Pushing Household Panel Survey Participants from CAPI to Web

Peter Lynn, ISER, University of Essex

Paper presented at the 28th International Workshop on Household Survey Nonresponse, Utrecht, August-September 2017

Abstract

Understanding Society: the UK Household Panel Study began life as a CAPI-only survey. Mainly for budgetary reasons, the study has introduced online interviewing (web questionnaires) as part of a mixed-mode design, with contractual targets to increase at each wave the proportion of sample members initially invited to take part online and the proportion of households in which all sample members participate online. The objective is to move as many sample households as possible from CAPI to web, without damaging retention rates or representativity (or measurement - but that's another story). Online interviewing was first introduced at wave 7, but only for individuals in households that did not participate at wave 6. The proportion to be initially invited to participate online was 40% at wave 8 and will be 60% at wave 9 and 70% at wave 10. The invitation to participate online has been targeted at sample households who are predicted to be relatively likely to participate online, and avoiding individuals whose participation propensity is predicted to be substantially lowered by an online invitation. The targeting strategy relies on a series of statistical models of participation, derived from experimental Innovation Panel data, and information provided by respondents at previous waves. This paper explains the objectives of the targeting, describes how it is being implemented, and illustrates the resultant characteristics of sample members invited to participate online and how these contrast with the remaining sample members.

1. Context: Understanding Society

Understanding Society is a large panel survey with a sample size of around 100,000 persons (at wave 1);

Data are collected annually from each sample member;

The sample is divided into 24 monthly samples, so field work for each wave takes a little over two years;

Wave 1 was collected entirely by in-home CAPI interviews;

From waves 2 to 6, around 1.8% of interviews were carried out by telephone, with the remaining 98.2% by CAPI. The telephone interviews were with sample members who most likely would not otherwise have participated.

2. Introduction of Web

Reflecting uncertainties about the possible impacts of introducing web as a main survey mode, and responding to the wishes of the survey's governmental co-sponsors, it was decided that a random 20% of the sample would be designated, indefinitely, as a ring-fenced "CAPI-only" sample.

To maximise cost savings, it was decided that web, when used, would be used as the first phase in a sequential design, with CAPI follow-up.

Amongst the other 80% of the sample, web was introduced for the first time at wave 7, but only in a very limited way. Sample members in households in which no-one had participate at wave 6 (about 8% of the sample) were administered a "web-first" sequential protocol, in which non-respondents following an initial 2-week web-only period were followed up by CAPI.

At wave 7, around 4% of all completed household enumeration grids, 3% of all completed household interviews, and 5% of individual interviews, were completed by web.

3. Large-scale targeted use of Web

At wave 8, the objective was to administer a web-first protocol to 50% of the non-ring-fenced households (i.e. 40% of the overall sample),

This was to be done in a targeted way (Lynn 2017), with the main objectives being to maximise cost savings and response rates. These objectives were articulated as follows:

- Maximise the proportion of households in which all survey instruments (household questionnaire and all individual questionnaires) are completed (promptly) online, thereby avoiding the need to send an interviewer to the address;
- Maximise the proportion of sample members (persons) who complete the individual questionnaire (regardless of mode of completion).

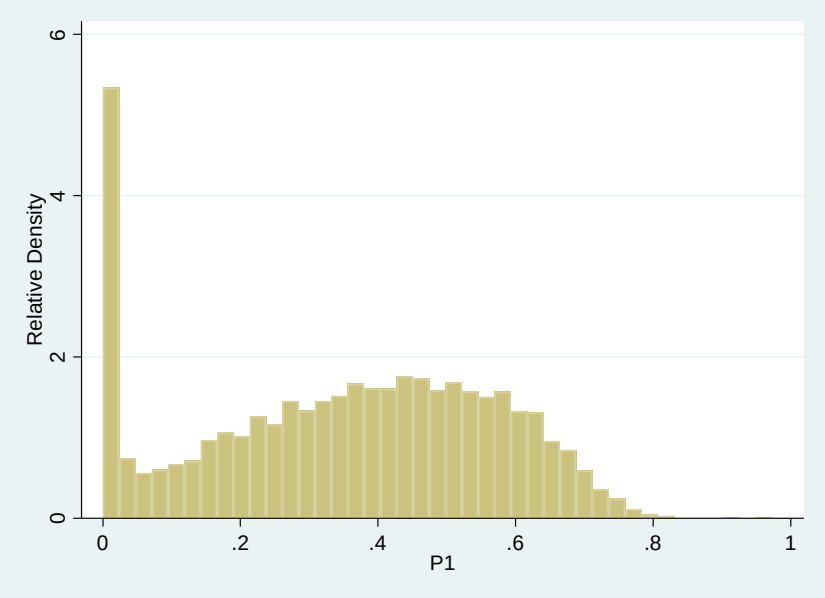
To inform the targeting we therefore build statistical models to predict:

- The probability of a household completing all survey instruments online ($P1$);
- The probability of an individual completing the individual questionnaire with the CAPI-only design ($P2$);
- The probability of an individual completing the individual questionnaire with the sequential mixed-mode design ($P3$).

