Improving between-wave mailings on longitudinal surveys: A randomised experiment on the UK Millennium Cohort Study

Lisa Calderwood
Centre for Longitudinal Studies
Institute of Education
London, UK

A between-wave mailing to keep in touch with sample members is one of the most commonly used tracking procedures on longitudinal surveys. 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 treatment, 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. It also involved changing the style and signatory, motivated by the psychological concepts of reciprocity, liking and helping tendencies. Our main finding was that the design of the covering letter has a minimal impact on the overall return rate of these mailings. However, we also found that the re-design of the letter can positively influence the return rate for particular sub-groups. These findings 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.

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

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. Most large-scale longitudinal surveys have developed highly successful tracking procedures for locating movers. But there is little evidence on the relative success, and cost-effectiveness, of different tracking methods (Couper & 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, which is a large-scale birth cohort study in the UK.

Our main finding was that the design of the covering letter has a minimal impact on the overall return rate of these mailings. However, we also found that the re-design of the letter can positively influence the return rate for particular sub-groups. These findings will help guide further research 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.

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 which are outside the control of the survey. However, 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 min-

imising 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, Smith, & Scott, 1999) and that fewer fieldwork resources are required to achieve an interview with them (Couper & 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 US Panel Study for Income Dynamics (PSID) (McGonagle, Couper, & Schoeni, 2011; McGonagle, Schoeni, & Couper, 2013) and the British Household Panel Survey (BHPS) (Fumagalli, Laurie, & Lynn, 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, Callegaro, Hox, Korendijk, & Lensvelt-Mulders, 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, Lepkowski, & Garabrant, 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. They note that liking can be related to how similar to themselves the person receiving the request perceives the person making the request to be. 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 & 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. On PSID, McGonagle et al. (2011) found that 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, Smyth, & Christian, 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 & Lynn, 1998). Increasing the proportion of returns to the initial mailing would reduce costs associated with the reminder mailings and hence improve cost-effectiveness, even if no overall increase in the return rate was achieved.

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. The positive association between responding to these mailings and taking part in the next round of the survey means that 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 (Office for National Statistics, 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%) (Plewis, Ketende, Joshi, & Hughes, 2008; Calderwood, 2013).

However, in common with all longitudinal studies, MCS has experienced attrition as families drop out over time despite the success of its tracking procedures. 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 & 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 a covering letter and feedback leaflet for parents, a feedback leaflet for the cohort child, a return form and a freepost envelope. As well as the family's address and home phone number, the return form also includes mobile and work phone numbers, email addresses and details of contact persons. An address confirmation approach is used. That is, study members are asked to return the form whether or not their details have changed. 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 from 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 difficultly 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. 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.

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.

Language complexity was measured using the Flesch-Kincaid readability score (see e.g. www.read-able.com). This 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 between half (0.49) and two-thirds

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

(0.67) of the length of the control versions. The reminder letters were also 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 signatory on treatment letter was 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. It was not anticipated that this change of signatory would not undermine the authority of the request as sample members are familiar with the study and are likely to perceive requests from the study as having legitimate authority, even if the signatory of the letter was not the study director. 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 who were expected to be issued for subsequent waves.

Study families were randomly assigned with equal probability to the treatment group or the control 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. All cases were included in the experiment, regardless of their prior participation history. 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 aca-

demic 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.

We first 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 method to correct for this.[1] This involves dividing the chosen threshold for statistical significance for each comparison according the number of comparisons made. In this way 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 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 first 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

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.

56% for the treatment group, p-value >0.3). The experiment therefore showed that 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).

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 experimental treatment 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.

In relation to education, Table 3 shows a clear gradient in the return rate by education level for both the treatment and control groups. The lower educated 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 treatment 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% compared with 34%, a difference of 8 percentage points. This difference emerges at the initial mailing as the early return rate in the treatment group (22%) is 7 percentage points 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

---

[1] As recommended by Williams, Jones, and Tukey (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.

