Optimizing quality of response through adaptive survey designs

Barry Schouten, Melania Calinescu and Annemieke Luiten

Discussion paper (201118)

Statistics Netherlands

The Hague/Heerlen, 2011

Explanation of symbols

.	= data not available
*****	= provisional figure
******	= revised provisional figure
x	= publication prohibited (confidential figure)
–	= nil or less than half of unit concerned
–	= (between two figures) inclusive
0 (0,0)	= less than half of unit concerned
blank	= not applicable
2010–2011	= 2010 to 2011 inclusive
2010/2011	= average of 2010 up to and including 2011
2010/'11	= crop year, financial year, school year etc. beginning in 2010 and ending in 2011
2008/'09–2010/'11	= crop year, financial year, etc. 2008/'09 to 2010/'11 inclusive

Due to rounding, some totals may not correspond with the sum of the separate figures.

Publisher
Statistics Netherlands
Henri Faasdreef 312
2492 JP The Hague

Prepress
Statistics Netherlands
Grafimedia

Cover
TelDesign, Rotterdam

Information
Telephone +31 88 570 70 70
Telefax +31 70 337 59 94
Via contact form:
www.cbs.nl/information

Where to order
E-mail: verkoop@cbs.nl
Telefax +31 45 570 62 68

Internet
www.cbs.nl

ISSN: 1572-0314

60083201118 X-10

Optimizing quality of response through adaptive survey designs

Barry Schouten, Melania Calinescu and Annemieke Luiten

Summary: In most surveys all sample units receive the same treatment and the same design features apply to all selected people and households. In this paper, it is explained how survey designs may be tailored to optimize response rates and to reduce nonresponse selectivity. Such designs are called adaptive survey designs. The basic ingredients of such designs are introduced and discussed and illustrated with a number of examples including a pilot study.

Keywords: Representativeness; R-indicators; Nonresponse; Responsive survey design

1. Introduction

In most surveys all sample units receive the same treatment and the same design features apply to all selected people and households. When auxiliary information is available from registry data or interviewer observations, then survey designs may be tailored to optimize response rates and to reduce nonresponse selectivity. Although a general terminology is lacking in the literature such designs are sometimes referred to as adaptive survey designs.

With this paper we aim to describe the basic ingredients of adaptive survey designs, to systematize these designs by providing a mathematical framework, and to illustrate their potential to improve efficiency of survey data collection.

Adaptive survey designs assume that different people or households may receive different treatments. These treatments are defined before the survey starts, but may also be updated via data that is observed during data collection. In other words, allocation of treatments is based on data that is linked to the survey sample and on paradata. Paradata are data about the survey data collection process, e.g. observations of interviewers about the neighborhood, the dwelling or the respondents, or the performance of interviewers themselves. In this paper, paradata are used in the widest sense as data that are observed during data collection and that are informative about the response behavior of sampled people and households. A general introduction to adaptive survey designs is given by Wagner (2008).

Adaptive survey designs find their origin in the literature on medical statistics where treatments are varied beforehand over patient groups but also depend on the responses of patients, i.e. depend on measurements during data collection. See for example Heyd and Carlin (1999) or Murphy (2003). Generally, in the treatments a number of stages is identified at which the patient status is evaluated and treatment regimes may change. Such designs are called dynamic. Here static and dynamic survey designs are distinguished. Static survey designs are set before data collection starts and do not depend on observations collected during data collection. Dynamic survey designs are partly based on observations made by interviewers or data collection staff. To illustrate the difference, offering different levels of pre-paid incentives is a static design, while offering different maximum levels of incentives given an interviewer assessment of the propensity to respond is a dynamic design.

In medical statistics, a surrogate or marker for the target quantity of interest, e.g. blood cell count or blood oxygen content, is usually employed to derive the effectiveness of treatments. Such surrogates are needed because the target quantity, e.g. the cure from a disease, cannot be measured directly or timely. There is again a close parallel to survey designs where the error due to nonresponse is of interest. The error itself cannot be measured exactly and is substituted by quality indicators like the response rate. Hence, adaptive survey designs need indicators of survey quality.

There is an extensive set of literature on assumptions about treatments and treatment responses that enable statistical inference from treatment data. One important assumption is about unobserved covariates in the allocation of treatments; the observed data must be complete and no unobserved data may affect the decisions to assign treatments. If such covariates would exist then conclusions that are drawn may not be valid. Treatment must not be confounded with the final outcome. In interviewer-administered surveys, the interviewer is taught to tailor the approach to the respondent and to maintain interaction. Interviewers often base their timing of subsequent visits on the result of the first visit. Based on the appearance of the dwelling, the apartment building and the mail box, or based on contact with neighbours, they may decide to adapt their contact strategy. Although the strategy of the interviewer may be effective and efficient, from the point of view of the survey organization such changes employ unobserved data, and assumptions about the impact of contact timing protocols are violated when data underlying the decisions of interviewers are missing or incomplete. For this reason, most survey organizations ask interviewers to keep records of the sampled units that are assigned to them. These records should enable data collection staff to recreate the history of the sample. Adaptive survey designs, therefore, must restrict focus to design features that are largely under control and that employ observed data only.

