: Millennium Cohort Study

Acknowledgements

The experimental intervention reported in this paper was carried out as part of a research project entitled ‘Predicting and Preventing Non-Response in Cohort Studies’ (Ref: RES-175-25-0010). This project was funded by the UK’s Economic and Social Research Council (ESRC) as part of their Survey Design and Measurement Initiative (SDMI). The UK Millennium Cohort Study (MCS) is funded by the ESRC and a consortium of UK government departments. The National Centre for Social Research (NatCen) carried out the data collection in Great Britain for the fourth wave of MCS. We would like to thank the research and operational staff who were responsible for the implementation of this experiment. We would also like to thank John Micklewright for helpful comments on earlier drafts of this manuscript.

1. Introduction

Most surveys typically devote considerable resources to maximising response rates because non-response is the main component of non-observational error in sample surveys (Groves, 1989). In the context of declining response rates over recent years (de Leeuw and de Heer, 2002), survey organisations have had to make increased efforts to maintain response rates (Stoop, 2005) and there has been a growth in methodological research into the effectiveness of fieldwork strategies to minimise non-response.

Higher response rates do not, however, necessarily imply lower levels of non-response bias. In their meta-analysis, Groves and Peytcheva (2008) point out that the extent to which higher response rates generate less bias depends in part on the correlation between the predictors of survey participation and the substantive survey variables and conclude that the level of bias can differ between surveys with similar response rates and between variables of interest within the same survey.

Refusal conversion is one of the fieldwork strategies commonly used to minimise non-response. This paper evaluates the effectiveness of refusal conversion in an experimental framework within an ongoing longitudinal study by assessing its impact on response rate and non-response bias at the implementation wave, and participation at the subsequent wave.

The next section reviews the literature in relation to refusal conversion and non-response bias. Section 3 presents the design of the experiment to convert refusals and provides details of the implementation of the experiment and the study, the UK Millennium Cohort Study, on which it was carried out. Section 4 provides results from the experiment. Section 5 concludes and reflects on the implications of the findings for fieldwork strategies in longitudinal studies.

2. Refusals, refusal conversion and non-response bias

There is a large body of theoretical and empirical literature about refusals (e.g. Groves and Couper, 1998; Groves et al., 1992) and it is well established that survey design factors such as respondent incentives (e.g. Singer, 2002; Laurie and Lynn, 2009) and advance letters (de Leeuw et al. 2007) can successfully reduce refusal rates. Refusal conversion is another well-established fieldwork strategy for reducing refusal rates (e.g. Stoop, 2004). This practice involves re-approaching a sampled person who has initially refused to take part and attempting to get them to reconsider their decision, i.e. to

‘convert’ them from a refusal to a successful interview. Conversion is usually, but not necessarily, attempted by a different and often a more experienced interviewer. The theoretical rationale for attempting refusal conversion is that a sample member’s decision to co-operate is influenced by their interaction with the interviewer. Hence sending a different interviewer will lead to a different interaction and, hopefully, a positive decision about participating. It is also based on evidence that refusals often result from the particular circumstances of the respondent at the time the interviewer called and that the same or a different interviewer calling back on a different occasion may mean that the circumstances which led to the refusal are no longer pertinent. It is well-established (and evidenced by non-monotonic or arbitrary response patterns) that refusal to participate at a particular wave of a longitudinal survey, conditional on participation at the baseline wave, does not necessarily mean that the sample member will not take part in future waves. The theoretical rationale for using more experienced interviewers for refusal conversion is that it is well known that interviewer characteristics can have a considerable impact on refusal rates (Hox and de Leeuw, 2002). Converted refusals constitute a significant minority of completed interviews in many surveys. Lynn et al. (2002) report that converted refusals accounted for between 1.2 per cent and 8 per cent of all completed interviews on six UK face-to-face surveys conducted between 1995 and 1998. Curtin et al. (2000) report that on the Survey of Consumer Attitudes, a long-running repeated cross-sectional telephone survey in the US, the proportion of interviews from refusal conversions doubled from 7.4 per cent in 1979 to 14.6 per cent in 1996.

Higher response rates increase precision but, as noted earlier, they do not necessarily result in less non-response bias. There are several examples in the literature from cross-sectional telephone surveys in the US which demonstrate that, although refusal conversion (and other extended field efforts to maximise response) can have a positive impact on response rates, there is little or no evidence that this is beneficial in terms of reducing non-response bias (Curtin et al., 2000; Keeter et al., 2000). One implication of these papers is that the additional resources devoted to response maximisation on these surveys may not be justified as they appear to bring little or no benefit in terms of bias reduction.

A similar approach to assessing the impact of refusal conversion (and other extended interviewer efforts) on non-response bias was taken by Lynn et al. (2002). They found that refusal conversion did appear to bring some benefit for the six cross-sectional

surveys they examined in terms of bias reduction for survey estimates relating to financial variables as statistically significant differences were found between those initially interviewed and converted refusals. This conclusion was not, however, replicated for health or attitude variables.

Longitudinal surveys are typically better placed to assess the impact of refusal conversion on non-response bias as, unlike cross-sectional surveys, information about most if not all sample members is available from prior waves. However, the context of refusal conversion is different for a longitudinal study compared with a cross-sectional survey as refusal conversion can take place both within and across waves of data collection i.e. refusals can be re-approached at subsequent waves as well as (or instead of) during the current wave of data collection. It is therefore important to be able to take informed decisions about when it is most cost-effective to devote resources to refusal conversion. The balance of risks and rewards in relation to refusal conversion in the current wave of data collection is different in longitudinal surveys when the objective of securing participation needs to be repeated at each wave of data collection and there will be another opportunity to convert refusals at a future wave. So, while maximising response at a particular wave is important, this short-term aim needs to be balanced against the desire not to jeopardise participation in future waves. Similarly, the long-term impact of refusal conversion on non-response bias and response rates in longitudinal surveys cannot be fully assessed until future waves of data collection have taken place.

Burton et al. (2006) evaluate the long-term effectiveness of within-wave refusal conversion procedures on a household panel survey. They use pooled data from waves 4-13 of the British Household Panel Survey (BHPS) and report that within-wave refusal conversion was attempted for around 36 per cent of refusals and 37 per cent of these were converted to a face-to-face, proxy or telephone interview giving an unconditional conversion rate of 13 per cent. They also show that the majority of converted refusals went on to participate in future waves. The paper shows that refusal conversion on the BHPS improved the representation in the sample of certain groups (such as the geographically mobile, self-employed and local authority renters) and, to the extent that these variables are correlated with other variables of interest, may be expected to reduce non-response bias. Other than this BHPS paper, there appears to be no published evidence about the effectiveness of refusal conversion attempts in longitudinal studies.

A limitation of almost all of the literature assessing the impact of refusal conversion on bias reduction is that it is based on observation of fieldwork procedures rather than randomised experimental interventions. There are only a few studies which experimentally evaluate the effectiveness of different refusal conversion techniques and these have been for telephone surveys (Basson and Chronister, 2006; Keeter et al., 2000). This is problematic because estimates of the effectiveness of refusal conversion attempts, including their impact on non-response bias, will be influenced by selection effects. In particular, there is a concern that allowing field staff discretion over which refusals to re-issue is likely to mean that a minority of cases are re-issued and the field staff are likely to re-issue cases with similar characteristics to those who have already been interviewed. Although this may result in an increase in response rate, it may not result in a reduction in bias. The randomised experiment described in the following section was specifically designed to address this limitation of the existing evidence.

3. Design and implementation of the experiment

Survey context

The experiment described in this paper was developed for and carried out on the fourth wave of the Millennium Cohort Study (MCS). The MCS, which is following over 19,000 children born in 2000/1, is one of four national birth cohort studies in the UK. The sample, which was recruited through records of those in receipt of a universal benefit paid to parents called Child Benefit, is disproportionately stratified and clustered at the level of electoral ward. The stratification was based on UK country and the characteristics of the sampled wards. In Scotland, Wales and Northern Ireland, there were two strata in each country: disadvantaged and advantaged. In England, there were three strata: minority ethnic, disadvantaged and advantaged. The probabilities of selection differed by strata. More details of the MCS sample design can be found in Plewis (2007). The data collection for the study takes place in the home and involves face-to-face interviews with multiple informants in each family. Interviews have been sought with up to two co-resident parents at every wave and, from wave two onwards, the child has also been asked to participate directly though the nature and extent of their participation has changed at each wave as the children get older. There have been five waves of the study so far: at nine months (2001-2), age three years (2003-4), age five (2006), age seven (2008) and age 11 (2012).

The MCS employs a range of response maximisation techniques, including between-wave mailings, advance letters and leaflets, and multiple call-backs. All interviewers

working on the project have received the fieldwork agency's standard training on probability surveys which covers refusal avoidance and additional project-specific training. Although interviewers are encouraged to make first contact by telephone for families who took part in the most recent wave, they are trained to withdraw if the family appears reluctant to take part on the phone and they then make a personal visit to the family at a later date. More generally, interviewers are trained not to return cases as refusals until reasonable efforts have been made to persuade the sample member to take part. These would typically include visiting the household on more than one occasion (with the exception of extremely firm refusals), making multiple attempts to re-arrange broken appointments and attempting to speak to all household members eligible for interview.

Design of the experiment

The experiment consisted of two interventions which aimed to increase the proportion of refusals who were converted to productive interviews. The first intervention (I1 in Table 1), designed to maximise the proportion of sample members for whom refusal conversion was attempted, involved the re-issuing of all refusals i.e. intensive re-issuing. The control condition for this intervention (C1) is the standard re-issuing strategy of the fieldwork agency i.e. re-issuing a non-random sub-set of refusals. In relation to interviewer allocation, the fieldwork agency's standard practice is that, where possible, re-issues should be allocated to a different and generally a more experienced interviewer. Overall, around nine in ten re-issued refusals were allocated to a different interviewer in this experiment. Interviewers were not aware that the experiment was taking place and experimental group was not taken into account in the allocation of cases to interviewers at any stage. Although interviewer-level variation in initial refusal rate (and re-issued conversion rate) is likely, this is not a concern for the experiment as there is no reason to expect this to be independently associated with experimental group.

The second intervention (I2), designed to maximise the proportion of refusals that were converted, involved sending a leaflet to families who had refused. (A copy of the leaflet is available on request from the first author.) The control condition for this intervention (C2) was an interviewer visit without any advance materials which is the fieldwork agency's standard approach for re-issues. The leaflet was designed to encourage sample members to take part by addressing reasons for refusal that are commonly reported on

the study. It was hypothesised that this leaflet would lead to higher conversion rates for refusal cases who were re-issued than those who had not received a leaflet.

As the experiment involved two interventions, each with two levels, a 2² factorial design was used as summarised in Table 1 below. The main benefit of this crossed design is that the interventions can be evaluated independently and in combination with each other. Cases were randomly assigned to one of these four groups. Further details about the randomisation process can be found in Appendix A.

Table 1: Experimental groups and interventions

<p>Group A: Intensive re-issue and leaflet I1: All refusals re-issued (to a different interviewer) I2: Leaflet sent to all refusals</p>	<p>Group C: Intensive re-issue, no leaflet I1: All refusals re-issued (to a different interviewer) C2: No leaflet sent to refusals</p>
<p>Group B: Standard re-issue and leaflet C1: Standard proportion of refusals re-issued (to a different interviewer) I2: Leaflet sent to all refusals</p>	<p>Group D: Standard re-issue, no leaflet C1: Standard proportion of refusals re-issued (to a different interviewer) C2: No leaflet sent to refusals</p>

Although the intention of the intensive re-issuing treatment was to re-issue all refusals, some refusals, in both the treatment and control groups, were classified as ‘hard refusals’ and therefore not considered for re-issuing. Refusals given by the sample member directly to the fieldwork agency (or to the office of the study’s principal investigator) rather than to the interviewer, refusals during tracking and refusals during the interview were automatically excluded from re-issuing based on their survey outcome code. As the leaflet was relatively non-invasive, these types of refusals were not excluded from the leaflet treatment. This meant that some families in groups A and B were sent a leaflet but were not considered for re-issue. However, as the leaflet was not expected to have much impact on its own, all families with these outcome codes have been treated as exclusions from the experiment. In addition, field staff also excluded some cases from both interventions as ‘hard refusals’ based on other information, usually from interviewer notes, about the nature and circumstances of the refusal. It was felt to be important that field staff retained the ability to do this to avoid interviewers re-approaching families, or sending a leaflet to families, where it would not have been appropriate to do so.

Implementation of the experiment

In total there were 1660 refusals (11% of the issued sample of 15, 350 cases). These are household rather than individual level refusals i.e. no interviews were conducted.

Typically the refusal would be given by one or both of the parents on behalf of the family. Families in which some but not all individual interviews were conducted are classified as partially productive and were not considered for re-issuing. Table 2 below shows the number of refusals in each experimental group and summarises exclusions and actual treatment by experimental group.

Table 2: Refusals, exclusions and actual treatment by experimental group

	Group A: Intensive re- issue, leaflet	Group B: Standard re- issue, leaflet	Group C: Intensive re- issue, no leaflet	Group D: Standard re- issue, no leaflet
Number of Refusals⁽¹⁾	414 (24.9%)	437 (26.3%)	389 (23.4%)	420 (25.3%)
Excluded as ‘hard’ refusal⁽²⁾	100 (24.2%)	61 (13.9%)	80 (20.6%)	67 (15.9%)
Outcome code ⁽²⁾	55 (13.3%)	56 (12.8%)	52 (13.4%)	67 (15.9%)
Field staff decision ⁽²⁾	45 (10.9%)	5 (1.1%)	28 (7.2%)	0 (0.0%)
Excluded for other reason	8	9	5	0
Number of treated refusals⁽²⁾	306 (73.9%)	367 (83.9%)	304 (78.1%)	353 (84.0%)
Type of treatment				
Re-issue and leaflet	300	4 ⁽³⁾	0	0
Re-issue only	6	0	304	3
Leaflet only	0	363	0	0
None	0	0	0	350

⁽¹⁾ Percentage of refusals (overall) ⁽²⁾ Percentage of refusals (within experimental group) ⁽³⁾ These four cases include one family which was re-issued after phoning in to request an interview on receipt of the leaflet.

Overall, refusals were roughly equally distributed by experimental group and the proportion of refusals excluded based on their outcome code was also roughly equal (between 13 and 16 per cent). The proportion of cases excluded due to field staff decisions, and therefore the overall proportion of cases excluded as ‘hard refusals’ did, however, vary by experimental group. The exclusion rate by field staff was higher in group A (11%) and group C (7%) than in groups B (1%) and D (0%). This was expected as groups A and C received the intensive re-issuing treatment and, as a result, field staff had to review all refusals eligible for re-issuing. It is highest in group A as field staff also had to review refusals which were eligible for the leaflet treatment but

not for re-issue. A small number of cases were also excluded from the experiment because the fieldwork agency was unable to supply an interviewer to work on the re-issue in the time available or because the refusal was received too late in the fieldwork period to apply the treatment. In group A, there were six cases which were re-issued without a leaflet as they were received too late for the leaflet to be sent.

Overall, the exclusions reduced the number of cases treated by 20 per cent, from 1660 to 1330. The non-random exclusion of cases from the treatment groups (Groups A, B and C) introduced an element of non-random selection into the actual treatment given and led to differential exclusion rates between groups. Although this has the potential to jeopardise the validity of the conclusions based on the initial randomisation, the standard solution to this problem is to estimate treatment effects for all cases for which there was an intention to treat as well as to estimate the effects of treatment on the treated (Shadish and Cook, 2009) i.e. compare conversion rates for all refusals as well as treated refusals.

In the groups in which the intensive re-issuing treatment was given, a very high proportion of refusals were re-issued (74% of the refusals in group A and 78% in group C). It is clear that the experimental intervention succeeded in maximising the proportion of refusals that were re-issued. A much higher proportion of refusals were reissued in these experimental groups than on previous waves of the MCS and in other comparable surveys. For example, in wave two of MCS, around 13 per cent of refusals were re-issued and on the BHPS waves 4-13, 36 per cent of refusals were re-issued (Burton et al., 2006).

However, it was surprising that in the groups in which the intensive re-issuing treatment was not given, an extremely low proportion of refusals were re-issued by the 'standard' procedure (<1% in groups B and D). This is a much lower proportion than would typically be re-issued on many large-scale surveys. Furthermore, as noted earlier, it is much lower than the proportion of refusals that were re-issued at a previous wave of MCS and on the BHPS. It is not clear why such a low proportion of refusals were re-issued under the 'standard' treatment. This appears to have been an unintended side-effect of increasing the proportion of re-issues in groups A and C. Unfortunately this means that it is not possible to make comparisons between the standard and intensive re-issuing experimental groups. For this reason, all subsequent analysis, including the result of the experiment, will omit the experimental groups with the standard re-issuing procedure i.e. B and D. The results from the experiment can still be used to evaluate the

impact of the second intervention i.e. the leaflet and it is also possible to evaluate the impact of intensive re-issuing on sample composition. We can also evaluate the impact of intensive re-issuing on participation at the subsequent wave.

4. Results

What proportion of refusals was converted to productive interviews in each of the experimental groups?

Table 3: Final survey outcome for all (treated) refusals by experimental group

	Group A: Intensive re- issue, leaflet	Group C: Intensive re- issue, no leaflet
Fully productive	54	44
Partially productive	17	24
Unproductive	235	236
Total	306	304
Productive (%) ⁽¹⁾	23.2	22.4
<i>Productive</i> (%) ⁽²⁾	<i>17.1</i>	<i>17.9</i>

⁽¹⁾ Percentage based on all treated refusals. Productive includes partially productive.

⁽²⁾ Percentage based on all refusals.

Table 3 shows the final survey outcome for all treated refusals by experimental group. Fully productive cases are those in which all individual interviews were conducted, partially productive where some but not all interviews were carried out, and unproductive cases are those in which no interviews were done. Treated refusals are defined, as explained in Table 2, as those refusals which were not excluded. Table 3 shows that in the experimental groups in which the intensive re-issuing intervention was administered (A and C) almost a quarter (23 % and 22% respectively) of treated (i.e. re-issued) refusals (and 17% and 18% of all refusals) were successfully converted to a productive interview. Compared with Burton et al. (2006), the conversion rate for treated refusals (23% and 22%) in the intensive re-issue groups is lower than the equivalent proportion reported for the BHPS (37%). However, the unconditional conversion rate i.e. the overall proportion of refusals converted (17% and 18%) is higher than the equivalent unconditional conversion rate reported by Burton et al. (2006) of 13.5 per cent.

Table 3 also shows that the leaflet had no effect on the proportion of refusals successfully converted. The expectation that the leaflet would have an additional effect in combination with re-issuing was not borne out. The expectation that the leaflet was

unlikely to have an effect in isolation was borne out with only one family responding to the leaflet to request an interview without first being re-approached by an interviewer.

4.2 What impact did re-issuing have on the overall achieved sample size and refusal rate?

In the experimental groups in which the intensive re-issuing was carried out, the refusal rate was reduced by around two percentage points (from 11% to 9% in group A and 10% to 8% in group C). This indicates that, if the intensive re-issuing treatment had been carried out on the whole sample, it would be reasonable to expect that the increase in achieved sample size would have been double what was actually observed i.e. 278 cases rather than 139.

Were converted refusals less likely than those initially interviewed to complete all of the survey elements?

The survey consisted of several different data collection elements. Families who were interviewed initially were much more likely to have completed all of the survey elements they were eligible for (i.e. to be fully productive) than families who were converted refusals (87% compared with 72%). The ratio of fully productive to partially productive families was 2.4 for converted refusals compared with 7.7 for those who did not refuse initially.

The most striking difference between the two groups was in the proportion of partner respondents who were interviewed. In productive families who were initially interviewed, almost 85 per cent of eligible partners were interviewed compared with 63 per cent among productive families who were converted refusals.

How were the refusal, re-issuing and conversion rates related to prior response history?

Prior response history is strongly associated with both initial refusal and refusal conversion rates at wave four. Families who did not take part at wave three were more likely to refuse (33% compared with 7%) and less likely to be converted to a productive interview (10% of all refusals compared with 18%) at wave four than those who did take part at wave three. Similarly, families who had taken part in all prior waves they were eligible for were less likely to refuse (6% compared with 27%) and more likely to be converted to a productive interview (20% of all refusals compared with 11%) at wave four than those who had missed at least one wave.

