Experiments with methods to reduce attrition in longitudinal surveys

Laura Fumagalli, Heather Laurie and Peter Lynn

University of Essex, Colchester, UK

Summary. We report findings from two large-scale randomized experiments, carried out on the British Household Panel Survey, with survey features designed to reduce sample attrition. The first experiment compares strategies for obtaining updated contact information from sample members. We find that the propensity to locate a sample member at wave *t* successfully is maximized by a between-wave mailing of a change-of-address card, rather than an address confirmation card or no card. The second experiment compares standardized and tailored respondent reports for young people and busy people. We find that tailored reports have a modest positive effect on rates of co-operation for both groups, though the effect for busy people depends on providing the option of a shorter telephone interview instead of the full face-to-face interview.

Keywords: Attrition; Co-operation; Experiments; Longitudinal surveys; Non-response; Respondent incentives; Tailoring

1. Introduction and background

Household panel surveys—and longitudinal surveys more generally—provide an important and popular resource for studying society and its dynamic changes. However, the validity of estimates that are derived from these surveys depends on the extent to which the sample remains representative of the study population over time. Two important elements are the success of the survey at avoiding non-response at the first wave (initial non-response) and at avoiding subsequent loss of sample members due to non-response at latter stages (sample attrition). In this paper we focus on the second of these elements: attrition.

Though much has been written about attrition, the literature has mainly been limited to descriptions of how it arises and how surveys attempt to prevent it (Kalton, 2009; Kalton and Citro, 1993; Laurie *et al.*, 1999) and to empirical studies of known correlates (Groves, 2006; Watson and Wooden, 2009; Fitzgerald *et al.*, 1998; Uhrig, 2008). There have been few attempts to compare the effects of alternative survey design features on attrition. Though much of the knowledge about effects of non-response on cross-sectional surveys may apply also to longitudinal surveys, there are at least two reasons why specific studies of attrition are warranted. First, the processes leading to attrition, particularly in the later waves of a survey, may be rather different from the processes leading to non-response in cross-sectional surveys. Consequently, the effects of design features could be different. Second, there are additional design features that can only be implemented in longitudinal surveys, owing to the extensive information that is obtained about sample members at previous waves and the additional opportunities for contacting them.

Address for correspondence: Laura Fumagalli, Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, UK.
E-mail: lfumag@essex.ac.uk

There are several reasons why it is important to study efficient ways of reducing attrition. First is the need to maintain the sample size for analysis, including analysis of relatively small subgroups of the population. Second, attrition can be a source of non-response bias, as those dropping out of the study may have different characteristics from the stayers, potentially leading to biased parameter estimates and misleading estimates of change measures. Third, understanding the cost effectiveness of alternative attrition reduction measures can help survey researchers to use resources wisely.

This paper provides some rare experimental evidence on methods to minimize attrition. The three components of attrition that were identified by Lepkowski and Couper (2002) are a failure to locate sample members, a failure to make contact with sample members with a known address and a failure to gain co-operation conditional on contact. Our experiments address two of these three components, namely failure to locate sample members and failure to gain co-operation.

Failure to locate sample members arises because of geographical mobility when movers cannot be traced and are lost to the sample (for an analysis of attrition among those with complex migration histories and multiple moves see Fitzgerald *et al*. (1998) and Zabel (1998)). Previous work has proposed and discussed tracing methods (Cohen *et al*., 1996; Freedman *et al*., 1980; Ribisl *et al*., 1996; Scott, 2004; Couper and Ofstedal, 2009), but little or no experimental evidence is available to assess the effectiveness of alternative strategies. Our first experiment compares strategies which may have an effect on the likelihood of locating a sample member given a move. Specifically, we compare three fundamentally different ways of asking sample members to supply address updates. Additionally, we test the role of conditional *versus* unconditional incentives and the effect of different amounts of monetary incentives in encouraging people to confirm or update their address details.

Failure to gain co-operation can arise when sample members are insufficiently motivated or interested in the survey (Groves *et al*., 2000, 2006), something which may occur after several waves of participation owing to 'panel fatigue'. Although longitudinal surveys contain detailed information on sample members, there is little or no research to suggest how this information might be used to motivate respondents better. Indeed, since tailoring methods were introduced in the context of interviewers' doorstep behaviour (Groves *et al*., 1992; Groves and Couper, 1998), the idea of tailoring has not spread to other aspects of the survey process affecting response. Our second experiment uses survey data from previous waves to tailor the content of a brochure mailed between waves with the intention of heightening interest in the survey. Specifically, we tailor a report of relevant research findings to the characteristics of two sample subgroups with relatively low co-operation propensities and compare subsequent co-operation rates with those achieved with a standard (not tailored) findings report.

Our study makes some original contributions to the literature and has some important strengths. We provide empirical evidence of the relative effectiveness of alternative strategies for persuading longitudinal sample members to provide updated contact details. We describe an original attempt to tailor materials designed to foster interest in the survey, and we provide empirical evidence of the effect of these materials on subsequent co-operation, relative to standard (non-tailored) materials. We also provide evidence on the role of incentives in determining response to between-wave contact exercises as distinct from response to the main survey request itself (see Laurie and Lynn (2009) and Petrolia and Bhattacharjee (2009)). Additionally, our findings contribute to understanding of the role of mixed mode designs in tackling non-response. Our study has good statistical power as it is based on a sample of 12500 individuals. It also benefits from good external validity as it is based on a nationally representative household sample.

2. Previous research

We know of only one other study that provides experimental evidence of the effectiveness of alternative methods to persuade longitudinal sample members to provide address updates. McGonagle *et al.* (2011) reported a study using the Panel Study of Income Dynamics (PSID) that was carried out at the same time as our experiment on the British Household Panel Survey (BHPS). Methods for keeping track of sample members are particularly important in the case of the PSID as, since 1997, the survey has shifted from annual to biennial interviews, with the longer interval increasing the likelihood of changes of address and therefore of non-response due to a failure to locate sample members (Couper and Ofstedal, 2009; Duncan and Kalton, 1987).