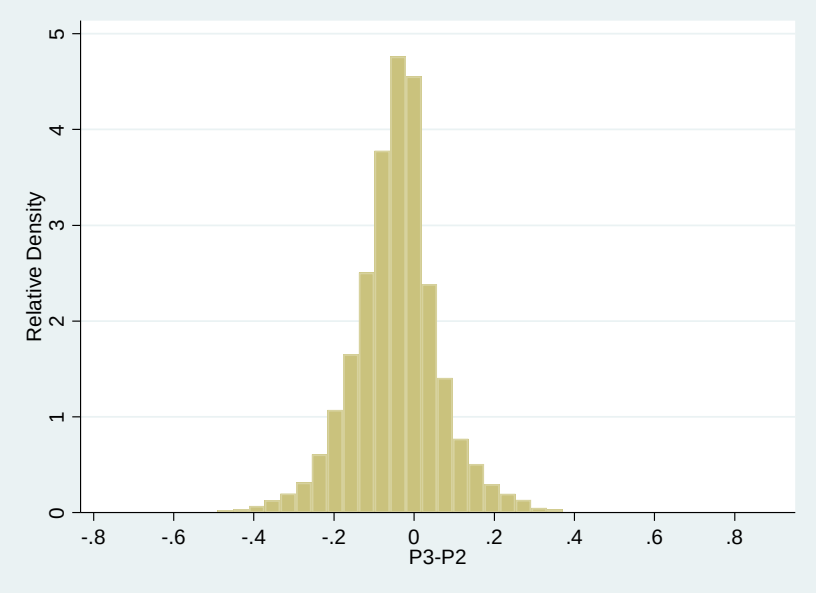
For each household in the wave 8 sample, we then computed the model-based estimate of a) $P1$ and b) $Min(P3 - P2)$, as the targeting objectives correspond (in some sense) to maximising both $\overline{P1}$ and $\overline{Min}(P3 - P2)$ amongst households administered the mixed-mode protocol.

The models were based on an experimental allocation to the two mode protocols under consideration at wave 5 of the Understanding Society Innovation Panel (n=2,100 persons). Predictor variables were taken from wave 4 data but restricted to questionnaire items that were also included in the main survey wave 5, so that the model could be fitted to the sample to be issued to wave 8 (wave 5 was the most recent wave for which complete data was available at the time that the targeted allocation had to begin).

$P1$ has an interesting distribution, with a sizeable number of households have a predicted probability close to zero:



The distribution of $(P3 - P2)$ shows that considerable numbers of sample members are predicted to be substantially more likely to respond with one protocol than the other and that this is true of both protocols (exactly what one would hope for if one believes that mixed-mode designs can harness the strengths of each mode to bring about improved outcomes):



Some of the strongest predictors in the models were indicators of whether or not the sample members had provided email addresses at previous waves, whether they were regular web users, housing tenure and household type.

The continuous nature of the survey fieldwork necessitates that allocation must take place each calendar quarter. Thus, the first quarterly sample was allocated to mode protocols in November 2015 (for fieldwork start in January 2016), while the eighth and final quarterly sample was allocated in August 2017 (for fieldwork start in October 2017).

The allocation process for the first four quarterly samples involved the following steps:

1. Ring-fenced sample (20%, on ave) → CAPI-only;
2. Remaining wave 7 non-responding hholds (6.5%, on ave) → Web-first;
3. Remaining households split into "high" and "low" web-propensity strata based on the modelled values:
4. All "low-propensity" hholds → CAPI-first;
5. A random subset of "high-propensity" hholds → web-first (proportion such that 40% in total web-first), remainder CAPI-only.