A special case of an adaptive survey design is the responsive survey design. Responsive survey designs were introduced by Groves and Heeringa (2006). Like general adaptive survey designs, responsive survey designs may apply differential design features to sample units. However, the main restriction is that responsive survey designs identify promising and effective treatments or design features during data collection. In order to do so, the data collection is divided into multiple design phases. A new phase employs the outcomes of randomized contrasts between sample units in previous phases to distinguish effective from ineffective treatments and to identify costs associated with the treatments. Randomized contrasts are differences in response rates between subpopulations for randomly assigned design features. See for example Mohl and Laflamme (2007), Karaganis and Laflamme (2009), Tabuchi et al (2010) and Peytchev et al (2010). The allocation of design features must be done in such a way that each phase reaches its phase capacity, which is the optimal trade off between quality and costs. Responsive designs are motivated by survey settings where little is known about the sample beforehand and/or little information about the effectiveness of treatments is available from historic data. In these settings multiple phases are needed and responsive designs are very practical. If the second and higher design phases of responsive designs are considered, however, then the starting point is similar to survey settings where substantial prior information about sample units is available or where a survey is repeated many times. The only distinction being that in previous design phases part of the sample responded already. It is assumed that historic data is available and that effective treatments are identified beforehand.

Although survey literature has devoted considerable attention to trade offs in survey designs between the various surveys errors, e.g. Lyberg et al (1997) and Dillman

(2007), in survey practice there are surprisingly few cases where differential design features are investigated and implemented. Some literature is devoted to the timing of call attempts, visits and call backs, e.g. Greenberg and Stokes (1990), Kalsbeek et al (1994) and Kulka and Weeks (1988). Optimal timing protocols focus on two causes of nonresponse, unknown non-eligibility and non-contact. Such protocols may be regarded as dynamic adaptive survey designs as they change timing of call attempts or visits based on the contact history. However, in most cases the protocol does not depend on person or household characteristics that are known prior to the start of data collection and that are available from linked data. Research into adaptive survey designs is still in its infancy.

Section 2 describes theory and concepts behind adaptive survey designs. Section 3 presents two examples based on virtual survey data. Section 4 discusses a simulation study based on real survey data. Section 5 gives a brief summary of a pilot study conducted within the 7th EU Research Framework Programme project RISQ. Finally, section 6 presents a summary and discussion.

2. What are adaptive survey designs?

2.1 Adaptive survey designs in general

In this section a mathematical framework is set out for adaptive survey designs. In subsequent sections, components of this framework are highlighted and elaborated.

Let the population consist of units $k = 1,2,\ldots,N$. The population of interest may consist of all units in a population but also of all respondents to a previous wave in a panel study. To each unit will be assigned a strategy s from the set of candidate strategies $S = \{\phi, s_1, s_2, \ldots, s_M\}$. In the *survey strategy set* S the empty strategy ϕ is explicitly included. The empty strategy means that no action is undertaken, i.e. the population unit is not sampled. This is the most general framework. In practice, one will often separate the sampling design from the strategy allocation and view the sample as given and fixed. However, one may include the decision to sample a unit explicitly in the overall allocation of resources.

In general a strategy s is a specified set of design features and may involve a sequence of treatments where treatments are only followed when all previous treatments failed. Some of those features may be sequential like the type of contact mode and the type of survey mode, but the features may also describe different aspect of a survey design. Examples of strategies are

s_1 = (advance letter 1, web questionnaire, one reminder);
s_2 = (advance letter 1, web questionnaire, no reminder);
s_3 = (advance letter 2, CATI administered, maximum of six call attempts);
s_4 = (advance letter 2, CATI administered, maximum of 15 call attempts).

Section 2.2 discusses the strategies and design features in more detail. It is assumed in this paper that the set of strategies S is known and fixed when strategy allocation is started. The set of strategies may be identified based on historical survey data, experience and pilot studies. We refer to Schouten et al (2010) for guidelines and examples on how to construct strategy sets.

The notation for strategy sets adopted here is different from notation that is used in adaptive designs for medical treatments as for example in Wagner (2008). We do not distinguish available actions and states as a function of time. A strategy in our notation is a sequence of actions, and we implicitly assume a sequence of states occurred when such a strategy is entirely conducted. However, both notations are easily translated to each other.

With each population unit k a vector of covariates $X_k = (X_{k1}, X_{k2}, \ldots, X_{kp})'$ is associated. X_k contains characteristics that are known before data collection starts and before strategies are allocated. The covariates must, therefore, be available in registrations or administrative data that can be linked to the sampling frame or in the sampling frame itself. Next to these general characteristics a second vector of covariates $\widetilde{X}_k = (\widetilde{X}_{k1}, \widetilde{X}_{k2}, \ldots, \widetilde{X}_{kq})'$ may exist for unit k that reflects characteristics observed during data collection for all sampled population units. These characteristics are termed paradata or process data because they are collected during the process of data collection by interviewers and data collection staff. However, other than the more traditional view on paradata as information about the process, in the adaptive survey design context $\widetilde{X}_k$ contains observations about the sampled person or household. Examples of X_k are gender, age, type of household or educational level. Examples of $\widetilde{X}_k$ are the interviewer assessment of the propensity to respond or the propensity to be contacted, the state of the dwelling or the neighborhood, and the presence of an intercom. $\widetilde{X}_k$ is deliberately restricted to observations about the sample that allow for differentiation of survey design features. It does not contain the values of the design features themselves like the interviewer that was assigned to the address.

The important distinction between X_k and $\widetilde{X}_k$ is the level of availability. $\widetilde{X}_k$ is known only for those units that are sampled and cannot be used in distinguishing subpopulations a priori. Let $q(x)$ represent the distribution of X_k in the population and $q(\widetilde{x}, x)$ the joint distribution of X_k and $\widetilde{X}_k$ in the sample. Furthermore, $q(\widetilde{x} \mid x)$ denotes the conditional sample distribution. It is assumed that $q(x)$ and $q(\widetilde{x}, x)$ are known in advance. In settings where no or little data can be linked, strategy allocation must be based fully on observations made during data collection.

