The role of email addresses and email contact in encouraging web response in a mixed mode design

Alexandru Cernat and Peter Lynn

Institute for Social and Economic Research, University of Essex, UK

Abstract

As response rates are decreasing and researchers and survey agencies are pressed to save costs, mixed modes designs, especially those including CAWI, are becoming increasingly used. This is also true for ongoing longitudinal surveys such as the British Birth Cohort Studies and Understanding Society. Sending an email invitation to participate by CAWI, as opposed to a mail invite, may make CAWI participation easier for the respondent and reduce costs for the survey organisation. But an email invitation is only possible when the sample member has already provided a valid email address. In this paper we assess the importance of the invite being sent by email, and hence of collecting valid email addresses, in the context of a longitudinal survey in which the sample members have already participated face-to-face at previous waves. Using the Innovation Panel, a methodological survey that is part of Understanding Society, we examine the influence of having provided an email address on a sample member's propensity to respond by CAWI. We also assess the extent to which this influence varies over sample subgroups, and the extent to which it is moderated by how recently the email address was provided, by whether other household members had provided an email address, and by the presence at home of internet and broadband access.

Introduction

This paper is concerned with the extent to which the propensity to participate in a CAWI-CAPI mixed-mode survey is influenced by the ability to contact the sample member by email. The context is a panel survey in which researchers have the opportunity at previous waves to ask sample members to supply an email address. There is a lack of evidence regarding the value of respondent-supplied email addresses. This makes it difficult to make informed decisions about the efforts that should be made to try to collect such email addresses or about how these email addresses should subsequently be used in the process of contacting sample members and encouraging them to take part in the survey. Our paper attempts to fill that knowledge gap by providing information that will be of practical use to researchers planning mixed mode panel surveys including web data collection.

Data

Our data come from the first five waves of the Understanding Society Innovation Panel (Uhrig, 2011). At wave 5 an experiment was carried out, in which two-thirds of sample persons were allocated to a mixed mode protocol. With this protocol, they were first invited to take part in a web survey. After two weeks, each sample person who had not yet responded to the web survey was issued to a face-to-face interviewer who visited the respondent's home address to attempt a CAPI interview. It is the outcome of the attempts to gain response from this two-thirds of the sample at wave 5 that is the focus of this paper. The other one-third of sample persons were allocated to a single-mode CAPI protocol.

At each wave, in the course of the survey interview each sample person was given the opportunity to provide an email address. Our analysis draws upon indicators of whether or not an email address was provided at each wave and also upon a range of classificatory variables collected at waves prior to wave 5.

Two other features of the survey design are pertinent to our analysis. First, the intention is to interview all members of the sample person's household at each wave. Consequently, even if a particular sample person did not supply an email address it is possible that another member of the household might have supplied an email address. We are therefore able to take this into account in our analysis and to consider intra-household dynamics in the supply of email addresses and the effect on subsequent response. Second, the Innovation Panel sample consists of two components: the original sample, for which this was their fifth wave, and a refreshment sample introduced at wave 4, for which this was their second wave. We can therefore compare effects between sample persons at an earlier and later stage of panel membership.

Research Questions

There are two separate but related outcomes of interest:

Whether a sample person participates in any mode. The overall response rate and response composition is of great importance, so the effect of a design feature on overall response propensity is of central interest.

Whether a sample person participates by web. A greater propensity to participate by web indicates a greater potential for savings in data collection costs. Any design feature, such as collecting an email address, that increases the propensity to participate by web will be of interest to survey managers;

To understand the distinction between the two outcomes, and why we must be concerned with both, consider the following scenario. Suppose that amongst one sample subgroup a treatment (e.g. obtaining an email address) increases the propensity of completing by web conditional on participating in any mode but does not affect the propensity of participating in any mode. Meanwhile, amongst another subgroup the treatment reduces the propensity of participating in any mode. In this scenario it would be possible for the treatment to have a positive effect on the proportion of sample members participating by web but a negative effect on the overall response rate. Such a treatment may therefore be undesirable despite the positive effect on the first outcome.

The treatments of interest are the following:

An attempt to collect an email address from each sample person;

Repeated attempts at subsequent waves to collect an email address from sample persons;

Attempts to collect email addresses from other members of the sample person's household.

For each of these treatments, if an email address is collected successfully, it is then used to contact the sample person to invite them to participate in the wave 5 web survey and, if necessary, for up to three reminder emailings. It is the overall effect of collecting the email address and using it in this way in which we are interested.

We are also interested in the effect of how recently prior to the wave 5 invitation the email address was collected and in interactions with other characteristics of the sample persons (including the presence at home of internet and broadband access). Estimation of interactions will allow us to identify whether there are particular sample subgroups for whom the treatments are either effective, on the one hand, or detrimental, on the other hand.

Thus, we can summarise our research questions as follows:

Do each of the treatments affect the propensity of sample persons to participate in any mode?

Do each of the treatments affect the propensity of sample persons to participate by web?

Are any of the effects on either outcome moderated by characteristics of sample persons or by the nature of the sample persons' response to the request for an email address (how recently they supplied an email address, whether other members of their household did so)?