McGonagle *et al.* (2011) experimented with four factors: incentives for updating the address (unconditional *versus* conditional $10), the design of the return card (traditional black-and-white design *versus* contemporary colour design) and receiving or not a study newsletter and timing of the mailing (June, October or both). They found that neither incentives nor the newsletter affected replies to the mailing. The traditional design was more effective than the updated design and people receiving two mailings were more likely to reply than those receiving just one. McGonagle *et al.* (2011) noted, however, that the reason for the effect of card design cannot be identified. The traditional design was perhaps less burdensome (instructions easier to follow) but the effect could simply have been due to respondents reacting to a familiar format.

McGonagle *et al.* (2011) also examined the effects of their treatments on two measures of the cost of field effort at the subsequent wave, namely whether tracking was needed and the total number of telephone calls needed to resolve the case. They found no difference between treatments in the probability of entering tracking, but they found that a prepaid incentive, the newsletter, the traditional design and the October mailing all have a modest effect of reducing the total number of calls, after controlling for other characteristics of the sample households. They reported no differences in interview rates to the subsequent wave, due to the high overall interview rates. They did not report effects on location rates.

Our experiment provides evidence in a context that is different from that of McGonagle *et al.* (2011). The BHPS, which started in 1991, is younger than the PSID, which started in 1968, so respondents are less 'seasoned', on average. BHPS fieldwork is conducted face to face whereas the PSID is primarily a telephone survey. Also, as noted above, the PSID has biennial waves, whereas the BHPS is annual.

Research has demonstrated that monetary incentives can increase the propensity of sample members to take part in surveys, both for mail surveys (Church, 1993) and interviewer-administered surveys (Singer *et al.*, 1999). However, in our case the task for which sample members are being incentivized, supplying updated address information, is rather different from survey participation and incentives will not necessarily work in the same way. Also, there is experimental evidence that monetary incentives that are used at the point of contacting sample members can reduce the number of calls that interviewers need to make before making contact (James, 1997; Lynn *et al.*, 1998; Rodgers, 2002) and can therefore save field costs. This result provides hope that incentives could also save field costs in our context, but the context is different as our mailings are targeted at mobile sample members with the aim of improving the efficiency with which sample members can be located, rather than the efficiency of making contact conditional on successful location.

3. Context of the study

Our study was carried out on the BHPS, which is a national survey of Great Britain which started in 1991 with an original sample of 5500 households and 10300 individuals. The sample

consists of all people who were resident in 1991 at a stratified probability sample of addresses in England, Scotland and Wales drawn from the small users *Postcode Address File*. Annual waves of interviewing have been carried out since then (for details of the BHPS design see www.iser.essex.ac.uk/bhps). Additional samples of 1500 households in each of Wales and Scotland were added in 1999 and 2000 households in Northern Ireland were added in 2001. The sample includes people of all ages, but sample members are only asked for a full individual interview from age 16 years upwards with a self-completion youth interview for those aged 11–15 years. At each wave, interviews are carried out with all sample members and all other members of their household. Sample members who move are followed to their new address. The initial household response rate was around 70% at wave 1, with over 80% of respondents successfully reinterviewed at wave 2. Since wave 5 the response rate among people who had been interviewed at the previous wave has been between 95% and 97% at each wave (Lynn, 2006).

The BHPS uses computer-assisted personal interviewing, but telephone interviews are carried out within the refusal conversion process as a last resort to avoid complete non-response by an individual sample member or household (see Burton *et al.* (2006)). At around 10 min in length, the telephone interview is much shorter than the 40-min face-to-face interview and contains the core longitudinal items from the main individual questionnaire. Telephone interviews are not used as a primary mode of data collection for reasons of data quality and response rate.

In addition to telephone interviews, the BHPS uses a variety of strategies to reduce non-response (see Laurie *et al.* (1999)). These strategies can be divided into those intended

- (a) to aid location of mobile sample members,
- (b) to increase the likelihood of contact and
- (c) to keep respondents interested in the survey and motivated to participate.

The strategies in categories (a) and (c) are those most relevant to our study as these are the components of attrition at which our interventions are targeted. To aid location, at each interview BHPS respondents are provided with a freepost change-of-address card preprinted with their known address details and between waves they are mailed a new confirmation-of-address card (the latter mailing corresponds to treatment 3 in our experiment—see Section 4.1 and Table 1 later). On both occasions they are informed that they will receive a £5 gift voucher if they return the card with details of a new address. To keep respondents interested and motivated they are mailed a short report of survey findings each year between waves corresponding to the control treatment in our tailored materials experiment—see Section 4.2. Each sample member aged 16 years or over also receives a mailing in advance of each wave's fieldwork, enclosing an unconditional £10 voucher. Further details of BHPS field procedures can be found in Taylor (2010).

4. Study design

Our study consists of two experiments carried out simultaneously on the sample of all wave 17 BHPS respondents, using a randomized design. The *address updating experiment* is inspired by Couper and Ofstedal (2009). The aim is to understand and evaluate alternative strategies for reducing attrition due to a failure to locate sample members. The experiment involves seven treatment groups, which are described in Section 4.1. The *tailored materials experiment* aimed to test the tailoring of content for the between-wave respondent report mailing to stimulate interest, loyalty and co-operation. It consisted of two treatment groups, as described in Section 4.2. To avoid confounding, the two experiments are crossed in a 7×2 design. To identify the most effective treatment for address updating, we carry out pairwise comparisons of treatments. Thus, none of the treatments is considered a control group.

Mailing units were either individuals or couples. A single mailing was sent to co-resident married or cohabiting couples, whereas respondents who were not in a couple received an individual mailing. This approach was used to keep treatment groups as small as possible. Separately allocating individuals and couples to treatments, rather than whole households, also maximizes the scope for tailoring in the tailored materials experiment and, arguably, increases the chances of obtaining new address information in cases where a subset of the household has changed address. Where information was provided by the couple in response to the mailing, this included details for each person in the couple. Mailing units within households received the same incentive treatment.

The experiments were carried out in the period between waves 17 and 18 of the BHPS. The mailing was sent out in June 2008 with replies received until the time of the wave 18 fieldwork which took place between September 2008 and February 2009. The sample for the experimental study consisted of about 12 500 people, 8877 mailing units and 7011 households. For the tailored material experiment randomization was done at the mailing unit level and, for the address updating experiment, at the household level.

To check that the random allocation was implemented correctly, the treatment groups were compared in terms of a set of demographic and behavioural measures from BHPS wave 17 (sex, age, marital status, employment status, education, income and intention to move). As expected, no systematic statistically significant differences were found.

