Comparing response rate maximisation to alternative allocation criteria in adaptive design for panel surveys

Olena Kaminska, ISER, University of Essex

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

1. Introduction

In recent years, researchers have begun to allocate survey samples to data collection protocols in a targeted way (Lynn 2017);

The criteria for targeting have not yet moved much beyond response rates, on the one hand, and costs/effort on the other (e.g. Freedman et al 2017);

We consider three alternative criteria: R-indicator, coefficient of variation of response probabilities, and effective sample size.

2. Research Questions

1) How can survey samples be allocated to data collection protocols according to different allocation criteria?

2) How different will the allocation be depending on which criterion is applied?

3) How will outcomes vary between the allocation scenarios?

4) Can adaptive designs out-perform one-protocol-for-all designs in terms of these outcomes?

3. Data: UK Household Longitudinal Study Innovation Panel

Randomised experiment implemented in wave 5 of IP in 2012:

- CAPI single-mode (one-third of sample): Advance letter, followed by 14 weeks of CAPI fieldwork;
- Web-CAPI sequential mixed-mode (two-thirds of sample): letter with invitation to web survey (for those with email addresses, same letter was sent by email); reminder letter (or two email reminders). After two weeks, 14 weeks of CAPI fieldwork with all remaining non-respondents.

Response rate (amongst wave 4 participants) was 84.6% with the CAPI single-mode protocol and 79.4% (of which 39.9% responded via web mode) with the mixed-mode protocol (Jäckle et al., 2015).

4. Our Study

We simulate six different allocation scenarios, using the IP experimental data. Four scenarios involve targeted allocation to one of the two data collection protocols (static adaptive designs), while the other two are single-protocol scenarios.

Each allocation scenario draws on predicted probabilities of participation under each data collection protocol. The first step is therefore to estimate probabilities of participating in each protocol for each sample element. To do this we use demographic and substantive variables from IP wave 4 as predictors (wave 4 was CAPI single-mode), selected as being either theoretically related to non-response, empirically associated with panel attrition or observed to distinguish between web and CAPI response in a mixed-mode context. We use the same set of covariates for all three models.

Using logistic regression we predict three propensities: the probability of participating,

(a) in the CAPI single-mode protocol (based on the CAPI experimental sample),

(b) in the mixed-mode protocol (based on the mixed-mode experimental sample), and

(c) by web (based on the mixed mode experimental sample).

Model-predicted values are then calculated for all three models for both experimental groups. Thus, for all sample elements, regardless of their experimental assignment, we obtain three predicted propensities:

p_{ij} is the probability of element i participating if allocated to protocol j ($j = 1$ indicates CAPI single-mode protocol; $j = 2$ indicates mixed-mode protocol);

q_{i2} is the probability of element i participating in web mode if allocated to protocol 2.

The $\{p_{ij}\}$ are used to determine the allocation of each sample element to a data collection protocol under each scenario, and all three probabilities are used subsequently to predict the outcomes of each scenario.

We denote $I_{ijk} = 1$ if element i is allocated to protocol j under scenario k, for $k = 1$ to 7.

Consequently, the response propensity of element i in scenario k will be $\rho_{ik} = \sum_{j=1}^{2} I_{ijk} p_{ij}$.

5. The Allocation Scenarios

Two single-protocol scenarios:

"CAPI-for-all" (k=1): $I_{i11} = 1$ and $I_{i21} = 0 \ \forall\, i$;

"MM-for-all" (k=2): $I_{i12} = 0$ and $I_{i22} = 1 \ \forall\, i$.

Random allocation scenario (k=3) reflects the actual experiment at IP wave 5. For this scenario, outcomes such as sample composition are observed (rather than being simulated using model-predicted probabilities).

Response-rate (RR) optimization (k=4):

Aim is to maximise the response rate. Hence:

$I_{i14} = 1$ and $I_{i24} = 0$ if $p_{i1} > p_{i2}$; $I_{i14} = 0$ and $I_{i24} = 1$ otherwise.

R-indicator optimization (k=5):

Aim is to maximize the R-indicator, $R(\rho) = 1 - 2S(\rho_i)$, where $S(\rho_i)$ is the standard deviation of the response probabilities of responding sample elements under the implemented design. Thus, $\rho_i = \rho_{i5} = \sum_{j=1}^{2} I_{ij5} p_{ij}$.

To obtain the optimal allocation we use an iterative process to minimise $S(\rho_i)$. We simulate the expected responding sample by weighting each element in the gross sample by ρ_i.

Coefficient of Variation (CV) optimization (k=6):

Aim is to minimise the coefficient of variation of the response propensities, $CV = \frac{S(\rho_i)}{\bar{\rho}_i}$. Again, an iterative procedure is used.