The random allocation at step 5 enabled assessment of the success of the targeted allocation in meeting the objectives. Analysis of early outcome data from the first two quarterly samples suggested a monotonic relationship between $P1$and the proportion of households responding fully by web, and no negative effect of targeting (compared to CAPI-only) on response rates, at least amongst this sub-stratum. Consequently, for the second four quarterly samples the allocation process was revised to remove random allocation and thereby further improve (hopefully)the extent to which the objectives were met (revisions in red font):

1. Ring-fenced sample (20%, on ave) → CAPI-only;
2. Remaining wave 7 non-responding hholds (8.4%, on ave) → Web-first;
3. Remaining households split into (different) "high" and "low" web-propensity strata based on the modelled values:
4. All "low-propensity" households → CAPI-first;
5. All "high-propensity" households → web-first

This involved changing the definition of "high web propensity", thus:

Year 1: $P1 > 0.019$ & $(P3 - P2) > -0.14$

Year 2: $P1 > 0.416$ & $(P3 - P2) > -0.10$

4. Characteristics of sample allocated to each protocol

Comparing the "high web propensity" (web-first protocol) to the "low web propensity" (CAPI-only protocol) we find (see graphs in annex) that the high propensity group contains:

- a *slightly higher* proportion of men (but the difference is small);
- a *much higher* proportion of 50-69 year-olds and much *lower* proportions of 16-29 year-olds and those aged 70 or older;
- *very few* of those who say they rarely or never use the internet;
- a *much higher* proportion of people with a first degree or higher qualification.

5. Outcomes

Based on partial data (wave 8 will still be in the field until Spring 2018), current estimates are:

- Proportion of web-first households completing fully online is around 39%;
- Proportion of wave 7 respondents responding at wave 8 is around 88% amongst the CAPI-first sample and 91% in the web-first sample;
- Of the web-first sample, around 64% respond by web, 25% by CAPI, and 2% CATI;
- Of the CAPI-first sample, 84% respond by CAPI, 4% by web, and 1% by CATI.
- Response rate amongst wave 7 non-respondents notably higher in the web-first sample than in the ring-fenced CAPI-only sample.

Though there are several caveats and uncertainties regarding the above figures, we are fairly confident that the objectives of the targeting are being achieved!

6. Discussion points

Ways to improve the targeting?

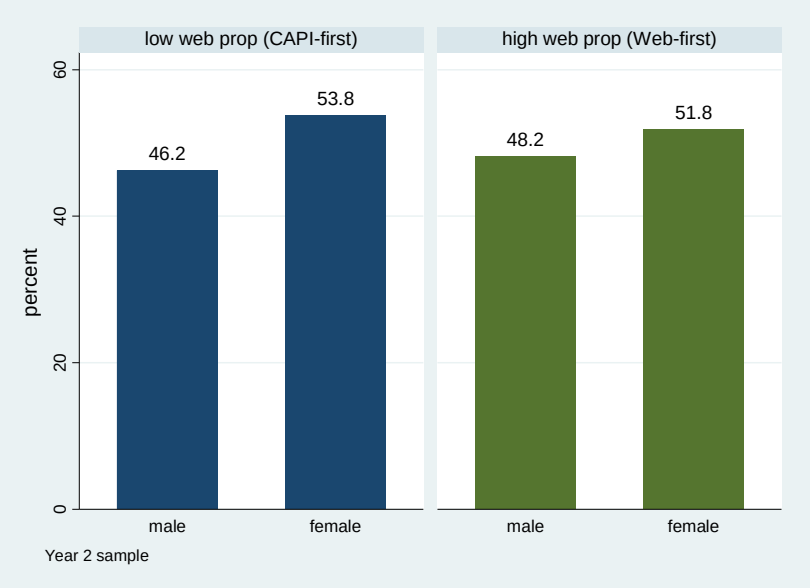
How should the allocation models / process be adapted for future waves?