Table 3 Return rates by experimental group and education and language

	Control %	Treatment %	Sample size	p-value
Returned				
Main respondents educational qualifications				
No qualifications	34.1	42.4	1,558	0.000*†
Overseas qualifications only	45.2	46.7	385	0.787
Level 1 (lowest)	46.6	48.2	951	0.642
Level 2	59.2	57.2	3,622	0.216
Level 3	63.2	61.1	2,082	0.285
Level 4	73.1	75.9	4,162	0.036*
Level 5 (highest)	75.7	77.0	930	0.653
Languages spoken at home				
English only	62.8	63.1	11,280	0.693
Other languages	50.8	55.4	1,876	0.007*†
All	61.2	62.1	13,696	0.242
Returned early				
Main respondents educational qualifications				
No qualifications	15.3	22.0	1,558	0.000*†
Overseas qualifications only	22.9	22.3	385	0.901
Level 1 (lowest)	25.5	26.0	951	0.850
Level 2	32.7	33.3	3,622	0.676
Level 3	36.5	36.5	2,082	0.980
Level 4	43.6	50.1	4,162	0.000*†
Level 5 (highest)	46.0	49.6	930	0.262
Languages spoken at home				
English only	36.5	39.9	11,280	0.002*†
Other languages	24.4	29.2	1,876	0.024*†
All	34.8	37.8	13,696	0.000*

*statistically significant at the unadjusted critical value i.e. 0.05.
†statistically significant at the Bonferroni adjusted critical value i.e. 0.007 for education and 0.025 for languages spoken at home.
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).

less impact. However, the limited impact that the reminder stage does have is in the same direction: a higher return rate in the treatment group leading to the gap between the treatment and control group increasing by 1 percentage point. 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 treatment 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 postgraduate degree or equivalent, though these differences are not statistically significant.

In summary, there is clear evidence that the treatment 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. The reason that the overall difference between the control and treatment group in the return rate disappears after the reminder mailings appears to be as follows: 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 and this leads 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 hypothesised. 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.

These findings show that the secondary aim of the experimental treatment, to increase the return rate 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.

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 treat-

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

	Returners		Overall sample
	Control %	Treatment %	%
Main respondents educational qualifications			
No qualifications	6.2	7.9	11.4
Overseas qual. 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
Sample size	4188	4248	13,690
Languages spoken at home			
English only	88.7	87.7	86.3
Other languages	11.3	12.3	13.7
Sample size	4188	4250	13,696

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

ment 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 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 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. In particular, our experimental treatment aimed to 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 & Lynn, 1998).

It is not possible to attribute the increase in the early return rate to a particular component of re-design. However, a plausible interpretation is that the informal and reciprocal nature version of the letter used in the treatment group would have greater impact on early returns (as was observed), and that the formal version from a more authoritative signatory used in the control group 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. 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. It is not possible to attribute the increase in the return rate from these groups to a particular component of re-design. But it is plausible that letters that are shorter and that 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.

Acknowledgements

References

Budowski, M., & 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., & Ofstedal, M. (2009). Keeping in contact with mobile sample members. In P. Lynn (Ed.), *Methodology of longitudinal surveys* (p. 183-203). Chichester: John Wiley & Sons.

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

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

Dillman, D., Smyth, J., & Christian, L. (2009). *Internet, mail and mixed-mode surveys: The tailored design method.* New York: Wiley.

Fumagalli, L., Laurie, H., & 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., Cialdini, R., & Couper, M. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly*, *56*, 475-495.

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

McGonagle, K., Couper, M., & Schoeni, R. (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., Schoeni, R., & Couper, M. (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., & Garabrant, D. (2011). An experimental examination of the content of persuasion letters on non-response 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., 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.

Taylor, S., & 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., Jones, L., & Tukey, J. (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: 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 (http://www.childnc .netwww.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

Appendix B: Key variable by
experimental Group

	Control		Treatment	
	n	*%*	*n*	*%*
Experimental group[1]	7,826	50.0	7,827	50.0
Particip. at most recent wave[2]	6,848	87.5	6,848	87.5
Main respondents educational qualifications[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
Sample size	6,847		6,843	
Languages spoken at home[3]				
English only	5,918	86.4	5,902	86.2
Other languages	930	13.6	946	13.8
Sample size	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.