

Understanding Society

Working Paper Series

No. 2017 – 02

February 2017

The Implications of Alternative Allocation Criteria in Adaptive Design for Panel Surveys

Olena Kaminska & Peter Lynn

Institute for Social and Economic Research, University of Essex



The Implications of Alternative Allocation Criteria in Adaptive Design for Panel Surveys

Olena Kaminska & Peter Lynn

Non-Technical Summary

Instead of using the same method to contact each survey sample member and request their participation, it may be more effective to use different methods for different sample members. This is known as an *adaptive design* (as the method is adapted to the characteristics of each sample member). The question then is how to choose which method to use for which sample members, in other words how to allocate sample members to methods.

In this article we compare four alternative ways of doing this allocation, in other words four different adaptive designs, and three non-adaptive designs. The four adaptive designs each involve attempting to optimize a measure of survey quality and they differ from one another in terms of which survey quality measure is used. All seven designs involve allocating sample members to one of two methods of data collection. The first method involves only *face-to-face interviewing*, in which survey interviewers visit the homes of sample members and attempt to interview them in person. The second method – a *mixed mode* design – involves first mailing each sample member an invite to participate in a web survey and then following up with face-to-face interview attempts only if the sample member fails to complete the web survey. The three non-adaptive designs involve allocating everyone to face-to-face interviewing, allocating everyone to mixed mode, or randomly allocating sample members to one of the two methods.

We find that all of the adaptive designs out-perform all of the non-adaptive designs in terms of how representative the responding sample is. Two of the adaptive designs result in higher response rates than any of the non-adaptive designs. All the adaptive designs cost less to implement than allocating everyone to face-to-face interviewing, but more than allocating everyone to mixed mode (which is the design that performs worst on all the survey quality measures). There are some differences in quality performance between the adaptive designs.

The Implications of Alternative Allocation Criteria in Adaptive Design for Panel Surveys

Olena Kaminska & Peter Lynn

Abstract

We compare four methods for allocating sample members to one of two data collection protocols, in an adaptive survey design framework. The methods aim to optimise, respectively, response rate, R-indicator, coefficient of variation of participation propensities, and effective sample size. The data collection protocols are CAPI single-mode and web-CAPI sequential mixed-mode. We find that two of the targeted allocations achieve a better response rate than single-mode CAPI at a lower cost. All four targeted designs out-perform both single-protocol designs in terms of representativity and effective sample size. For all but the smallest sample sizes, the adaptive designs bring modest cost savings relative to CAPI-only.

Key words: R-indicator, coefficient of variation, effective sample size, non-response error, optimal allocation, mixed mode survey

JEL classifications: C81, C83

Author contact details: olena@essex.ac.uk

Acknowledgements: The authors are funded by a research award from the UK Economic and Social Research Council (award no. ES/K005146/1) for “Understanding Society: the UK Household Longitudinal Study, phase 3.” The successful design and implementation of wave 5 of the Understanding Society Innovation Panel, our source of data, was due to the diligence of many colleagues at ISER, University of Essex, and at NatCen Social Research, who carried out the data collection.

The Implications of Alternative Allocation Criteria in Adaptive Design for Panel Surveys

Olena Kaminska & Peter Lynn

1. Introduction

At the heart of adaptive design lies the idea that survey method can be adapted to best suit different subgroups of sampled persons (elements). This perspective brings new opportunities for exploring methods for improving survey quality. A large body of previous methodological research has addressed questions of whether one survey strategy (or protocol) is better than another in terms of some desired outcome(s) such as response rate, sample composition, survey costs, measurement validity, measurement precision, and so on. Such studies focus on the average effect, across the sample, of the protocol on the desired outcome. These studies are therefore informative regarding which protocol is likely to perform better if it is applied to everyone in a sample. However, it can be the case that even better outcomes could be achieved if a protocol with a 'worse' average effect were applied to some subgroups of the sample. That would be the case if effects were heterogeneous across sample subgroups in such a way that a protocol with a worse average effect was nevertheless the optimum strategy for one or more subgroups. A close look at the survey methodology literature suggests that this may often be the case. For example, studies have found effects on response rates to be heterogeneous across sample subgroups for the following design features: respondent incentives (Singer and Ye, 2013), design of survey invitation mailings (Kaplowitz, Lupi, Couper, & Thorp, 2012), mode of approach (Kaplowitz et al., 2004), survey topic (Sheehan, 2001), and interviewer characteristics (Durrant et al, 2010). In other words, in the same way that advertisers benefit from targeting online adverts based on personal web browsing information, survey

researchers may be able to benefit from targeting survey protocols based on some relevant information about survey sample members.

Improving desirable survey outcomes through adaptive designs depends on the ability to effectively allocate sample members to alternative protocols. Important questions therefore concern the criteria that should be used to allocate sample members to different available protocols and the methods that should be used to implement these criteria. The choice of criterion will depend on which survey outcome(s) the researcher wants to optimize. If suitable information is available about sample members before data collection starts, static adaptive designs (Bethlehem et al., 2011) or targeted designs (Lynn, 2014) can be employed. If such information is not available it may be possible to learn about sample members during data collection (Durrant et al., 2015) and adapt the design using the initial information during the fieldwork in a dynamic adaptive design (Bethlehem et al., 2011) or responsive design (Groves and Heeringa, 2006). This article considers static adaptive designs in which sample allocation is based on information available before fieldwork, in the manner of Lynn (in press).

Most of the literature on adaptive design considers nonresponse error as the outcome of interest, with one notable exception which looks at measurement error (Calinescu and Schouten, 2015). Yet there is no one agreed measure of quality with respect to nonresponse error, and currently research explores different quality indicators that can be optimized through adaptive design. A broad review of such indicators is presented by Groves et al (2008), and a typology is suggested by Wagner (2012). Many quality indicators are item-dependent, i.e. their value differs across survey items and one or a few important items need to be selected for optimization. These include estimated nonresponse bias (Groves and Heeringa, 2006) and the fraction of missing information (Wagner, 2008). Furthermore, balance indicators which measure the balance of the response set with regard to means of some important items (Särndal, 2011; Lundquist and Särndal, 2013) also depend on the selection of survey items.