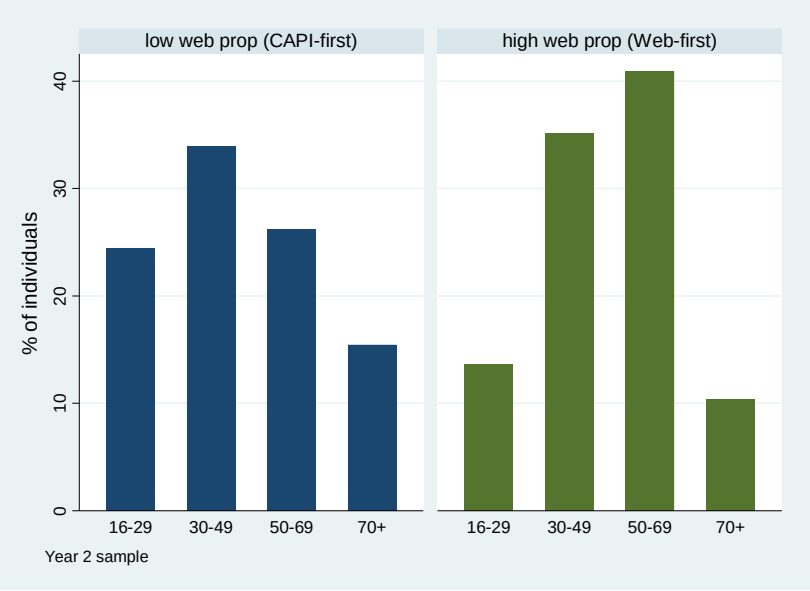
Specifically, how should mode of response (and other outcomes?) at wave 8 affect allocation at wave 9, etc?

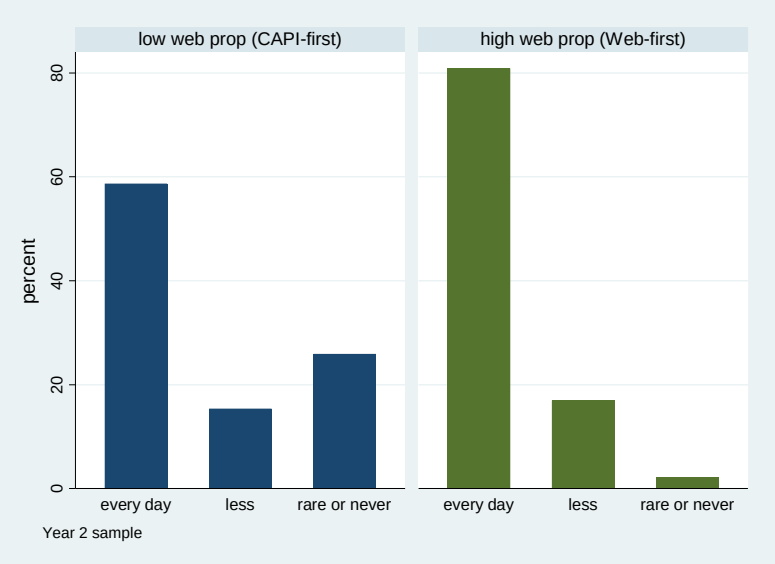
What generalisable knowledge could usefully be generated from our random allocation in year 1 of wave 8?

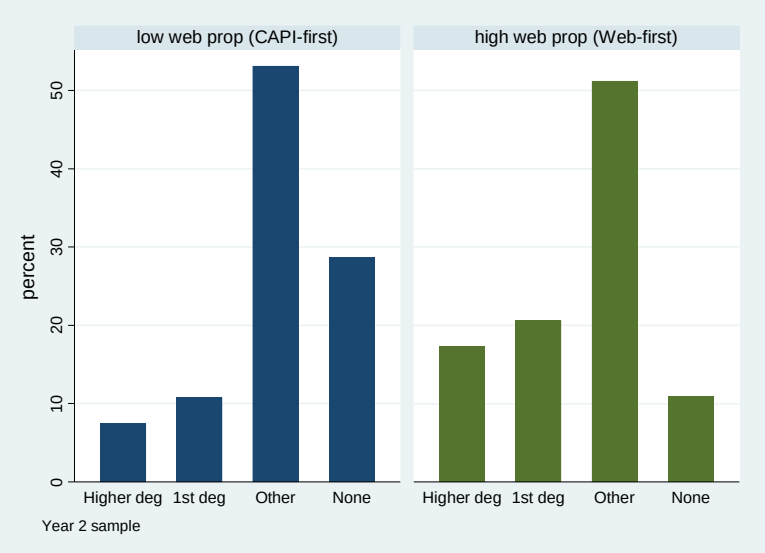
Any other comments about future directions for targeted allocation to mode protocols?

Annex: Graphs comparing CAPI-first and web-first groups






low web prop (CAPI-first)
high web prop (Web-first)
percent
0
20
40
60
80
every day
less
rare or never
every day
less
rare or never
Year 2 sample


low web prop (CAPI-first)
high web prop (Web-first)
percent
0
10
20
30
40
50
Higher deg
1st deg
Other
None
Higher deg
1st deg
Other
None
Year 2 sample