4.1. Address updating experiment

The address updating experiment is designed to test multiple hypotheses. Specifically, we manipulate three factors:

  (a) the form of request made to respondents to collect information on their address,
  (b) the amount of a monetary incentive that is associated with the mailing and
  (c) whether the incentive was offered unconditionally or conditionally on reply.

The design allows separation of the effects of type of incentive (unconditional or conditional) and amount of the incentive, unlike some studies which confound these two elements (e.g. Petrolia and Bhattacharjee (2009)).

Three alternative forms of information request were compared:

  (i) asking all sample members to confirm their address details and providing a freepost address confirmation card (hereafter denoted by AC),
  (ii) asking only those whose details have changed to inform us of their new address and providing a freepost change-of-address card (hereafter COA) and
  (iii) asking only those whose details have changed to inform us of the details and *not* providing a reply card (hereafter NOC).

Within group (i) (AC card), four alternative incentive treatments were compared, consisting of a cross-classification of amount (£5 *versus* £2) and conditionality (unconditional *versus* conditional on the return of the card). Within group (ii) (COA card), two alternative levels of incentive were compared, but both were offered only conditionally on return of the card (as in this case the incentive was aimed only at those who had moved, and we did not know in advance which cases these were). No incentives were offered within group (iii). This design gives a total of seven treatment groups.

Before wave 17 BHPS sample members had received with each between-wave mailing an AC card with a £5 incentive conditional on its return with new address details. In addition, they were handed a COA card at each wave and so are attuned to information being requested in a combination of forms (i) and (ii).

The sample was assigned randomly to the seven treatments, with a quarter of the sample assigned to group (iii) (treatment NOC) and an eighth to each of the other six treatments. The larger sample in group (iii) is for budgetary reasons as this is the least costly treatment. The number of treatment groups was set so that each group would contain at least 1000 mailing units, a sample size that should yield sufficient power for the detection of effects that would be both plausible and of practical significance. Mailing units were systematically assigned to treatments, after stratifying the sample by region at wave 17 and then by interviewer area, so the treatments are approximately evenly represented in regions and interviewer areas. This aspect of the design effectively controls for stratum and interviewer area effects. Table 1 summarizes the treatments and sample sizes for each treatment group.

4.2. Tailored materials experiment

The tailored materials experiment aims to assess whether the use of reports which contain information that is directly relevant to the respondent's circumstances can increase the level of co-operation for groups that are characterized by lower response rates. This contrasts with the experiment of McGonagle *et al*. (2011) with a respondent newsletter, in which they did not exploit information about sample members to tailor the newsletter to specific groups of respondents.

The two target groups for our tailored reports are people aged 16–24 years ('young people') and people who work long hours or who work full time and have long commutes ('busy people'). It has been frequently observed in household panel surveys that younger people exhibit lower response rates and higher attrition rates (Behr *et al*., 2005; Lillard and Panis, 1998; Stoop, 2005; Uhrig, 2008; Watson and Wooden, 2009). Also, young people aged under 25 years in the BHPS sample have been eligible for a full individual interview for between 1 and 9 waves, whereas most other sample members have been eligible for longer, many for 18 waves. Young people may therefore have a lesser sense of loyalty or commitment to the survey. Busy people have similarly been observed to exhibit lower response propensities (Abraham *et al*., 2006; Groves

Table 1.   Treatment groups for the address updating experiment

Group type and number		Type of card	Type of incentive	Amount	Abbreviation†	Sample size (number of mailing units)
(i)	1	AC	Unconditional	£5	ACu5	1124
(i)	2	AC	Unconditional	£2	ACu2	1111
(i)	3	AC	Conditional on return	£5	ACc5	1125
(i)	4	AC	Conditional on return	£2	ACc2	1104
(ii)	5	COA	Conditional on moving and return	£5	COA5	1104
(ii)	6	COA	Conditional on moving and return	£2	COA2	1096
(iii)	7	NOC	No incentive	None	NOC	2213
Total mailing units						8877
Total individuals						12500
Total households						7011

†For analysis, we shall also combine pairs of groups with different amounts of incentive, but where the treatment is otherwise identical. We shall denote these combinations by ACu (combination of ACu5 and ACu2), ACc (ACc5 and ACc2) and COA (COA5 and COA2).

and Couper, 1998; Lynn and Clarke, 2002; Watson and Wooden, 2009). As well as exhibiting lower response rates, both groups are distinctive in terms of important survey measures (with attendant potential to introduce non-response bias) and share common characteristics that should make tailoring of the respondent report possible.

The target groups are defined by information collected at wave 17. Young people are those who were aged 16–24 years at the scheduled time of the start of wave 18 fieldwork. Busy people are those who reported working more than 40 h per week, or commuting for more than 10 h per week in addition to working full time, or being self-employed.

Two types of reports were used. The standard report (Fig. 1) followed the same format as the usual report that is sent every year to BHPS sample members, consisting of eight A4 paper pages containing five pages of text describing general findings with simple graphics. Two tailored reports were prepared: one for young people (Fig. 2) and the other for busy people (Fig. 3). Compared with the standard report, the tailored reports were of a smaller format, more colourful, with less text, simpler sentence construction and extensive use of colour photographs. Not only were the findings and topics selected to be of high salience for each subgroup, but the figures, colours, photographs, layout and even the fount were designed to appeal to members of the relevant subgroup. The tailored reports were designed to have high leverage (Groves *et al.*, 2000).
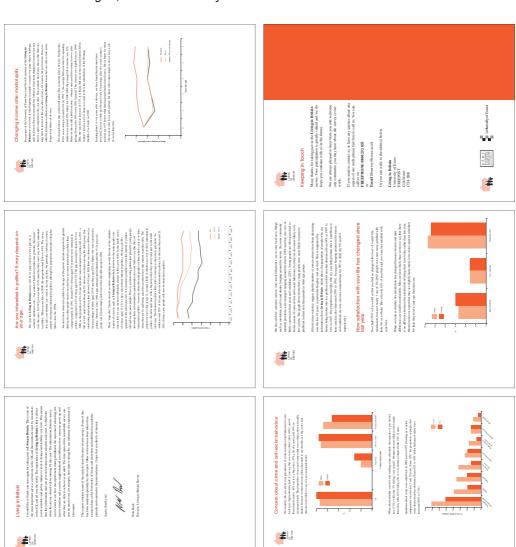
Half of the mailing units were randomly allocated to the treatment group and the other half to the control group. All those in the control group received the standard report. In the treatment group, mailing units containing young people received the young person's report, other units containing busy people (but no young people) received the busy person's report and the remainder received the standard report. The random allocation was carried out within each of the seven treatment groups described in Section 4.1, resulting in 14 treatments overall. One report only was sent to both members of a couple. If either member of the couple belonged to one of the target groups and the couple was allocated to the tailored treatment, they received the tailored report for that group.