How were the refusal, re-issuing and conversion rates related to sample characteristics?

Table 4 shows refusal rates, conversion attempt rates and conversion success rates for all cases in experimental groups A and C who were interviewed at wave three by a range of characteristics observed at wave three. The wave three characteristics chosen have been shown to be related to attrition by Plewis et al. (2008), and to key substantive variables on MCS. They are comparable to those used by Burton et al. (2006).

The aim of the intensive re-issuing treatment was to increase the proportion of refusals which were reissued - the conversion attempt rate - with the intention that the observable demographic characteristics which are related to initial refusal would not be taken into account in the decisions about whether or not to re-issue the case. In terms of bias reduction, it is preferable for sample members with a higher propensity to refuse also to have higher conversion rates as this is likely to lead to a reduction in bias on this variable. Conversely, if the conversion rate is highest for sample members who have a low propensity to refuse this can lead to an increase in bias in the achieved sample.

Table 4: Refusal and conversion rates at wave four by wave three characteristics

	<i>Number of refusals (wave four)</i>	<i>Refusal rate⁽¹⁾</i>	<i>Conversion attempt rate⁽²⁾</i>	<i>Conditional conversion rate⁽³⁾</i>	<i>Unconditional conversion rate⁽⁴⁾</i>
	<i>N</i>	<i>%</i>	<i>%</i>	<i>%</i>	<i>%</i>
	531	6.9	72.9	25.3	18.4
<i>Main respondent's educational qualifications</i>					
No qualifications	108	13.5	75.3	16.4	12.4
Level 1	49	8.4	83.1	28.1	23.3
Level 2	153	6.9	78.9	28.5	22.5
Level 3	66	5.8	71.6	20.0	14.3
Level 4	111	5.3	62.5	32.2	20.1
Level 5	26	5.6	61.8	27.3	16.9
Overseas qualifications only	17	6.6	82.8	8.4	6.9
<i>F-statistic</i>		<i>8.64</i>	<i>1.89</i>	<i>1.16</i>	<i>0.97</i>
<i>p-value</i>		<i><0.001</i>	<i><0.1</i>	<i>>0.3</i>	<i>>0.4</i>
<i>Whether main respondent is in work (including on leave)</i>					
Yes	255	5.9	71.2	29.8	21.2
No	273	8.2	74.5	20.9	15.5
<i>F-statistic</i>		<i>14.15</i>	<i>0.49</i>	<i>3.19</i>	<i>2.18</i>
<i>p-value</i>		<i><0.001</i>	<i>>0.4</i>	<i><0.1</i>	<i>>0.1</i>

<i>Cohort member's ethnic group</i>					
White	418	6.7	73.0	26.1	19.1
Non-white	113	8.1	72.2	20.4	14.7
<i>F-statistic</i>		2.29	0.02	0.85	0.91
<i>p-value</i>		>0.1	>0.8	>0.3	>0.3
<i>Whether main respondent voted in last general election</i>					
Yes	255	5.3	64.5	25.0	16.1
No	262	9.1	80.2	26.6	21.2
<i>F-statistic</i>		28.42	9.44	0.07	1.59
<i>p-value</i>		<0.001	<0.003	>0.7	>0.2
<i>Main respondent's general health</i>					
Excellent/Very good	286	6.3	70.9	26.3	18.7
Good	145	6.6	70.4	24.1	17.0
Fair/Poor	88	8.9	82.0	26.4	21.7
<i>F-statistic</i>		3.26	1.77	0.07	0.29
<i>p-value</i>		<0.1	>0.1	>0.9	>0.7
<i>Family Type</i>					
Married or cohabiting natural parents	378	6.4	69.7	27.0	18.8
Lone natural mother	111	8.1	84.5	21.1	17.9
Other family type	42	9.2	72.3	23.1	16.7
<i>F-statistic</i>		3.28	3.95	0.49	0.06
<i>p-value</i>		<0.05	<0.03	>0.6	>0.9
<i>Housing Tenure</i>					
Own	278	5.4	67.0	26.7	17.9
Rent/Other	240	9.7	80.4	24.7	19.9
<i>F-statistic</i>		44.94	10.6	0.13	0.24
<i>p-value</i>		<0.001	<0.002	>0.7	>0.6

Notes:

- (1) This is based on the total sample responding at wave three in groups A and C (n = 6,806).
- (2) The conversion attempt rate is the proportion of refusals which are re-issued.
- (3) The conditional conversion rate is the proportion of re-issued refusals which are interviewed.
- (3) The unconditional conversion rate is the proportion of all refusals which are interviewed.
- (4) Design-based F tests (degrees of freedom omitted) were used to test the null hypothesis of no relationship between each of the variables and the refusal rate, conversion attempt rate, conditional conversion rate and unconditional conversion rate. The analysis was carried out using the *svy* commands in STATA to adjust for the sample design.

Table 4 shows, as Plewis et al. (2008) did for wave two, that the main respondent's education level, employment status, health, voting behaviour as well as family type and housing tenure are all statistically significantly related to the refusal rate. Refusals were more common among those with no or lower (levels 1 and 2) educational qualifications, those who were not in work, in fair or poor health, those who did not vote in the last election, lone parent and other family types and those in rented or other tenure accommodation. The refusal rate did not vary significantly with the ethnic group of the cohort member suggesting a relative lack of bias in this variable.

The table also shows that for all of the characteristics associated with refusal rates there were no statistically significant differences in the unconditional conversion rate. This clearly shows that, in relation to these characteristics, comparable conversion rates can be achieved among cases with high initial refusal rates as among those with low initial refusal rates and indicates that re-issuing may have led to a reduction in bias in these variables.

Overall, the evidence from Table 4 lends support to the fieldwork strategy of re-issuing a high proportion of refusals, including groups with high refusal rates, as it is clear that high conversion rates can be achieved among some of these groups.

What impact did re-issuing have on bias in the achieved sample?

The discussion in the previous section provides some indicative evidence on the impact that re-issuing may have had on bias by examining the correlates of refusal rates and conversion rates. This section aims to assess this more directly by comparing the distributions of those initially interviewed, converted refusals and unconverted refusals on the same wave three characteristics as reported in Table 4. Clearly, this is a relative assessment of bias i.e. how much bias there is at wave four compared with wave three and does not account for any bias existing at wave three. As for Table 4, this analysis is confined to cases in experimental groups A and C who were interviewed at wave three.

The approach taken to assessing bias reduction is similar to the approach taken by Burton et al. (2006) and involves comparing the characteristics of both converted refusals and unconverted refusals with initially interviewed cases. If (i) the distributions for converted refusals and the initially interviewed are different and the distributions of unconverted refusals and the initially interviewed are the same or (ii) the distributions of both converted and unconverted refusals are different from the distribution for those initially interviewed, this indicates that there is potential bias in the survey estimates for

this variable and that the refusal conversion attempts are likely to have led to a reduction in this bias. Conversely, if the distributions of converted refusals and the initially interviewed are similar, and the distributions of unconverted refusals and the initially interviewed are different, this indicates that there is bias in the survey estimates for this variable and that refusal conversion attempts are unlikely to have led to a reduction in this bias. If the distributions of both converted and unconverted refusals are similar to the distribution of the initially interviewed, this indicates that there is little or no bias in the survey estimates for this variable.

Bias reduction was assessed by a series of binary, unordered and ordered logistic regression models with the characteristics of interest as the dependent variables and the three interview outcomes handled as dummy explanatory variables with initially interviewed as the reference category. The results are given in Table 5 as estimated parameters from the models along with their standard errors.

Table 5: Wave three characteristics by outcome at wave four

	<i>Initially interviewed vs. Converted refusals</i>	<i>Initially interviewed vs. Unconverted refusals</i>
	Estimate (s.e.)	Estimate (s.e.)
<i>Main respondent's educational qualifications (unordered)</i>		
No qualifications	0.21 (0.37)	0.89 (0.16)
Level 1	0.26 (0.45)	0.24 (0.20)
Level 2 (ref.)	-	-
Level 3	-0.66 (0.39)	-0.15 (0.18)
Level 4	-0.42 (0.32)	-0.31 (0.15)
Level 5	-0.54 (0.58)	-0.12 (0.25)
Overseas qualifications only	-1.18 (0.68)	0.32 (0.32)
<i>F-statistic</i>	<i>F(6,328)=1.56</i>	<i>F(6,328)=9.73</i>
<i>p-value</i>	<i>>0.1</i>	<i>p<0.001</i>
<i>Whether main respondent is in work (including on leave)</i>		
Yes (ref.)	-	-
No	0.07 (0.23)	0.45 (0.10)
<i>t-statistic</i>	<i>0.31</i>	<i>4.40</i>
<i>p-value</i>	<i>>0.7</i>	<i><0.001</i>
<i>Cohort member's ethnic group</i>		
White (ref.)	-	-
Non-white	0.007 (0.31)	0.38 (0.14)
<i>t-statistic</i>	<i>0.02</i>	<i>2.73</i>
<i>p-value</i>	<i>>0.9</i>	<i><0.01</i>

Whether main respondent voted in last general election		
Yes (ref.)	-	-
No	0.89 (0.24)	0.58 (0.12)
<i>t</i> -statistic	3.70	4.92
<i>p</i> -value	<0.001	<0.001
Main respondent's general health (ordered)	0.26 (0.24)	0.16 (0.11)
<i>t</i> -statistic	1.07	1.43
<i>p</i> -value	>0.2	>0.1
Family Type		
Married or cohabiting natural parents (ref.)	-	-
Lone natural mother	0.27 (0.28)	0.40 (0.14)
Other family type	0.33 (0.46)	0.56 (0.21)
<i>F</i> -statistic	$F(2,332)=0.65$	$F(2,332)=6.34$
<i>p</i> -value	>0.5	<0.01
Housing Tenure		
Own (ref.)	-	-
Rent/Other	0.79 (0.24)	0.65 (0.10)
<i>t</i> -statistic	3.28	6.26
<i>p</i> -value	<0.01	<0.001

Note: Models fitted using *svy mlogit*, *svy logit* and *svy ologit* in STATA for the unordered, binary and ordered dependent variables.

Table 5 shows that the results are mixed and differ between variables. For employment status, education level, ethnic group and family type, the distributions of unconverted refusals are significantly different from the initially interviewed cases but there was no significant difference between converted refusals and initially interviewed cases for these variables. This indicates that there is bias in this variable at wave four which refusal conversion has done little or nothing to remove. For health, the distributions for converted refusals and unconverted refusals were not significantly different from the distribution for the initially interviewed cases. This indicates that there is little or no bias in this variable at wave four. However, for voting and housing tenure, there is evidence that refusal conversion has led to a removal of bias. Both converted refusals and unconverted refusals have significantly different distributions from the initially interviewed cases and similar distributions to each other: tests of the equality of the coefficients in the two columns of estimates in Table 5 gave *p*-values > 0.2 and > 0.6 for voting and tenure. This indicates that the refusal conversion process has removed some bias in these variables.

Overall, these results show that refusal conversion has led to a statistically significant reduction in bias in the survey estimates for two of the seven characteristics considered i.e. voting and housing tenure. This is supported by the evidence in the previous section (4.5) showing that for these variables the groups with the highest refusal rates, i.e. non-voters and those living in rented accommodation, also had high conversion rates. Although the refusal rate did not vary significantly with ethnic group, the analysis in this section shows that the converted refusals were similar to the initially interviewed and unconverted refusals were different from the initially interviewed for this variable. This indicates that the refusal conversion process may have introduced bias for this variable. For health, employment status and education level, there was indicative evidence in the previous section that refusal conversion may have led to a reduction in bias as the groups with higher refusal rates also had high conversion rates. However, the analysis in this section reveals that this was not the case for health, as there was no difference in the distributions for both converted refusals and unconverted refusals and the initially interviewed indicating a lack of bias in the variable, or for employment status or education level, as the converted refusals were similar to those initially interviewed and unconverted refusals were different from those initially interviewed.

Burton et al. (2006) also consider the impact of refusal conversion in relation to some similar variables: employment status, housing tenure, health and political preference. In relation to housing tenure, they also found that refusal conversion led to a reduction in bias. They found that refusal conversion led to a reduction in bias in relation to employment status, which we do not find. However, we use a binary indicator of whether the sample member is in work or not and they use a more detailed employment status variable which also distinguishes self-employed and retired. They find no clear pattern in relation to health which is not inconsistent with our finding of a lack of bias for this variable. We find evidence of bias reduction in relation to voting behaviour and they do not find any evidence of this in relation to political preference, although again the variables used are not directly comparable. It should also be borne in mind that BHPS is a study of all households whereas MCS is a study of families with young children.

What impact did re-issuing have on participation at the subsequent wave?

In order to assess the longer-term effectiveness of refusal conversion we examined the fieldwork outcomes at wave five (in 2012) for both converted refusals and unconverted refusals (for groups A and C only) in wave four (in 2008).

Table 6: Wave five outcome by wave four re-issue outcome

Wave five outcome	Converted Refusals (Wave four)			Unconverted Refusals (Wave four)		
	Group A: Intensive re-issue, leaflet	Group C: Intensive re-issue, no leaflet	Total (Group A and C combined)	Group A: Intensive re-issue, leaflet	Group C: Intensive re-issue, no leaflet	Total (Group A and C combined)
Productive	48	37	85	54	68	122
Unproductive	20	28	48	125	116	241
Not issued	3	3	6	56	52	108
<i>Productive (%)</i>	<i>67.6</i>	<i>54.4</i>	<i>61.2</i>	<i>23.0</i>	<i>28.8</i>	<i>25.9</i>

The results, in Table 6 above, show that around 61 per cent of the converted refusals took part again at wave five compared with around 26 per cent of unconverted refusals. This demonstrates that, for the majority of converted refusals, this was not just a short-lived effect. However, it also shows that a significant minority of unconverted refusals also went on to take part again at the next wave. Burton et al. (2006) show that a similar proportion of converted refusals took part at the subsequent wave; around 60 per cent for refusals converted at waves 4-7, dropping to around 40 per cent for waves 8-9 and around 30 per cent for waves 10-13.

Comparing the wave five outcomes for converted refusals by experimental group shows that converted refusals in Group A were more likely to take part at wave five than converted refusals in Group C (68 per cent compared with 54 per cent). Although the sample sizes are small, one possible interpretation of this difference is that the leaflet which was given to cases in Group A at wave four may have had an impact in the longer-term and led to higher participation rates among this group at the subsequent wave. This may indicate that sending tailored between-wave mailings to prior wave refusals may help to boost participation at subsequent waves.

5. Conclusions

In this paper we have clearly shown that devoting additional field resources to converting refusals on the fourth wave of the MCS brought some benefits to the study in terms of an increased achieved sample size and a lower refusal rate. It also implies that if intensive re-issuing had been applied to the study as a whole, rather than being carried out on half of the sample only, the magnitude of these benefits would be roughly

doubled. Despite the small number of converted refusals, there is also some evidence that re-issuing refusals may have led to a reduction in bias in the cross-sectional survey estimates on a few key variables. There was also a longer-term benefit, as the majority of the converted refusals went on to take part at the subsequent wave.

We have made two important contributions to the methodological evidence on refusal conversion. Firstly, by providing evidence from an experiment, we have addressed a limitation in the existing research in this area which was largely based on observational studies. Secondly, we have also made a significant contribution to improving knowledge in an area of methodological research in which there is relatively little existing evidence i.e. the immediate and longer-term effectiveness of refusal conversion strategies in longitudinal surveys.

As well as showing the important role refusal conversion can play in response maximisation and bias reduction, this paper has also brought into focus the fact that these two aims may sometimes conflict and that the approach taken to re-issuing may differ depending on which of these two aims survey managers wish to prioritise.

If the focus is solely on response maximisation, the aim of refusal re-issuing should be to maximise the unconditional conversion rate in the most cost-effective way. As demonstrated in section 4, the unconditional conversion rate is a product of the proportion of refusals that are re-issued and the conversion rate among re-issued refusals. It is clear that there is an inverse relationship between the proportion of refusals that are re-issued and the proportion that will be converted. In general, field staff tend to re-issue a relatively low proportion of refusals and to re-issue cases which they feel are most likely to be converted. This tends to be because there is a relatively high fixed cost associated with re-issuing cases, regardless of the outcome. However, this paper has clearly shown that it is possible to achieve relatively high conversion rates for groups with high refusal rates. The clear implication of this finding is that, where response maximisation is the objective, decisions about which cases to re-issue for refusal conversion should be based on empirical data on conversion rates and information about the marginal cost of re-issuing in order to find the optimal balance between maximising the unconditional conversion rate and minimising the field costs associated with re-issuing.

However, if the main objective of refusal re-issuing is to minimise response bias, the optimal strategy will be to re-issue higher proportions of refusals from groups who are

known to be under-represented in the achieved sample (which will usually be those with higher refusal rates) and lower proportions of refusals from groups who are known to be over-represented in the achieved sample (which will usually be those with lower refusal rates). In practice, this is likely to mean that refusals from groups with lower than average refusal rates should not be re-issued (even though there may be a relatively high chance that they will be converted to a successful interview). This paper has shown that some groups with higher than average refusal rates also have higher than average conversion rates i.e. it can be worthwhile to re-issue these cases. It is also clear that it is the conversion of these refusals which has the strongest impact on bias reduction (and conversely that, for some variables, the re-issuing of cases from groups with low refusal rates may actually serve to increase bias). As well as reducing the proportion of refusals re-issued among groups that are over-represented in the achieved sample, it is also necessary to increase the proportion of re-issued refusals that are converted to an interview among groups with high refusal rates and low conversion rates. It is clear from the experimental results that the leaflet did not have an effect on the conversion rate. However, it may be that enhancing re-issuing procedures through the production of letters and leaflets which are tailored specifically to the groups with high refusal rates and low conversion rates would be a cost-effective way to improve conversion rates among these groups.

As noted by Groves and Peytcheva (2008), the level of bias can differ between variables of interest in the same survey so, although it is generally acknowledged that focusing on bias reduction may be methodologically preferable to focusing on response maximisation, it can be difficult to do this in practice, particularly for multi-purpose studies which collect data on many different domains such as the MCS.

It is clear that refusal reissuing is an expensive way to obtain an interview. Data from the fieldwork agency shows that the cost of achieving an interview was over three times as high for converted refusals as it was for families who did not refuse initially.

However, it should also be borne in mind that this cost was incurred for a very small proportion of the achieved sample and a small number of cases overall. So, in absolute terms, the additional total cost of the extra interviews achieved via refusal conversion is small compared with the fieldwork costs for the study as a whole. More generally, targeted interventions such as refusal conversion are likely to be more cost-effective than universal ones, such as incentives given to all respondents, as resources are not 'wasted' on sample members who would participate without this intervention. It should

also be noted that the approach taken for the experiment i.e. re-issuing all refusals, is expensive as there is a high fixed cost associated with re-issuing a case regardless of outcome and, as discussed earlier, this approach may not be the most cost-effective way either to maximise response or to reduce bias. For this reason, it is expected that a more refined approach to refusal re-issuing, which takes into account all available information, may lead to a reduction in the cost of an interview achieved through refusal re-issuing.

In addition, for a longitudinal study such as MCS, the evaluation of the cost-effectiveness of refusal conversion should also take into account whether or not converted refusals continue to participate more than unconverted refusals at later waves of the study. Longitudinal studies are concerned about response maximisation over the long-term rather than at only one point in time, and about bias reduction in estimates of change rather than in cross-sectional estimates. Decisions about resource allocation should also be taken with a long-term perspective. This implies that decisions about refusal conversion on a particular wave should, if possible, be informed by evidence about the likely impact on survey costs, response rates and bias at future waves.