This article considers quality indicators which have one value for the whole survey and measure “the degree to which the respondents to a survey resemble the complete sample” (Schouten, Shlomo and Skinner, 2011, p.2). As alternatives to the response rate different quality indicators have been developed to measure the general representativity of survey participants. These include the R-indicator (Schouten, Cobben and Bethlehem, 2009; Schouten et al., 2012; Shlomo, Schouten and de Heij, 2013); variance functions of nonresponse weights (Groves et al., 2008; Bethlehem et al., 2011); maximal bias and maximal mean square error (Bethlehem et al., 2011), among others. More complex optimization exercises include constraining minimum sample size and/or cost while simultaneously optimizing one of the above quality indicators (Schouten, Calinescu and Luiten, 2013; Schouten and Shlomo, 2015; Calinescu and Schouten, 2015).

While the debate on different quality indicators continues this article explores related 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? Specifically, we look at an allocation to one of two protocols (CAPI single-mode and web-CAPI sequential mixed-mode) in the context of a panel survey. This context is useful as there is rich information on panel participants at any point after wave 1. This information can be used to inform the allocation procedures, as we show below, and can also serve as auxiliary data to evaluate the outcomes of an adaptive design. However, our approach should be more broadly applicable to any situation where suitable information is available before the fieldwork and two or more protocols are available to choose from.

We structure our paper in the following way. First we describe our dataset and the context for the need of allocation. We follow with description of the allocation criteria based on alternative quality indicators and of the allocation methods. The next section presents the resultant allocation to protocols for each scenario and evaluates the scenarios using

different outcome indicators, including both quality indicators and costs. We conclude with discussion of the potential for adaptive design in practice and the role of different quality indicators in this context.

2. Data: UK Household Longitudinal Study Innovation

Panel

Our data for this study comes from Innovation Panel (IP) which is a standalone small-scale probability-based panel designed specifically to inform through experimentation a large UK Household Longitudinal Study (UKHLS) (Burton, 2013). Such a set up is especially advantageous for adaptive design as it allows measurement of participants' response to different strategies before they are implemented on the mainstage UKHLS. IP started in 2008, one year before UKHLS, and included many experiments that have already informed the mainstage procedures (Burton, 2013; Al Baghal, 2014; Al Baghal, 2015).

In this article we explore data from a mixed mode experiment implemented in wave 5 of IP, with field work carried out from 11 May to 5 September 2012. While IP is mostly CAPI in previous waves, at wave 5 one random third of the sample was allocated to CAPI mode protocol and the remaining two random thirds were allocated to mixed mode protocol (web followed by CAPI mode for nonrespondents). The mixed mode(MM) protocol included a postal letter with an invitation to complete a survey via web, for those with email addresses such letter was also sent by email with two subsequent reminders (or one postal reminder if no email address was known). After two weeks all adults who had not yet filled the questionnaire online were issued into CAPI mode (Burton, 2013), meaning that interviewers began visiting the sample addresses to seek face-to-face interviews. The CAPI phase of the field work lasted 14 weeks. The total response rate conditional on participation in wave 4 was 81.2%, with the mixed-mode protocol obtaining 79.4% (of

which 39.9% responded via web mode) and CAPI protocol obtaining 84.6% (Jäckle et al., 2015). Further details of the survey design and implementation can be found in White et al (2011).

The main aim of the experiment was to inform the mainstage UKHLS on how to best allocate participants to these protocols as part of the study is switching from CAPI single-mode to mixed-mode protocol. While the study interviews everyone aged over ten in each household and the allocation is at the household level (and thus more complex) and has budget restrictions, for the purpose of this article we treat individual sample members as the units to be allocated to protocols and do not explicitly take into account cost in the allocation criteria. Our study is restricted to sample members who participated in the survey at wave 4 and the objective is to shed light on some of issues involved in implementing allocation procedures to optimize different quality criteria for nonresponse.

3. Allocation to Optimise Quality Criteria

In this section we describe the four different quality criteria that we use to inform sample allocation and the method that we use to implement the allocation in each case. In all cases the allocation method draws on predicted probabilities of participation under each data collection protocol. The first step is therefore to estimate probabilities to participate in CAPI protocol and separately in mixed mode protocol for each panel member conditional on participation in the previous wave. To do this we use general demographic and substantive variables from IP wave 4 as predictors, selected as being either theoretically related to non-response (Groves and Couper, 1998), empirically associated with panel attrition (Uhrig 2008; Watson and Wooden 2009) or observed to distinguish between web and CAPI response in a mixed-mode context (Jäckle et al, 2015). These include age, gender, area type, ethnicity, whether born in the UK or outside, marital status, religion, employment status, personal income, savings, self-rated health, BMI, household size,

government region, presence of children under 10 in the household, ownership of the house, number of rooms per person, amenities, whether lone parent, presence of parents in the household, whether mobile phone owner, whether a carer, whether respondent has a driving licence, whether likely to move in the upcoming year, whether others were present at last interview, influence of others on previous interview, interviewer assessment of whether respondent was cooperative at the last interview, whether there is a computer in the house, number of cars owned, whether up to date with all bills, whether can afford holiday, and internet usage. Note that the wave 4 data collection protocol was CAPI-only for all sample members, so the variables regarding the interview context have a consistent interpretation across both wave 5 experimental groups, and there is no mode-induced differential measurement. We use the same set of covariates for all three models in order to avoid differences between protocols in predicted probabilities being due to omitted variables. We note, however, that some variables may play an important role in one model while having little influence in another. The advantage for adaptive design is that not only do we have very rich information about our panel members but also it is highly related to the variables of interest as mostly the predictors are lagged versions of the variables measured in the following wave.

With the above predictors, using logistic regression we predict three propensities: probability of participation in the CAPI single-mode protocol (based on the CAPI experimental sample); probability to participate in mixed-mode protocol (based on the mixed-mode experimental sample); and probability to participate on the web (based on the mixed mode experimental sample). All three models are conditional on participation in the previous wave. Although web participation is part of mixed mode total participation, it is of particular interest as high web participation is desired for cutting costs of data collection (see section 4 below). Model-predicted values are then calculated for all three models for both experimental groups. Thus, for all participants, regardless of their experimental

assignment, we obtain three predicted propensities: web probability and total probability for mixed mode protocol, and probability for CAPI protocol, denoted as follows:

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.