5. Hypotheses

For the address updating experiment we expect to observe the highest rate of return with the AC card, given that the respondents in the COA group are expected to return the card only if they move. The lowest rate of return is expected for the no-incentive group.

We expect the proportion of sample members left untraced at wave 18 to be negatively correlated with the proportion of movers who return address information in response to the address update mailing. We expect the proportion returning address information to be lowest in the no-incentive group. The COA cards are explicitly targeted at movers so they may have greater leverage on that group. However, people whose contact details have changed only slightly, or who may not realize that their address is different from the address that is known to the survey researchers (e.g. because their mail is forwarded to their new address) may not be motivated to return the COA card, whereas they should return the AC card, as all sample members are requested to do so.

As regards the effects of conditional *versus* unconditional incentives, the literature has described two channels through which incentives work (Singer, 2002): the first is loyalty, when the respondent who receives the unconditional incentive feels an obligation to co-operate with the request irrespective of any additional benefit that the co-operative behaviour might provide. The second channel is based on economic exchange theory, according to which the respondent considers the advantages and disadvantages of responding to the request and puts into practice the action that is associated with the biggest difference between the former and the latter. Thus, if loyalty prevails unconditional incentives will be more effective, whereas conditional incentives

Fig. 1. Standard report

Fig. 2. Report tailored to young people

will be more effective if economic exchange considerations dominate. The dominant finding regarding survey co-operation is that unconditional incentives are more effective than conditional incentives, both for mail surveys (Church, 1993) and interviewer-administered surveys (Singer *et al.*, 1999), so we expect that to be the case here, even though the request is of a different type (returning address information by mail). We expect the £5 incentive to be more effective than the £2 incentive.

We have no *a priori* expectation regarding the effect of the treatments on costs, given that five cost components may be affected:

(a) the direct cost of providing incentives;
(b) the cost of printing materials and mailing them to sample members;
(c) the postage cost of return mailings;
(d) the office costs of tracing activities, including telephone calls;
(e) field costs of tracing activities, including visits by interviewers to sample addresses.

The first component is unambiguously highest with unconditional incentives and lowest with

Fig. 3.   Report tailored to busy people

no incentives. The second is lower with no incentives, as there is no card to print, but does not differ between COA and AC cards. However, the relative costs of the other three components depend on the rate of return of address information and the quality of that information.

For the tailored materials experiment we hypothesize that the tailored reports should have a positive effect on interview co-operation among the subgroups to whom they are tailored. This should result from the relevance of the survey having been given increased saliency (Groves *et al.*, 2000).

6. Methods

We compare the address updating treatments in terms of three categories of outcomes. The first relate to the outcomes of the updating process itself; the second relate to survey outcomes at the subsequent field wave; the third relate to costs. In this section we motivate our interest in these outcome measures and describe the measures and methods that we use for comparing treatments.

The first set of outcomes indicates the success of the mailing in achieving its immediate aims of encouraging sample members to reply and to provide updated details. These outcomes are primarily of interest because they could shed light on the channels by which the various treatments impact on the subsequent wave field outcomes. Knowledge of the outcomes may also assist with survey planning. We examine the proportion of sample units who return the card or otherwise communicate information about their address in response to the mailing, and the nature of the information that is reported (confirmation of existing details, supply of new details or a return indicating that the addressee is not known at the address).

The second type of outcome relates to the ultimate goal of the address updating exercise. We examine the proportion of sample members who are successfully located at the subsequent field wave, regardless of whether or not an interview is achieved.

The third set of outcomes relate to costs. There is clearly a cost to implement an updating exercise, but if successful the exercise should reduce the cost of field effort at the next wave, as interviewers should be less likely to find themselves visiting out-of-date incorrect addresses and less effort should be needed to trace sample members. To indicate implementational cost we use estimates of the direct cost of incentives plus postage costs, based on an analysis of the proportion of returns received in each treatment. To indicate field costs, we use measures of the number of telephone calls made from the office in the course of tracing sample members and the number of personal visits made by an interviewer in the field. Unlike monetary values, these are portable statistics which can be interpreted by survey researchers who may be working in an environment with very different field cost structures from those of the BHPS.

To assess the effectiveness of the tailored materials treatments, we examine the survey response rate at the subsequent wave. We make the assessment separately for the two sample subgroups that are the target of the tailored materials as the design and content of the materials were different for the two groups and the effects could differ.

When assessing the outcomes of the tailored materials experiment and the first two types of outcomes of the address updating experiment, we compare sample proportions by using one-tailed t-tests. We believe that the normal approximation to the binomial distribution is reasonable as very few of the relevant observed proportions approach 0.0 or 1.0. In fact, the number of observations in the numerator of the proportion ($n \times p$) is usually considerably more than 10, as is $n \times (1 - p)$. We use one-sided tests as our hypotheses specify the direction of the effect, as described in Section 5. We do not make any formal adjustment for multiple comparisons as the number of comparisons presented is to some degree arbitrary and reflects our desire to present the reader with the broadest possible information about the effect of the treatments on the survey process. It is clear that the key outcomes and comparisons are few and that other comparisons are secondary. The informed reader will be able to make her or his own judgement about the effects of multiple comparisons. We highlight test results at the 0.10 level of significance to help to identify potentially useful treatments that may warrant further investigation, but actual p-values are presented. All analyses were performed in Stata version 11.1; estimates allow for sample clustering via specification of postcode sectors, the clustering units in the BHPS sample design, as primary sampling units with the SVYSET command.