References

- Basson, D. and Chronister, M. (2006). Recordings of prior refusals: do they improve later conversion attempts?. Presented at Methodology of Longitudinal Surveys conference, Essex, England, July 2006.
- Burton, J., Laurie, H. and Lynn, P. (2006). The long-term effectiveness of refusal conversion procedures on longitudinal surveys. *Journal of the Royal Statistical Society Series A*, Royal Statistical Society, 169(3), 459-478.
- Curtin, R., Presser, S. and Singer, E. (2000). The effects of response rate changes on the Index of Consumer Sentiment. *Public Opinion Quarterly*, 64, 413–428.
- Groves, R.M. (1989). *Survey Errors and Survey Costs*. New York: Wiley.
- Groves, R.M., Cialdini, R.B. and Couper, M.P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly*, 56, 475-495.
- Groves, R.M. and Couper, M.P. (1998). *Nonresponse in Household Interview Surveys*. New York: Wiley.
- Groves, R. M. and Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis, *Public Opinion Quarterly*, 72(2), 167-189.
- Hox, J. and de Leeuw, E. (2002). The Influence of Interviewers' Attitude and Behavior on Household Survey Nonresponse: An International Comparison. In *Survey Nonresponse*, eds. R.M. Groves, D.A. Dillman, J.L Eltinge and R. J. A. Little , New York: Wiley.
- Keeter, S., Miller, C., Hohut, A., Groves, R.M., and Presser, S. (2000). Consequences of Reducing Nonresponse in a National Telephone Survey. *Public Opinion Quarterly*, 64, 125-148.
- Laurie, H. and Lynn, P. (2009). The Use of Respondent Incentives on Longitudinal Surveys. In *Methodology of Longitudinal Surveys*, ed. P.Lynn, Chichester: Wiley.

de Leeuw, E. and de Heer, W. (2002). Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. In *Survey Nonresponse*, eds. R.M. Groves, D.A. Dillman, J.L Eltinge and R. J. A. Little , New York: Wiley.

de Leeuw, E., Callegaro, M., Hox, J., Korendijk, E. and Lensvelt-Mulders, G. (2007). The Influence of Advance Letters on Response in Telephone Surveys: A Meta-Analysis. *Public Opinion Quarterly*, 71(3), 413-443.

Lynn, P., Clarke, P., Martin, J. and Sturgis, P. (2002). The effects of extended interviewer efforts on nonresponse bias. In *Survey Nonresponse*, eds. R.M. Groves, D.A. Dillman, J.L Eltinge and R. J. A. Little , New York: Wiley.

Plewis, I. (ed.) (2007). *Millennium Cohort Study: Technical Report on Sampling* (4th edition). London: Centre for Longitudinal Studies.

Plewis, I., Ketende, S.C., Joshi, H. and Hughes, G. (2008). The Contribution of Residential Mobility to Sample Loss in a British Birth Cohort Study: Evidence from the first two waves of the Millennium Cohort Study. *Journal of Official Statistics*, 24, 364-385.

Shadish, W. R. and Cook, T.D. (2009). The Renaissance of Field Experimentation in Evaluating Interventions. *Annual Review of Psychology*, 60, 607-629.

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

Stoop, I.A.L. (2004). Surveying Nonrespondents. *Field Methods*, 16, 23-54.

Stoop, I. A. L. (2005). *The Hunt for the Last Respondent: Nonresponse in sample surveys*. The Hague: Social and Cultural Planning Office.

Appendix A: Randomisation

The random assignment of cases took place prior to the start of the wave four fieldwork. All 19,244 cases in the study were randomly assigned to one of the four treatment groups. The decision to carry out this assignment in advance was taken primarily for operational reasons as the fieldwork agency believed that this would enable the experimental treatment to be applied more efficiently and accurately. Prior to the randomisation, the cases were stratified by two variables which were expected to be related to participation – (i) the nine original sampling strata and (ii) whether or not the family participated in the third wave of MCS – in order to ensure that these variables were distributed equally in each of the experimental groups. In addition, randomisation took place within each of the two fieldwork phases within each country. Implicit stratification (the cases ordered by the values of the selected variables) rather than explicit stratification (the cases split into groups) was used. After the cases were ordered by the stratification variables, a random start was generated and cases were allocated sequentially to each of the four experimental groups. Not all of the 19,244 cases were issued to the field for the fourth wave of the study. Cases which were known to be ineligible (due to death or emigration), permanent refusals and permanently untraced were excluded from the issued sample. However, the issued sample for the fourth wave did include cases which had not participated at wave two and/or wave three due to non-location, non-contact or non-co-operation.

Table A1 below shows the number of issued cases in Great Britain by experimental group along with descriptive statistics for the stratification variables and a selection of background variables – lone parent status, non-white ethnic groups and mother's education – which have been shown by Plewis et al. (2008) to be predictors of attrition in MCS. As expected, families are distributed roughly equally by experimental group and the distribution of these stratification and background variables is roughly the same for each experimental group.

The MCS is a UK wide study and the experiment was intended to cover all four countries in the UK. However, data collection in Northern Ireland was sub-contracted to a different fieldwork agency. Despite extensive efforts by the agency in Great Britain to ensure the same procedures for the experiment were followed in Northern Ireland, the experiment was inadequately implemented there. Hence, Northern Ireland has been excluded from the analyses presented in this paper.

Table 1A: Stratification and background variables from wave three by experimental group for the issued sample in Great Britain for wave four

	Group A: Intensive re- issue, leaflet	Group B: Standard re- issue, leaflet	Group C: Intensive re- issue, no leaflet	Group D: Standard re- issue, no leaflet
Experimental Group⁽¹⁾	3,824 (24.9%)	3,811 (24.8%)	3,840 (25.0%)	3,875 (25.1%)
Original Stratum				
England- advantaged	1085 (28.4%)	1083 (28.4%)	1088 (28.3%)	1095 (28.3%)
England- disadvantaged	1050 (27.5%)	1054 (27.7%)	1069 (27.8%)	1084 (28.0%)
England-ethnic	562 (14.7%)	557 (14.6%)	559 (14.6%)	567 (14.6%)
Wales- advantaged	187 (4.9%)	181 (4.7%)	179 (4.7%)	188 (4.9%)
Wales- disadvantaged	429 (11.2%)	425 (11.2%)	426 (11.1%)	420 (10.8%)
Scotland- advantaged	245 (6.4%)	245 (6.4%)	248 (6.5%)	254 (6.6%)
Scotland- disadvantaged	262 (6.9%)	260 (6.8%)	264 (6.9%)	262 (6.8%)
Northern Ireland- advantaged ⁽²⁾	1 (0.0%)	2 (0.1%)	4 (0.1%)	2 (0.1%)
Northern Ireland- disadvantaged ⁽²⁾	3 (0.1%)	4 (0.1%)	3 (0.1%)	3 (0.1%)
Refusal at wave three	162 (4.2%)	159 (4.2%)	164 (4.3%)	185 (4.8%)
Wave four Survey Phase				
England, P1	1343 (35.1%)	1337 (35.1%)	1349 (35.1%)	1376 (35.5%)
England, P2	1363 (35.6%)	1369 (35.9%)	1372 (35.7%)	1383 (35.7%)
Scotland, P1	123 (3.2%)	134 (3.5%)	129 (3.4%)	118 (3.3%)
Scotland, P2	384 (10%)	367 (9.6%)	383 (10.0%)	394 (10.2%)
Wales, P1	314 (8.2%)	305 (8.0%)	307 (8.0%)	305 (7.9%)
Wales, P2	294 (7.7%)	294 (7.7%)	297 (7.7%)	297 (7.7%)
Northern	2	5	2	1

Ireland, P1 ⁽³⁾	(0.1%)	(0.1%)	(0.1%)	(0.0%)
Northern Ireland, P2 ⁽³⁾	1 (0.0%)	0 (0.0%)	1 (0.0%)	1 (0.0%)
Lone natural mother at baseline	655 (17.2%)	650 (17.1%)	648 (16.9%)	672 (17.4%)
Cohort child from non-white ethnic group	748 (19.7%)	748 (19.8%)	735 (19.3%)	737 (19.1%)
Mother has no educational qualifications	564 (15.4%)	571 (15.6%)	594 (16.2%)	594 (16.1%)

⁽¹⁾ Percentage of all issued cases; all other percentages are percentage within experimental group.

⁽²⁾ These are families who were sampled in Northern Ireland and subsequently moved to Great Britain.

⁽³⁾ These are families who were living in Northern Ireland at the time fieldwork phases were assigned but subsequently moved to Great Britain.

Paper 2:

Calderwood, L. (2014). Improving between-wave mailings on longitudinal surveys: a randomised experiment on the UK Millennium Cohort Study. *Survey Research Methods*. 8 (2), pp. 99-108.

Abstract

One of the most commonly used tracking procedures on longitudinal surveys is a between-wave mailing to keep in touch with sample members. This is a prospective tracking method designed to minimise sample loss through failure to locate sample members. This paper reports the results from a randomised experiment to improve the effectiveness of the between-wave mailing on the Millennium Cohort Study, a large-scale birth cohort study in the UK. Our experimental intervention, which involved revising the content of the covering letters used in the 2010 mailing, aimed to increase the proportion of sample members responding to the mailing, particularly among lower educated sample members and minority ethnic groups who have higher attrition rates. The re-design involved making the letter easier to read, due to a concern that poor literacy or English may be a barrier to returning the form for these groups, and changing the style and signatory, motivated by the psychological concepts of reciprocity, liking and helping tendencies. Our main findings, that the design of the covering letter has a minimal impact on the overall return rate of these mailings, but that re-design of the letter can positively influence the return rate for particular sub-groups, will help guide further research in this area and help to inform practice on longitudinal surveys. The results also make an important contribution to the existing evidence on the content of advance letters more generally and thereby have broader applicability for survey research and practice.

Key words: longitudinal; tracking; nonresponse; attrition; survey methods; between-wave mailing; covering letters; advance letters

Acknowledgements

The UK Millennium Cohort Study (MCS) is funded by the Economic and Social Research Council (ESRC) and a consortium of UK government departments and run by the Centre for Longitudinal Studies (CLS). This research was carried out as part the Resource Centre funding for CLS from the ESRC (Ref: RES-579-47-001). I would like to thank the Cohort Maintenance Team at CLS for their input into the design of the experiment reported in this paper and Professor John Micklewright for helpful comments on drafts of this manuscript.

1. Introduction

Attrition is an important concern for longitudinal surveys as it can lead to biased estimates if sample members who drop out over time are systematically different to those who remain.

Failure to locate sample members who move is a major component of attrition.

Although most large-scale longitudinal surveys have developed highly successful tracking procedures for locating movers, there is relatively little evidence on the relative success, and cost-effectiveness, of different tracking methods (Couper and Ofstedal, 2009).

Between-wave mailings, designed primarily to prevent loss of contact by keeping contact details up to date, are one of the most commonly used tracking procedures. This paper reports results from a randomised experiment to improve the effectiveness of the between-wave mailing on the Millennium Cohort Study, a large-scale birth cohort study in the UK.

Our main findings, that the design of the covering letter has a minimal impact on the overall return rate, but that it can positively influence the return rate for particular sub-groups, will help guide further research and inform practice on longitudinal surveys. The results also make an important contribution to the existing evidence on the content of advance letters more generally and thereby have broader applicability for survey research and practice.

2. Background

The problem of locating sample members in longitudinal surveys is related to an individual's propensity to move and, conditional on moving, to be located. Couper and Ofstedal (2009) show that the main factors affecting propensity to move are person-level and societal-level factors outside the control of the survey, the propensity to be located can be influenced by survey design factors, such as the interval between waves and the tracking procedures used.

Tracking procedures can be either retrospective, which involves trying to find sample members known to have moved, or prospective, which aims to prevent loss of contact with sample members by ensuring that contact information is updated frequently. Between-wave mailings are a widely-used, prospective tracking method and their

effectiveness in minimising attrition at future waves is well-established. Previous research has shown that sample members who respond to these mailings are more likely to take part at subsequent waves (Laurie et al, 1999) and that fewer fieldwork resources are required to achieve an interview with them (Couper and Ofstedal, 2009).

The focus of this paper is on the optimal design of between-wave mailings. There is variation between surveys in their practice, and until recently, little or no methodological literature to inform the choice of design. This gap has recently been addressed by randomised experiments carried out on between-wave mailings using the Panel Study for Income Dynamics (PSID) in the US (McGonagle et al., 2011, 2013) and using the British Household Panel Survey (BHPS) (Fumagalli et al., 2013). These studies provide evidence on many important design considerations including the number and spacing of mailings and how this relates to the interval between waves, whether an address confirmation or change of address approach is more effective, the design of the change of address form, the use of reminders and incentives to encourage returns and the inclusion of a newsletter or findings leaflet.

Another key feature of between-wave mailings, not included in these experiments, is the design and content of the covering letter sent as part of the mailing. Covering letters in these mailings perform a similar function to advance letters; that is they are designed to promote compliance with a request. Advance letters are standard practice on most surveys and there is strong evidence that they are associated with higher response rates (e.g. de Leeuw et al. 2007). Design principles regarding advance letters include using headed paper to demonstrate the authority of the request, explaining how sample members were chosen and the purpose of the study. However, the empirical literature on the design of advance letters is relatively limited, and recent evidence suggests that the content of letters may not make a difference to response rates (Olson et al., 2011). In their meta-analysis exploring the influence of advance letters on response in telephone surveys, de Leeuw et al. (2007) found that letters which appealed to the psychological concept of reciprocity were most effective.

The concept of reciprocity in the context of surveys was discussed by Groves, Cialdini and Couper (1992). They argued that the reciprocal nature of human relationships leads individuals to respond positively to requests from other people and that compliance with requests is more likely if it is viewed as part of a reciprocal exchange. The reciprocal

nature of this exchange is enhanced when the request is given by someone who is known to the sample member and when it emphasises how their co-operation will be of direct benefit to the person making the request. Groves et al. also discuss the importance of authority and the role of 'liking' in relation to compliance with requests, and note that liking can be related to how similar to themselves the person receiving the request perceives the person making the request. They also discuss the helping tendency of human beings which implies that sample members will be more willing to respond to a request if they are told that by doing so they will be helping someone else or other people in general. The applicability of these concepts to between-wave mailings on longitudinal studies has previously been discussed in relation to the Swiss Household Panel study (Budowski and Scherpenzeel, 2005).

The primary measure of the effectiveness of between-wave mailings is the overall proportion of returns to the mailing, either to confirm or update addresses. Longitudinal surveys which use an address confirmation approach in their between-wave mailings often use reminder mailings to boost the proportion of returns. McGonagle et al. (2011) found that on PSID sending a reminder mailing significantly increased the proportion of forms that were returned, and this is consistent with the literature on postal surveys (e.g. Dillman et al., 2009). In common with postal surveys, it is typically observed that response is highest to the initial mailing, with proportionately fewer returns at each reminder. For this reason, any increase in the overall proportion of returns is expected to have the most impact on the return rate to the initial mailing. However, there is also some experimental evidence showing that, in the context of a postal survey, measures designed to boost response led to an increased rate of early returns but not an overall increase in response rates (Taylor and Lynn, 1998). In relation to the cost-effectiveness of between-wave mailings, increasing the proportion of returns to the initial mailing would be beneficial, even if no overall increase in the return rate was achieved as this would reduce the costs associated with the reminder mailings, and hence the overall cost of the mailing.

Another important measure of the effectiveness of between-wave mailings is whether there is differential response according to observable characteristics related to future participation and survey measures. Given the positive association between responding to these mailings and taking part in the next round of the survey, increasing return rates from groups which are under-represented due to differential attrition may lead to a

reduction in bias at future waves. Fumagalli et al. (2013) experimented with a tailored leaflet design on the BHPS mailing, and found that this was effective at boosting response at the subsequent wave among the targeted groups.

3. Data and methods

Survey context

This experiment was developed for and carried out on the Millennium Cohort Study (MCS); a large-scale longitudinal survey following over 19,000 children in the UK born in 2000/1. The sample was recruited through Child Benefit records, a universal benefit paid to parents, and is disproportionately stratified and clustered at the level of electoral ward. More details of the sample design can be found in Plewis (2007). Data collection takes place in the home and involves face-to-face interviews with multiple informants in each family. There have been five waves of the study so far: at 9 months (2001-2), age 3 (2003-4), age 5 (2006), age 7 (2008) and age 11 (2012). Interviews were sought with up to two co-resident parents at every wave and, from wave two onwards, the child has also participated directly, with the nature of their participation changing as they get older.

Unlike on household panel surveys, which usually have annual or biennial interviews, the interval between waves on birth cohort studies is not fixed, rather it depends on the age of the study members. The MCS has carried out a ‘keep in touch’ mailing annually between waves. As families with children tend to have high mobility rates (ONS, 2004) and the interval between waves is often longer than on panel surveys, prospective tracking between waves is particularly important on (child) cohort studies. These mailings have typically achieved return rates of around 55%-65%, which have contributed to high wave-on-wave tracking rates (over 90 per cent) (Plewis et al., 2008; Calderwood, 2013).

However, despite the success of its tracking procedures, in common with all longitudinal studies, MCS has experienced attrition as families drop out over time. A range of socioeconomic and demographic characteristics are associated with higher attrition rates. In particular, minority ethnic groups and lower educated parents have among the highest drop-out rates (Plewis et al., 2008; Dex and Rosenberg, 2008). This is a particular concern as MCS is specifically designed to over-represent families from deprived areas, including those with lower education levels, and minority ethnic groups.

Design of the experiment

The MCS mailings usually comprise of a covering letter and feedback leaflet for parents, a feedback leaflet for the cohort child, a return form and freepost envelope. As well as the family's address and home phone number, the return form, which is pre-printed on A4 coloured paper, also includes mobile and work phone numbers, email addresses and details of contact persons. An address confirmation approach is used i.e. study members are asked to return the form whether or not their details have changed, and up to two postal reminders are sent at roughly six-week intervals. The covering letter, and reminder letters, are sent on study headed paper and the signatory is the principal investigator (referred to as 'study director').

The main objective of the randomised experiment was to increase the proportion of forms that were returned. This is referred to as the 'return rate'. We expected that the experimental treatment would have most effect on the proportion of forms returned in response to the initial mailing. This is referred to as the 'early return rate'. A secondary aim was to increase the return rate from particular groups of sample members who are known to have higher rates of attrition, as this may reduce bias at future waves.

Specifically, we wanted to boost the return rate less well educated sample members and those from minority ethnic groups.

We hypothesised that low literacy and poor English-language reading skills may mean that parents with lower levels of education and those who spoke languages other than English may have difficulty reading and understanding the covering letter, which are not provided in translation. We also felt that the relatively formal and complex language, the length of the letters and the fact that the signatory is a senior academic at professorial grade might be disengaging for sample members from these groups.

Our experimental treatment involved re-designing the content of the covering letters for both the initial and the reminder mailings. The four components of the re-design are shown in Table 1 below. The first two components – simplifying the language and reducing the length – were intended to make the letter easier to read for those with lower levels of education and who speak languages other than English at home. We hypothesised that this would lead to higher return rates from these groups. Assuming that the return rates from other groups would not be negatively affected by this

treatment, higher return rates from the targeted groups would also imply a higher return rate overall. The second two components – changing the style from formal to informal and changing the signatory from the study’s principal investigator to the study’s cohort maintenance officer - were intended to enhance the likelihood of compliance with the request to return the form for all sample members and particularly those in the targeted groups. We hypothesised that this would lead to higher return rates from all sample members and, in particular, higher return rates from those in the targeted groups. Other elements of the design and content of the covering letters - the study branding, the broad content and order - were held constant as far as possible between the control and treatment versions of the letters.

Overall, our hypotheses implied that that these changes would lead to a higher return rate overall and a higher return rate among those with lower levels of education and who speak languages other than English at home. An increase in the overall return rate was also expected to imply an increase in the early return rate. An increase in the return rate among sample members with lower levels of education and those who speak languages other than English at home would also imply a change in the composition of returners, and specifically an increase in the proportion of returners from these groups. As discussed earlier, even if no increase in the overall return rate was achieved, an increase in the early return rate would make the mailing more cost-effective and a higher return rate among the two targeted groups would make the returners more representative of the sample as a whole and potentially lead to reduced bias at future waves.

Table 1: Design of experiment

Design of letter	Experimental group	
	Control	Treatment
Simplicity: Flesch-Kincaid score (Reading Ease Age)		
- Initial Mailing	74.5 (13-14 years)	91.9 (9-10 years)
- Reminder 1	74.0 (13-14 years)	93.2 (9-10 years)
- Reminder 2	73.9 (13-14 years)	92.3 (9-10 years)
Length: Number of words (proportion)		
- Initial Mailing	473 (1)	315 (0.67)
- Reminder 1	274 (1)	133 (0.49)
- Reminder 2	191 (1)	104 (0.55)

Style	Formal	Informal
Signatory	Principal Investigator	Cohort Maintenance Officer

Language complexity was measured using the Flesch-Kincaid readability score (see e.g. www.read-able.com). It is one of the most commonly used scores and is designed to measure comprehension difficulty. The score is based on word length and sentence length. The higher the score, the easier the text is to read, and the maximum possible score is 120. Age-equivalents are given to indicate the age at which most people would be expected to read and understand the text.