Note that $q_{i2} \leq p_{i2} \forall i$.

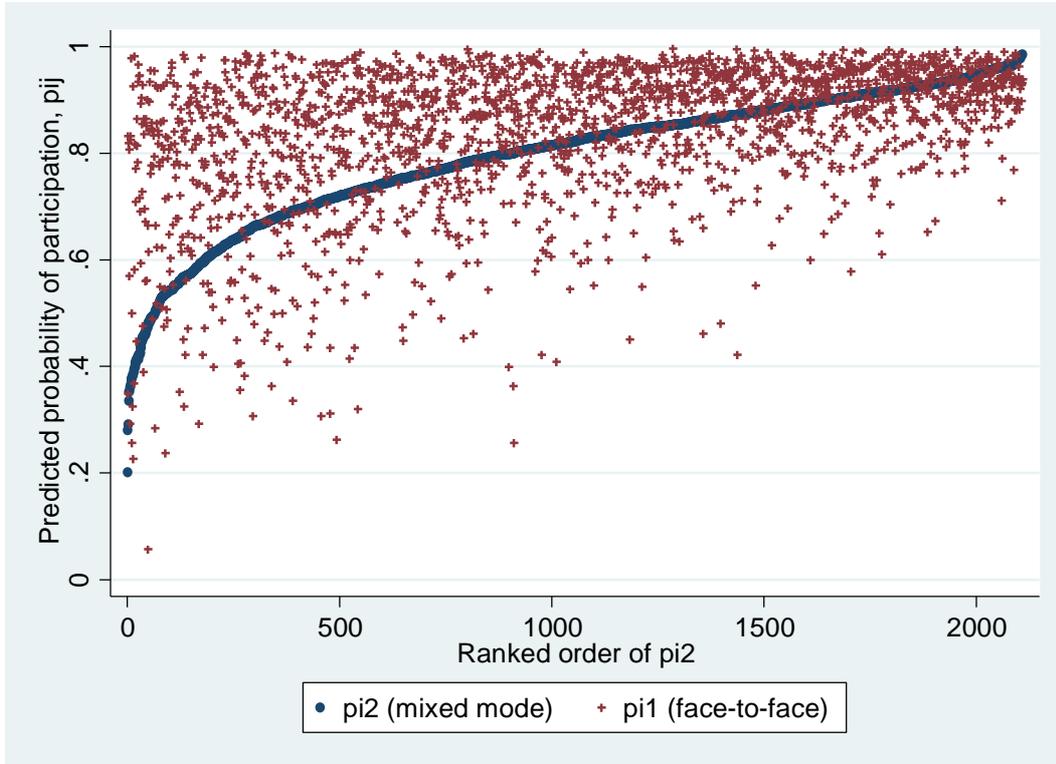
The probabilities $\{p_{ij}\}$ will be used at the next step to determine the allocation of each sample member to a data collection protocol under each allocation scenario, and all three probabilities will be used at the subsequent step (section 5) to predict the outcomes of each allocation scenario. Details of the three fitted models are summarised in Table 1. It can be observed that some variables appear to perform a similar role in each model, suggesting that they are broadly predictive of survey participation regardless of the data collection mode (for example age, wealth and body mass index), while others have very different coefficients in each model, suggesting a mode-specific association with participation (for example, rurality, marital status and the presence of children in the household). Figure 1 shows the association between p_{i1} and p_{i2} and demonstrates that the two models provide considerable discrimination: for any value of p_{i2} there is considerable variation in the values of p_{i1} .

Table 1. Predictive models of participation under CAPI protocol, mixed mode protocol and web-only protocol; Coefficients and standard errors

	CAPI prob (p_{i1})	MM prob (p_{i2})	Web prob (q_{i2})
Age	0.085 (0.051)	0.042 (0.028)	0.043 (0.030)
Age squared	-0.001 (0.001)	-0.000 (0.000)	-0.001 (0.003)
Rural area	0.604 (0.345)	-0.522 (0.180)	-0.451 (0.160)
Carer	1.056 (0.541)	0.095 (0.267)	-0.025 (0.251)
Has driving licence	-0.142 (0.299)	-0.464 (0.183)	-0.337 (0.178)
Government Office Region (Ref=North):			
Midlands	0.488 (0.370)	-0.094 (0.228)	-0.156 (0.200)
East Anglia	0.298 (0.446)	-0.011 (0.290)	0.298 (0.242)
London	0.068 (0.453)	-0.812 (0.277)	-0.140 (0.270)
South	0.384 (0.336)	-0.230 (0.201)	0.013 (0.179)
Wales	0.126 (0.516)	0.606 (0.398)	-0.050 (0.336)
Scotland	-0.453 (0.453)	-0.389 (0.269)	0.105 (0.256)
Marital status (Ref=never married):			
Married/civil partnership	-0.572 (0.410)	0.129 (0.246)	0.639 (0.232)
Widowed/divorced/separated	0.264 (0.492)	-0.349 (0.265)	-0.166 (0.259)
Belongs to a religion	0.058 (0.255)	0.273 (0.158)	0.309 (0.138)
Interest from savings (Ref=none):			
Up to £500	0.228 (0.343)	0.249 (0.207)	0.389 (0.167)
Over £500	0.178 (0.605)	0.737 (0.340)	0.726 (0.278)
Others influenced interview	-0.466 (0.420)	-0.527 (0.266)	-0.085 (0.262)
Co-operative with interview (Ref=v good):			
Good	0.239 (0.330)	-0.237 (0.184)	-0.155 (0.174)
Fair/poor/very poor	0.204 (0.633)	-0.716 (0.370)	-0.884 (0.491)
Body mass index ≥ 28	-0.304 (0.257)	-0.353 (0.160)	-0.245 (0.145)
Number of children aged under 10	0.378 (0.354)	-0.273 (0.226)	-0.483 (0.210)
Internet access at home	0.421 (0.460)	0.446 (0.286)	1.645 (0.349)
Number of cars in household (Ref=0):			
One	0.857 (0.412)	-0.300 (0.250)	-0.194 (0.270)
Two or more	0.583 (0.498)	0.093 (0.300)	0.262 (0.297)
Behind with paying bills	-0.792 (0.426)	0.084 (0.307)	-0.353 (0.340)
Material deprivation indicator:			
Cannot afford holiday	0.319 (0.329)	0.046 (0.197)	-0.347 (0.189)
Does not want holiday	-0.382 (0.381)	-0.020 (0.240)	-0.446 (0.224)
In paid employment	0.496 (0.301)	0.261 (0.197)	0.266 (0.181)
Regularly uses internet	-0.450 (0.404)	0.237 (0.254)	0.888 (0.235)
R^2	0.130	0.093	0.201
n	735	1,374	1,374
Mean predicted probability	0.85	0.80	0.41
Std. Dev.	0.13	0.13	0.24
Skewness	-1.73	-1.06	-0.11
Kurtosis	6.56	4.00	2.00