7. Results and discussion

7.1. Address updating

We first focus on whether treatment affects the propensity to respond to the address updating mailing. Table 2 shows for each treatment group the share of mailing units returning the card or otherwise responding, along with the results of mean comparison tests. Of those sent the AC

Table 2. Comparisons between treatment groups in percentage of mailing units returning address information†

Treatment‡	% returned	n	Significance of pairwise differences for the following treatments:							
			ACu2	*ACc5*	*ACc2*	*ACc*	*COA5*	*COA2*	*COA*	*NOC*
ACu5	40.9	1124	0.318	**0.008**	**0.000**		**0.000**	**0.000**		**0.000**
ACu2	39.8	1111		**0.025**	**0.001**		**0.000**	**0.000**		**0.000**
ACu	*40.4*	*2235*				*0.000*			*0.000*	*0.000*
ACc5	35.2	1125			**0.077**		**0.000**	**0.000**		**0.000**
ACc2	32.0	1104					**0.000**	**0.000**		**0.000**
ACc	*33.6*	*2229*							*0.000*	*0.000*
COA5	13.2	1104						**0.081**		**0.000**
COA2	15.5	1096								**0.000**
COA	*14.4*	*2200*								*0.000*
NOC	6.8	2213								

†Standard error estimates take into account the clustering of mailing units within households through use of the `svy` commands in Stata 11.1. Rows in italics are summary rows (i.e. each is a combination of the two preceding rows); numbers in bold have $p < 0.10$.
‡See Table 1 for a definition of treatments.

card, 40% replied if sent an unconditional incentive and 34% did so with a conditional incentive, compared with 14% of those mailed the COA card and 7% of those who were not sent either a card or an incentive. Some of the last group may have kept a card from an earlier contact, such as the wave 17 interviewer visit, and used that to respond to the between-wave mailing.

Unconditional incentives performed better than conditional incentives, regardless of the amount: all four pairwise differences between a conditional and unconditional incentive are statistically significant at the 0.01 level and in the direction of higher return rates with the unconditional incentive. This points towards a prevalence of loyalty over maximizing behaviour, which is consistent with research on the effect of incentives on survey participation (Church, 1993; Singer *et al.*, 1999). The experiment provides no clear evidence that the amount of the incentive affects the rate of return: of the three pairwise comparisons in which the other treatment factors are held constant, one is not statistically significant (ACu5 *versus* ACu2), whereas the other two are of borderline significance ($p = 0.08$ in both cases) and are in opposite directions.

Responding to a request for address information will, however, only improve subsequent survey participation rates if the response provides new information that proves helpful. The majority of responses to the AC mailing merely confirmed that address details were unchanged (the results are not shown). The proportion of mailing units for which new details were received—including notifications that the addressee is not known at the address—is not statistically significantly different by using the COA card (3.1%) compared with either of the AC card treatments, though the proportion is higher ($p < 0.05$) with ACu (4.2%) than with ACc (2.4%). For each of the three incentivized treatments, the amount of the incentive made no statistically significant difference to the propensity for new address information to be reported (the results are not shown), though in all cases the direction of the difference was consistent with a higher incentive encouraging more reporting.

7.2. Location propensity
We now consider our main outcome of interest, namely the proportion of sample units that could not be located at the following wave because they had moved. Table 3 summarizes survey

Table 3. Household level interview outcome† at the subsequent wave, and comparison between treatment groups in the percentage of untraced movers‡

Treatment§	Outcomes (%) at subsequent survey wave						n	Significance of pairwise differences in percentage of untraced movers for the following treatments:							
	All interviewed	Some interviewed	Telephone interview	Untraced movers	Resident non-contact	Refusals		ACu2	ACc5	ACc2	ACc	COA5	COA2	COA	NOC
ACu5	77.8	13.3	5.93	0.19	0.39	2.32	1552	0.495	0.100	0.496		0.159	0.165		0.488
ACu2	76.5	14.7	5.94	0.200	0.52	2.15	1533		0.117	0.499		0.205	0.210		0.496
ACu	*14.0*	*14.0*	*5.93*	*0.19*	*0.45*	*2.24*	*3085*				*0.151*				*0.491*
ACc5	78.0	12.9	5.19	0.51	0.38	2.94	1562			0.101		**0.026**	**0.027**		**0.091**
ACc2	78.8	11.8	5.92	0.20	0.13	3.12	1538					0.157	0.163		0.493
ACc	*12.4*	*12.4*	*5.55*	*0.35*	*0.26*	*3.03*	*3100*							*0.105*	*0.145*
COA5	78.4	13.5	6.31	0.06	0.39	1.29	1554						0.491		**0.099**
COA2	76.0	14.8	6.39	0.07	0.60	2.13	1503							0.015	0.106
COA	*14.2*	*14.2*	*6.35*	*0.07*	*0.49*	*1.70*	*3057*								*0.078*
NOC	*76.8*	*14.0*	*6.25*	*0.20*	*0.79*	*1.97*	*3039*								

†Units of analysis are individuals, but the outcome is that of the household to which the individual belongs.
‡Rows in italics are summary rows (i.e. each is a combination of the two preceding rows); numbers in bold have $p < 0.10$.
§See Table 1 for a definition of treatments.

outcomes at BHPS wave 18 for each treatment group and presents the results of tests of pairwise differences in the proportion of sample members who could not be located at wave 18. The proportion of untraced movers is lowest with treatment COA: it is lower than with either treatment ACu ($p = 0.10$) or ACc ($p = 0.02$). Moreover, COA is the only treatment which does significantly better ($p = 0.08$) than NOC. The proportion of untraced movers is highest with treatment ACc. The amount of the incentive does not affect the rate of untraced movers for either treatment ACu or COA.

The overall interview rate is no higher with treatment COA than with AC. There is a suggestion that although COA reduces the proportion of untraced cases (and of refusals) it produces a slightly higher proportion of telephone interviews rather than full face-to-face interviews (Table 3).