Table 1 shows that the treatment group letters have a much higher readability score than the control group letters (92-92 compared with 73-74). These scores indicate that the letters in the treatment group should be easily understood by most 9-10 year olds compared with a reading ease age of 13-14 for the letters in the control group. Table 1 also shows that the treatment letters were much shorter than the control versions; between half (0.49) and two-thirds (0.67) the length and reminder letters were considerably shorter than the initial mailing letter in both groups. However, it is notable that although the treatment letter was easier to read and shorter, the control letter had a relatively low reading ease age (13-14 years) and was not particularly long (equivalent to less than one side of A4). This indicates that, even in the control group, the majority of the parents should have been able to read it easily.

The other components to the re-design – changes of signatory and style - were motivated by the psychological concepts of reciprocity, liking and helping tendencies. Both the treatment and control letters contain a request from the signatory to send the form back. However the change of signatory meant that on treatment letter this request came from the person who will actually receive and process the forms when they are returned. This is explained in the letter. The intention of this change was to enhance the reciprocal nature of the request and to strengthen the appeal to helping tendencies by emphasising the direct benefit to the signatory of the recipient returning the form. The change of style of the letter to be more informal and conversational was intended to make the recipient feel that they liked the signatory and, for most recipients, to feel that the signatory was similar to them. We felt that this change of signatory and style would

be particularly appealing to sample members with lower education levels and poorer English-language skills, who may feel put off by the formal style of the standard letter and by the standard signatory. We felt that, in the context of an ongoing longitudinal study, this change of signatory would not undermine the authority of the request as sample members are familiar with the study and would perceive requests from the study as having legitimate authority even if from a different signatory. Furthermore, as they were sent on study headed paper it was clear that they were ‘official’ letters. The full text of the covering letter is included in Appendix A. Reminder letters were similar. In order to retain the internal consistency of the treatment letter and to maximise its potential impact, we decided to integrate all four components into a single treatment letter, rather than have a number of different treatment letters each with different re-design components. We felt that the different components of the re-design would be more likely to lead to an increase in return rates in combination than in isolation and that they went well together. However, as a consequence, it is not possible to evaluate the impact of these different components independently of each other.

Implementation of the experiment

The experiment was implemented on the between-wave mailing carried out in 2010, which was the second annual mailing since the last wave of data collection in 2008. In total, 15,653 families were included in the mailing, which was sent in three separate batches during the course of the year. All families were included in the mailing except deaths, emigrants, permanent refusals and permanently or temporarily untraced cases. This included some families who had not taken part in the most recent wave but were expected to be issued for subsequent waves.

Study families were randomly assigned to an experimental group prior to the mailing. The randomisation was carried out in SPSS and involved generating a random value between 0 and 1 for each case and assigning cases to one of two groups depending on this value. The experimental treatment was applied to all cases, regardless of their prior participation history, and to the reminder mailings as well as the initial mailing. Appendix B shows the characteristics of families by experimental group to demonstrate that the randomisation was implemented robustly. Overall, there were almost 8,000 cases in each group, giving the experiment a high level of statistical power to detect differences in return rates between the groups.

Measures and analysis

The main outcome measures used to evaluate the impact of the experiment - the overall return rate and the early return rate - are taken from the tracking database used on the MCS to record the receipt of forms from the mailing.

Education and languages spoken at home are taken from the most recent wave of data collection in 2008. As MCS interviews up to two co-resident parents, we use the education of the 'main respondent' who is almost always the natural mother of the cohort member. Education is measured using qualifications. The respondent's highest academic or vocational qualifications (reported across all prior waves) are mapped to an equivalent level on a standard scale which is used in the UK for National Vocational Qualifications (NVQs). The highest level, Level 5, is equivalent to a postgraduate degree and the lowest level, Level 1, is equivalent to the minimum high school leaving qualification which is at least one General Certificate of Secondary Education (GCSE) awarded at grade D-G. Language spoken at home is a self-reported household-level measure usually given by the main respondent on behalf of the family.

In order to evaluate the impact of the experiment, we firstly examined whether there were significant differences by experimental group in the overall return rate, and in the early return rate. We then explored sub-group differences to assess whether there was evidence to support our hypothesis that treatment would lead to higher return rates among those with lower levels of education and those who spoke languages other than English at home. As the comparisons of the return rates are made for multiple groups, this increases the probability of finding a significant difference by chance and thereby making a Type 1 error. We use the Bonferroni method¹ to correct for this. This involves dividing the chosen threshold for statistical significance for each comparison according to the number of comparisons made in order that the familywise error rate for the group of comparisons remains at the chosen level of statistical significance (typically 5%). Lastly, we compared the final composition of returners in each experimental group to the overall sample in order to evaluate whether the experimental treatment led to improved representation of sub-groups with lower levels of education and who speak

¹ As recommended by Williams et al. (1999), we also tried the Benjamini and Hochberg correction method for multiple comparisons, which is less conservative than the Bonferroni method. In our case, we found that both methods led to the same conclusions regarding which sub-group differences were statistically significant.

languages other than English at home. This could potentially lead to a reduction in bias in the achieved sample at future waves.

All of the 15, 653 cases that were included in the mailing are used in the first part of the analysis examining the overall and early return rates. The sub-group analysis comparing the return rates by education and languages, and the analysis of the composition of returners, is restricted to the 13,696 cases which took part at the most recent wave in 2008. Due to stratified and clustered design of the MCS sample, the analysis is carried out using *svy* commands in stata software which makes appropriate statistical adjustments for the complex design.

4. Results

In this section we firstly examine whether there are significant differences by experimental group in the overall return rate and the early return rate. We then explore variations in the return rate by education level and languages spoken at home between experimental groups and examine the composition of those who returned their forms in each experimental group in relation to education and languages spoken.

How did the return rate vary by experimental group?

Table 2 shows that there was no statistically significant difference between the experimental groups in the proportion returning their forms (55% for the control group and 56% for the treatment group, p-value >0.3). The primary aim of the experiment, to increase the return rate, was not achieved as the treatment did not succeed in increasing the proportion of sample members returning their form. However, there was a statistically significant difference in the proportion returning their forms early. A higher proportion of those in the treatment group returned their forms early compared with the control group (34% compared with 31%, p-value <0.005).

Table 2: Return rates by experimental group

	Control %	Treatment %	p-value
Returned	54.8	55.5	>0.3
Returned early	31.2	33.7	<0.005
Sample size	7826	7827	

Note: Design-based F tests were used to test the null hypothesis of no relationship between experimental group and each of the return rates. The analysis was carried out using the *svy* commands in STATA to adjust for the sample design.

What was the relationship between the return rate and education and languages spoken at home? Did this vary by experimental group?

Table 3 shows the return rate, both overall and for early returners, by experimental group and by education and language. The secondary aim of the experiment was to increase the return rate among lower educated groups and families who do not speak English at home. As the analysis sample is restricted to families who took part in the most recent wave, Table 3 also shows the overall and early return rate for all cases in the analysis.

Table 3: Return rates by experimental group and education and language

	<i>Control</i> %	<i>Treatment</i> %	<i>Sample size</i>	<i>p-value</i>
				<u><i>Returned</i></u>
<i>Main respondent's educational qualifications</i>				
No qualifications	34.1	42.4	1,558	<i>0.000*†</i>
Overseas qualifications only	45.2	46.7	385	<i>0.787</i>
Level 1 (lowest)	46.6	48.2	951	<i>0.642</i>
Level 2	59.2	57.2	3,622	<i>0.216</i>
Level 3	63.2	61.1	2,082	<i>0.285</i>
Level 4	73.1	75.9	4,162	<i>0.036*</i>
Level 5 (highest)	75.7	77.0	930	<i>0.653</i>
<i>Languages spoken at home</i>				
English only	62.8	63.1	11,280	<i>0.693</i>
Other languages	50.8	55.4	1,876	<i>0.007*†</i>
All	61.2	62.1	13,696	<i>0.242</i>
				<u><i>Returned early</i></u>
<i>Main respondent's educational qualifications</i>				
No qualifications	15.3	22.0	1,558	<i>0.000*†</i>
Overseas qualifications only	22.9	22.3	385	<i>0.901</i>
Level 1 (lowest)	25.5	26.0	951	<i>0.850</i>
Level 2	32.7	33.3	3,622	<i>0.676</i>
Level 3	36.5	36.5	2,082	<i>0.980</i>
Level 4	43.6	50.1	4,162	<i>0.000*†</i>
Level 5 (highest)	46.0	49.6	930	<i>0.262</i>
<i>Languages spoken at home</i>				
English only	36.5	39.9	11,280	<i>0.002*†</i>
Other languages	24.4	29.2	1,876	<i>0.024*†</i>

Note: Design-based F tests were used to test the null hypothesis of no relationship between experimental group and each of the return rates within education level and language groups. The analysis was carried out using the *svy* commands in STATA to adjust for the sample design. Bonferroni corrections for multiple comparisons have been made to the critical value for statistical significance for education and languages spoke at home i.e. it has been divided by the number of comparisons (7 for education and 2 for languages spoken at home). * denotes statistically significant at the unadjusted critical value i.e. 0.05. † denotes statistically significant at the Bonferroni adjusted critical value i.e. 0.007 for education and 0.025 for languages spoken at home.

In relation to education, Table 3 shows a clear gradient in the return rate by education level for both the treatment and control group. Those in the lower educated groups were much less likely to return the form than those with higher levels of education. This pattern is apparent in both the treatment and control groups and for both overall and early return rates. However, the experimental intervention did have a positive impact on the return rates among the lower educated groups. Among those with no qualifications, the overall return rate is much higher in the treatment group than the control group: 42 per cent compared with 34 per cent overall; 8 per cent higher. This difference emerges at the initial mailing as the early return rate in the treatment group (22%) is 7 per cent higher than in the control group (15%). Both of these differences are statistically significant (p-values <0.007). This shows that the impact of the treatment was strongest at the initial mailing with the reminder mailings having relatively less impact, though the impact they do have is in the same direction, that is a higher return rate in the treatment group leading to the gap between the treatment and control group increasing by 1 per cent after the reminders. A similar pattern is also observed among those with overseas qualifications and Level 1 qualifications, though the differences between treatment and control groups are not statistically significant.

Table 3 also shows that the experimental intervention had the unintended effect of boosting return rates among the higher educated groups, particularly to the initial mailing. Those with Level 4 qualifications, which is a degree or equivalent vocational qualification, have higher early return rates in the treatment group than in the control group (50% compared with 44%, p-value <0.007). However, unlike for those with no qualifications, the gap between treatment and control groups in the early return rate narrows after the reminder mailings, and the difference is no longer statistically significant for the overall return rate. The same pattern is observed for the highest

educated group, Level 5 which is a post-graduate degree or equivalent, though these differences are not statistically significant.

In summary, there is clear evidence that the treatment has had a positive and statistically significant impact on the overall and early return rate for those with no qualifications and, unexpectedly, the early return rate for those with degree-level qualifications. It appears that the reason that the overall difference between the control and experimental group in the return rate disappears after the reminder mailings is because those with degree-level qualifications, which are the largest group, respond more favourably to the reminders in the control group than in the treatment group leading to a narrowing of the gap for this group, and for the sample as a whole. Among the no qualifications group, there remains a large and statistically significant gap in the return rate after the reminders, but no significant difference in the overall return rate is observed for the sample as a whole.

In relation to languages spoken at home, Table 3 shows that sample members who speak languages other than English at home were more likely to return the form in the treatment group than the control group. This was true both after the initial mailing only (29% compared with 24%, p-value <0.025) and after the reminder mailings (55% compared with 51%, p-value <0.025). These differences are significant and in the direction predicted by the experiment. They show that the treatment letter has led to higher return rates among those who speak a language other than English at home. Among those who speak English only at home, the early return rate is significantly higher in the treatment group compared with the control group (39% compared with 37%, p-value <0.025) but this difference disappears after the reminder mailings (63% for both treatment and control groups). This mirrors the pattern observed for the overall sample. As with education, although there is a significant difference in the overall return rate for those who speak languages other than English at home, because this group make up only a small proportion of the sample, this difference has a negligible impact on the overall return rate.

Overall, these findings show that the secondary aim of the experiment, to increase the return rate in the treatment group among those with lower levels of education and those who do not speak English at home, was achieved.

What impact do these different rates have on the composition of returners? How does this vary by experimental group?

Table 4 shows the impact of the different return rates on the composition of returners in relation to education and language spoken at home. It compares the distribution of education and language among returners to the distribution for everyone who was included in the mailing. This shows what the distribution of education and language among returners would look like if everyone returned their forms or if there was no difference in return rates by education and language. The intention is to show the extent to which those with lower education and who speak a language other than English at home are under-represented among returners and to evaluate whether the experimental treatment succeeded in improving the representation of these groups.

Table 4: Education and language for returners by experimental group compared with the overall sample

	Returners		Overall sample
	<u>Control</u> %	<u>Treatment</u> %	%
<i>Main respondent's educational qualifications</i>			
No qualifications	6.2	7.9	11.4
Overseas qualifications only	2.0	2.2	2.8
Level 1 (lowest)	5.2	5.5	7.0
Level 2	25.8	24.2	26.5
Level 3	15.7	15.0	15.2
Level 4	36.5	37.0	30.4
Level 5 (highest)	8.6	8.3	6.8
<i>Sample size</i>	4188	4248	13,690
<i>Languages spoken at home</i>			
English only	88.7	87.7	86.3
Other languages	11.3	12.3	13.7
<i>Sample size</i>	4188	4250	13,696

Note: The analysis was carried out using the *svy* commands in STATA to adjust for the sample design.

In relation to education, Table 4 shows that higher educated groups (Level 4 and 5) are over-represented and the lowest educated group (no qualifications) are under-represented among returners in both the treatment and control groups. For the lowest educated, this difference is smaller in the treatment group compared with the control

group showing that the treatment has improved the representation of the lowest educated group among the returners. For example, those with no qualifications comprise 6 per cent of returners in the control group, 8 per cent of returners in the treatment group and around 11 per cent overall.

In relation to language spoken at home, in both the treatment and control groups, those who speak English only at home are over-represented among returners and those who speak other languages at home are under-represented. The proportion of returners who speak languages other than English at home is slightly higher among the treatment group (12%) than the control group (11%) and thereby closer to the overall proportion in the sample (14%).

These differences show that the treatment has led to an improvement in the representation of those with lower educational qualifications and those who speak languages other than English among the returners. Given the positive association between responding to between-wave mailings and taking part in subsequent waves of data collection, this indicates that the treatment may help to reduce bias at future waves.

5. Discussion and conclusions

This paper reported the results from a randomised experiment on MCS, a large-scale birth cohort study in the UK, designed to explore whether the re-design of covering letters used on the between-wave mailing would lead to an increase in the overall proportion of returns, and in particular, an increase in the return rate among sample members with lower levels of education and minority ethnic groups, who have higher levels of attrition on the study. The results showed that the experimental treatment did not succeed in boosting the overall return rate. However, it did lead to an increase in the proportion of sample members returning their form without the need for a reminder, and it did succeed in increasing the return rate among the lowest educated sample members and those who speak languages other than English at home.

The increase in the proportion of forms returned in response to the initial mailing meant that the cost-effectiveness of the mailing was improved, as reminders had to be sent to fewer sample members. In the context of a large-scale study like MCS, these cost savings were relatively trivial. However, our results may have useful implications for other surveys, particularly those with limited resources for reminder mailings. This increase in the early return rate, without a corresponding increase in the overall return

rate, is also consistent with previous experimental evidence from the postal surveys literature (Taylor and Lynn, 1998). Although it is not possible to attribute the increase in the early return rate to a particular component of re-design, it is plausible to argue that the informal and reciprocal nature of the request to return the form used in the treatment letter would have greater impact on early returns (as was observed), and that the more formal approach from a more authoritative signatory used in the control letter would be more effective for reminder mailings (as was observed).

The increase in the return rate from the lowest educated sample members and those who speak languages other than English at home meant that the treatment led to an improved representation of these groups among the returners. This shows that the treatment improved the effectiveness of the mailing for these important groups, and given the positive association between responding to between-wave mailings and taking part in subsequent waves of data collection, this may lead to a reduction of bias at future waves. Although it is not possible to attribute the increase in the return rate from these groups to a particular component of re-design, it is plausible that letters which are shorter and use simpler language may increase compliance with survey requests among these groups of sample members, who are often under-represented in surveys. We therefore recommend that survey managers incorporate readability testing into the development of advance letters in order to help ensure that those with literacy and language problems are able to read and understand them. These findings also imply that different versions of letters tailored for specific sub-groups can improve compliance with survey-related requests, and that greater use of tailored content on advance letters generally and covering letters on between-wave mailings could be beneficial.

The experimental treatment also led to an increase in the early return rate among one of the more highly educated groups. This was unexpected, and is more difficult to explain. It may be that as this group are more co-operative in general with requests to participate in the survey they were more susceptible to the changes designed to promote greater compliance with requests.

Our main finding, that the re-design of the covering letters had no impact on the overall return rate, implies that return rates on these mailings are unlikely to be strongly influenced by the design of the covering letter included in the mailing. It also provides indicative evidence that the willingness of sample members to take part in surveys more

generally may not be strongly influenced by the content of advance letters, which is consistent with other recent experimental literature (Olson et al., 2011). However, it is important to bear in mind the context of this experiment when considering the implications of these findings. Crucially the control letter used in the experiment was well-designed and followed best practice guidelines. It is possible that if the control letter was poorly designed, the re-design treatment may have led to an increase in return rates. Additionally, although significant revisions were made to the treatment letter, the recipients were sample members in an established longitudinal survey who were used to receiving such letters annually. This may have meant that they were less easily influenced by these changes than respondents on other surveys may be.

References

Budowski, M. and Scherpenzeel, A. (2005). Encouraging and Maintaining Participation in Household Surveys: The Case of the Swiss Household Panel. *ZUMA-Nachrichten*, 56(29), 10-36.

Calderwood, L. (2013). The role of respondent characteristics in tracking on longitudinal surveys: evidence from the UK Millennium Cohort Study. *Lifecourse and Longitudinal Studies*, 4(2), 105-118.

Couper, M.P. and Ofstedal, M.B. (2009). Keeping in Contact with Mobile Sample Members. In *Methodology of Longitudinal Surveys* (ed. P.Lynn), pp. 183-203. Chichester (UK): John Wiley & Sons, Inc.

de Leuw, E., Callegaro, M., Hox, J., Korendijk, E. and Lensvelt-Mulders, G. (2007). The influence of advance letters on response in telephone surveys: a meta-analysis. *Public Opinion Quarterly*, 71, 413-443

Dex, S. and Rosenberg, R. (2008). Ethnic minorities and nonresponse in the Millennium Cohort Study. *CLS Working Paper*, 2008/11. London: Centre for Longitudinal Studies

Dillman, D.A., Smyth, J.D., and Christian, L.M. (2009) *Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method*. Wiley: New York

Fumagalli, L., Laurie, H. and Lynn, P. (2013). Experiments with Methods to Reduce Attrition in Longitudinal Surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(2), 499-519.

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

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

McGonagle, K.A., Couper, M.P., and Schoeni, R.F. (2011). Keeping Track of Panel Members: An Experimental Test of a Between-Wave Contact Strategy. *Journal of Official Statistics*, 27(2), 319-338

McGonagle, K.A., Schoeni, R.F., and Couper, M.P. (2013). The Effects of a Between-Wave Incentive Experiment on Contact Update and Production Outcomes in a Panel Study. *Journal of Official Statistics*, 29(2), 261-276.

Office for National Statistics (2004). Census 2001 National Report for England and Wales, Part 2. London:TSO

Olson, K., Lepkowski, J.M. and Garabrant, D.H. (2011). An Experimental Examination of the Content of Persuasion Letters on Nonresponse Rates and Survey Estimates in a Nonresponse Follow-Up Study. *Survey Research Methods*, 5, 21-26

Plewis, I. (ed.) (2007). Millennium Cohort Study: Technical Report on Sampling (4th edition). London: Centre for Longitudinal Studies.

Plewis, I., Ketende, S.C., Joshi, H. and Hughes, G. (2008). The Contribution of Residential Mobility to Sample Loss in a British Birth Cohort Study: Evidence from the first two waves of the Millennium Cohort Study. *Journal of Official Statistics*, 24, 364-385.