Note: Each model included same set of 32 variables, as explained in the text; table shows only variables with $P < 0.05$ for at least one model

Figure 1: Association of predicted probabilities of participation under two data collection protocols



We consider seven different allocation scenarios. Three of these can be considered as benchmark designs and are described in the next paragraph. The other four are static adaptive designs, each involving allocation to data collection protocols based on a different criterion. The allocation criteria for the adaptive designs and the method used to implement them are described in sub-sections 3.1 to 3.4 respectively. For each of the allocation scenarios the allocated protocol can be indicated as follows:

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

Note that: $\sum_{j=1}^2 I_{ijk} = 1 \forall (i, k)$, i.e. under each scenario, each element is allocated uniquely to one protocol.

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

For comparative purposes we consider three benchmark allocation scenarios. Two of these assume that all participants are assigned either to the CAPI single-mode protocol ('F2F-for-all') or to the mixed-mode protocol ('MM-for-all'). Thus, under F2F-for-all ($k=1$), $I_{i11} = 1$ and $I_{i21} = 0 \forall i$, while under MM-for-all ($k=2$), $I_{i12} = 0$ and $I_{i22} = 1 \forall i$. The third scenario ($k=3$) reflects the random allocation from the actual experiment which assigned two-thirds of the sample to the mixed-mode protocol and one-third to the CAPI single-mode protocol ('experimental allocation'). Unlike for other allocations the composition of the responding sample is observed (rather than being simulated using model-predicted probabilities).

3.1 Response-rate (RR) optimization

Survey researchers have for decades striven to achieve the highest possible response rates (Kanuk and Berenson 1975; Sheehan 2001; Yu and Cooper 1983). Many of the early static adaptive designs allocated sample elements to treatments with the objective of maximising response rate (Lynn, in press). We therefore include amongst the set of adaptive designs that we study, one ($k=4$) which aims to maximise the response rate. To do this, we allocate each person according to their highest predicted response propensity. Thus for each person we compare response propensity for CAPI protocol and for mixed mode protocol and choose the higher one between them. For example, if a sample member is predicted to have higher chance of response in the CAPI protocol than in the mixed mode protocol, the sample member is allocated to the CAPI protocol:

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

$$I_{i14} = 0 \text{ and } I_{i24} = 1 \text{ otherwise.}$$

3.2 R-indicator optimization

Increasing response rate does not necessarily decrease nonresponse error (Groves and Peytcheva, 2008). Thus we explore an allocation ($k=5$) that aims to improve

representativity. In other words we aim to allocate sample members to protocols such that the expected distribution of selected variables among respondents is as close to their distributions for the gross sample as possible. We optimize to the R-indicator in its simplest form (Schouten et al., 2009):

$$R(\rho) = 1 - 2S(\rho_i),$$

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

As the probabilities used in the R-indicator calculation are those of respondents only, we should simulate the expected responding sample by weighting each element in the gross sample by their predicted probability of response, ρ_i . Thus, all statistics that would be based on the responding sample are estimated based on the gross sample weighted by ρ_i .

To optimize the R-indicator we need to minimise $S(\rho_i)$. However, ρ_i depends on the allocation I_{ij5} , so to obtain the optimal allocation that will maximise the R-indicator we use an iterative process. We first calculate a starting value for the mean response propensity under scenario 5 as the mean of the probabilities in each protocol:

$$\bar{\rho}_5^{(1)} = \frac{\sum_{i=1}^n \sum_{j=1}^2 p_{ij}}{2n}$$

Then, for each sample element we compare the distance between $\bar{\rho}_5^{(1)}$ and each of the protocol-specific response propensities, p_{i1} and p_{i2} . The panel member is assigned to the protocol with the smallest distance and this way we obtain allocation for the first iteration:

$$I_{i15}^{(1)} = 1 \text{ and } I_{i25}^{(1)} = 0 \text{ if } |p_{i1} - \bar{\rho}_5^{(1)}| < |p_{i2} - \bar{\rho}_5^{(1)}|;$$

$$I_{i15}^{(1)} = 0 \text{ and } I_{i25}^{(1)} = 1 \text{ otherwise.}$$

Each subsequent iteration uses the average probability of participation according to the current allocation as a starting point. Again, distances are compared between this average

and the element-specific probabilities to participate in each protocol to see whether some reallocation may improve the R-indicator. In other words, for iteration (h):

$$\bar{\rho}_5^{(h)} = \frac{\sum_{i=1}^n \sum_{j=1}^2 I_{ij5}^{(h-1)} p_{ij}}{n};$$

$$I_{i15}^{(h)} = 1 \text{ and } I_{i25}^{(h)} = 0 \text{ if } |p_{i1} - \bar{\rho}_5^{(h)}| < |p_{i2} - \bar{\rho}_5^{(h)}|;$$

$$I_{i15}^{(h)} = 0 \text{ and } I_{i25}^{(h)} = 1 \text{ otherwise.}$$

Iterations cease when $R(\rho)$ changes very little and very few sample elements are re-allocated to a different protocol. Specifically, the Stopping rule is: $0.999 < R_h/R_{h-1} < 1.001$ and fewer than 1% of elements switch allocation.