7.3. Costs of locating sample members

The first component of cost that we compare is the direct cost of mailings, incentives and the associated returns, for which the survey organization must pay postage. We ignore the costs of printing the materials and outward postage, as the cost per mailing unit for these components would be the same for each treatment. We estimate the cost of providing the incentive on the basis of, for the conditional incentives, the proportion of sample units in our experiment who returned a card. We then estimate the cost of return postage based on the same proportion, and assuming a unit postage cost of 46p. For example, with treatment ACc5, 35.2% of units returned a card (Table 2), so the mean unit cost of the incentive is £5.00 × 0.352 = £1.76, whereas the return postage cost is £0.46 × 0.352 = £0.16, giving a total unit cost for these variable cost components of £1.92. Similarly, treatment ACu5 costs £5.19 and COA5 costs 71p. With a £2 incentive, the respective cost components are £2.18 for treatment ACu2, 79p for ACc2 and 37p for COA2. It is clear that for a given incentive amount the COA cards are considerably cheaper to administer than either treatment with AC cards. Given that few effects of amount of incentive were found (Section 7.2), we might restrict attention to the £2 level, noting that the unit cost of treatment COA at this level, 37p, is less than half the cost of ACc and about a sixth the cost of ACu.

We turn next to field costs, looking first at the number of visits made by interviewers to sample addresses. We decompose the total number of interviewer visits into two parts: the 'number of calls at issued address' and the 'number of calls at a new address'. Successful address updating strategies should reduce the number of the latter as the survey organization will, in most cases, be issuing the case to the current address (so there will be no additional 'new address' identified during fieldwork). In contrast, there could be an increase in the number of visits if it becomes possible to contact more mobile sample members, who may require greater effort to contact.

Table 4 shows the average number of visits of interviewers for each treatment group. The COA card treatments seem the most effective in reducing the number of calls to the issued address (mean 1.76, compared with 1.82 for ACu, 1.83 for ACc and 1.82 for NOC). Calls to an additional new address were most likely with ACu so, when these calls are added to the calls to the issued address, it is this treatment that appears to require the largest number of calls overall and the COA card again comes out best with a mean of 1.83 calls in total (compared with 1.92 for ACu and 1.88 for ACc and for NOC). In the case of ACu and COA, the difference reaches borderline significance ($p = 0.09$).

The amount of the incentive has a positive and statistically significant effect ($p = 0.008$) in reducing the total number of visits of interviewers only in the case of treatment ACu (a mean of 1.81 visits with ACu5 and 2.02 with ACu2).

Table 4. Comparisons between treatment groups in numbers of interviewer visits per sample member†

Treatment‡	(1), calls at issued address		(2), calls at new address		(3), total number of calls		Significance of pairwise differences in total calls for the following treatments:							
	Group size	Average number of calls	Group size	Average number of calls	Group size	Average number of calls	ACu2	ACc5	ACc2	ACc	COA5	COA2	COA	NOC
ACu5	1581	1.73	90	1.31	1581	1.81	**0.008**	0.167	0.233		0.269	0.488		0.159
ACu2	1560	1.90	93	2.12	1560	2.03		0.091	**0.035**		**0.034**	**0.007**		**0.033**
ACu	*3141*	*1.82*	*183*	*1.72*	*3141*	*1.92*				0.283			**0.087**	0.263
ACc5	1586	1.86	55	1.09	1586	1.89			0.370		0.348	0.166		0.416
ACc2	1566	1.79	81	1.38	1566	1.87					0.468	0.233		0.429
ACc	*3152*	*1.83*	*136*	*1.26*	*3152*	*1.88*							0.223	0.483
COA5	1570	1.78	77	1.62	1570	1.86						0.271		0.398
COA2	1548	1.74	87	1.21	1548	1.81								0.156
COA	*3118*	*1.76*	*164*	*1.40*	*3118*	*1.83*								0.227
NOC	3078	1.82	143	1.21	3078	1.88								

†Rows in italics are summary rows (i.e. each is a combination of the two preceding rows); numbers in bold have $p < 0.10$.

‡See Table 1 for a definition of treatments.

In the BHPS, tracing during fieldwork takes place in two ways. Interviewers first attempt to trace anyone who they find has moved when they call at the issued address for a household. If an interviewer finds a new address, she or he may simply visit the new address to seek an interview if the move is local. If the new address is out of their area, it will be reallocated through the field office to another interviewer. If the initial interviewer fails to obtain a new address, the case is returned to the office where additional tracing attempts are made using other information held about the respondent. This information includes, for example, details of all current and past stable contact names, various alternative telephone numbers for the respondent and details of people with whom the untraced respondent has co-resided in prior waves.

A second component of field costs is therefore telephone calls made by field office staff during the tracing process. Table 5 shows that treatment NOC is associated with the lowest average number of phone calls during office tracing: just 0.009 calls per sample member, compared with 0.030 with COA ($p = 0.018$), 0.026 with ACu ($p = 0.037$) and 0.019 with ACc ($p = 0.093$). Differences between the means for the other three methods are not statistically significant ($p > 0.10$). Successful address updating strategies could either increase the number of calls from the office (due to providing extra contact information for mobile sample members that would not otherwise have been available) or reduce the number of calls (by averting the need for office tracing in a proportion of cases where the correct updated address would not otherwise have been the address that was issued to the field). The first of these two phenomena may be dominant in our experiment. The NOC group may include sample members who had moved and not provided updated address details and who either did not enter office tracing (because they were simply a 'no contact' in the field with no indication that they had moved) or who entered tracing but could not be called as no telephone number was on the survey records.

COA cards therefore result in lower unit costs than AC cards as they cost less in direct implementational costs, result in slightly fewer interviewer field visits and are not different in terms of telephone calls by office staff. The NOC treatment is cheaper still. Considering that

(a) the proportion of sample members remaining untraced at the subsequent wave was lowest for treatment COA and

(b) COA is the only treatment which is significantly better than NOC in terms of the number of untraced movers,

we conclude that this treatment is the most cost effective of those compared.

7.4. Effects of tailored materials on response rates

We now analyse whether tailoring the between-wave respondent reports affects response rates at the subsequent interview. The increase in the cost of our improved reports was negligible both in terms of additional hours of work required and in terms of printing costs. Moreover, the tailored leaflets were smaller and lighter than the standard report, leading to a reduction in postage costs. Overall, the per-unit cost of the tailored respondent reports was slightly less than that of the standard report. In consequence, we focus on the effect of the treatment on response rate without further consideration of costs.

Table 6 compares the response rate at wave 18 for the two target groups at whom the tailored materials were aimed. We consider two alternative definitions of response rate. Column (1) reports the results when considering responders as just those completing a full face-to-face interview. Column (2) reports the results that are obtained when we include also those completing a telephone interview (see Section 3). Respondents were not offered the option of a telephone interview until all the usual means for obtaining a face-to-face interview had been exhausted.