Taylor, S. and Lynn, P. (1998). The effect of a preliminary notification letter on response to a postal survey of young people. *Journal of the Market Research Society*, 40, 165-173.

Williams, V.S.L., Jones, L.V., and Tukey, J.W. (1999). Controlling Error in Multiple Comparisons, with Examples from State-to-State Differences in Educational Achievement. *Journal of Educational and Behavioural Statistics*, 24, 42-69.

Appendix A: Full text of covering letters

Initial mailing: Control

Dear Parent or Guardian,

Your child is one of the 19,000 special children born in the UK in 2000/2001 whose lives are being followed by the Child of the New Century Study. The study is continuing to build up a unique picture of modern childhood.

Keeping your contact details up-to-date

We want to keep the contact details we have for your family up-to-date and complete so that we can get in touch with you about future surveys and send you findings from previous surveys. We have enclosed a yellow form containing your family's contact details. I would be very grateful if you could take a moment to check (and if necessary correct) your existing details, add any additional information and return the form in the Freepost envelope provided **even if all the information is correct and complete**. If you don't send back the form, we will write to you again as we won't know whether or not you've received it.

Findings from the Age 5 and Age 7 Surveys

The enclosed leaflets contain some findings from the recent surveys. The leaflet for parents looks at how children's lives have changed over the past 30 years and the leaflet for children covers some of the things they told us about themselves when they were aged 7. I hope that you and your child find them interesting. Additional copies can be downloaded from the study website (www.childnc.net).

Children of the 21st Century: the first 5 years

The second book documenting the lives of the children of the 21st century has now been published. Covering the first five years of life, it examines children's home and family backgrounds and stages of development as they start school. The authors also consider the implications of their findings for family policy and health and social services. A summary of each of the 14 chapters in the book is available on the study website.

The publishers, Policy Press, have kindly agreed to offer a 30% discount on the price of this book for study families, which means that you can buy it for £17.99 instead of £24.99. If you wish to buy a copy, please use the enclosed order form to get this special price. Alternatively, you may wish to ask your local library to purchase a copy of the book. We are sorry that we are not able to provide you with a free copy but due to the large number of families in the study we cannot afford to do this.

Thank you for the help you have given us so far with this important study. We very much hope that you will be willing to help us again in the future. The next survey will take place in 2012. We'll write to you again next year.

With kind regards,



Professor Heather Joshi OBE
Study Director

Initial mailing: Treatment

Hello again from the Child of the New Century!

My name's Peter Deane. I'm the cohort maintenance officer for the study. It's my job to try to find you if you move house and to keep our records of your name and address up-to-date. I hope that you and your family are keeping well.

Please will you check your details on the **yellow form** and send it back to me? Just write on any changes or additions you want to make. I will type them into our database when I get the form back. Even if all of your details are correct, please send the form back anyway so we know that you've checked them. If you don't send the form back, we'll send you another letter as we won't know whether or not you've got this one.

Kate Smith and Lisa Calderwood are the survey managers who work on the study. They've written the **leaflets** that we've sent to you. We hope you and your child like them! You can print off extra copies from the study website (www.childnc.net).

Professor Heather Joshi is the director of the study. Along with some other academics, she's been busy writing a **book all about your children**. It's called 'Children of the 21st Century: the first 5 years'. We can't send you a copy for free I'm afraid. We can offer a 30% discount so it costs £17.99 instead of £24.99. If you want to buy a copy, there is an order form enclosed. You can also ask your local library to order a copy. A summary of each of the 14 chapters in the book is available on the study website for you to read and print off if you want.

That's all from me for now. The next survey will be 2012 but we'll write to you again next year.

Take care!



Peter Deane

Table B1: Key variables by experimental group

	Control	Treatment
<i>Experimental group</i> ¹	7,826 (50.0%)	7,827 (50.0%)
<i>Participated at most recent wave</i> ²	6,848 (87.5%)	6,848 (87.5%)
<i>Main respondent's educational qualifications</i> ³		
No qualifications	766 (11.2%)	792 (11.6%)
Overseas qualifications only	188 (2.8%)	197 (2.9%)
Level 1 (lowest)	463 (6.8%)	488 (7.1%)
Level 2	1,825 (26.7%)	1,797 (26.3%)
Level 3	1,039 (15.2%)	1,043 (15.2%)
Level 4	2,092 (30.6%)	2,070 (30.3%)
Level 5 (highest)	474 (6.9%)	456 (6.7%)
<i>Sample size</i>	6,847	6,843
<i>Languages spoken at home</i> ³		
English only	5,918 (86.4%)	5,902 (86.2%)
Other languages	930 (13.6%)	946 (13.8%)
<i>Sample size</i>	6,848	6,848

¹ Percentage of all cases.

² Percentage of all cases within experimental group.

³ Percentage of cases which took part at most recent wave within experimental group.

Paper 3:

Calderwood, L. (2013). The role of respondent characteristics in tracking on longitudinal surveys: evidence from the UK Millennium Cohort Study. *Lifecourse and Longitudinal Studies*, 4(2), 105-118. DOI: <http://dx.DOI.org/10.14301/lcs.v4i2.231>

Abstract

Longitudinal surveys typically devote considerable resources to tracking procedures designed to minimise attrition through failure to locate sample members who move. Although these tracking procedures are often very successful, there is relatively little methodological evidence about the relative success, and cost-effectiveness, of different tracking procedures (Couper & Ofstedal, 2009). This paper extends the existing literature by exploring the effectiveness of office tracking and field tracking separately, and by examining the role of respondent characteristics as a determinant of tracking success rates. These issues are explored using the Millennium Cohort Study, a large-scale birth cohort study in the UK. The existing research on tracking procedures has been based on household panel surveys, but in the context of a birth cohort study with relatively high mobility rates among the study population and longer intervals between waves, the effectiveness of office tracking procedures is particularly important. Our main finding, that respondent characteristics are related to overall tracking success rate, implies that survey practitioners should consider ways of improving their tracking procedures for certain groups of respondents who are the least likely to be located through existing methods.

Key words:

tracking; attrition; mobility; non-response; Millennium Cohort Study

Acknowledgements:

The UK Millennium Cohort Study (MCS) is funded by the Economic and Social Research Council (ESRC) and a consortium of UK government departments, and run by the Centre for Longitudinal Studies (CLS). This research was carried out as part the Resource Centre funding for CLS from the ESRC (Ref: RES-579-47-001). I would like to thank Professor John Micklewright for helpful comments on drafts of this manuscript.

1. Introduction

As longitudinal surveys aim to follow sample members over time, they employ a range of procedures designed to minimise sample loss due to failure to locate those who move. From a scientific perspective, this is crucially important. As residential mobility is driven by social processes such as relationship and employment changes, failure to locate mobile sample members can lead to biased estimates of change in these and other important domains of substantive interest to data users.

Many longitudinal surveys have developed highly successful procedures to minimise sample loss through failure to locate. For example, the Panel Survey of Income Dynamics (PSID) and the Health and Retirement Study (HRS) successfully located 97-98 per cent of sample members who moved between the 2003-2005 and 2002-2004 waves of these US panel studies, and the German Socio-Economic Panel (GSOEP) and the British Household Panel Survey (BHPS) have tracking rates of 96 per cent and 94 per cent respectively (Couper & Ofstedal, 2009). These authors also show that in PSID and HRS, around 90 per cent of sample members who move and were located, took part in the next wave of data collection. Similarly, research on the Millennium Cohort Study (MCS) in the UK has shown that conditional on location, families who moved after wave 1 were as likely as non-mobile families to take part wave 2 (Plewis, Ketende, Joshi, & Hughes, 2008). This combination of high rates of tracking and a high likelihood that located sample members will be interviewed, means that the resources involved in tracking on longitudinal surveys are generally viewed as ‘money well spent’ by survey practitioners.

However, as Couper and Ofstedal (2009) point out, there is very little methodological evidence on the relative success, and cost-effectiveness, of different tracking procedures. They argue that the survey design features, such as the interval between waves and the tracking procedures used, are a major determinant of tracking success and as these are under the control of survey practitioners, research should focus on optimising their design and evaluating their cost-effectiveness. This has led to increasing interest in improving the effectiveness of tracking procedures and in particular, on the optimal design of between-wave mailings, with randomised experiments being carried out on these mailings on the BHPS (Fumagalli, Laurie, & Lynn, 2010), PSID (McGonagle, Couper, & Schoeni, 2011) and the MCS (Calderwood,

2012). Although this research has undoubtedly improved knowledge in relation to between-wave mailings, there has been little attention in the literature to other tracking procedures.

This paper extends the existing literature by exploring the effectiveness of tracking carried out remotely in the office prior to the start of data collection and tracking carried out by interviewers in the field during data collection, and investigating how respondent characteristics are related to tracking success. By examining both office and field tracking, we provide indicative evidence on cost-effectiveness; as office tracking is remote, it is less expensive than field tracking, and therefore increasing the proportion of movers that are located using office-based methods should lead to an improvement in cost-effectiveness. For longitudinal surveys with high mobility rates, long intervals between waves and without the resources to carry out field tracking, improving the effectiveness of office tracking will be particularly important. This paper also addresses the role of respondent characteristics as a determinant of tracking success rates. We examine how the office tracking rate, field tracking rate and overall tracking rate are related to a range of respondent characteristics which we hypothesise may be related to tracking success. If certain types of sample members are more difficult to locate than others, or more likely to be located through different tracking methods, this may have implications for survey practice and the design of tracking procedures. For example, it may be that tracking procedures should be tailored for different types of respondents. To our knowledge the relationship between the characteristics of sample members and tracking success has not been explored directly before.

We examine these different tracking success rates using data from the Millennium Cohort Study, a large-scale birth cohort study in the UK. We explore mobility over a two-year period between wave 2 (at age 3) and wave 3 (at age 5), for families who took part at wave 2. As this is a study of families with young children, the between-wave mobility rate is relatively high. Around one in five (21%) of the co-operating families had moved in this two-year period. The interval between waves on cohort studies is not fixed, rather it varies with the age of the sample member, so unlike panel surveys, which tend to have relatively short, fixed intervals between waves, cohort studies can have much longer intervals between waves. In this context, increasing the proportion of movers located using office tracking procedures is particularly important.

The next section reviews the existing evidence in this area and develops our hypotheses, the third section describes the data and methods that we use, the fourth and fifth sections present and discuss our results and the final section concludes with some implications for survey practice and recommendations for further research.

2. Background

The extensive range of tracking procedures that can be employed by longitudinal surveys is well-documented and there are several examples in the literature reporting the procedures used on different surveys. The earliest examples of published research on this topic are from the US in the 1960s and 1970s, published by the American Association of Public Opinion Research. Eckland (1968) presents tracking rates from several US longitudinal surveys carried out in the 1960s and reviews the tracking procedures used on these studies (which include postal and telephone services and directories, other public records, and employing local co-ordinators to search at ‘grass-roots’ level). Crider, Willits, & Bealer, (1971) and McAllister, Goe, & Butler, (1973) extend this literature by reviewing the tracking procedures used on a particular large-scale longitudinal survey and emphasising the need to collect extensive contact information, including full names and date of birth of study participants and the details of contact persons i.e. friends and/or relatives of sample members who may know where they are if they moved.

More recently, Laurie, Smith, & Scott, (1999) reviewed the tracking procedures used on the British Household Panel Study. They make a distinction between prospective and retrospective tracking procedures. *Prospective tracking* aims to prevent loss of contact in the event of a change of address by ensuring that multiple alternative methods of contacting sample members are collected and kept as up-to-date as possible. This includes collecting email addresses and multiple phone numbers from sample members as well as the contact details of one or more contact persons. Most surveys encourage sample members to get in touch with changes to their contact information by providing toll-free telephone numbers, email addresses and websites. ‘Keep-in-touch’ mailings are also often sent between waves of data collection to prompt sample members to confirm or update their contact details. These mailings can also lead to the discovery of a move, if they are returned to sender by the current occupiers. *Retrospective tracking* involves trying to find sample members who are known to have moved. This includes attempting to contact the new occupiers and neighbours of the sample member’s last known

address and the contact persons given by sample members. This can be done by post, telephone, email or face-to-face. Retrospective tracking can also include seeking new addresses and other contact information in public records such as electoral roles, phone and postal directories as well as administrative data sources.

Couper and Ofstedal (2009) extend this classification of tracking methods by making a further distinction between office tracking and field tracking. Office tracking is often prospective e.g. sending out between-wave mailings and processing returns, receiving updates to contact information from sample members. It can also be retrospective, involving active attempts to locate sample members who are known to have moved through e.g. post office returns from between-wave mailings. Sometimes office-based tracking can involve automated processing of large numbers of cases at the same time e.g. matching to administrative records, which is a more efficient use of resources than case-by-case review. However, often office-based tracking, for example sending emails, letters, making phone calls and searching directories, does require staff to review cases individually. This can be resource intensive, particularly for large-scale surveys. As this is done remotely, usually from a centralised location, it is less expensive than field tracking, which involves interviewers making personal visits to the last known addresses of sample members, their neighbours and contact persons. This incurs additional direct costs of travel in addition to the labour costs associated with making these tracking attempts. Field tracking is usually only carried out on longitudinal surveys which use face-to-face data collection. It is relatively uncommon for surveys which use only remote methods of data collection to use field-based tracking, though this is not unheard of, particularly for local area surveys. However, face-to-face surveys typically only carry out field tracking during data collection waves, rather than between waves.

The difference in the relative costs of field tracking compared with office tracking implies that face-to-face longitudinal surveys should aim to locate as many movers as possible using office-based methods prior to the start of data collection for a wave, in order that resources are only used on more expensive interviewer tracking in the field for sample members who cannot be located through office tracking. However, office tracking can only be carried out prior to data collection for known movers, and it is not usually possible to identify all movers prior to the start of fieldwork i.e. some sample members may not be identified as movers until the interviewer visits the issued address.

In addition, very recent moves, which take place during the data collection period, cannot, by their nature, be identified before the start of fieldwork. For this reason, office and field tracking are often carried out iteratively during the data collection period. Movers identified by interviewers during fieldwork, but for whom a new address cannot be found through field tracking, are returned for office tracking during the data collection period, and subsequently re-issued to interviewers if a new address is found.

Overall tracking success will depend largely on the range of tracking procedures adopted by the study and the amount of resources devoted to tracking. Many large-scale longitudinal surveys employ a wide range of tracking procedures, both prospectively and retrospectively and in the office as well as in the field, and for this reason have very high rates of tracking success. For example, the Panel Survey of Income Dynamics (PSID) and the Health and Retirement Study (HRS) successfully located 97-98 per cent of sample members who moved between the 2002-2004 and 2003-2005 waves of these US panel studies, and the German Socio-Economic Panel (GSOEP) and the British Household Panel Survey (BHPS) have tracking rates of 96 per cent and 94 per cent respectively (Couper and Ofstedal, 2009).

However, as well as being a function of the amount of tracking effort made by the study, tracking success is also related to the behaviours of the study members themselves. Tracking is sometimes characterised as something that is 'done to' sample members, but it is important to remember that they can influence how easy or difficult it is for the study to locate them if they move. An obvious example of this is by notifying the study team when they move. Other behaviours likely to be associated with tracking success include leaving a forwarding address with neighbours/new occupiers, responding to keep-in-touch exercises and providing multiple sources of contact information. It is plausible to hypothesise that different types of respondents, living in different circumstances, may be more or less likely to exhibit these behaviours, and hence that respondent characteristics may be related to tracking success. In general, very little attention has been given in the literature to how the characteristics of movers are related to the likelihood of successful tracking; to our knowledge the relationship between the characteristics of sample members and tracking success has not been explored directly before.

Couper and Ofstedal (2009) discuss the role of respondent characteristics in tracking success and argue that individuals with large family and social networks will be easier to locate than individual who are socially isolated. However, due to the high overall tracking rate in the HRS and PSID, they do not examine how individual characteristics are related to tracking success. Call (1990) explores how individual characteristics are related to the number of contact persons given by respondents on the National Survey of Families and Households in the US. They find that younger and older respondents, ethnic minorities and single persons provided fewer contact persons. Although they acknowledge that this may be because such individuals have fewer potential contacts to provide, rather than as a result of unwillingness, it is nevertheless likely that respondents who provided fewer contact persons will be less easy to locate if they move. The literature on between-wave mailings also shows that younger sample members (McGonagle et al., 2011), lower-educated sample members and those who speak languages other than English at home (Calderwood, 2012) are less likely to return these mailings, which also makes successful tracking less likely. We hypothesise that a range of individual, family and housing characteristics will be associated with tracking success.

In relation to individual characteristics, we expect age, ethnic group and education to be related to tracking success. Specifically, it is hypothesised that younger, non-white and lower-educated sample members are less likely to be successfully tracked. Previous research has established that younger sample members, those from non-white ethnic groups and lower-educated sample members tend to have lower response rates on many longitudinal surveys, and we would expect these characteristics to also be negatively associated with tracking success. For some minority groups, poor English language skills may also make tracking more difficult, particularly tracking through members of the extended family e.g. grandparents.

In relation to family characteristics, we expect family type and family employment situation to be associated with tracking success. We also hypothesise that lone parents will be less likely to be successfully tracked than couple families. In part, this is because lone parent families tend to have lower response rates in general than couple families and we expect this to also be related to tracking success. Additionally, for lone parent families, it is likely that less contact information will be available to use for tracking i.e. we only have contact details for one respondent and one contact person, whereas in

couple families, we collect contact details for two respondents and, for many families, two contact persons i.e. one for each parent. We expect that sample members in paid employment will be more likely to be tracked than those who are not in work. This is primarily because employment tends to be positively associated with taking part overall.

In relation to housing characteristics, we hypothesise that tenure and accommodation type will be associated with tracking success. Specifically, we expect that sample members living in rented accommodation and those living in flats will be less likely to be successfully tracked than families living in owner-occupied accommodation and in houses. Living in rented accommodation is associated with less stability in terms of residential moves, which in turn is likely to be associated with lower rates of tracking success, as renters are, in general, less likely to develop social ties with their neighbours and less likely to leave forwarding addresses for new occupiers. New tenants of rented properties may also be less likely to return mailings for previous occupants to their sender, meaning that moves may remain undiscovered for longer. This is also true of the new occupants of flats, rather than houses, particularly flats which are part of multiple-occupancy blocks. Flats can also be more difficult for interviewers to gain access to than houses, which makes it more difficult to speak to new occupants and neighbours, and is another reason to expect that tracking is less successful for sample members living in flats.

We will examine how these characteristics are associated with office tracking success, field tracking success and overall tracking success. We hypothesise that the same individual, family and housing characteristics that are associated with successful tracking overall will be associated with successful office and field tracking, and that the direction of the relationships between these characteristics will be the same for both office and field tracking. Moreover, we would expect that individual characteristics will be more strongly associated with office tracking success than field and overall tracking success, because office tracking is more dependent on the proactive behaviour of respondents than field tracking.

3. Data and Methods

The Millennium Cohort Study (MCS) is a longitudinal birth cohort study following the lives of over 19,000 children in the UK who were born in 2000 and 2001. The sample was drawn from records of recipients of a universal benefit for children, and was

initially geographically clustered by electoral ward with an over-representation of areas with high proportions of Black or Asian families, disadvantaged areas and areas in the three smaller UK countries. There have been five waves of the study so far, when the cohort member was aged 9 months, then at 3, 5, 7 and 11 years of age. At all waves, interviews were conducted with both resident parents, and from the second wave onwards, data has been collected directly from the cohort member. The study has also collected data from siblings and teachers as well as consents to link to administrative data for cohort member, parents and siblings. More information about the design of the study can be found in Plewis (2007).

The MCS employs both office and field tracking. The study provides a Freephone number, email address and a website through which cohort families can inform the study's cohort maintenance team if they change their address or contact details. Contact details for study members are updated annually between-waves through the mailing of a reply-slip which is pre-printed with all of the families' contact details i.e. address, names, phone numbers, email address and contact person details. Undelivered mail, usually indicating that the family has moved, is returned to the study by the post office, which triggers retrospective office-based tracking by the cohort maintenance team. Multiple attempts are made to contact sample members, their nominated contact person and the current occupiers of the address previously occupied by sample members through telephone, mail, email and text messaging. Specialist tracing software which combines publicly available Post Office, electoral and phone records is also used routinely in the office for individual searches. As the fieldwork for the study is carried out face-to-face, interviewers also attempt to track families who have moved by making personal visits to the last known addresses of cohort members, neighbours and, if local, their contact persons, in addition to attempting contact through phone and mail. During the fieldwork period, movers who cannot be located by interviewers are returned for office tracking.