Adaptive survey designs that allocate strategies beforehand based on population characteristics are termed *static*, while adaptive survey designs that allocate strategies that depend (also) on paradata are termed *dynamic*. It is important to remark that both static and dynamic designs have a strategy set that is fixed before data collection starts. However, for dynamic designs it is not known beforehand which strategies are going to be used because the choice of strategy depends on data that is observed during data collection.

Let $\rho(x,s)$ be the *response propensity* of a unit carrying characteristic $X = x$ and that is assigned strategy s. It is assumed that $\rho(x,s)$ is available from historic data, i.e. from previous versions of the same survey, from surveys with similar topics and designs or from initial design phases. Obviously, the anticipated response propensity must be a close estimate of the true propensity. Section 2.5 returns to this essential component of adaptive survey designs.

The *expected costs* of the assignment of strategy s to a unit with $X = x$ is denoted as $c(x,s)$. It is an individual cost component. Literature tells us that survey costs consist of many components of which some are overhead and others are individual, e.g. Groves (1989). Section 2.3 discusses cost functions.

Let $p(s \mid x)$ be the *allocation probabilit*y of a population unit with characteristics x for strategy s, and let $p(s \mid x, \tilde{x})$ be the allocation probability to that strategy given that also paradata $\tilde{x}$ is observed. The following must hold

$$0 \le p(s \mid x) \le 1, \quad 0 \le p(s \mid x, \tilde{x}) \le 1 \tag{1}$$

$$\sum_s p(s \mid x) = 1, \quad \sum_s p(s \mid x, \tilde{x}) = 1, \tag{2}$$

i.e. all units are assigned a strategy. In general, allocation probabilities may have values between 0 and 1. In other words subpopulations with the same scores on x and $\tilde{x}$ may be (randomly) assigned to different strategies. For instance, only part of the non-respondents may be re-approached in a follow-up. Allowing for allocation probabilities between 0 and 1 increases the flexibility in meeting quality levels or cost constraints. In the following, p denotes the matrix of allocation probabilities, i.e. $p = \{p(s_j \mid x, \tilde{x})\}_{1 \le j \le M, x, \tilde{x}}$ and contains the decision variables in the optimization.

The response propensities ρ_X can be derived from the strategy response propensities and the allocation probabilities by

$$\rho_X(x) = \sum_{s \in S} \sum_{\tilde{x}} q(x \mid \tilde{x}) p(s \mid x, \tilde{x}) \rho(s, x). \tag{3}$$

The strategies, covariates, response propensities, cost functions and allocation probabilities form the ingredients to adaptive survey designs. With these building blocks the adaptive survey design optimization problem can be formulated. Two ingredients are still missing, however, a quality function and an overall cost function. Let $Q(p)$ be some indicator of quality and $C(p)$ be an evaluation of total costs. The dependence on the allocation probabilities in both functions is stressed as the probabilities are the decision variables in the optimization.

The optimization can now be formulated as

$$\max_p Q(p) \text{ given that } C(p) \le C_{\max} \tag{4}$$

or as

$$\min_p C(p) \text{ given that } Q(p) \geq Q_{\min}, \tag{5}$$

where $C_{\max}$ represents the budget for a survey and $Q_{\min}$ minimum quality constraints. Problems (4) and (5) are called dual optimization problems, although the solutions to both problems may be different depending on the quality and cost constraints. Sections 2.3 proposes a number of quality functions.

Responsive survey designs (Groves and Heeringa 2006) show much similarity to adaptive survey design, but they are different on a number of key features. Responsive designs introduce design phases in data collection. Strategies may be identified at the end one phase and applied to subsequent phases. In responsive designs paradata about the effectiveness and costs of strategies play an important role for that reason and are motivated by survey settings where little or nothing is known in advance. Individual response propensities and cost functions are to some extent identified and assessed during data collection based on randomized contrasts, i.e. differences in response behavior between subpopulations with respect to design features that are randomly assigned.

2.2 Survey strategies and survey design features

A survey strategy is a specified list of survey design features. In the literature many such design features are suggested and evaluated, e.g. Groves and Couper (1998) and Groves et al (2002). Here, only a brief summary of commonly used features is presented:

- Length of the period of data collection
- Survey mode or sequence of survey modes in mixed-mode surveys
- Contact mode
- Invitation or advance letter
    - Type of letter;
    - Type of reminder
    - Number of reminders;
- Contact protocol
    - Number of attempts per address;
    - Timing of attempts per address;
- Incentives
    - Type of incentive
    - Level of incentive
- Questionnaire type
    - Full questionnaire
    - Basic (short) questions
- Proxy or self report
- Assignment and re-assignment to interviewers

The length of the data collection can be varied for different people or households, for instance by lengthening data collection for groups that are harder to contact or by performing follow-ups for refusal conversion for specific groups. The main restriction is the publication date of the survey results. It makes a considerable difference whether questions are asked with respect to a fixed date or with respect to a fixed time lag.

Common survey modes are paper, web, telephone and face-to-face. The different survey modes generally lead to different survey errors like coverage errors, nonresponse and measurement errors. Furthermore, different costs are associated with the modes. Interviewer-administered modes are generally more expensive than self-administered modes. Hence, it may be fruitful both from the quality and costs perspective to allocate different modes to different groups, thus, moving to a mixed-mode survey design. For example, one may not assign elderly people to web. One may also decide to re-assign nonrespondents to other modes after a specified time period. See De Leeuw (2008).