Methods

In order to answer our research questions, three logistic regression models are developed:

1. The first model predicts survey participation conditional on being in the mixed mode design. This allows us to test the effect of having an email address, and interactions between this and other respondent characteristics, in the mixed mode context in which we are interested.
2. The second model will explain participation based on the full sample (single mode and mixed mode groups). This allows us to test interactions between having an email address and mode treatment, which should help us to understand whether having an email address is simply an indicator of a generally co-operative sample member or whether it actually aids the response process (in which case the effect should only be observed for the mixed mode group).
3. The third model predicts response mode conditional on participation, based on the mixed mode group alone. This allows us to address the research questions regarding data collection costs.

Results

<u>Model 1 – propensity to respond (be it CAPI or CAWI) within a mixed mode design (Table 3):</u>

- Giving the email address in a previous wave increases your chances of participating in a mixed mode design. This effect is also moderated by the tendency of the partner to give the email address. Thus, if both partners give their emails it increases the odds of giving a full response in a mixed mode design by 2 compared to when none of them provides it.

- Urban dwellers, white British, people who own their homes, people who have broadband at their home and those part of the refreshment sample[1] have higher propensity to give full response in a mixed mode design. People that are part of households with 4 or more members are less likely to give a full response.
- Those who indicated preference for CAWI or no mode preference in the previous waves have a smaller propensity of giving a full interview.
- Giving an email in previous waves moderates the effects of home ownership on participation. The difference to those that rent their home is reduced if the respondent gives the email address.
- There is no significant interaction between giving an email address and any of the other significant predictors listed above, suggesting that the effect of the email address is fairly uniform across the sample

Model 2 – propensity to respond in any mode design (be it single mode or mixed mode) and in any mode (be it CAPI or CAWI) (Table 4):

- While giving an email in a previous wave increases your chances of giving a full interview overall (odds ratios of approximately 3) it does not have an additional effect on propensity to participate in a mixed mode design. Actually the opposite effect is present, giving an email in a previous wave decreases your propensity to participate in a mixed mode design, although the effect is not statistically significant. This suggests that giving an email address is simply an indicator of being a co-operative respondent, as it is equally associated with increased response propensity in a design that makes no use of the email address (single-mode face-to-face) as in the mixed mode design involving emailed invitations to a web survey.
- Having email addresses from the partner or from other members of the household does not increase the likelihood of giving a full interview in either mode design.
- Indicating a preference for CAWI in a previous wave decreases your overall propensity to participate in the survey in either single mode or mixed mode (and interaction with mode design is not significant).

Model 3 – propensity to respond by CAWI conditional on full response within a mixed mode design (Table 5):

- People who gave the email in wave 4 are 3 times more likely to answer CAWI than the rest. This effect is smaller for those in the original sample and for those for whom the partner did not provide an email.
- For people who didn't give an email in a previous wave, the probability of participating in CAWI doubles if their partner gave an email address.

[1] The refreshment sample received higher-value incentives, so this result could be picking up an incentive effect rather than a time-in-sample effect. We will attempt to separate these two effects in further analysis.

- Using the internet and having access to internet at home increases the chances of participating with CAWI.
- Those that indicated preference for CAWI and those that said they are likely to respond using the web have higher probabilities of answering using CAWI.
- People in couples are more likely to answer CAWI.

Using predicted probability for targeting

One of the aims of the analyses was to find a mechanism of targeting the respondents that are most likely to answer using CAWI. We divided the predicted probabilities from model 3 into deciles and looked at the proportion of respondents responding by CAWI in each decile. Thus, Table 1 highlights that out of the top 10% most likely CAWI respondents 92% participated using web. At the other extreme only 13.5% respond using web.

Table 1. Percentage of responses in each mode given the deciles of the predicted probabilities

	Observed response	
Deciles	**F2F**	**CAWI**
1	86.5	13.5
2	71.9	28.1
3	65.2	34.8
4	46.1	53.9
5	39.8	60.2
6	30.3	69.7
7	25.8	74.2
8	32.6	67.4
9	16.9	83.2
10	8.0	92.1
Total	42.3	57.7

Extending this targeting mechanism we can see how choosing different deciles to include in a mixed mode design influences both money saving (approximated by the percentage of respondents that answer using CAWI) and overall response rates.

The last row of Table 2 shows the effect of selecting the most likely 10% to respond by CAWI (based on model 3) for a mixed mode design (and using a single mode CAPI design for the other 90% of sample members). Using this strategy would lead to response rates of 84.2% and a proportion of the total sample using CAWI of 9.3%. Including the top two quintiles in a mixed mode design would result in a slightly smaller response rate, 83.7%, but a bigger proportion of respondents using CAWI (17.8%). As we include more and more deciles in the mixed mode design the relative proportion of people who answer using CAWI increases but the overall response rates decrease reaching a response rate of 80.5% when the entire sample is in mixed mode and 58.6% of these responding using CAWI.

We must caution that these predictions are based on the results of our survey, where the mixed mode design showed lower response rates than the single mode. In addition, they are based on our regression models which are not perfect predictions. Nevertheless, our results indicate that targeting can lead to cost saving but at the expense of decreasing response rates and other potential problems such as measurement issues, which we ignore here.