This paper examines tracking between wave 2, carried out at age 3 in 2003-4 and wave 3, which took place at age 5 in 2006. We restrict the sample to the families who took part in wave 2 and use respondent characteristics measured at this survey in our analysis. As wave 2 was the first follow-up wave, and only families who were interviewed at baseline (wave 1), were followed up, almost all of the families in our

sample had taken part in both wave 1 and wave 2² i.e. there was little variation in terms of their participation history. In total, 15,590 families took part in wave 2, and 3, 278 (21%) of them had moved by wave 3. For a very small number of families (1%) which were not issued to the field at wave 3 it is not possible to know with certainty whether or not they moved. For all other cases, it is possible to know with a very high degree of certainty whether or not they moved because, even if they didn't participate in the survey, an interviewer visited their address and established whether or not they were still resident. This mobility rate is lower than the proportion of families who moved (over a slightly longer period) between wave 1 and wave 2 of MCS, which was 38 per cent (Plewis et al., 2008). It is very similar to the mobility rates observed in a two-year period in PSID, around 21-22 per cent (McGonagle et al., 2011) and comparable to rates observed in BHPS, around 10 per cent each year (Laurie et al., 1999).

We define the office tracking rate to be the percentage of all movers who were located using office tracking methods alone prior to the start of data collection, the field tracking rate as the percentage of movers *not located by office tracking* who were located by interviewers in the field, and the overall tracking rate as the percentage of all movers who were located by either office or field tracking. Under these definitions, office tracking necessarily takes place prior to the start of fieldwork and therefore prior to field tracking; so they are sequential processes. Importantly, the field tracking rate, defined in this way, is a conditional success rate i.e. we choose to analyse the outcome of field tracking only for those who were not found by office tracking. This means that the office and field tracking rates are not directly comparable. It should also be noted that some of the cases defined as being located using field tracking may have also been tracked in the office during fieldwork, and so may have been tracked through a combination of field and office tracking. It is unclear whether these cases could have been located using office tracking alone, and we therefore include them in the field tracking rate. Additionally, as it is not possible to identify all movers prior to the start of fieldwork, office tracking was not attempted for all movers prior to the start of fieldwork i.e. some of the 'movers not located through office tracking' did not actually receive office tracking (prior to the start of fieldwork).

² Families interviewed at wave 2 included a small number (692) of 'new' families, who were first recruited to the study at wave 2 and who had not been approached at wave 1.

Figure 1 gives the office and field tracking rates and Figure 2 gives the overall tracking rate.

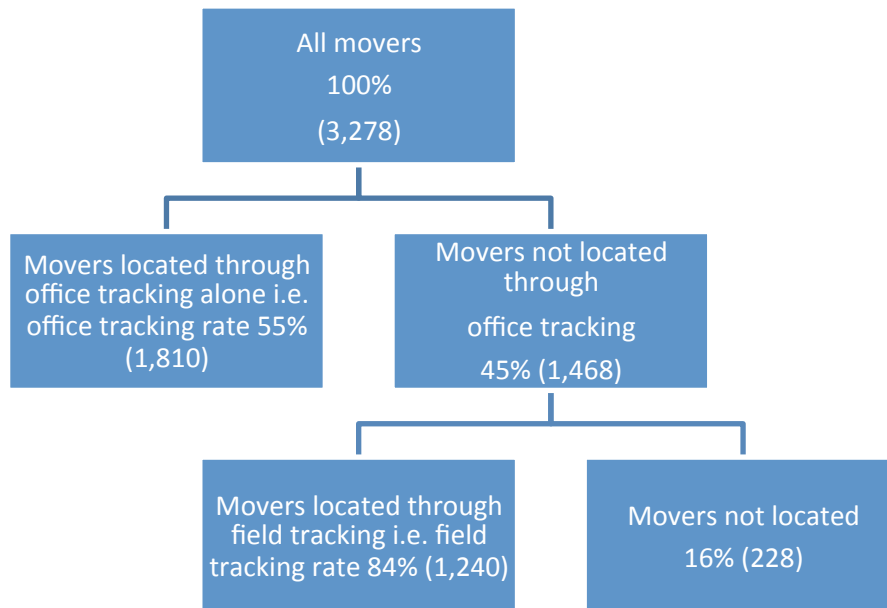


Figure 1. Office and field tracking rates between wave 2 and wave 3 of MCS

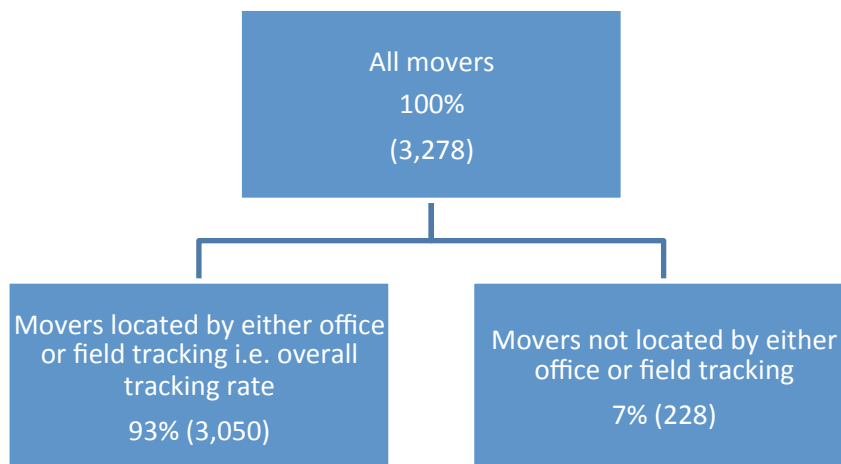


Figure 2. Overall tracking rates between wave 2 and wave 3 of MCS

Figure 1 shows that 55 per cent of all movers were successfully located using office tracking alone prior to the start of fieldwork. This is the office tracking rate. Figure 1 also shows that, of the movers not located prior to the start of fieldwork using office tracking, 84 per cent were successfully located during fieldwork using field tracking. This is the conditional field tracking rate. Figure 2 shows that 93 per cent of all movers were located by either office tracking prior to the start of fieldwork or by field tracking. This is the overall tracking rate.

The overall level of tracking success is comparable with other major longitudinal surveys. It is encouraging that the majority of movers are located prior to the start of data collection, but it is difficult to know how this proportion compares with other surveys as there is limited published research in this area. One exception is the paper by Laurie et al. (1999) which reports that around half of movers on BHPS are found using office tracking prior to the start of fieldwork.

Our analyses will test the hypothesis that sample members who are younger, non-white, lower-educated, lone parents, not working, living in flats or living in rented accommodation will all be less likely to be located by comparing how these characteristics, measured at wave 2, are associated with the office, field and overall tracking rate. As noted earlier, the MCS involves interviews with up to two resident parents. The individual characteristics used, i.e. age, ethnic group and education, are those of the main respondent, who is almost always the child's natural mother. Age and ethnic group are self-reported by the main respondent. The education measure used is based on qualifications. The main respondent's highest academic or vocational qualifications are mapped to an equivalent level on a standard scale which is used in the UK for National Vocational Qualifications (NVQs). The highest level, Level 5, is equivalent to a postgraduate degree and the lowest level, Level 1, is equivalent to a General Certificate of Secondary Education (GCSE), usually taken at age 16, in at least one subject, awarded at the lower range of grades i.e. D-G. Family type is derived from household composition, and household employment status uses both household composition and the working status of main and partner respondents. Most partner respondents are the natural or step-father of the child. Tenure and accommodation type are reported by the main respondent on behalf of the household.

4. Results

Table 1 gives the office tracking rate prior to the start of fieldwork for all movers, the field tracking rate for movers not located by office tracking prior to the start of fieldwork, and the overall tracking rate for all movers, by the selected individual, family and housing characteristics of sample members at wave 2.

Table 1. Office, field and overall tracking rates for movers by wave 2 characteristics

Wave 2 characteristics	Office tracking rate (% of all movers)	Field tracking rate (% of movers not located by office)	Overall tracking rate (% of all movers)	Sample size (all movers)	Sample size (movers not located by office)
Individual characteristics (main respondent)					
Age					
16-24	51.4	81.8	91.2	698	351
25-29	55.2	84.0	92.8	768	358
30-34	58.4	90.4	96.0	967	416
35-39	61.0	88.8	95.6	602	237
40+	59.6	89.9	95.9	207	89
<i>F-statistic</i>	2.19	2.33	3.63		
<i>p-value</i>	>0.05	>0.05	<0.01		
Ethnic group					
White	58.7	90.2	96.1	2652	1153
Mixed	48.1	76.0	87.5	40	23
Indian	51.4	75.2	87.9	63	30
Pakistani and Bangladeshi	54.9	72.7	87.7	180	84
Black or Black British	43.9	67.4	81.7	98	48
Other	30.4	61.5	73.2	55	33
<i>F-statistic</i>	4.07	8.67	15.35		
<i>p-value</i>	<0.01	<0.001	<0.001		
Educational qualifications					
No qualifications	48.3	77.3	88.3	498	246
Overseas qualifications only	46.1	79.3	89.0	109	55
Level 1 (lowest)	49.8	80.0	89.9	288	150
Level 2	58.1	90.4	96.0	936	414
Level 3	57.9	85.2	93.7	461	206
Level 4	61.3	92.2	97.0	848	345
Level 5 (highest)	62.5	91.7	96.9	120	42
<i>F-statistic</i>	3.39	4.84	8.00		
<i>p-value</i>	<0.01	<0.001	<0.001		
Family characteristics					
Family Type					
Married or cohabiting natural parents	56.9	90.0	95.7	2311	1021
Lone natural mother	58.1	79.1	91.3	789	358
Other family type	51.9	72.6	86.8	178	89
<i>F-statistic</i>	0.86	13.72	14.73		
<i>p-value</i>	>0.4	<0.001	<0.001		
Household employment status					
Main (and/or partner) in work	57.2	89.9	95.7	2456	1105
Main (and partner) not in work	56.2	75.3	89.2	780	345
<i>F-statistic</i>	0.13	33.48	32.80		
<i>p-value</i>	<0.7	<0.001	<0.001		
Housing characteristics					
Tenure					
Own	58.8	92.5	96.9	1562	680

	Rent	54.4	80.3	91.0	1462	681
	Other	59.3	87.6	94.9	211	88
<i>F-statistic</i>		<i>1.65</i>	<i>16.11</i>	<i>19.65</i>		
<i>p-value</i>		<i>>0.1</i>	<i><0.001</i>	<i><0.001</i>		
Accommodation type						
	House	57.5	89.6	95.6	2607	1167
	Flat	55.4	74.7	88.6	628	282
<i>F-statistic</i>		<i>0.61</i>	<i>22.64</i>	<i>23.72</i>		
<i>p-value</i>		<i>>0.4</i>	<i><0.001</i>	<i><0.001</i>		
Sample size		3278	1468	3278		

Notes. Design-based F tests (degrees of freedom omitted) were used to test the null hypothesis of no relationship between each of the characteristics and the tracking rates. The analysis was carried out using the *svy* commands in STATA to adjust for the sample design.

Table 1 shows that ethnic group and education had a statistically significant relationship with the office tracking rate and all of the characteristics, except for age, had a statistically significant relationship with the field tracking rate, and all of the characteristics, including age, had a statistically significant relationship with the overall tracking rate. The observed relationships between the tracking rates and the characteristics chosen were in the hypothesised direction.

Looking firstly at the office tracking rate, there is a strong relationship with both ethnic group and education. For example, almost 59 per cent of white movers were located using office tracking prior to the start of fieldwork compared with around 44 per cent of Black or Black British movers. Similarly, over 60 per cent of movers with relatively high education levels (Level 4 or 5 qualifications) were located using office tracking compared with less than half, around 48/49 per cent, of those with the lowest level of education (Level 1) or no qualifications. There is also a clear age gradient in office tracking success, with older respondents more likely to be located using this method, though this relationship is not statistically significant. The office tracking rate was not related to family and housing characteristics. Although overall these results do not support our hypotheses that all of the chosen characteristics would be related to the office tracking rate, it is interesting that the only statistically significant relationships observed are with individual characteristics i.e. ethnic group and education. This supports our secondary hypothesis that individual characteristics would be most strongly related to office tracking, as this method is more dependent on the proactive behaviour of respondents than field tracking. In addition, prior research on MCS has shown that ethnic group and education are both significantly related to responding to the

'keeping-in-touch' mailing, which is one of the main office tracking methods used (Calderwood, 2012).

Now turning to the field tracking rate, which, as noted earlier, is conditional on not being located using office tracking alone prior to the start of fieldwork. As hypothesised, this shows a statistically significant relationship with all of the chosen characteristics, with the exception of age. The higher value of the F-statistic for ethnic group and education, compared with the comparable F-statistic for the office tracking rate, shows that the relationship between these characteristics is even stronger for field tracking compared with office tracking. This is particularly notable given the smaller sample size for field tracking, though, as noted earlier, the field tracking rate is conditional on not being located through office tracking. As with the office tracking rate, there is a clear age gradient in the field tracking rate, though it is not statistically significant. In relation to family characteristics, as hypothesised, couple families had higher field tracking rates than lone parent families (90 per cent compared with 79 per cent) and families with at least one parent employed had higher field tracking rates than those without parental employment (90 per cent compared with 75 per cent). Housing characteristics also showed the hypothesised relationships with the field tracking rate. Movers living in rented accommodation were more difficult to locate in the field than those living in owner-occupied accommodation (80 per cent compared with 93 per cent) and those living in flats were more difficult to locate in the field than those in houses (75 per cent compared with 90 per cent).

In relation to the overall tracking rate, as hypothesised, all of the chosen characteristics were statistically significantly associated with the overall tracking rate, including age, which was not significant for either the office or the field tracking rate when examined separately. Overall, these results provide strong evidence that respondent characteristics are important determinants of tracking success and provide support for our hypotheses regarding the relationship between tracking and age, ethnic group, education, family type, employment status, housing tenure and accommodation type.

In order to further investigate these relationships, we carried out multiple logistic regression analysis for each of the three tracking rates, in order to ascertain whether these relationships remain statistically significant when controlling for other characteristics i.e. the rest of our chosen characteristics. Table 2 gives the odds ratios

associated with the office tracking rate, conditional field tracking rate and overall tracking rate from the logistic regression models, which included all of our chosen characteristics.

Table 2: Odds ratios of being located through office tracking, being located through field tracking and being located through either office or field tracking from logistic regression models, by wave 2 characteristics

Wave 2 characteristics		Office tracking (all movers)	Field tracking (movers not located by office)	Overall tracking (all movers)
		Odds ratios (95% CI)	Odds ratios (95% CI)	Odds ratios (95% CI)
Individual characteristics (main respondent)				
Age				
	16-24	1	1	1
	25-29	1.24 (0.96,1.61)	0.94 (0.57,1.54)	1.02 (0.64, 1.63)
	30-34	1.38 (1.03,1.85)	1.12 (0.59,2.13)	1.28 (0.68,2.40)
	35-39	1.69 (1.23,2.34)	1.27 (0.55,2.93)	1.79 (0.86,3.72)
	40+	1.77 (1.13,2.77)	1.17 (0.44,3.12)	1.61 (0.59,4.35)
	<i>p-value</i>	<0.05	>0.9	>0.4
Ethnic group				
	White	1	1	1
	Mixed	0.68 (0.32,1.44)	0.60 (0.18,2.06)	0.47 (0.16,1.34)
	Indian	0.75 (0.39,1.44)	0.21 (0.88,0.51)	0.22 (0.09,0.54)
	Pakistani and Bangladeshi	0.99 (0.67,1.47)	0.23 (0.09,0.57)	0.31 (0.14,0.67)
	Black or Black British	0.53 (0.31,0.89)	0.25 (0.10,0.63)	0.23 (0.12,0.45)
	Other	0.29 (0.15,0.56)	0.11 (0.04,0.31)	0.08 (0.03,0.20)
	<i>p-value</i>	<0.001	<0.001	<0.001
Educational qualifications				
	No qualifications	1	1	1
	Overseas qualifications only	0.92 (0.52,1.62)	1.31 (0.44,3.95)	1.27 (0.48,3.37)
	Level 1 (lowest)	0.99 (0.67,1.46)	0.82 (0.43,1.57)	0.73 (0.42,1.26)
	Level 2	1.34 (1.01,1.78)	1.51 (0.81,2.81)	1.70 (0.89,3.24)
	Level 3	1.32 (0.95,1.83)	0.78 (0.45,1.35)	0.89 (0.54,1.47)
	Level 4	1.42 (1.01,1.99)	1.67 (0.85,3.29)	1.64 (0.88,3.06)
	Level 5 (highest)	1.36 (0.74,2.52)	1.19 (0.29,4.93)	1.37 (0.36,5.14)
	<i>p-value</i>	>0.1	>0.9	>0.1
Family characteristics				
Family Type				
	Married or cohabiting natural parents	1	1	1
	Lone natural mother	1.14 (0.88,1.47)	0.83 (0.47,1.48)	1.01 (0.57, 1.79)
	Other family type	1.00 (0.68,1.48)	0.39 (0.18,0.86)	0.39 (0.19,0.79)
	<i>p-value</i>	>0.5	>0.05	<0.05
Household employment status				
	Main (and/or partner) in work	1	1	1
	Main (and partner) not in work	1.10 (0.85,1.43)	0.46 (0.30,0.70)	0.52 (0.34,0.81)
	<i>p-value</i>	>0.4	<0.001	<0.001
Housing characteristics				
Tenure				
	Own	1	1	1

	Rent	0.94 (0.76,1.17)	0.69 (0.41,1.16)	0.71 (0.44,1.16)
<i>p-value</i>	Other	1.12 (0.71,1.75)	1.42 (0.53,3.80)	1.45 (0.59,3.58)
		>0.6	>0.2	>0.1
Accommodation type				
	House	1	1	1
<i>p-value</i>	Flat	1.03 (0.81,1.32)	0.53 (0.32,0.88)	0.59 (0.36,0.94)
		>0.7	<0.05	<0.05
-2log likelihood		-2058.87	-489.29	-648.97
Sample size		3,040	1,346	3,040

Notes. Design-based F tests (degrees of freedom omitted) were used to test the null hypothesis of no relationship between each of the characteristics and the tracking rates. The analysis was carried out using the *svy* commands in STATA to adjust for the sample design. Note that the -2log likelihood values are not available using the *svy* commands and were obtained by re-running the models without the *svy* commands.

In relation to office tracking, Table 2 shows that, controlling for other characteristics, age and ethnic group are significantly related to the office tracking rate. Compared with the results from Table 1, this shows that, controlling for other factors, education is not significantly related to office tracking, whereas age is now shown to have a significant relationship with office tracking. The direction of the relationship between age and ethnic group is as expected and as we previously observed in Table 1 i.e. older respondents and white movers are more likely to be located than younger respondents and non-white movers. None of the other characteristics are significantly associated with respondent tracking.

In relation to conditional field tracking, Table 2 shows that, after controlling for other characteristics, many fewer characteristics are significantly related to the field tracking rate. In Table 1, all of the chosen characteristics except age were significantly related to field tracking, whereas, controlling for other characteristics, only ethnic group, employment status and accommodation type remain significant. The direction of the relationship between these characteristics and field tracking is as expected and as observed in Table 1 i.e. non-white, non-employed and flat-dwellers are less likely to be tracked in the field. Education, family type and tenure are no longer significantly related to the field tracking rate, controlling for other characteristics. The relationship between age and field tracking remains non-significant, controlling for other characteristics.

Similarly for overall tracking, Table 2 shows that, after controlling for other characteristics, many fewer characteristics are significantly related to the overall

tracking rate. Ethnic group, family type, employment status and accommodation type are the only characteristics which remain significantly related to the overall tracking rate, controlling for other characteristics. The direction of the relationship is as expected and as observed in Table 1 i.e. non-white, other family types, non-employed, flat-dwellers are less likely to be tracked overall. Age, education and tenure are no longer significantly related to overall tracking, controlling for other characteristics.