3.3 Coefficient of Variation (CV) optimization

Variance functions of the nonresponse weights – and therefore of the response propensities – have been proposed as useful measures of the representativity of responding survey samples (Groves et al, 2008; Bethlehem et al, 2011). More recently, Schouten et al. (2013) have suggested that the most useful amongst these may be the coefficient of variation of the response propensities. This has the following general form:

$$CV = \frac{S(\rho_i)}{\bar{\rho}_i}$$

where again $S(\rho_i)$ is the standard deviation of response probabilities amongst the responding sample, and $\bar{\rho}_i$ is the mean of these probabilities. A smaller value of CV indicates a more representative sample. To minimize CV ($k=6$) we implement the following iterative procedure. For each sample element we first calculate the average of their probabilities to respond in each protocol: $\rho_{i6}^{(1)} = (p_{i1} + p_{i2})/2$. This unrealistic probability

serves as a starting value. Based on these average probabilities we simulate the CV of the

responding sample by weighting each element i in the gross sample by $\rho_{i6}^{(1)}$. Now for each element we estimate two new weighted CVs: keeping the average probabilities for all other sample members for person i we substitute $\rho_{i6}^{(1)}$ with p_{i1} (CAPI) to estimate $CV_{i1}^{(1)}$ and then separately we substitute $\rho_{i6}^{(1)}$ with p_{i2} (mixed mode) to estimate $CV_{i2}^{(1)}$. Thus for each person we have two estimates of CV – one corresponding to allocation of this person to CAPI protocol and the other to mixed mode protocol while keeping average probabilities of all others. We aim to lower CV and therefore at the first iteration we allocate element i to the CAPI single-mode protocol if $CV_{i1}^{(1)} < CV_{i2}^{(1)}$ and *vice versa*.

As a starting point for the second iteration we use the response probabilities based on the protocol allocation after the first iteration. Keeping these probabilities for all other people, for each person we again estimate CV under each of the allocation options, $CV_{i1}^{(2)}$ and $CV_{i2}^{(2)}$. Element i will again be allocated to the protocol corresponding to the smaller CV. More iterations can be implemented until there is little or no change in allocation between iterations. Our Stopping rule is that $0.999 < \frac{CV_h}{CV_{h-1}} < 1.001$ and fewer than 1% of elements switch allocation.

3.4 ESS (Effective sample size) optimization

To the best of the authors' knowledge, ESS has not been used as an allocation criterion in the context of adaptive design until now. Yet it has the advantage of reflecting both variance in inclusion probabilities and the final sample size, both of which influence the precision of survey estimates. Theoretically ESS varies across different estimates, but an approximation proposed by Kish (1965) can have one value for the whole survey and takes the general form:

$$ESS = \frac{(\sum_{i=1}^m w_i)^2}{\sum_{i=1}^m w_i^2},$$

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. The ESS value indicates a sample size of simple random sample which would be needed to provide the same precision as the design in question. Our scenario 7 therefore involves finding a design which gives the highest value of ESS and therefore the best precision for estimation.

To maximize ESS we again use an iterative procedure. The first step is to use the ‘average’ response probability for each sample element, $\rho_{i7}^{(1)} = (p_{i1} + p_{i2})/2$ as a starting value to derive ‘average’ weights, $w_{i7}^{(1)} = 1/\rho_{i7}^{(1)}$. For each sample element i we then calculate two weighted ESS values: for CAPI protocol and for mixed mode protocol: $ESS_{i1}^{(1)}$ uses $1/p_{i1}$ for person i and $w_{i7}^{(1)}$ for each other sample element; and $ESS_{i2}^{(1)}$ uses $1/p_{i2}$ for person i and $w_{i7}^{(1)}$ for each other sample element. As the ESS depends on the inclusion probabilities of the responding sample we simulate it by weighting each of the n elements in the gross sample by their probability to respond (p_{i1} or p_{i2} for person i and $\rho_{i7}^{(1)}$ for everyone else). Element i is allocated to the CAPI protocol if $ESS_{i1} > ESS_{i2}$, and *vice versa*.

Each subsequent iteration, (h), uses the weights from the previous iteration as starting values and analogously calculate $ESS_{i1}^{(h)}$ and $ESS_{i2}^{(h)}$ for each sample element i , reallocating the element if their current assignment achieves lower ESS than assignment to the other protocol. Similarly to the previously-described procedures, our Stopping rule is that $0.999 < ESS_h/ESS_{h-1} < 1.001$ and fewer than 1% of elements switch allocation.

All optimization and analysis is implemented in Stata 14.1.

4. Data Collection Costs

When the alternative data collection protocols differ in terms of costs, it is important to take this into account when selecting an allocation method. Broadly speaking, there are two ways to do this. The first is to explicitly incorporate data collection costs into the allocation criteria, perhaps either by applying a cost constraint or by extending the criteria themselves to include a cost element. The second way is to include a prediction of data collection costs in the evaluation of alternative allocation scenarios. Here, we do the latter.

Data collection costs for any survey can be expressed in terms of the fixed and variables costs associated with various aspects of the design. The traditional model (Groves, 1989, p.51) can be extended to allow any of the fixed or variable cost components to vary between design options and hence to facilitate the comparison of designs. Costs may vary between the scenarios we consider here due to differences in the number of sample elements attempted, and achieved, in each mode, and (large) differences between modes in the unit costs of data collection. We can express the data collection cost as follows:

$$C_k = C_0 + \sum_{l=1}^2 [C_{0l} + C_{1l}n_{kl} + C_{2l}m_{kl}] ,$$

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 ($l = 1$ 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 .

Note that 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 member, 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.

5. Prediction of Outcomes

To assess the relative performance of each of the allocation scenarios, we predict several quality measures for each scenario as well as estimating the data collection cost. The quality measures correspond to the four measures that were optimised under the adaptive scenarios - overall response rate, R-indicator, CV and ESS – plus the proportion of the total sample responding by web, as this indicates the relative cost of data collection.