The contact mode is the mode in which the sample unit is informed about the survey and is invited to participate. In some cases the contact and survey mode may coincide as the sample unit is invited to participate right away. The set of contact modes is the same as the set of survey modes. However, in surveys where samples of addresses are drawn, the most common contact mode is paper. These surveys use advance or invitation letters. In surveys based on random digit dialing, telephone naturally is the contact mode. In many surveys, there is little opportunity to vary contact modes. This, however, changes for subsequent waves in multi-wave surveys or panels. For instance, the Dutch Labour Force Survey asks respondents to the first wave for their telephone numbers for the other waves. Alternatively, one may ask for e-mail addresses or may ask respondents for their favorite survey mode.

When advance letters or advance e-mails are sent, the content may be personalized. Examples are different language options or tailored wording and style. Furthermore, reminders may be sent. Both the number and form of reminders may be varied over subpopulations. For instance, groups with low response rates may receive more reminders or a higher frequency of reminders. See Dillman (2007) and De Leeuw et al (2007).

One of the design features that is investigated for interviewer-administered survey modes is the contact protocol. Usually, some protocol is defined by survey organisations. They may enforce a minimum number of contact attempts, spreading of visits and calls over week and weekend days, and spreading over different time slices during the day. Generally, telephone surveys allow for more detailed contact protocols than face-to-face surveys due to the travelling distances. It is common for telephone surveys that calls and appointments for calls are operated by a management system. Contact protocols may be different for different groups in the population, either by allowing for different numbers of calls or visits or by tailoring timing based on linked data and paradata. See for instance Wagner (2008).

Some surveys employ incentives to increase response rates, especially for people not intrinsically interested in the survey. Apart from the type of incentive (voucher, money, small presents or stamps), also the form (pre-paid, promised or conditional) may be varied. One may decide to offer higher incentives to people or households that have high refusal rates. See for instance Barón et al (2009).

Next to the advance letter also the questionnaire itself may be differentiated in content and length. Different versions of the questionnaire may be available in split questionnaire designs or shorter, condensed versions of the questionnaire may be developed for follow-up. Examples are the basic question approach, (Kersten and Bethlehem 1984, Cobben 2009) and the PEDAKSI approach (Lynn 2003). Groups that have high refusal rates may be offered short questionnaires.

A design feature that is often used to increase response rates, especially in household surveys, is proxy reporting. Proxy reporting as opposed to self-reporting may be offered in order to reduce the number of calls and visits and to reduce travelling costs. Proxy reporting may, however, lead to an increase in measurement error. One may, therefore, decide to offer the option of proxy reporting to certain subpopulations only. See for instance Moore (1988).

Finally, in interviewer-assisted contact and survey modes one may assign interviewers to different sample units based on their historic performance or their match to the person or household. Examples are specialized market research companies that have multilingual interviewers or interviewers that are trained for city districts that are generally viewed as difficult. Apart from the prior assignment of interviewers, one may also choose to re-assign specific nonresponse cases to special interviewers. An example is presented in Cobben (2009).

Theoretically, all or a number of the design features in this section and other design features not mentioned here may be varied simultaneously. However, it is important to understand that the more design features are optional the more decision variables are constructed in the optimization and the more reliance is attached to historic data or initial design phases.

2.3 Quality objective functions

It would go beyond the scope of this paper to give a full and elaborate display of quality of surveys. Nevertheless, it is important to keep in mind that other survey errors may sometimes play a dominant role and that design features that have a high risk of survey errors other than nonresponse must be avoided. Examples are survey modes with low coverage of some subpopulations or a higher risk of response bias, increased item nonresponse in proxy reporting or increased response bias in follow-ups.

When quality is optimized according to (4), then quality functions map the survey sample with linked data, paradata and answers to survey items to a single value which can be interpreted and optimized. When costs are minimized subject to constraints on quality as in (5), then quality may be multi-dimensional (but cost functions should be one-dimensional). In the following, quality functions that allow for optimization of quality are discussed.

In general two types of quality functions can be distinguished; quality functions that employ covariates from linked data and paradata only, and quality functions that also employ the answers to the survey target variables. We refer to them as *covariate-based* and *item-based*, respectively. An item-based quality function is a function of the response distribution of a survey item and the anticipated, estimated full population distribution given the available linked data and paradata. The main distinction between covariate-based and item-based quality functions is that item-based quality requires assumptions. Evidently, the answers of nonrespondents are missing. Hence, quality evaluation must be purely based on relations between target variables and covariates as observed in the response. As a consequence, there is a risk attached to item-based quality functions that originates directly from the phenomenon it attempts to measure. Relations between target variables and covariates may be different for nonrespondents and item-based quality may pose an incomplete image. Furthermore, in surveys with many survey target variables, different target variables may lead to different decisions and optimal survey designs. However, contrary to covariate-based quality functions, item-based quality functions tailor survey designs specifically to the topics of the survey. For two different surveys with similar response patterns, covariate-based quality functions will lead to the same optimal survey designs.

The most obvious covariate-based quality function is the response rate. It is not a true covariate-based quality function in the sense that it depends on linked data or paradata. However, since the 0-1 response indicator may be viewed as the simplest form of paradata, it is termed a covariate-based quality function. There are various definitions of response rates, and one would have to pick a suitable one for the optimization. The response rate is represented as the mean response propensity

Response rate: $$Q(p) = \hat{\bar{\rho}}_X \tag{6}$$

Other candidate covariate-based quality functions are given in Schouten et al (2009) and Särndal and Lundström (2008):

R-indicator: $$Q(p) = 1 - 2\hat{S}(\hat{\rho}_X) \tag{7}$$

Maximal bias: $$Q(p) = \hat{B}_m(\hat{\rho}_X) = \frac{\hat{S}(\hat{\rho}_X)}{\hat{\bar{\rho}}_X} \tag{8}$$

Variance of calibration weights:

$$Q(p) = q^2 = [\sum\nolimits_{k=1}^{N} a_k R_k d_k]^{-1}[\sum\nolimits_{k=1}^{N} a_k R_k d_k (\hat{\phi}_k - \hat{\bar{\phi}})^2] \tag{9}$$

In (7) to (9) $\hat{S}(\hat{\rho}_X)$ is the standard deviation of response propensities estimated with covariates X. In (9), d_k is the inclusion weight, a_k the 0-1 sample indicator, R_k the 0-1 response indicator and $\hat{\phi}_k = \hat{\rho}_{X_k}^{-1}$ is the inverse response propensity, also referred to as response influence.