ESS (Effective sample size) optimization (k=7):

ESS reflects both variance in inclusion probabilities (selection probabilities and response probabilities) and the final sample size, both of which influence the precision of survey estimates. Theoretically ESS varies across different estimates, but we maximise an approximation which takes one value for the whole survey: $ESS = \frac{\left(\sum_{i=1}^{m} w_i\right)^2}{\sum_{i=1}^{m} w_i^2}$,

where w_i is the weight of the respondent i (reciprocal of inclusion probability) and both of the sums are over the m elements in the responding sample. ESS indicates the size of a fully-responding simple random sample which would provide the same precision as the design in question. We again use an iterative procedure.

7. Results (Table 1)

'CAPI-for-all' appears superior to 'MM-for-all' on all quality indicators, but more costly;

'RR-optimised' and 'ESS-optimised' result in identical allocation, in which 65.8% of sample elements are allocated to the CAPI single-mode protocol and 34.2% to the mixed-mode protocol;

The proportion allocated to the mixed-mode protocol is substantially higher in the other two targeted scenarios: 53.0% with 'CV-optimised' and 54.0% with 'R-indicator-optimised';

Performance on the four quality measures varies only slightly between the four targeted designs. While the two scenarios that aim to maximise representativity do indeed perform better on both representativity measures than the RR/ESS scenario, differences are small. The RR/ESS scenario, on the other hand, appears to perform slightly better in terms of both response rate and effective sample size;

For each of the four outcome indicators, the design that performs best is one of the targeted designs. Indeed, for three of the four indicators all four targeted designs outperform both single- protocol designs. The sole exception is response rate, where the CV-optimised and R-indicator scenarios perform similarly to the CAPI single-mode protocol. Maximised performance in terms of one measure tends to coincide with improved performance on other measures too.

Compared to CAPI-only, cost savings with the targeted designs are minimal or non-existent when the gross sample sample is 2,000, but larger, though still modest, with larger sample sizes. With the larger sample sizes, the relative costs are largely driven by the proportion of the total sample who respond by web (row 6 of Table 1) as in a sequential mixed-mode design such people will not need a visit by an interviewer (reflected in the size of C_{12} relative to C_{11} and C_{22} relative to C_{21} – see annex). But with the smaller sample size, much of this saving is eroded by the additional fixed cost associated with offering two modes (C_{02}).

The MM-for-all scenario results in the highest overall web response rate (40.6%) and hence the lowest data collection costs. Yet this is the allocation scenario that performs worst on all four of the outcome indicators of interest (lowest total response rate, lowest R-indicator, highest CV and lowest ESS).

8. Discussion / Questions

Through targeted allocation, response rate, representativity (R-indicator and CV), and efficiency (ESS and deff) can all be improved. **A promising future for adaptive design?** Even the best non-adaptive designs might benefit from a design review.

Many researchers have long believed that CAPI-for-all achieves the highest response rate. We found it possible to improve response rate from 84.5% with CAPI-for-all to 87.5% with RR/ESS-optimized targeted design. Current best practice in survey design often points to use of the best one-protocol-for-all design without considering whether such a design may be outperformed by a targeted design.

We did not find a single allocation scenario that was optimal for all outcome criteria. Some designs are predicted to achieve a higher response rate (RR-optimized) while others should achieve higher representativity (R-indicator and CV optimized). But variation in outcomes is small between the targeted designs considered in this study. **While it is clear that targeted designs achieve better quality outcomes than non- targeted designs, how do we choose among them?**

Limitation: our allocation models and outcomes are not based on independent samples. **Will we ever have large enough experiments to randomly split into test and control samples? Does it matter?**

To implement targeted allocation it is necessary, a) to have informative auxiliary data available for all elements in the gross sample and, b) to have an applicable model of predicted response propensity in terms of these auxiliary data. Many surveys, especially panels, can meet criterion a). But requirement b) implies that the researcher must have confidence that models which, by necessity, must be based on prior studies, are likely to be applicable to the current survey. Such models need not necessarily be based on randomised methodological experiments such as the one we have presented here. They could instead be based on similar surveys each using a different protocol of interest. However, in either case the surveys must have broad external validity. **How can we accumulate/up-date/test such models?**

Extensions to our allocation criteria are possible, for example to include cost constraints, or to simultaneously meet multiple criteria. **What are the most desirable objectives for allocation criteria in the future?**

9. References

Freedman V.A., K.A. McGonagle, and M.P. Couper. 2017. "Use of a targeted sequential mixed mode protocol in a nationally representative panel study." *Journal of Survey Statistics and Methodology.* Advance articles (published online 6 July 2017).https://doi.org/10.1093/jssam/smx012.

Groves, R.M. 1989. *Survey Errors and Survey Costs*. New York: Wiley.