Overall, the results from the multiple logistic regressions provide further evidence that respondent characteristics are important determinants of tracking success and show that, controlling for other characteristics, ethnic group, family type, employment status and accommodation type are strongly related to overall tracking success. However, it should be noted that the $-2\log$ likelihood values of the models show that their overall goodness of fit is relatively low, indicating that other factors are also important.

5. Discussion

Overall, these results show that several respondent characteristics show statistically significant relationships with office tracking success, field tracking success and overall tracking success between wave 2 and wave 3 of the UK Millennium Cohort Study.

Controlling for other characteristics using multiple logistic regression showed that ethnic group was related to office tracking, conditional field tracking and overall tracking, age was related to office tracking but not field tracking or overall tracking, employment status and accommodation type were related to both field tracking and overall tracking but not office tracking, and family type was related to overall tracking but not office or field tracking.

Reflecting on our hypotheses in the light of these findings, it is perhaps unsurprising that ethnic group was strongly associated with tracking success, though for surveys like MCS which incorporate over-sampling of minority ethnic groups, this is particularly worrying. Further research is needed to establish the mechanisms through which ethnic group impacts negatively on tracking success, but it seems likely that our hypothesised mechanisms i.e. language barriers, affecting both office and field tracking, are part of the explanation.

It was interesting that age was the only other characteristic which was significantly related to office tracking success, and that it was not related to field or overall tracking success. This gives some support to our hypothesis that individual characteristics would be more strongly related to office tracking than field and overall tracking, as office tracking is more reliant on proactive behaviours by sample members e.g. responding to keeping in touch mailings.

Family type showed an unusual relationship with tracking success; it was not significantly related to either office or field tracking but it was related to overall tracking success, controlling for other characteristics. We hypothesised that the mechanism between tracking success and family type was the additional contact information provided by partners in couple families. However, there was no difference between tracking success rates for lone mother families and couple families once other characteristics were controlled for. Rather the reason why family type was significant related to overall tracking was due to a much lower overall tracking rate among 'other family types'. It is unclear why the tracking rate should be much lower for this group and it is surprising that tracking success rates were no different for couple parents and lone mothers, as this implies that the additional contact information collected from partner respondents is of limited value in relation to tracking. However, further research is needed before concluding this. For example, it may be that the quantity and quality of information collected from partners at wave 2 was limited.

It was surprising that employment status was significantly associated with both field and overall tracking success, controlling for other characteristics, as we did not have a clear hypothesis about the mechanism through which this was likely to have an impact on tracking success. This may be reflection of a lower level of commitment to the study among workless families. However, further research is needed to better understand the process through which employment status is related to tracking success.

It was unsurprising that housing characteristics i.e. accommodation type and tenure, both showed the strong relationships with both field and overall tracking success in the bivariate analysis, as the hypothesised mechanisms through which these characteristics were expected to influence tracking were clear and direct. However, only accommodation type, and not tenure, remained significantly associated with field and overall tracking once other characteristics were controlled for. This implies that

compositional differences between owners and renters explains the relationship between tenure and field and overall tracking observed in the bivariate analysis. Conversely, the fact that accommodation type remained significantly associated with both field and overall tracking, controlling for other characteristics, shows that this relationship cannot be explained by compositional differences between the types of people living in flats and those living in houses.

6. Conclusions

Overall, this paper has clearly shown that respondent characteristics are related to the successful office, field and overall tracking. In doing so, this paper makes a significant contribution to the survey research literature in this area. Although there are examples in the literature which show how individual characteristics are associated with returns to between-wave mailings and collection of contact information, to our knowledge, the relationship between the characteristics of sample members and tracking success has not been explored directly before. As noted earlier, sample members are sometimes viewed as passive in discussions of tracking procedures, though it is clear that by their actions e.g. notifying the study of a change of address or inaction e.g. failure to leave a forwarding address when they move, they can influence how likely it is that they can be located when they move. Further research is needed to examine this issue in more detail, and in particular, to directly test the hypothesised mechanisms through which these characteristics are related to tracking success. However, in terms of implications for survey practice arising from this paper, given the difficulty associated with tracking families living in flats and families from non-white ethnic groups, it is clear that longitudinal surveys should consider implementing additional and/or tailored tracking methods for these groups, including the collection of additional contact information for those living in flats and use of translated tracking materials and office tracking staff/interviewers who speak minority languages for non-white ethnic groups.

This paper also explored the effectiveness of both office and field tracking, and compared how individual, family and housing characteristics were associated with both office and field tracking. Our aim was to provide evidence which would help other longitudinal surveys to increase the proportion of movers which are found using office tracking methods, which are less expensive than field tracking, and thereby improve the cost-effectiveness of their tracking procedures.

Overall, we found that over half of all movers were located by office tracking prior to the start of fieldwork. We thereby demonstrated that it is possible to locate a relatively high proportion of movers using office tracking, prior to the start of fieldwork. Overall, only two of the individual characteristics i.e. age and ethnic group, were significantly related to office tracking success, providing some support for our hypothesis that individual characteristics would be more strongly associated with office tracking than field and overall tracking, because office tracking is more dependent on the proactive behaviour of respondents. The fact that the other characteristics were not significantly related to office tracking success implies that office tracking prior to fieldwork is not differentially effective for these different types of sample members. In some ways, this is a reassuring finding as it shows that, with the exception of younger and non-white respondents, office tracking procedures are not systematically failing to locate certain types of sample members. Attempting to improve office tracking procedures to make them more effective for younger and non-white respondents e.g. by using tailored and/or translated materials during keeping-in-touch mailings, would seem to be worth exploring.

However, we have also clearly shown the importance of field tracking to achieving high overall tracking rates. In addition, our analysis demonstrated that overall tracking success was significantly related, in the bivariate analysis, to all of the respondent characteristics. This implies that improving the effectiveness of tracking, both in the office and in the field, for these 'hard-to-locate' groups should be the primary aim of further research and improvements to survey practice.

References

- Calderwood, L. (2012). Improving between-wave mailings on longitudinal surveys: a randomised experiment on the UK Millennium Cohort Study. *CLS Working Paper*, 2012/4. London: Centre for Longitudinal Studies.
- Call, V.R.A. (1990). Respondent Cooperation and Requests for Contacts in Longitudinal Research. *National Survey of Families and Households Working Paper No.35*. Madison: University of Wisconsin-Madison.
- Crider, D.M., Willits, F.K., & Bealer, R.B. (1971). Tracking Respondents in Longitudinal Surveys. *Public Opinion Quarterly*, 35, 613-620.
- Couper, M.P., & Ofstedal, M.B. (2009). Keeping in Contact with Mobile Sample Members. In P. Lynn (ed.) *Methodology of Longitudinal Surveys* (pp. 183-203). Chichester: John Wiley & Sons.
- Eckland, B.K. (1968). Retrieving Mobile Cases in Longitudinal Surveys. *Public Opinion Quarterly*, 32, 51-64.
- Fumagalli, L., Laurie, H., & Lynn, P. (2010). Experiments with Methods to Reduce Attrition in Longitudinal Surveys. *ISER Working Paper Series*, No. 2010-04.
- Laurie, H., Smith, R., & Scott, L. (1999). Strategies for reducing nonresponse in a longitudinal panel survey. *Journal of Official Statistics*, 15, 269-282.
- McAllister, R.J., Goe, S.T., & Butler, E.W. (1973). Tracking Respondents in Longitudinal Surveys: Some preliminary considerations. *Public Opinion Quarterly*, 37, 413-416.
- McGonagle, K.A., Couper, M.P., & Schoeni, R.F. (2011). Keeping Track of Panel Members: An Experimental Test of a Between-Wave Contact Strategy. *Journal of Official Statistics*, 27, 319-338.
- Plewis, I. (ed.) (2007). *The Millennium Cohort Study: Technical Report on Sampling (4th. ed.)*. London: Institute of Education, University of London.
- Plewis, I., Ketende, S.C., Joshi, H., & Hughes, G. (2008). The Contribution of Residential Mobility to Sample Loss in a British Birth Cohort Study: Evidence from the

first two waves of the Millennium Cohort Study. *Journal of Official Statistics*, 24, 364-385.

Paper 4:

Brown, M., and Calderwood, L. (2014). Can encouraging respondents to contact interviewers to make appointments reduce fieldwork effort? Evidence from a randomized experiment in the UK. *Journal of Survey Statistics and Methodology*. 2 (4), pp. 484-498. DOI: 10.1093/jssam/smu017.

Abstract

One of the main components of survey costs is interviewer call attempts associated with making contact. This paper describes an experiment conducted on a UK household panel study, which sought to evaluate whether an ‘Early Bird’ approach whereby participants are encouraged to contact their interviewer before fieldwork began in order to set up an appointment, could increase fieldwork efficiency by reducing the number of calls required. This approach has been successfully used on the National Longitudinal Study of Youth 1979 cohort for some time but has not been evaluated experimentally. Our experiment involved two treatment groups, one group were promised a modest financial incentive (£5 per participant) to take up the ‘Early Bird’ offer and went on to complete an interview, the other received an appeal to their altruistic tendencies which emphasized that being an ‘Early Bird’ would make their interviewer’s life easier. A parallel experiment, sought to evaluate the impact of differential standard incentives on response. The ‘Early-Bird’ take-up rate was higher for the incentivized group (10% compared with 6% for the non-incentivized group) and was highest when combined with the higher standard incentive rate (17%). Offering both an early-bird incentive and the higher standard incentive did increase overall fieldwork efficiency, as measured by calls required per complete case, but the modest take-up rates meant that the overall impact was fairly minimal. The paper also finds indicative evidence that that the early-bird offer, if sufficiently incentivized could potentially have a beneficial impact on response rates.

Key words:

longitudinal; non-response; incentives; call attempts; randomized experiment; appointments

Acknowledgements:

This research was supported by the Resource Centre funding for the Centre for Longitudinal Studies from the Economic and Social Research Council (ESRC) [RES-579-47-001] and the ESRC funding for Innovation Panel of *Understanding Society*: the UK Household Longitudinal Study (UKHLS) [RES-586-47-0002]. We would like to thank Professor John Micklewright at the Institute of Education for helpful comments on drafts of this manuscript and Jonathan Burton and Noah Uhrig at the Institute of Social and Economic Research, University of Essex, and researchers and field staff at

the National Centre for Social Research (NatCen), for their work on implementing the experiment described in this paper on Wave 4 of the Innovation Panel in 2011.

1. Introduction

This paper describes a randomized experiment, conducted on the Innovation Panel of *Understanding Society*: the UK Household Longitudinal Study (UKHLS) in 2011, which sought to evaluate whether using a financial incentive or an appeal to altruism to encourage respondents to contact their interviewer to arrange an appointment for their interview could increase fieldwork efficiency.

Interviewer call attempts associated with making contact with sample members and persuading them to take part are one of the main components of fieldwork effort, and hence overall survey costs. The number of calls required to achieve an interview has been increasing over time on major cross-sectional surveys in the UK and US (Lynn and Clarke 2002; Curtin et al. 2000).

Respondent incentives are another major component of survey costs, and they are increasingly required to maintain adequate response rates. There is widespread evidence that incentives boost response rates on mail, telephone and face-to-face surveys (Singer and Ye, 2012) and a large literature on how they should be administered and their impact on bias (Singer 2002; Bruderl et al. 2008).

In 2002 the 20th wave of the National Longitudinal Study 1979 cohort (NLSY79) introduced an innovative fieldwork strategy, the ‘early-bird’ approach, whereby sample members are sent a letter two weeks before the start of fieldwork which offers an incentive to those that call a toll-free number to set up an appointment for their interview within four weeks. The main aim was to reduce the survey costs associated with the multiple call attempts required by interviewers to contact respondents. In 2002 the offered early-bird incentive was \$40, and in waves 21 and 22 in 2004 and 2006, two different incentive values, \$60 and \$80, were tested experimentally. These incentives were in addition to the standard \$40 incentive for participation. In these waves, it was only offered to sample members who had been mostly co-operative for the previous four rounds, though it was subsequently rolled out to everyone in the sample at later waves (Kochanek et al, 2010).

At waves 21 and 22, almost half (48% and 49% respectively) of those offered the early-bird incentive took it up and most importantly, early-birds took on average three hours of interviewer time to complete, compared with five hours for non-early birds. In addition, at wave 22, the response rate was slightly higher among those who were

offered the early-bird than the sample overall (83% compared with 80% overall). Kochanek et al. concluded therefore that the 'early-bird' innovation had lowered fieldwork costs and reduced field effort for the most cooperative respondents, freeing up field resources for more difficult cases.

However, the innovation was not evaluated experimentally. As the early-bird offer was provided to only a sub-sample comprising the most co-operative respondents, and those taking it up were likely to be the most willing respondents, one might expect that they would require less fieldwork effort, and have a higher response rate, regardless of the early bird intervention. It is therefore not possible to definitively attribute the lower level of fieldwork effort observed for early-bird cases, or the higher response rate at wave 22, to the intervention.

This paper seeks to evaluate whether an 'early bird' approach could reduce fieldwork effort in the UK context. We will also investigate the impact on response. To our knowledge, this was the first time that an 'early-bird' approach had been used on a major UK longitudinal survey, and the first time anywhere that this approach has been evaluated using a randomized experimental approach.

Like many US surveys the NLSY is conducted primarily via telephone. As UK surveys are typically carried out face-to-face, there is potential for this 'early-bird' approach to deliver even greater cost-savings in the UK than in the US.

Another major difference between the US and the UK is that although incentives are becoming more widespread in the UK, and are used on the UKHLS, they tend to be much lower value than those used in the US (Laurie & Lynn, 2009). There are also several major longitudinal surveys in the UK, such as the British Birth Cohorts, which do not use incentives at all.

There is an extensive empirical and theoretical literature about non-monetary reasons for taking part in surveys. For example, Groves, Cialdini and Couper (1992) discuss the applicability of a number of concepts from social psychology to the decision to participate in a survey. In particular, they argue that compliance with requests to participate in a survey include appeals to 'helping tendencies'.

For these reasons, our experiment also included a treatment group that was not offered an incentive and were instead encouraged to become ‘early-birds’ by an appeal to altruism. Clearly, if this were possible, the potential cost savings would be greater still.

2. Design and implementation of the experiment

The experiment described in this paper was carried out in 2011 on the fourth wave of the Innovation Panel of the UK Household Longitudinal Study (UKHLS), a major household panel study involving annual interviews with 40,000 households. The Innovation Panel involves an additional sample of around 1,500 households and is used to develop and evaluate methodologies for longitudinal survey data collection.

There are some important differences between UKHLS and NLSY79. Firstly UKHLS is a household panel survey involving multiple respondents per household whereas the NLSY79 survey is a sample of individuals. UKHLS is conducted face-to-face whereas NLSY is primarily conducted by telephone. Finally NLSY respondents have been taking part since 1979 with the early-bird innovation being introduced at wave 20, whereas UKHLS started more recently, in 2008, and the experiment was carried out on only the fourth time respondents were contacted.

Experimental design and hypotheses

The main aims were to assess the impact of the early-bird innovation on fieldwork effort, as measured by interviewer call attempts, and to compare the effectiveness of financial incentives and an appeal to altruism. For this reason, there were two experimental groups who both received the early bird offer, but with different encouragements to take it up. One group was offered a financial incentive for being an ‘early-bird’ and the other received an appeal to altruism. A control group received no early-bird offer.

Of the 1,390 cases that had participated in Wave 3 and were issued for Wave 4 a total of 1,088 (78.3%) were included in the experiment. Those excluded were those who had not participated at Wave 3 (n=294) and those in which one or more of the household members were known or suspected to have left the household since the last wave were excluded from the experiment, as it would have been difficult operationally to ensure the correct treatment was applied in these circumstances (n=8).

Households were randomly allocated to one of three equally sized experimental groups. Those in the ‘incentivized’ group were offered an extra £5 each for pre-booking an appointment. Early-bird incentives were promised at the individual level, so, for example, in a four-person household this would amount to an extra £20 in total. Advance materials made clear that the incentive was paid to all household members who were subsequently interviewed if *any* household member contacted the interviewer to pre-book an appointment.

As in NLSY, the early-bird incentive was additional to the standard incentive amount for taking part in the survey. The UKHLS early-bird incentive was conditional and paid after the interview while standard incentives were unconditional and sent in advance. Early bird and standard incentives both took the form of high street vouchers which could be spent in most leading stores. The NLSY early-bird incentive was paid upon booking an appointment and both the early-bird and standard incentives were paid by check. The UKHLS standard incentives were considerably smaller than those used by NLSY. Wave 4 of the innovation panel involved a parallel experiment examining the impact of differential standard incentives. Households included in the Early Bird experiment were randomly allocated to seven groups who received either £5 or £10 and an eighth group where individuals were given £5 initially, rising to £10 if all eligible adults in the household completed an interview. For some groups the amount received was more than they had been offered in prior waves, for others the amount was the same and for some the wave 4 incentive was less than they had received previously. The early bird incentive therefore presented the opportunity to increase the total payment received by either 100% or 50%. Although differential standard incentives were not explicitly part of our experimental design, the parallel experiment allows us to examine the impact that varying the standard incentive has on the take-up rate of the early-bird offer and on response rates. A full description of the parallel incentive experiment and its impact on response is provided by Budd et al. (2012). When the two experiments are crossed the cell sizes become very small (see Table 1) so in this paper, when examining the impact of the standard incentive we will focus on the difference between a £5 unconditional incentive and a £10 unconditional incentive at this wave only. Cases allocated to the group receiving a £5 unconditional incentive which increased to £10 if all eligible adults were interviewed will be treated as having received a £5 unconditional incentive.

Table 1: Number of Cases Receiving Early-Bird Treatment, by Standard Incentive

	Early bird with incentive	Early bird with no incentive	Control group	Total
1. £5 (same at W1, W2, W3)	46	50	55	151
2. £10 (was £5 at W1, W2, W3)	63	62	64	189
3. £10 (same at W1, W2, W3)	31	38	35	104
4. £5 (was £10 at W1, W2)	30	33	33	96
5. £5 (was £10 at W1)	28	30	38	96
6. £10 (was £10 at W1 and £5 at W2 & W3)	32	32	29	93
7. £5 to £10 for complete cooperation (same at W1, W2, W3)	60	63	64	187
8. £5 (was £5 to £10 at W1)	56	58	57	171
Total	346	366	375	1087

Sample members in the ‘no incentive’ treatment group were told that contacting the interviewer to pre-book their appointment will ‘make your interviewer’s life much easier as they will not have to make repeated telephone calls or visits to your home in order to try and reach you’. The incentive and control groups were not told this. The ‘no incentive’ group, and the control group, still received the standard incentives for participation in the survey.

We hypothesized that those incentivized to take-up the early-bird offer would be more likely to do so than those in the ‘no incentive’ group and that among those in the incentivized group, those receiving a larger standard incentive would be more likely still to take up the offer.

The primary outcome measure was the number of calls required to complete interviewing in productive households - meaning households where at least the household interview was completed³. We hypothesized that the average number of interviewer calls required to complete interviewing in a productive household would be lower for ‘early-birds’ than for cases who did not take-up (or were not given) the offer,

³ The survey is comprised of an initial household interview followed by individual interviews for each household member.

and that the take-up rate would be sufficient to translate into a lower average number of calls among the treatment groups overall, resulting in a reduction in total fieldwork effort. Clearly the higher the take-up rate of the early-bird offer, the greater the potential for reducing fieldwork effort overall. As we expected the greatest take-up of the offer to be among those offered both the early-bird incentive and the higher standard incentive, we hypothesized that this group would require the lowest number of calls overall.

We examine the impact of the intervention on response as a secondary outcome. We hypothesized that, as per NLSY, the early-bird intervention, when incentivized would lead to a modest boost in response rates, due to the increase in the total amount of the incentive on offer. We hypothesized that the boost to response would be most significant among those receiving a larger standard incentive. We did not anticipate that the non-incentivized approach would increase the overall response rate.

Implementation of the experiment

Advance letters were posted to the two ‘early bird’ treatment groups approximately three weeks before fieldwork began. Letters were sent to all individual sample members (rather than one per household) and explained that their household was being provided with “the opportunity to request an “Early Bird Appointment” by contacting your interviewer on their mobile phone before anyone else to arrange your interview at a time that best suits you”. The mailing also contained a “Want to be an Early Bird?” leaflet providing more information about the offer. The design was identical for the incentivized and non-incentivized groups but the text varied slightly in line with the treatment. Copies are included in the Appendix.