The maximal bias attempts to estimate the standardized maximal absolute bias of response means. It assumes maximal correlation between response and survey target variable. The variance of calibration weights was originally designed to serve as a tool in the selection of weighting variables. However, when fixing X it may also serve as a quality function. A smaller q^2 is preferred as it implies that one has to rely less on the nonresponse adjustment.

Quality functions (6) to (9) do not account for the impact of nonresponse on precision. An alternative that does may be the extension of (8) to the maximal mean square error, i.e. the sum of squares of the maximal bias plus the standard error

Maximal mean square error:

$$Q(p) = MSE_m(\hat{\rho}_X) = \left(\frac{\hat{S}(\hat{\rho}_X)}{\hat{\bar{\rho}}_X} \right)^2 + \lambda \frac{1}{n\hat{\bar{\rho}}_X}. \tag{10}$$

The quality function MSE_m represents the standardized maximal mean square error in a simple random sample. The standard error in such a sampling design is $S^2(y)/(n\bar{\rho}_X)$, ignoring the finite sample correction. Standardization by $S^2(y)$ leads to the second term on the righthandside of (10) except for λ. Parameter λ is included in order to be able to vary the weight that is assigned to precision as opposed to bias. The larger λ the more importance is attached to precision. Since the standardized variance is proportional to the inverse of the response rate, large λ conform to maximizing the response rate.

Examples of item-based quality functions are presented by Groves and Heeringa (2006), Wagner (2008) and Rao et al (2008). For a specific target variable *Y*, Groves and Heeringa (2006) suggest the estimated nonresponse bias of the response mean

Estimated nonresponse bias: $Q(p) = \hat{B}(\bar{y}_R) = \dfrac{\mathrm{cov}(Y, \hat{\rho}_x)}{\hat{\bar{\rho}}}$, (11)

with $\mathrm{cov}(Y, \hat{\rho}_X)$ the response covariance between the target variable and the estimated response propensities given covariates *X*. Wagner (2008) and Andridge and Little (2010) propose the fraction of missing information as a quality criterion

Fraction of missing information: $Q(p) = (1 + \dfrac{1}{D}) \dfrac{DB_D}{DW_D + (D+1)B_D}$, (12)

where B_D and W_D are, respectively, the between-imputation variance and the within-imputation variance of the estimated population parameter in D multiply imputed data sets. The obvious population parameter is the population mean of a survey target variable which may be estimated by the sample mean of the imputed target variable. Multiple imputation means that the nonresponse for the target variable is fully imputed in D independent runs. For relatively modest D, (12) is equal to the relative size of the between-imputation variance. A small fraction of missing information is preferable as it implies that the imputed data sets lead to stable estimates for the population mean.

Quality functions (6) to (12) clearly depend on the allocation probabilities p, although this is not directly apparent from their definitions. The dependence is illustrated for the response rate. The response rate can be written in terms of a weighted response rate

$$Q(p) = \overline{\rho}_X = \sum\nolimits_{x,s} q(x)p(s \mid x)\rho(x,s) \,. \tag{13}$$

Hence, (13) indicates that the response rate is a linear function of the allocation probabilities, the decision variables in the optimization. Similar derivations can be made for the other quality functions, but in general the quality functions lead to nonlinear objective functions in the optimization.

Currently, adaptive and responsive survey designs focus on the nonresponse error and completely ignore the response error. It is well known, however, that survey design features, e.g. the survey mode, may have a strong impact on response error and, consequently, on the total survey error. Also, for some survey topics, e.g. voting behaviour, response error and nonresponse error may have comparable magnitudes. Quality functions incorporating other survey errors are an important topic for future research.

2.4 Cost functions

Cost functions are the counterpart of the quality functions. There are several components to cost functions. It is important to restrict specification of costs relative to the design features that are varied in the adaptive survey design. For example, when it is the incentive that is differentiated with respect to different subpopulations, then costs need not be specified and detailed for interviewer traveling times. When it is the contact timing protocol that is the design feature that may be tailored, then, obviously, traveling times and costs play a dominant role.

If all features of section 2.2 are optional, then the cost functions have complex forms with many overhead cost and variable cost components. Generally, variable costs depend on the sample size while overhead costs do not. Overhead cost components may come from data collection staff, sampling and processing of samples. Variable costs arise for example from training and instruction of interviewers, mail and print of questionnaires and reminders, processing of paper questionnaires, interviewer hour rates and travel expenses, incentives and telephone number linkage, telephone usage and computer servers.

Before starting the optimization one has to distinguish between permanent cost components and starting up cost components. For example, once survey questionnaires are developed for all survey modes, the costs for questionnaire design do not have to be accounted for as long as the questionnaires remain unchanged. Starting up cost components may be ignored or spread over time when it is anticipated that the survey will be conducted frequently in the future without major redesigns.