Table 5. Comparisons between treatment groups in the number of tracing calls made per sample member†

Treatment‡	Sample members with tracing calls	Total sample members	Proportion with tracing calls (%)	Mean number of calls: those with calls	Mean number of calls: all	Significance of pairwise differences in total calls for the following treatments:							
						ACu2	ACc5	ACc2	ACc	COA5	COA2	COA	NOC
ACu5	17	1582	1.08	3.41	0.037	0.114				0.243	0.479		**0.037**
ACu2	5	1566	0.32	4.80	0.015		**0.072**	0.277		0.258	0.135		0.272
ACu	*22*	*3148*	*0.70*	*3.73*	*0.026*		*0.407*	*0.246*	*0.262*			*0.385*	*0.037*
ACc5	6	1587	0.38	3.33	0.013			0.167		0.168	**0.091**		0.335
ACc2	12	1567	0.77	3.33	0.026					0.464	0.304		**0.086**
ACc	*18*	*3154*	*0.57*	*3.33*	*0.019*							*0.172*	*0.093*
COA5	15	1573	0.95	2.53	0.024						0.270		**0.075**
COA2	12	1550	0.77	4.58	0.035								**0.051**
COA	*27*	*3123*	*0.87*	*3.44*	*0.030*								*0.018*
NOC	14	3097	0.45	2.07	0.009								

†Rows in italics are summary rows (i.e. each is a combination of the two preceding rows); numbers in bold have $p < 0.10$.
‡See Table 1 for a definition of treatments.

Table 6. Response rate, by tailored materials treatment and sample subgroup†

Tailored group		(1), face-to-face interviews only		(2), face-to-face and telephone interviews	
		n	*%*	*n*	*%*
Young	Tailored	843	93.24	843	94.07
	Standard	856	91.59	856	94.15
	Difference		1.65		−0.08
			p = 0.11		*p* = 0.53
Busy	Tailored	1205	90.29	1205	97.51
	Standard	1157	90.06	1157	96.54
	Difference		0.23		0.97
			p = 0.44		*p* = 0.08

†Standard error estimates take into account the clustering of mailing units within households through use of the `svy` commands in Stata 11.1.

For busy people—19.68% of the sample—the tailored materials did not affect the response rate when we consider only full face-to-face interviews. However, when we include telephone interviews, we see a statistically significant boost to the response rate, from 96.5% to 97.5% ($p = 0.08$). In other words, the tailored report appears to encourage some sample members who would not otherwise have responded at all at least to complete the telephone interview. As the telephone interview is shorter and can be scheduled at almost any time, it makes sense that this might be a relatively attractive option for busy people.

For young people—14.96% of the sample—the tailored materials did not affect the response rate when including telephone interviews, but the proportion of sample members completing the full face-to-face interview may have been positively affected: the response rate for the full face-to-face interview was 93.2% with the tailored report and 91.6% with the standard report ($p = 0.11$). In other words, the tailored report may encourage some young people who would otherwise have completed only the telephone interview instead to complete the full face-to-face interview.

8. Conclusion and discussion

Despite being one of the cheapest methods among those analysed, mailing a change-of-address card with an incentive conditional on return if a move occurs was most effective. This has been the standard BHPS strategy, and the experiment confirms that it is the most effective strategy both in collecting information on new addresses and in reducing the number of cases left untraced at the following wave. This strategy was more effective than mailing an address confirmation card to be returned by all sample members, regardless of the level or conditional nature of the associated incentive. The COA card treatment may be more likely than the AC treatment to be seen as relevant by sample members who have moved or are intending to move. Consequently, they are more likely to return updated address information, which in turn increases the chances of an interviewer locating them in the field.

We found no evidence that the amount of the incentive (£2 or £5) made a difference to the rate of response at the following wave. This was so for all three mailing strategies with which levels of incentive were tested, suggesting that the choice of strategy is more important than the value of the incentive.

As well as finding that the COA card was the most effective in reducing the number of people who remain untraced, we also find it to be the least costly treatment aside from the no-incentive treatment. We believe that the COA strategy is preferable to the no-incentive strategy for at least three reasons. First, given the maturity and relative compliance of our sample the relative advantage of the COA card could be greater with a sample of younger people. Second, in longitudinal surveys attrition rather than simple wave non-response is what really matters. A small effect at each wave could produce a large cumulative effect over several waves. Third, although effects on response rates at the following wave may be small, they are generally statistically significant and any effect on non-response bias could be important, given that the additional respondents are likely all to be movers, a subgroup which is of substantive importance. Movers have a high value to the survey as they have both distinct substantive characteristics and high non-response propensity, so a modest resource investment to retain them is justified.

Tailoring respondent reports could be a successful strategy to keep specific categories of lower response propensity respondents interested in the survey. Even in the context of a compliant sample and high wave-by-wave response rates, tailored materials increased the share of full face-to-face interviews among young people, and increased the overall rate of response among busy people. The effect, although positive for both target groups, is of a different nature in each case, reflecting the ways in which different sample members can change their response behaviour as a consequence of the treatment. The effects are sufficiently large to be worthwhile in practice.

Our findings have important practical implications for researchers designing and running longitudinal surveys; they shed new light on aspects of the survey non-response process and suggest promising avenues for further research. AC cards are simply not an efficient strategy. The relative success of COA cards at obtaining new addresses suggests that the emphasis of the message is important. New address information is of much greater value to the survey organization than a simple confirmation of an existing address; this message may have been diluted in the AC treatment.

It is interesting that manipulation of a relatively minor part of the survey process—the design and content of a respondent report—can have an effect on response rates, especially so when we consider that our experiment was carried out on a fairly seasoned sample with high wave-by-wave response rates, leaving limited scope for improvement. This is very encouraging, suggesting that even greater gains from such strategies may be possible in other circumstances, such as at the early stages of a longitudinal survey when respondents are not yet so committed or engaged. Further work is warranted in identifying the most promising subgroups for tailored materials and the most appropriate nature of the materials. In particular, we do not know which specific features of the tailored reports prompted response to the next wave. The tailored reports were smaller in size, with more colour, more pictures, more graphics and fewer words, as well as having different content. All that can be said at present is that tailoring seems promising; we cannot say much about how best to do it.