Interviewers were all issued with a mobile telephone and these numbers, along with their names, were included in the advance letters and leaflets. Respondents were asked to send a text message or leave a voicemail message on the interviewer’s mobile phone within a two week window prior to the beginning of fieldwork. They were asked to leave their name, reference number (which was also mail-merged onto the leaflet), phone number and preferred interview date and time. The leaflet stated that interviewers would call back within 24 hours to confirm the appointment.

Interviewers were instructed not to attempt to contact sample members until these two weeks had passed and fieldwork had begun. Sample members were able to pre-book their interview for any day in the four-week data collection period.

The control group also received an advance letter which explained that the next wave of the study was about to begin but did not encourage sample members to contact the interviewer and did not include an early bird leaflet.

4. Results

What was the take-up rate for the early bird offer?

Table 2 shows that just under one in ten (7.7%) households in the early-bird groups took up the offer and made an appointment. Those in the incentivized group were more likely to do so than those in the non-incentivized group (10.1% compared with 5.5%). The take-up of the offer was also related to the size of the standard incentive, with those given £10 being more likely to do so than those given £5 (11.6% compared with 5.5%). The highest take-up rate by some degree, 16.7%, was among the group given a £10 standard incentive and offered the £5 early-bird incentive, suggesting that take-up rates are maximized when the combined total incentive is maximized. The results of a logistic regression⁴, with take-up of the early bird offer as the dependent variable and the offer of an early-bird incentive, the standard incentive and their interaction as covariates are presented in Table 3. The likelihood of taking up the offer was much reduced among those offered no early-bird incentive and those offered the smaller standard incentive; however the interaction between the offer of the early-bird incentive and the standard incentive was not significant.

Table 2: Take-up of the early bird offer by treatment group and standard incentive amount

	Standard Incentive		All
	£5	£10	
Early-Bird Incentive	6.4% (220)	16.7% (126)	10.1% (346)
Early-Bird No Incentive	4.7% (234)	6.8% (132)	5.5% (366)
All Early Bird	5.5% (454)	11.6% (258)	7.7% (712)

*Base sizes in brackets

⁴ This analysis and all analyses which follow were conducted using the complex samples procedure in SPSS which takes account of the clustered sample design.

Table 3: Logistic regression results: Take up of the early bird offer by treatment group and standard incentive

	Exp(B)	95% Confidence Interval for Exp(B)		Sig.
		Lower	Upper	
(Intercept)	0.200	0.112	0.357	0.000
Early bird incentive not offered (Ref: Incentive offered)	0.366	0.158	0.847	0.020
Standard incentive £5 (Ref: £10)	0.340	0.163	0.709	0.005
Early bird incentive not offered*£5 incentive offered	1.984	0.532	7.392	0.301

What was the impact on total fieldwork effort?

Table 4 shows the number of interviewer visits required to complete interviewing among productive households. In the context of a household panel study, the number of visits required will also depend to a large extent on the number of individuals within the household, that is, how many interviews are required, but there were no differences between the three groups in terms of the average number of individuals in the household eligible for interview.

Those in the early-bird treatment groups required slightly fewer visits on average than those not offered the early-bird option (3.2 compared with 3.4) and the difference between the incentivized group and the non-incentivized group was also small (3.2 compared with 3.3). Overall there was no difference at all between those given £5 and those given £10 as a standard incentive, with both groups requiring 3.3 calls. Those offered the early-bird incentive and the £10 incentive required the fewest calls (2.9). When the treatment group, the standard incentive and their interaction were used as covariates in a linear regression model (not shown) with number of calls required to achieve a productive interview as the dependent variable neither the main effects nor the interaction were found to be significant.

Table 4: Number of interviewer visits required to complete all interviewing in productive households

	Standard incentive		All
	£5	£10	
Early-Bird Incentive	3.4 (167)	2.9 (99)	3.2 (266)
Early-Bird No Incentive	3.3 (182)	3.3 (103)	3.3 (285)
<i>All Early Bird</i>	3.3 (349)	3.1 (202)	3.2 (551)
Control group	3.4 (247)	3.6 (89)	3.4 (274)
<i>All</i>	3.3 (534)	3.3 (291)	3.3 (825)

If we focus solely on productive households in which the early bird offer was taken up (analysis not shown) then as expected, the difference was considerable with around two fewer interviewer visits being required at early-bird households (1.4 calls on average) than at those not taking up the offer, either because they were not given the opportunity or chose not to (3.4 calls on average); the difference was statistically significant (p -value ≤ 0.001). Among early-bird households, there was no difference between the incentivized and non-incentivized groups. There were also no differences between those who were offered early-bird but did not take it up and those not offered early-bird, nor between those given £5 and those given £10 standard incentives.

In order to conclusively attribute the reduced number of calls among the early-bird households to the introduction of the offer we can compare the number of calls required at the current wave (wave 4) with the number required at the previous wave (wave 3), for both those who took up the offer and those who did not. Table 5 shows that the cooperative nature of those taking up the early bird offer at wave 4 had resulted in them requiring slightly fewer visits at the previous wave. However, the number of calls required at wave 4 for these households was significantly lower than the number required at the previous wave (1.4 compared with 3.2), demonstrating that the early-bird initiative had significantly reduced the fieldwork effort required for these particular households. Fewer calls were also required at wave 4 for those who were offered the early-bird but did not take it up, and those who were not offered the early-bird, but the differences for these groups were much smaller and were not significant.

Table 5: Number of interviewer visits required at wave 3 and wave 4 to complete interviewing in households in which a productive household interview was achieved at wave 4 (by whether Early Bird taken up).

	Mean number of calls at wave 3	Mean number of calls at wave 4	Number of cases
Early Bird offered and taken up	3.2	1.4	55
Early Bird offered and not taken up	3.7	3.4	484
Early Bird not offered	3.9	3.4	262

What was the impact on response rates?

It was hypothesized that the early-bird treatment would have a modest positive impact on response rates for the ‘incentive’ treatment group. Conducting a logistic regression (not shown) with a binary variable identifying productive households as the dependent variable found that neither the early bird treatment, the standard incentive nor the interaction between them had a significant impact on response. However, Table 6 provides indicative evidence that early-bird approach could potentially have a positive impact on response, given that the response rate was indeed slightly higher for both the incentivized early bird group (76.7%) and the non-incentivized group (77.9%) than for the control group (73.1%). There was some initial concern that providing respondents with mobile numbers for their interviewer would provide an easy avenue for refusals but there was no evidence of this. Somewhat surprisingly, response in the control group for those receiving a £5 standard incentive was actually slightly higher than for those receiving £10⁵ but the highest response rate, 78.6%, was achieved among the incentivized early bird group with a £10 standard incentive, although again small cell sizes resulted in the difference between this group and other groups not reaching statistical significance. Nevertheless, there is clear indicative evidence that again, maximising the combined total incentive is beneficial in terms of boosting response.

⁵ In this paper we define 'productive' as a household in which the household interview was completed. Budd et al. (2012) showed that larger incentives were effective at boosting response at the individual level.

Table 6: Response rate by experimental group and standard incentive

	Standard Incentive		All
	£5	£10	
Early-Bird Incentive	75.9% (220)	78.6% (126)	76.7 (346)
Early-Bird No Incentive	77.8% (234)	78.0% (132)	77.9% (366)
<i>All Early Bird</i>	76.9% (454)	78.3% (258)	77.3% (712)
Control group	74.9% (247)	69.5% (128)	73.1% (375)
<i>All</i>	76.2% (701)	75.4% (386)	75.8 (1087)

5. Discussion and conclusions

Our results provide robust experimental evidence that incentivising respondents to take up the early-bird offer, particularly when coupled with a higher standard incentive rate, leads to a higher take-up rate than an appeal to altruism. The highest take-up rate was found among the group offered both the higher standard incentive rate and the early-bird incentive. The main aim of the early bird intervention was to reduce survey costs by reducing the number of interviewer visits required to complete interviewing. The paper has shown that when taken up the early bird offer does result in fewer visits being required, and that the higher level of take-up among those receiving the early bird incentive and the higher standard incentive, led to a significant reduction in the number of calls required to complete interviewing with households in this group in comparison to the control group.

Although it was found that take-up rates were highest where the total potential incentive was maximized, take-up rates were fairly low among all groups when compared with those obtained on the NLSY. It seems likely that this stems mainly from the much lower value of the incentives used. Another key difference is that NLSY early-bird incentives were paid once an appointment was arranged, whereas in our experiment, the early bird incentives were conditional on participation. If the incentives had been larger and paid once an appointment was made then take-up rates might have been higher which would have led to more significant reductions in the fieldwork effort required.

It may also be possible to boost the take-up rate in non-monetary ways. Anecdotal evidence from interviewers suggested that the early bird offer was not sufficiently prominent within the advance materials; interviewers recorded comments about their

experience of contacting those allocated to the early bird treatment groups and many noted that respondents had little awareness of the offer. The advance letters made relatively little mention of the offer, and the accompanying leaflet may have been overlooked so better marketing of the offer may have lead to better take-up.

Appealing to the self-interest of sample members, either instead of or in addition to altruism, by placing more emphasis on the benefits of taking up the early bird offer for respondents themselves (rather simply to the interviewer) such as mentioning explicitly that taking up the offer would mean that they would not receive telephone calls or visits from interviewers at inconvenient times, may be a more successful strategy. Further experimentation could also examine the impact of combining these appeals with a financial incentive; an approach which has not yet been tested.

The experiment also showed that the combination of a higher standard incentive with the early-bird incentive lead to higher response rates and that even without an incentive the early- bird initiative appeared to have some beneficial impact on response rates. Relatively small cell sizes meant that differences in response were not found to be statistically significant. Nevertheless, our results provide indicative evidence that the early-bird offer, particularly if sufficiently incentivized could potentially have a beneficial impact on response rates. It could of course have been the case that a larger standard incentive would have had a similar (or larger) impact

Overall our experiment has clearly shown that the early-bird approach has the potential to be a successful way to reduce fieldwork effort and fieldwork costs, and also to potentially boost response rates in the context of a major household longitudinal study in the UK. However, in order to maximize the benefits we need further research to examine ways of encouraging a higher proportion of respondents to take-up the early-bird offer and set-up appointments for themselves.

References

Bruderl, J., Castiglioni, L., Krieger, U., and Pforr, K. (2008), “The effects of incentives on attrition bias: results from an experimental study”, Presentation to the 1st International Panel Survey Methods Workshop, July, University of Essex, UK.

Curtin, R., Presser, S. and Singer, E. (2000), “The effects of response rate changes on the Index of Consumer Sentiment”, *Public Opinion Quarterly*, 64, 413–428.

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

Kochanek, K., Krishnamurty, P., and Michael, R. (2010), “The ‘Early Bird’ Field Innovation on a 30-year-old Longitudinal Survey”, Presentation at the 2nd International Panel Survey Methods Workshop, July, Mannheim, Germany.

Laurie, H. and Lynn, P. (2009), “The Use of Respondent Incentives on Longitudinal Surveys”, in *Methodology of Longitudinal Surveys* (ed. P. Lynn), Chichester: Wiley.


Lynn, P. and Clarke, P. (2002), “Separating Refusal Bias and Non-Contact Bias: Evidence from UK National Surveys”. *The Statistician*, 51(3), 319-333.

Singer, E. (2002), “The Use of Incentives to Reduce Nonresponse in Household Surveys”, in *Survey Nonresponse* (eds. R.M. Groves, D.A. Dillman, J.L. Eltinge and R.J.A. Little), New York: Wiley.

Singer, E., and Ye, C. (2013). “The use and effects of incentives in surveys. *The Annals of the American Academy of Political and Social Science*”, 645, 112–141.

Appendix 1 – Participant leaflets

Incentive version



Want to be an Early Bird?

Understanding Society

IP4 Early Bird Information Leaflet_incentive_v10.indd 1 18/01/2011 16:20:32

Want to be an Early Bird? This time around we're giving you the opportunity to get in touch with your interviewer to arrange a time for your interview. By doing this you'll be able to pre-book an appointment at a time which is convenient for you. If you do get in touch to make an appointment, we'll give everyone in your household who is interviewed an extra £5 to say thank you.

How? Get in touch with your interviewer between 22nd February and 7th March 2011 (<<mobilenumber>>). Either send an SMS (text) or leave an answer phone message with the information listed on the right. Your interviewer will call you back within 24 hours.

When? You can pre-book your interview for any day between 8th March and 8th April 2011.

Then what? Your interviewer will call you back to confirm your choice of interview date and time. Everyone interviewed in your household will get an extra £5 so book an appointment now.

Just send us an SMS (text) or leave an answer phone message, with the following information:

1. Your name
2. Your telephone number
3. Your reference (<<EBRef>>)
4. Preferred date & time of interview that is convenient to you

www.understandingsociety.org.uk/participants
Freephone 0800 252 853
contact@understandingsociety.org.uk


F3025 IP4 CARD A
IP4 Early Bird Information Leaflet_incentive_v10.indd 2 18/01/2011 16:20:33

Printed on Forest Stewardship Council (FSC) accredited and 50:50 recycled paper

Non-incentive version



Want to be an Early Bird?



IP4 Early Bird Information Leaflet_non incentive_v10.indd 1

24/01/2011 10:43:15



Want to be an Early Bird? This time around we're giving you the opportunity to get in touch with your interviewer to arrange a time for your interview. By doing this you'll not only be able to pre-book an appointment at a time which is convenient for you, but will also make your interviewer's life much easier as they will not have to make repeated telephone calls or visits to your home in order to try and reach you.

Just send us an SMS (text) or leave an answer phone message, with the following information:

How? Get in touch with your interviewer between 22nd February and 7th March 2011 (<<Mob_Phone>>). Either send an SMS (text) or leave an answer phone message with the information listed on the right. Your interviewer will call you back within 24 hours.

1. Your name
2. Your telephone number
3. Your reference (<<EBRef>>)
4. Preferred date & time of interview that is convenient to you

When? You can pre-book your interview for any day between 8th March and 8th April 2011.

Then what? Your interviewer will call you back to confirm your choice of interview date and time. Be an Early Bird, and book an appointment now.



P3025 IP4 CARD B

www.understandingsociety.org.uk/participants
Freephone 0800 252 853
contact@understandingsociety.org.uk

Printed on Forest Stewardship Council (FSC) accredited and 50:50 recycled paper



IP4 Early Bird Information Leaflet_non incentive_v10.indd 2

24/01/2011 10:43:16

Bibliography

Journal articles

Brown, M., and Calderwood, L. (2014). Can encouraging respondents to contact interviewers to make appointments reduce fieldwork effort? Evidence from a randomized experiment in the UK. *Journal of Survey Statistics and Methodology*. 2 (4), pp. 484-498.

Calderwood, L. (2014). Improving between-wave mailings on longitudinal surveys: A Randomised experiment on the UK Millennium Cohort Study. *Survey Research Methods*. 8(2), pp99-108

Calderwood L., Rose N., Ring, S. and McArdle W. (2014). Collecting saliva samples for DNA Extraction from children and parents: findings from a pilot study using lay Interviewers in the UK, *Survey Methods: Insights from the Field*.

Calderwood, L. (2013). The role of respondent characteristics in tracking on longitudinal surveys: evidence from the UK Millennium Cohort Study. *Longitudinal and Life Course Studies*. 4(2), pp105-118.

Calderwood, L. (2013). The use of new technologies on the British Birth Cohort Studies. *International Journal of Market Research*. 55(2). pp. 590-595.

Calderwood, L. (2012). Tracking sample members in longitudinal studies. *Survey Practice*. 5(4).

Plewis, I., Ketende, S. and Calderwood, L. (2012). Assessing the accuracy of response propensity models in longitudinal studies. *Survey Methodology*, 38(2).

Hockey, C., Quigley, M.A., Hughes, G., Calderwood, L., Joshi, H., Davidson, L. (2008). Linking Millennium Cohort data to birth registration and hospital episode records. *Paediatric and Perinatal Epidemiology*, 22(1) 99-109.

Tate, R.A., Calderwood, L., Dezateux, C., Joshi, H. and The Millennium Cohort Study Child Health Group (2006). Mother's consent to linkage of survey data with her child's

birth records in a multi-ethnic national birth cohort, *International Journal of Epidemiology*, 35(2), 294-298.

Chapters in books

Calderwood, L. and Lessof, C. (2009). Enhancing longitudinal surveys by linking to administrative records. In P. Lynn (ed.). *Methodology of Longitudinal Surveys*, Chichester: Wiley.

Bartley, M., Calderwood, L., Jayaweera, H., Plewis, I. and Ward, K. (2005). *Children's origins* in 'Children of the 21st Century: from Birth to Nine Months. Bristol: The Policy Press.

Calderwood, L., Kiernan, K., Joshi, H., Smith, K. and Ward, K. (2005). Parenthood and Parenting. In S. Dex and H. Joshi (eds) *Children of the 21st Century: from Birth to Nine Months*, Bristol: The Policy Press.

Other outputs, including working papers, reports and report chapters

Nicolaas, G., Calderwood, L., Lynn, P., and Roberts, C. (2014). Web surveys for the general population: How, why and when?. National Centre for Research Methods Report.

Calderwood, L., Smith, K., Rose, N., and Williams, R. (2014). Training and accreditation of interviewers to carry out physical measurements on the UK Millennium Cohort Study. CLS Working Paper 2014/9. London: Centre for Longitudinal Studies.

Brown, M. and Calderwood, L. (2013). Can encouraging respondents to contact interviewers to make appointments boost co-operation rates and save costs? Evidence from a randomised experiment in the UK. CLS Working Paper 2013/9. London: Centre for Longitudinal Studies.

Calderwood, L. (2012). Improving between-wave mailings on longitudinal surveys: A randomised experiment on the UK Millennium Cohort Study. CLS Working Paper 2012/4. London: Centre for Longitudinal Studies.

Calderwood, L., Plewis, I., Ketende S.C. and Taylor, R. (2010). Experimental testing of refusal conversion strategies in a large-scale longitudinal study. CLS Working Paper 2010/9. London: Centre for Longitudinal Studies.

Calderwood, L. (2010). Keeping in touch with mobile families in the UK Millennium Cohort Study. CLS Working Paper 2010/7. London: Centre for Longitudinal Studies.

Calderwood, L. (2008). Family Demographics. In H. Joshi and K. Hansen (eds) *Millennium Cohort Study Third Survey: A User's Guide to Initial Findings*, London: Centre for Longitudinal Studies.

Calderwood, L. (2007). Family Demographics. In H. Joshi and K. Hansen (eds) *Millennium Cohort Study Second Survey: A User's Guide to Initial Findings*, London: Centre for Longitudinal Studies.

Calderwood, L., Kelly, Y. and Panico, L. (2007) 'Parent's Health' in 'Millennium Cohort Study Second Survey: A User's Guide to Initial Findings', H. Joshi and K. Hansen (eds) London: Centre for Longitudinal Studies.

Shaw, A. and Calderwood, L. (2004) 'Millennium Cohort Study First Survey: Technical Report on Questionnaire Development and Fieldwork' London: NatCen

Plewis, I., Calderwood, L., Hawkes, D, and Nathan, G (2004) 'National Child Development Study and 1970 British Cohort Study Technical Report: Changes in the NCDS and BCS70 Populations and Samples over Time, London: CLS.

Gjonca, E and Calderwood, L (2003) 'Socio-demographic characteristics'. In Health, Wealth and Lifestyles of the Older Population in England: the 2002 English Longitudinal Study of Ageing, M.Marmot, J.Banks, R.Blundell, C.Lessof and J.Nazroo. (eds), London: IFS.

Taylor, R., Conway, L., Calderwood, L and Lessof, C. (2003) 'Methodology'. In Health, Wealth and Lifestyles of the Older Population in England: the 2002 English Longitudinal Study of Ageing, M.Marmot, J.Banks, R.Blundell, C.Lessof and J.Nazroo. (eds), London: IFS.

Calderwood, L. and Park, A. (2000) 'General Health'. In Scottish Health Survey 1998, A.Shaw, A.McMunn and J.Field (eds.), Edinburgh, HMSO.

Calderwood, L. and Tait, C. (2000) 'Self-reported health and psychosocial well-being'. In Health Survey for England 1999, B.Erens, P.Primatesa and G.Prior (eds.) London, HMSO.

Calderwood, L. at al. (2000) 'Survey Methodology and Response'. In Health Survey for England 1999, B.Erens, P.Primatesa and G.Prior (eds.) London, HMSO.