In the optimization two cost components may be identified: a fixed and a variable component. The variable component depends on the allocation of population units to strategies while the fixed component consists of all remaining costs. It must be stressed that the fixed component is different for adaptive survey designs that focus on different design features. For example, when pre-paid incentives are varied in amount in a face-to-face survey and no attempts are made to convert refusing households, the variable component consists of the incentives themselves and the variable interviewer costs for conducting interviews. All other costs, like the travel expenses by interviewers, are fixed assuming that incentives do not affect the propensity to contact a household. Hence, the cost function $C(p)$ is the sum of two components

$$C(p) = C_F + C_V(p), \tag{14}$$

of which only the second, the variable component, depends on the allocation probabilities. Clearly, in the optimization the fixed costs may be subtracted from the budget in order to derive the budget for the variable costs.

In general, with a survey strategy s costs $c(x,s)$ are associated for population units from group x. The individual cost function may be a function of response propensities, or even more specifically of contact and participation propensities. For instance, the interviewer costs in different contact timing protocols depend on the contact rates of the selected subpopulations. The cost function $c(x,s)$ is a relative cost function as it describes only the contribution of the strategy to the variable cost component $C_V(p)$

$$C_V(p) = \sum_{x,s} q(x)p(s \mid x)c(x,s) . \tag{15}$$

Three additional remarks are in place. First, the derivation of fixed and variable cost components is complicated when a survey organization runs many surveys in parallel. Usually, interviewers are allowed to work on different surveys at the same time, thus, allowing for efficiency in traveling costs by clustering sample units. On one hand, the interaction between surveys makes it hard to separate costs per survey, especially when strategies are tailored. Simple heuristics must be used to disentangle costs. On the other hand, when multiple surveys are conducted some of the variable costs components may be labeled as fixed. For example, when only a relatively small number of population units are assigned to the face-to-face survey mode, then traveling costs may be assumed to remain unchanged as the addresses are clustered with addresses from other surveys. The second remark concerns the multidimensional aspect of costs. Apart from the overall budget it may be requested that interviewer occupation rates are close to one throughout time or that none of the interviewers has to work overtime more than a fixed amount of time. Such constraints can be incorporated in the cost function as well. As a consequence, the cost function becomes a vector and the constraint a vector of constraints. For the optimization this only affects the range of feasible solutions. The third remark concerns the validity of the cost functions. Since cost functions are hard to construct in practice, it may turn out that the optimization was too optimistic. This will result in incomplete or aborted strategies during data collection such as a smaller number of interviewer visits. It is important, therefore, to monitor data collection closely and to build in indicators for strategies; ideally this should be done for every decision variable in the survey design. This form of paradata is needed to evaluate the validity of the optimization and the adaptation of the design if necessary.

2.5 Estimating response probabilities

Next to cost functions, the other important ingredient of adaptive survey designs is the set of response propensities for the various strategies. Such propensities need to be known from past surveys, preferably the same survey or otherwise a similar survey. Alternatively, a Groves and Heeringa (2006) propose, one may use earlier phases of the data collection to learn and derive propensities. This will be at the expense of efficiency since part of the survey is already conducted. Nonetheless, the gathered information directly feeds back to the current survey. In either case, it is essential that response propensities are estimated from randomized contrasts, which means that historic data must carry randomized assignments to different design features. For example, from a face-to-face survey with a maximum of three visits it cannot be inferred what the response propensities would be when the maximum is raised to ten visits. Similarly, in telephone surveys with evening calls only, it cannot be deduced what contact propensities are for day time shifts.

As a consequence, one should be careful in selecting the number and type of design features to be optimized. Historic survey data may not have the appropriate randomizations to support the estimation of corresponding propensities without strong assumptions. Responsive survey designs to some extent account for this problem by explicitly allowing for data mining of design features in the initial design phase. However, still informative randomization is indispensable and obviously only part of the design features can be randomized.

A statistical institute that carries out adaptive survey designs on a regular basis, may always allocate a relatively small proportion of its samples to alternative design features. The small sample is used to pick up changes in contact and participation behaviour over time and to identify new and promising treatments. This clearly comes at a cost and, hence, it is very useful that statistical institutes disseminate and exchange findings.

Literature on household surveys gives an extensive list of models for response that include design features. The most traditional models are logistic regression models that, apart from covariates, contain a number of design features. Since some design features affect only certain types of nonresponse, i.e. incentives may have an impact on refusals mostly, the regression models may be extended to nested or sequential regression models. For example, the model may distinguish an equation for obtaining contact and an equation for obtaining cooperation. In the cooperation equation the incentive type and level may be included as independent variable. When design features are only partially randomized over subpopulations, then often multilevel models are used. The first level then corresponds to the design feature, e.g. the interviewer or the survey mode, and the second level to sample units.

The common denominator in all models is that response propensities are estimated based on a number of assumptions about the true nature of the nonresponse missing-data mechanism. In general such models are simplifications. Consequently, anticipated response propensities $\rho(x,s)$ have a standard error. In the optimization this uncertainty can be accounted for by allowing response propensities to be random variables rather than fixed quantities. The randomness allows for sensitivity analyses and evaluations of the robustness of the optimization. It must be realized that the optimization leads to an expected quality level and to expected costs. Sensitivity analyses provide insight into the variation of quality and costs when the survey is conducted multiple times (under the same circumstances).

3. Examples

In this section we provide examples of a static and a dynamic design: the choice of the survey mode in a household survey for two age groups and the re-assignment of interviewers based on observations of the propensity to cooperate. Both examples are based on hypothetical response propensities and cost functions.

3.1 An example of a static adaptive survey design: choosing the survey mode

As the objective function we take the response rate given by (6), i.e. we search for the assignment of survey modes to two age groups that leads to a maximal response rate.