For each of the six scenarios, the total response rate for scenario k is predicted as the mean of the predicted element response propensities:

$$\bar{p}_k = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^2 I_{ijk} p_{ij} = \frac{1}{n} \sum_{i=1}^n \rho_{ik}$$

For example, the predicted response rate for the F2F-for-all scenario is:

$$\bar{p}_1 = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^2 I_{ij1} p_{ij} = \frac{1}{n} \sum_{i=1}^n I_{i11} p_{i1}$$

The web response rate for the total sample is predicted for scenario k thus:

$$\bar{q}_k = \frac{1}{n} \sum_{i=1}^n I_{i2k} q_{i2}$$

The web response rate within the sample allocated to the mixed-mode protocol is predicted as follows:

$$\bar{q}_{C_k} = \frac{\sum_{i=1}^n I_{i2k} q_{i2}}{\sum_{i=1}^n I_{i2k}}$$

The other three outcome measures that we consider – R-indicator, CV, and ESS - are all characteristics of the responding sample and are predicted by weighting each member of the gross sample by the model-predicted response probability for the mode to which they are allocated under that scenario, ρ_{ik} . The method of calculation for each of the predicted quantities is presented in the annex.

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. We present the relative costs of each allocation scenario for three alternative gross sample sizes of $n = 2,000$, $10,000$ and $40,000$.

6. Results

Table 2 presents predicted outcomes for each of the four adaptive design scenarios and for each of the two scenarios in which all sample members are allocated to a single protocol. The final column presents the observed outcomes for the experimental allocation to protocols that was implemented at IP wave 5.

For each scenario, the first two rows of Table 2 present the percentages of sample members allocated to each of the two data collection protocols, $\frac{100}{n} \sum_{i=1}^n I_{i1k}$ and $\frac{100}{n} \sum_{i=1}^n I_{i2k}$ respectively. The third and fourth row present the response rate among those allocated to the CAPI protocol (i.e. their average CAPI probabilities to participate) and the response rate among those allocated to the mixed-mode protocol, respectively. The fifth row presents the web response rate amongst people allocated to the mixed-mode protocol, $\bar{q}c_k$, while the sixth row presents the proportion of the whole sample who respond by web, \bar{q}_k (which is the product of the proportion allocated to the mixed-mode protocol and the web response rate among them). Row 7 presents the total expected response rate (average of response probabilities for the specified allocation), \bar{p}_k , while the following three rows present the R-indicator (R_k); CV (CV_k); and ESS (ESS_k). Additionally we present the responding sample size (resulting n), which is a product of the 2,109 gross sample size and \bar{p}_k , and the estimated design effect, calculated as the responding sample size divided by ESS. Note that except for the last column ('experimental allocation') all these outcome values are predicted based on the model-predicted values of participation propensity, as explained in section 5 above. The final three rows of the table present estimated data collection costs relative to the benchmark CAPI-only scenario (scenario 1) for three alternative sample sizes.

Table 2. Allocation rate, response rates and quality indicators for different allocation scenarios

	Adaptive designs				Non-adaptive designs		
	RR- optimized (k=4)	R-ind- 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)
Allocated to F2F (%)	65.8	46.0	47.0	65.8	100.0	0.0	34.9
Allocated to MM (%)	34.2	54.0	53.0	34.2	0.0	100.0	65.2
F2F RR	89.3	85.2	85.5	89.3	84.5	0.0	84.6
MM RR	84.2	83.3	83.6	84.2	0.0	79.6	79.4
Web RR within MM	51.2	40.5	41.0	51.2	0.0	40.6	39.9
Web RR of total	17.5	21.8	21.7	17.5	0.0	40.6	26.0
Total RR	87.5	84.2	84.5	87.5	84.5	79.6	81.2
R-indicator	0.827	0.849	0.849	0.827	0.779	0.774	0.775
CV	0.098	0.089	0.088	0.098	0.128	0.138	0.135
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
Relative Costs (n=10,000)	0.93	0.86	0.87	0.93	1.00	0.61	0.80
Relative Costs (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. RR=Response rate; F2F=CAPi single-mode protocol; MM=mixed-mode protocol; CV=coefficient of variation of response propensities; ESS=effective sample size; Deff=design effect (due to variation in response propensities).

Our interest is primarily in comparing the four adaptive designs and in comparing each of them with the two single-protocol designs. The two single-protocol designs can in that sense be seen as benchmarks. We therefore describe the benchmark designs first. If all sample members were allocated to the CAPI protocol ('F2F-for-all' scenario, Table 2) the predicted response rate would be 84.5% with an R-indicator of 0.78 and an effective sample size of 1,708. If, on the other hand, all sample members were allocated to the mixed-mode protocol ('MM-for-all' scenario) the predicted response rate would be lower (79.6%), as would the effective sample size (1,624) but the R-indicator would be similar (R=0.77). If only single-protocol designs were to be considered, it would most likely be concluded that the CAPI protocol was slightly superior to the mixed mode protocol.

Comparing first the proportions allocated to each protocol, it can be seen that two scenarios – the scenario in which allocation is based on maximising response rate ('RR-optimised' scenario in Table 2) and the scenario in which allocation aims to maximise the effective sample size ('ESS-optimised') – result in an identical allocation, in which 65.8% of sample members are allocated to the CAPI single-mode protocol and 34.2% to the mixed-mode protocol. This does not have to be true in general but in our situation selecting respondents with higher probabilities was more important to ESS optimization. The proportion allocated to the mixed-mode protocol is substantially higher in the other two scenarios, where the aim is to maximise representativity: 53.0% when the aim is to minimise the coefficient of variation of the response propensities ('CV-optimised') and 54.0% when the aim is to maximise the R-indicator ('R-indicator').

In terms of the four quality measures of interest, it is striking that performance varies only slightly between the four adaptive 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 R-indicator is predicted to be 0.85 in both the CV-optimised and R-indicator scenarios, but 0.83 in the RR/ESS scenario, while the coefficient of variation of the response probabilities is 0.088 in the CV-optimised scenario

and 0.089 in the R-indicator scenario, compared to 0.098 in the RR/ESS scenario. The RR/ESS scenario, on the other hand, performs slightly better in terms of both response rate and effective sample size. The effective sample size is 1,818 under the RR/ESS scenario, compared to 1,753 and 1,749 under the R-indicator and CV-optimised scenarios respectively. Predicted response rate is 87.5% under the RR/ESS scenario, compared to response rates of 84.2% and 84.5% respectively for the R-indicator and CV-optimised scenarios (and 84.5% for the F2F-for-all scenario).

Comparing the adaptive designs with the single-protocol designs, we see that for each of the four outcome indicators, the design that performs best is one of the adaptive designs. Indeed, for three of the four indicators all four of the adaptive designs outperform both of the 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. It is therefore clear that adaptive designs can out-perform single-protocol designs and that in this study maximised performance in terms of one measure tends to coincide with improved performance on other measures too.