Table 2. Response rates and % of respondents using CAWI with different targeting strategies

	Predicted propensity to respond	
Deciles	**Response rates**	**% with CAWI**
>= 1	80.5%	58.6%
>= 2	80.5%	57.2%
>= 3	81.0%	54.2%
>= 4	81.5%	50.0%
>= 5	81.9%	45.1%
>= 6	82.4%	39.3%
>= 7	82.8%	32.8%
>= 8	83.2%	25.5%
>= 9	83.7%	17.8%
10	84.2%	9.3%

Points of discussion

- What other potential variables can help target the most likely CAWI respondents?
- Possible cut-off points for most likely respondents by CAWI
- Other recent evidence on effect of email vs. mail approach for web interview (especially in longitudinal context)?
- Experience of other surveys regarding collection of email addresses from respondents

References

Uhrig, S. C. N. (2011) Using experiments to guide decision making in *Understanding Society*: Introducing the Innovation Panel, chapter 13 in S. L. McFall & C. Garrington (ed.s), *Understanding Society: Early Findings from the First Wave of the UK's Household Longitudinal Study*, Colchester: University of Essex. At: http://research.understandingsociety.org.uk/findings/early-findings

Annex

Table 3. Propensity to give full response (conditional on a mixed mode design)

Variable		Odds ratios
	Email[2]	3.38**
Email * partner email	**No * Yes**	0.99
	Yes * Yes	2.21***
	Own home	3.15***
	Own home * email	0.36*
	Urban	1.50*
	Female	1.14
Education (ref. = Higher)	**Other higher**	1.44
	A levels	0.87
	GCSE or CSE	1.00
	Vocational	0.99
Age (ref. = 16-19)	**20-29**	0.55
	30-39	0.87
	40-49	1.05
	50-59	1.24
	60-69	1.97
	Over 70	1.33
	White British	1.65*
	Employed	0.96
	Has mobile	1.48
	Internet use	0.92
Mode preference (ref. = CAPI)	**CATI**	0.83
	Postal	0.90
	CAWI	0.54**
	No preference	0.33+
	How likely on web	1.04
HH size (ref. = 1)	**2**	0.92
	3	0.89
	4	0.46*
	5 or bigger	0.44*
	Broadband	5.63***
	In couple	1.08
	Refreshment sample	1.40+
R-squared: 0.13		
N. of cases: 1120		
+ .10 * .05 ** .01 * .001**		

[2] Includes also those that gave an email in wave 5

Table 4. Propensity to give full response (includes both single mode and mixed modes)

Variable		Odds ratios
	Mixed mode	0.62
HH emails (ref. = No emails)	**Some emails**	0.75
	All emails	0.89
	Partner email	1.15
	Email	2.99*
	Mixed mode * email	0.50
HH emails * Mixed mode (ref. = No emails * MM)	**Some emails * Mixed mode**	1.68
	All emails * Mixed mode	2.38
	Urban	1.24
	Female	1.20
Education (ref. = Higher)	**Other higher**	1.55
	A levels	0.94
	GCSE or CSE	1.13
	Vocational	1.04
Age (ref. = 16-19)	**20-29**	0.76
	30-39	1.07
	40-49	1.21
	50-59	1.46
	60-69	2.09*
	Over 70	1.72
	White British	1.65*
	Employed	0.86
	Has mobile	1.09
	Internet use	0.94
Mode preference (ref. = CAPI)	**CATI**	0.95
	Postal	0.67*
	CAWI	0.55***
	No preference	0.44+
	How likely on web	1.00
	Own home	1.64**
	HH size	0.86**
	Broadband	3.02**
	In couple	1.43+
	Refreshment sample	1.38*
R-squared: 0.12		
N. of cases: 1729		
+ .10 * .05 ** .01 * .001**		

Table 5. Propensity to respond in CAWI (conditional on giving a full interview in a mixed mode design)

Variable		Odds ratios
Email * partner email (ref. = No * No)	No * Yes	2.05+
	Yes * No	0.61+
	Yes * Yes	0.83
	Email in wave 4	3.33***
	Refreshment sample	1.44
	Original sample * email in w4	0.46+
	Urban	1.25
	Female	1.08
Education (ref. = Higher)	Other higher	0.77
	A levels	1.29
	GCSE or CSE	0.67+
	Vocational	0.70
Age (ref. = 16-19)	20-29	0.55
	30-39	0.66
	40-49	0.91
	50-59	0.86
	60-69	0.65
	Over 70	0.53
	White British	1.42
	Employed	0.92
	Has mobile	1.27
	Internet use	1.31***
Mode preference (ref. = CAPI)	CATI	1.75
	Postal	1.23
	CAWI	1.49+
	No preference	1.50
	How likely on web	1.11***
	Own home	2.37***
	HH size	0.93
	Internet at home	4.58+
	In couple	1.73*
Pseudo R-squared: 0.18		
N. of cases: 888		
+ .10 * .05 ** .01 *** .001		