Suppose the strategy set consists of four elements, $S = \{s_1, s_2, s_3, s_4\}$, where the elements are equal to

s_1 = *web administered, 1 reminder,*

s_2 = *web administered, no reminders, CAPI administered, maximum of 3 visits*,

s_3 = *CAPI administered, maximum of 6 visits,*

s_4 = *the empty strategy.*

Strategies s_1 and s_3 lead to a single mode design and s_2 describes a mixed-mode survey where CAPI is administered to the nonrespondents from an initial wave of web. The web survey is combined with one reminder when no follow-up is done with CAPI. The CAPI sample receives a maximum of six visits. The web-CAPI mixed-mode survey has no reminders and limits the number of visits to a maximum of three in order to reduce the overall length of the data collection period.

We assume individual hypothetical cost functions that are build using the following components:

- administration of web questionnaire = €5;
- web reminder = €2;
- one visit = €15;
- one interview = €20.

The maximal budget is varied in order to investigate the trade off between costs and quality. The budget range will be 0 to 50000 euro.

The population has a size of 2000 people and is divided into two groups based on age; $X \in \{\leq 35, > 35\}$. We let the two groups have equal size; $q(x) = (0.5, 0.5)'$ is assumed to be the distribution of X

Note that the design is static, since no paradata are employed. Since the empty strategy is part of the strategy set, we implicitly also choose the inclusion probabilities for both age groups.

Let $p(s \mid x)$ be the allocation probability to strategy $s \in S$ given $x \in X$. Consider the quality function Q given by the weighted response rate:

$$Q(p) = \sum_{x,s} q(x) p(s \mid x) \rho(x,s).$$

The expected costs are computed per subpopulation per strategy. The empty strategy obviously costs 0 euro and leads to a zero response probability. Assuming there is enough interviewer capacity to carry out the survey regardless of the chosen strategy[1], the problem at hand is how to assign strategies to subpopulations such that the response rate is maximized and there are no budget overruns.

Hypothetical response probabilities and total costs per strategy per subpopulation are given in table 3.1.1. Strategy s_1 is the cheapest and s_3 the most expensive. Strategy s_2 leads to the highest response rate. From these parameters we may already conclude that strategy s_3 will not be selected since it delivers less quality than s_2 and it is more expensive.

Table 3.1.1 - Response probabilities and cost for the four strategies. Costs are based on a full assignment of the group to the strategy.

	Younger or equal to 35 years				*Older than 35 years*			
Strategy	s_1	s_2	s_3	s_4	s_1	s_2	s_3	s_4
Response	0.432	0.789	0.684	0	0.574	0.806	0.663	0
Cost	6320	20094	20535	0	6120	20103	23132	0

Table 3.1.2 shows the response rate as a function of the available budget. The response probabilities were fixed. A budget larger than €40.200 does not lead to any further improvement to the response rate. For comparison the R-indicator given by (7) and the maximal bias given by (8) are also shown. Figure 3.1.1 displays the evolution of the response rate for various budget levels.

[1] This assumption is needed for the simple setting of the model.

Table 3.1.2:The response rate and R-indicator as a function of budget.

	Budget					
	5.000	*10.000*	*20.000*	*30.000*	*40.000*	*40.200*
Response rate	0.235	0.420	0.601	0.713	0.796	0.797
Remaining budget	0	0	0	0	0	2
R-indicator - age	0.435	0.571	0.743	0.790	0.959	0.983
Maximal bias - age	1.202	0.511	0.214	0.147	0.026	0.011

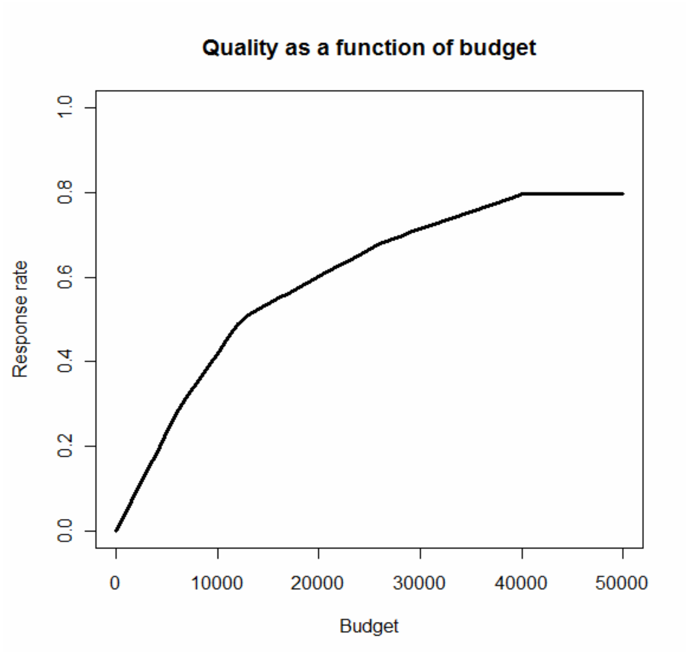

Fig. 3.1.1 – Response rate as a function of budget

Table 3.1.3 shows the optimal values of the allocations probabilities for different maximal budget levels. From the allocation probabilities we can conclude that when budget is increased the strategies assigned to the two age groups are:

- (s_4, s_4): With zero budget none of the sample units are approached.
- (s_4, s_1): The cheapest strategy is applied to older persons. Older persons are preferred to younger persons as their response probabilities are higher.
- (s_1, s_1): The cheapest strategy is now also applied to younger persons.
- (s_1, s_2): Older persons receive the mixed-mode strategy as their response probabilities are higher than for younger persons.
- (s_2, s_2). Younger persons also get the mixed-mode strategy. The full CAPI strategy is completely ignored as expected.