One other important aspect is the cost of data collection. Compared to the CAPI-only design, we see that cost savings with the adaptive designs are minimal or non-existent when the gross sample size 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 2) 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}). But with the smaller sample size, much of this saving is eroded by the additional fixed cost associated with offering two modes (C_{02}). Naturally, 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). The overall web response rate is lower in the

adaptive designs and hence data collection costs would be higher. Costs would be slightly lower with the R-indicator and CV-optimized scenarios (web response rate 21.7% - 21.9%) than with the RR/ESS-optimized scenario (17.5%).

7. Conclusions

A few important conclusions emerge from our findings. First and foremost we find that through adaptive design allocation, response rate, representativity (R-indicator and CV), and efficiency (ESS and deff) can all be improved. All adaptive designs resulted in almost the same or better total response rate, and consistently better R-indicator, CV, ESS and deff than either of the one-protocol-for-all designs or the random allocation observed in the experiment. This suggests a promising future for adaptive design within survey research: even the best non-adaptive designs might benefit from a design review.

For example, in many contexts for a long time researchers have believed that the F2F-for-all design achieves the highest response rate. Yet our study shows that it is possible to improve response rate from 84.5% for the F2F-for-all design to 87.5% for the RR/ESS-optimized adaptive 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 an adaptive design.

Second, 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 response rates, representativity measures (R-indicator and CV) and precision of estimation (ESS) is small between the adaptive designs considered in this study. While it is clear that adaptive designs achieve better quality outcomes than non-adaptive designs, it is harder to select the best scenario among them. Our original intuition was that ESS-optimized design should be preferable as it should come closest to the most efficient

design. It is interesting that in our study the ESS-optimized design achieved the same allocation as RR-optimized design. This may not necessarily be the case in general, but it could be the case that if the variation in response propensities does not differ greatly between allocation scenarios the responding sample size – which is determined by response rate – becomes the dominant influence on the effective sample size. We feel this should be explored further, especially given the relative ease of implementing RR-optimized design.

We have demonstrated how alternative allocation criteria can be applied, but to do this we have benefited from the rich context provided by the Understanding Society Innovation Panel. Specifically, to implement our methods 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. Most types of panel surveys, from wave 2 onwards, typically meet the requirement of rich auxiliary data. Cross-sectional studies may be able to draw upon sampling frame data, linked data, or interviewer observations. 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. An extension of our allocation criteria, for example to include cost constraints, would be possible. So, for example, a researcher may be interested in optimising one of the criteria that we have presented, but with a constraint that the data collection cost should not exceed a certain proportion of the costs of a F2F-for-all design. The cost model that we have introduced in section 4, or a similar one, could be used for this purpose. With the data collection protocols that we have considered in this article, such a constraint would have the effect of ensuring that a certain minimum proportion of respondents participate online (and this

proportion would depend on the overall sample size). Extensions of this kind to the allocation criteria, effectively involving simultaneously meeting multiple criteria, would seem a natural next step.

8. References

- Al Baghal, T. (ed.). 2014. "Understanding Society Innovation Panel Wave 6: Results from Methodological Experiments." *Understanding Society Working Paper Series* 2014-04. Available at <https://www.understandingsociety.ac.uk/research/publications/working-paper/understanding-society/2014-04.pdf> (accessed 24 February 2016).
- Al Baghal, T. (ed.). 2015. "Wave 7: Results from Methodological Experiments." *Understanding Society Working Paper Series* 2015-03. Available at <https://www.understandingsociety.ac.uk/research/publications/working-paper/understanding-society/2015-03.pdf> (accessed 24 February 2016).
- Bethlehem, J., F. Cobben, and B. Schouten. 2011. *Handbook of Nonresponse in Household Surveys*. New Jersey: Wiley.
- Burton, J. (ed.) 2013. "Understanding Society Innovation Panel Wave 5: Results from Methodological Experiments." *Understanding Society Working Paper Series* 2013-06. Available at <https://www.understandingsociety.ac.uk/research/publications/working-paper/understanding-society/2013-06.pdf> (accessed 24 February 2016).
- Calinescu, M., and B. Schouten. 2015. "Adaptive Survey Designs to Minimize Survey Mode Effects – a Case Study on the Dutch Labor Force Survey." *Survey Methodology* 41(2): 403-425.
- Durrant, G., Groves, R.M., Staetsky, L. & Steele, F. 2010. "Effects of Interviewer Attitudes and Behaviors on Refusal in Household Surveys." *Public Opinion Quarterly* 74(1): 1-36. <http://dx.doi.org/10.1093/poq/nfp098>.
- Durrant G.B., O. Maslovskaya, and P. Smith. 2015. "Modelling Final Outcome and Length of Call Sequence to Improve Efficiency in Interviewer Call Scheduling." *Journal of Survey Statistics and Methodology* 3: 397-424. <http://dx.doi.org/10.1093/jssam/smv008>.

- Groves, R.M. 1989. *Survey Errors and Survey Costs*. New York: Wiley.
- Groves, R.M. and Couper, M.P. *Nonresponse in Household Interview Surveys*. New York: Wiley.
- Groves, R.M. and S.G. Heeringa. 2006. "Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs." *Journal of the Royal Statistical Society Series A* 169: 439-457.
- Groves, R.M. and E. Peytcheva. 2008. "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis." *Public Opinion Quarterly* 72(2): 167-189. <http://dx.doi.org/10.1093/poq/nfn011>.
- Groves, R.M., J.M. Brick, W. Kalsbeek, B. Harris-Kojetin, F. Kreuter, B.-E. Pennell, T. Raghunathan, B. Schouten, T. Smith, R. Tourangeau, A. Bowers, M. Jans, C. Kennedy, R. Levenstein, K. Olson, E. Peytcheva, S. Ziniel, and J. Wagner. 2008. "Issues Facing the Field: Alternative Practical Measures of Representativeness of Survey Respondent Pools." *Survey Practice* 1.3: 1-6.
- 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>
- Kanuk, L. and Berenson, C. 1975. "Mail Surveys and Response Rates: A Literature Review." *Journal of Marketing Research* 12: 440-453.
- Kaplowitz, M. D., Hadlock, T. D., & Levine, R. 2004. "A Comparison of Web and Mail Survey Response Rates." *Public Opinion Quarterly* 68: 94-101.
- Kaplowitz, M.D., Lupi, F., Couper, M.P., & Thorp, L. 2012. "The Effect of Invitation Design on Web Survey Response Rates." *Social Science Computer Review* 30(3): 339-349. <http://dx.doi.org/10.1177/0894439311419084>.
- Lynn, P. in press. From standardised to targeted survey procedures for tackling non-response and attrition. *Survey Research Methods*.
- Lynn, P. 2014. "Targeted Response Inducement Strategies on Longitudinal Surveys." In *Improving Survey Methods: Lessons from Recent Research*, edited by U. Engel, B. Jann, P. Lynn, A. Scherpenzeel, and P. Sturgis, 322-338. Abingdon UK: Psychology Press.