One might also speculate whether there might be some crossover between our two experiments, in the sense that one might consider tailoring the approach that is used to collect updated address information, by targeting groups at heightened risk of changing address with more expensive or intensive methods (e.g. two mailings between each wave instead of just one), or by using different designs of letters and cards. Moreover, experiments like these could usefully be carried out at a much earlier stage of a longitudinal survey to test effects in a different context. There is plenty for researchers yet to learn about how best to maximize sample retention in longitudinal surveys.

Acknowledgements

References

Abraham, K. G., Maitland, A. and Bianchi, S. M. (2006) Nonresponse in the American time use survey: who is missing from the data and how much does it matter? *Publ. Opin. Q.*, **70**, 676–703.

Behr, A., Bellgardt, E. and Rendtel, U. (2005) Extent and determinants of panel attrition in the European Community Household Panel. *Eur. Sociol. Rev.*, **21**, 489–512.

Burton, J., Laurie, H. and Lynn, P. (2006) The long-term effectiveness of refusal conversion procedures on longitudinal surveys. *J. R. Statist. Soc.* A, **169**, 459–478.

Church, A. H. (1993) Estimating the effect of incentives on mail survey response rates: a meta-analysis. *Publ. Opin. Q.*, **57**, 62–79.

Cohen, A. S., Patrick, D. C. and Shern, D. L. (1996) Minimizing attrition in longitudinal studies of special populations: an integrated management approach. *Evaln Program Planng*, **19**, 309–319.

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. New York: Wiley.

Fitzgerald, J., Gottschalk, P. and Moffit, P. (1998) An analysis of sample attrition in panel data: the Michigan Panel Study of Income Dynamics. *J. Hum. Resour.*, **33**, 251–299.

Freedman, D., Thornton, A. and Camburn, D. (1980) Maintaining response rates in longitudinal studies. *Sociol. Meth. Res.*, **9**, 87–89.

Groves, R. M. (2006) Non-response rates and non-response bias in household surveys. *Publ. Opin. Q.*, **70**, 646–675.

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

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

Groves, R. M., Couper, M., Presser, S., Singer, E., Tourangeau, R., Piani Acosta, G. and Nelson, L. (2006) Experiments in producing non-response bias. *Publ. Opin. Q.*, **70**, 720–736.

Groves, R. M., Singer, E. and Corning, A. (2000) Leverage-saliency theory of survey participation: description and an illustration. *Publ. Opin. Q.*, **64**, 299–308.

James, T. L. (1997) Results of the wave 1 incentive experiment in the 1996 Survey of Income and Program participation. *Proc. Surv. Res. Meth. Sect. Am. Statist. Ass.*, 834–839.

Kalton, G. (2009) Surveys across time. In *Sample Surveys: Inference and Analysis* (eds C. R. Rao and D. Pfeffermann). New York: Elsevier.

Laurie, H. and Lynn, P. (2009) The use of respondent incentives on longitudinal surveys. In *Methodology of Longitudinal Surveys* (ed. P. Lynn), pp. 205–233. New York: Wiley.

Laurie, H., Smith, R. and Scott, L. (1999) Strategies for reducing nonresponse in a longitudinal panel survey. *J. Off. Statist.*, **15**, 269–282.

Lepkowski, J. M. and Couper, M. P. (2002) Nonresponse in longitudinal household surveys. In *Survey Nonresponse* (eds R. M. Groves, D. Dillman, J. Eltinge and R. J. A. Little), pp. 259–272. New York: Wiley.

Lillard, L. A. and Panis, C. W. A. (1998) Panel attrition from the Panel Study of Income Dynamics. *J. Polit. Econ.*, **94**, 489–506.

Lynn, P. (ed.) (2006) *Quality Profile: British Household Panel Survey, Version 2.0: Waves 1–13, 1991-2003*. Colchester: University of Essex. (Available from http://www.iser.essex.ac.uk/bhps/quality-profile.)

Lynn, P. and Clarke, P. (2002) Separating refusal bias and non-contact bias: evidence from UK national surveys. *Statistician*, **51**, 319–333.

Lynn, P., Thomson, K. and Brook, L. (1998) An experiment with incentives on the British Social Attitudes survey. *Surv. Meth. Cent. Newslett.*, **18**, 12–14.

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. *J. Off. Statist.*, **27**, 319–338.

Petrolia, D. R. and Bhattacharjee, S. (2009) Revisiting incentive effect: evidence from a random sample mail survey on consumer preferences for fuel ethanol. *Publ. Opin. Q.*, **73**, 537–550.

Ribisl, K. M., Walton, M. A., Mowbray, C. T., Luke, D. A., Davidson II, W. S. and Bootsmiller, B. J. (1996) Minimizing participant attrition in panel studies through the use of effective retention and tracing strategies: review and recommendations. *Evaln Program Planng*, **19**, 1–25.

Rodgers, W. (2002) Size of incentive effects in a longitudinal study. *Mimeo*. Survey Research Centre, University of Michigan, Ann Arbor.

Scott, C. K. (2004) A replicated model for achieving over 90% follow-up rates in longitudinal studies of substance abusers. *Drug Alc. Depend.*, **74**, 21–36.

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), pp. 163–177. New York: Wiley.

Singer, E., Gebler, N., Raghunathan, T., van Hoewyk, J. and McGonagle, K. (1999) The effect on incentives in telephone and face-to-face surveys. *J. Off. Statist.*, **15**, 217–230.

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

Taylor, M. F. (ed.) (2010) *British Household Panel Survey User Manual*, vol. A, *Introduction, Technical Report and Appendices*. Colchester: University of Essex. (Available from `http://www.iser.essex.ac.uk/bhps/documentation/pdf_versions/`.)

Uhrig, S. C. N. (2008) The nature and causes of attrition in the British Household Panel Study. *Working Paper 2008-05*. Institute for Social and Economic Research, University of Essex, Colchester. (Available from `http://www.iser.essex.ac.uk/publications/working-papers/iser/2008-05`.)

Watson, N. and Wooden, M. (2009) Identifying factors affecting longitudinal survey response. In *Methodology of Longitudinal Surveys* (ed. P. Lynn), pp. 157–181. New York: Wiley.

Zabel, J. E. (1998) An analysis of attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation with an application to a model of labor market behavior. *J. Hum. Resour.*, **33**, 479–506.