Jäckle, A., P. Lynn, and J. Burton. 2015. "Going Online with a Face-to-Face Household Panel: Effects of a Mixed Mode Design on Item and Unit Nonresponse". *Survey Research Methods* 9(1): 57-70. http://dx.doi.org/10.18148/srm/2015.v9i1.5475

Lynn, P. 2017. From standardised to targeted survey procedures for tackling non-response and attrition. *Survey Research Methods* 11(1): 93-103. http://dx.doi.org/10.18148/srm/2017.v11i1.6734

Annex: Data Collection Cost Model

Data collection costs for each scenario are predicted using an extension of the traditional model (Groves, 1989, p.51) to allow any of the fixed or variable cost components to vary between protocols:
$C_k = C_0 + \sum_{l=1}^{2}[C_{0l} + C_{1l}n_{kl} + C_{2l}m_{kl}]$,

where

C_0 represents fixed costs that are incurred regardless of which data collection protocols are implemented and regardless of sample sizes;

C_{0l} are fixed costs associated with including mode l in the survey protocol ($l = 1$ for CAPI, 2 for web);

C_{1l} is the unit cost per sample element attempted in mode l ;

C_{2l} is the additional unit cost per sample element that responds in mode l;

n_{kl} is the number of sample elements attempted in mode l under scenario k;

m_{kl} is the number of sample elements that respond in mode l under scenario k.

In our set-up $n_{k1} = \sum_{i=1}^{n}[I_{i1k} + I_{2k}(1 - q_{i2})]$, as elements allocated to the mixed mode protocol will be attempted by CAPI if they do not respond by web in the first fieldwork phase, while $n_{k2} = \sum_{i=1}^{n} I_{2k}$. The mode-specific responding numbers are as follows: $m_{k1} = \sum_{i=1}^{n}[I_{i1k}p_{i1} + I_{2k}(p_{i2} - q_{i2})]$; $m_{k2} = \sum_{i=1}^{n} I_{i2k}q_{i2}$. The unit cost per sample element attempted by CAPI (C_{11}) includes interviewer time and travel expenses to contact and liaise with the sample element, plus associated tasks and materials such as the mailing of advance letters. The unit cost per sample element attempted by web (C_{12}) includes the costs of invitation and reminder mailings and emailings. The additional unit cost per sample element responding by CAPI (C_{21}) includes interviewer time for carrying out the interview, plus costs of data editing and data management, while the additional unit cost per sample element responding by web (C_{22}) consists solely of the costs of data editing and data management. The fixed costs associated with a CAPI fieldwork operation apply to all seven of our scenarios, so we can subsume these within C_0 and set $C_{01} = 0$, leaving C_{02} to represent the margin additional cost of setting up and running a web survey option. For any given values of the cost parameters ($C_0, C_{02}, C_{11}, C_{12}, C_{21}, C_{22}$) and gross sample size n, C_k therefore depends on the allocation to protocols, the protocol-specific response rates and the proportion of mixed-mode respondents who respond by web.

Data collection costs are estimated by assuming the following relative values for the cost components. We believe these to be broadly realistic, at least for the UK context, assuming that fieldwork is carried out by an established survey agency: $C_0 = 20{,}000, C_{02} = 8{,}000, C_{11} = 20, C_{12} = 1, C_{21} = 20, C_{22} = 0.5$. In practice, the relative size of these cost components will depend on features such as the complexity of the survey instruments and the software and methods used for script writing and testing.

Table 1. Allocation rate, response rates, quality indicators and relative costs for different allocation scenarios

	Adaptive designs				Non-adaptive designs		
	RR-optimized (k=4)	R-indicator-optimized (k=5)	CV-optimized (k=6)	ESS-optimized (k=7)	CAPI-for-all (k=1)	MM-for-all (k=2)	Experimental allocation (k=3)
Allocation rate (%)							
Allocated to CAPI-only	65.8	46.0	47.0	65.8	100.0	0.0	34.9
Allocated to mixed-mode	34.2	54.0	53.0	34.2	0.0	100.0	65.2
Response rates by mode (%)							
CAPI-only protocol	89.3	85.2	85.5	89.3	84.5	0.0	84.6
MM protocol	84.2	83.3	83.6	84.2	0.0	79.6	79.4
Web (MM protocol)	51.2	40.5	41.0	51.2	0.0	40.6	39.9
Web (total sample)	17.5	21.8	21.7	17.5	0.0	40.6	26.0
Quality indicators							
Overall response rate	87.5	84.2	84.5	87.5	84.5	79.6	81.2
R-indicator	0.83	0.85	0.85	0.83	0.78	0.77	0.78
CV	0.10	0.09	0.09	0.10	0.13	0.14	0.14
ESS	1818	1753	1749	1818	1708	1624	1654
Resulting n	1846	1775	1782	1846	1782	1678	1713
Deff	1.016	1.012	1.019	1.016	1.044	1.034	1.036
Relative Costs							
n=2,000	1.01	0.96	0.96	1.01	1.00	0.74	0.91
n=10,000	0.93	0.86	0.87	0.93	1.00	0.61	0.80
n=40,000	0.91	0.84	0.85	0.91	1.00	0.58	0.78

Note: Costs are model-based estimates for all seven scenarios; all other estimates are also model-based, as described in the text, except for values in the "Experimental allocation" column (k=3), which are observed from the data. MM=mixed-mode; CV=coefficient of variation of response propensities; ESS=effective sample size; Deff=design effect (due to variation in response propensities).