- Lundquist, P., and C.-E. Sarndal. 2013. "Aspects of Responsive Design with Applications to the Swedish Living Conditions Survey." *Journal of Official Statistics* 29.4: 557-582.
- Sarndal, C.-E. 2011. "The 2010 Morris Hansen Lecture Dealing with Survey Nonresponse in Data Collection, in Estimation." *Journal of Official Statistics* 27.1: 1-21.
- Schouten, B., J. Bethlehem, K. Beullens, O. Kleven, G. Loosveldt, A. Luiten, K. Rutar, N. Shlomo, and Ch. Skinner. 2012. "Evaluating, Comparing, Monitoring, and Improving Representativeness of Survey Response Through R-Indicators and Partial R-Indicators." *International Statistical Review* 80.3: 382-399.
- Schouten B., F. Cobben and J. Bethlehem. 2009. "Indicators for the Representativeness of Survey Response." *Survey Methodology* 35.1: 101-113.
- Schouten, B., N. Shlomo, and C. Skinner. 2011. "Indicators for Monitoring and Improving Representativeness of Response." *Journal of Official Statistics* 27(2): 231-253.
- Schouten, B., M. Calinescu and A. Luiten. 2013. "Optimizing Quality of Response Through Adaptive Survey Designs." *Survey Methodology* 39.1: 29-58.
- Schouten, B., and N. Shlomo. 2015. "Selecting Adaptive Survey Design Strata with Partial R-indicators." *International Statistical Review*. <http://dx.doi.org/10.1111/insr.12159>
- Sheehan, K.B. 2001. "Email Survey Response Rates: A Review." *Journal of Computer-Mediated Communication* 6(2): 1-16.
- Shlomo, N., B. Schouten and V. de Heij. 2013. *Designing Adaptive Survey Designs with R-Indicators*. Available at: <http://www.risq-project.eu/papers/shlomo-schouten-heij-2013.pdf> (accessed February 24 2016).
- Singer, E., & Ye, C. 2013. The use and effects of incentives in surveys. *Annals of the American Academy of Political and Social Science*, 645: 112
- Uhrig, S. C. 2008. The nature and causes of attrition in the British Household Panel Study. *ISER Working Paper Series*, 2008(05). Colchester: Institute for Social and Economic Research, University of Essex.
- Wagner, J.R. 2008. *Adaptive Survey Design to Reduce Nonresponse Bias*, PhD Dissertation. University of Michigan.

- Wagner, J.R. 2012. "Research Synthesis: a Comparison of Alternative Indicators for the Risk of Nonresponse Bias." *Public Opinion Quarterly* 76(3): 555-575. <http://dx.doi.org/10.1093/poq/nfs032>.
- Watson, N. and Wooden, M. 2009. "Identifying factors affecting longitudinal survey response." In *Methodology of Longitudinal Surveys*, edited by P. Lynn, 157–181. Chichester: Wiley. <http://dx.doi.org/10.1002/9780470743874>
- White, N., Bourne, T. and Patterson, A. 2011. *Understanding Society Innovation Panel 2: Technical Report*. Available at: <https://www.understandingsociety.ac.uk/documentation/innovation-panel/technical-reports> (accessed January 24 2017).
- Yu, J. and Cooper, H. 1983. "A Quantitative Review of Research Design Effects on Response Rates to Questionnaires." *Journal of Marketing Research* 20(1): 36-44.

Annex: Prediction of Quality Measures

For each scenario, the R-indicator, CV and ESS are predicted by adapting the standard formula for the respective quantities to reflect the fact that we are using information from the gross sample rather than the net (responding) sample. In each case, wherever a summation is made over sample elements, instead of an unweighted sum of responding sample elements we substitute a weighted sum of gross sample elements, where the weight represents the probability of the element appearing in the responding sample, in other words ρ_{ik} . The predicted outcomes for scenario k are therefore as follows.

R-indicator:

$$\hat{R}_k = 1 - 2\hat{s}_k$$

where

$$\begin{aligned}\hat{s}_k &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n \rho_{ik} (\rho_{ik} - \bar{p}_k)^2} \\ &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n \sum_{j=1}^2 I_{ijk} p_{ik} (p_{ik} - \bar{p}_k)^2}\end{aligned}$$

Coefficient of variation of the response propensities ('CV'):

$$\widehat{CV}_k = \frac{\hat{s}_k}{\bar{p}_k}$$

Effective sample size ('ESS'):

$$\begin{aligned}\widehat{ESS}_k &= \frac{(\sum_{i=1}^n \rho_{ik} (1/\rho_{ik}))^2}{\sum_{i=1}^n \rho_{ik} (1/\rho_{ik})^2} \\ &= \frac{n^2}{\sum_{i=1}^n (1/\rho_{ik})}\end{aligned}$$

Data collection costs: $\hat{C}_k = C_0 + J_k C_{02} + C_{11} \sum_{i=1}^n [I_{i1k} + I_{2k}(1 - q_{i2})] + C_{21} \sum_{i=1}^n [I_{i1k} p_{i1} + I_{2k}(p_{i2} - q_{i2})] + C_{12} \sum_{i=1}^n I_{2k} + C_{22} \sum_{i=1}^n I_{i2k} q_{i2}$, where $J_k = 0$ if $k = 1$; $J_k = 1$ if $2 \leq k \leq 7$.