Table 3.1.3 - Allocation probabilities

		Budget					
		5.000	*10.000*	*20.000*	*30.000*	*40.000*	*40.200*
Young	$p(s_1 \mid x_1)$	0	0.614	0.451	0	0	0
	$p(s_2 \mid x_1)$	0	0	0.549	1	1	1
	$p(s_3 \mid x_1)$	0	0	0	0	0	0
	$p(s_4 \mid x_1)$	1	0.386	0	0	0	0
Old	$p(s_1 \mid x_2)$	0.817	1	1	0.729	0.014	0
	$p(s_2 \mid x_2)$	0	0	0	0.27	0.986	1
	$p(s_3 \mid x_2)$	0	0	0	0	0	0
	$p(s_4 \mid x_2)$	0.183	0	0	0	0	0

In order to analyze the dependence on the quality objective function that is used, figure 3.1.2 displays the evolution of another quality objective function, the maximum bias, for various budget levels. Although the maximal bias was not taken as the quality objective function it almost shows a constant decrease. Only when (s_1, s_1) is gradually replaced by (s_1, s_2), the maximal bias increases slightly.

This simple example also illustrates that different objective functions lead to different decisions and different costs. Optimization of the response rate in this case does not conform to optimization of the maximal bias or R-indicator.

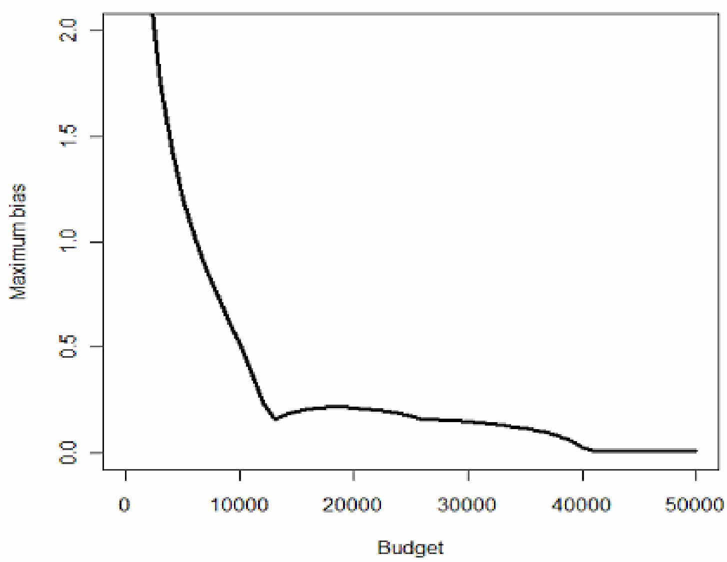

Fig. 3.1.2 - Maximum bias as a function of budget

3.2 An example of a dynamic adaptive survey design: re-assigning interviewers

In the second example we construct a hypothetical dynamic design. Interviewers are assigned to sample cases that have refused once, based on an assessment of the propensity to respond made during a first phase of the survey. The assessment is made for respondents and refusers, but is not available for people who were not contacted during the first phase. It provides a judgement on the propensity that the person participates in the survey when contacted again. The assessment is made on a three point scale: *easy, medium, difficult*. *Easy* means that there is a high probability that if contacted again the sample unit would respond.

After a first phase of data collection, the intermediate survey results are evaluated and people are divided into respondents, refusers and noncontacts. Obviously, respondents are not approached again. Refusers receive a different treatment. Interviewers are rated based on their historic performance and grouped in *good* and *less good* interviewers. Refusers are re-assigned to one of the two groups of interviewers. Since there is no assessment available for non-contacts, the treatment for this group is not altered.

We use the R-indicator given by (7) as the quality objective function. We split the sample using *X=(age)* into two groups, labelled as *young* and *old*. The goal in the second phase is to assign refusers to the two interviewer groups such that the R-indicator with respect to age is optimized.

Let N be the sample size of the survey. Given the available administrative data X, we split the sample into the two subpopulations, i.e. *young* and *old*, with q^1 and q^2 the respective population proportions. We let q_i^k be the probability that a sample unit from subpopulation k is of type i, where $i \in L = \{\texttt{easy, medium, difficult}\}$. Furthermore, let λ_i^k be the probability that a sample unit of type i from subpopulation k is a refusal. If a person is not a refuser, then μ_i^k is the probability that the person either was a respondent after the first phase or becomes a respondent when he/she was a noncontact after the first phase.

The total number of interviewers is M and p_j M represents the number of interviewers with skill $j \in J = \{\texttt{good, less good}\}$, $0 \le p_j \le 1$ and $p_{\texttt{good}} + p_{\texttt{less good}} = 1$. We assume that each interviewer can handle at most c cases in the second phase of the survey. The probability that a refusal of type i from subpopulation k will respond if contacted by an interviewer of skill j is denoted by r_{ij}^k and it is again assumed to be known from previous surveys.

Let $\{x_{ij}^k\}$ be the set of decision variables, where x_{ij}^k represents the probability that a sample unit of type i will be assigned to an interviewer of skill j given that he/she belongs to subpopulation k. In other words, we allow for a random assignment of sample units to the two interviewer groups.

Since decision variables $\{x_{ij}^k\}$ are probabilities, it must hold that $0 \le x_{ij}^k \le 1, \quad \forall i, j, k$. We demand that all refusers after the first phase are re-assigned in the second, i.e. restrict the decision variables to satisfy

$$\sum_j x_{ij}^k = 1, \quad \forall i, k .$$

In this example we express costs in terms of the overall interviewer occupation rates. Since interviewers can handle at most c cases, there are two constraints

$$N \sum_{i,k} q^k q_i^k x_{ij}^k \lambda_i^k \le M p_j c, \quad \forall j \in J .$$

In other words, the total number of cases that can be assigned to interviewers of skill j is restrained to the maximum possible workload.

Now, consider the following input data for the example:

- N = 2000, M = 80, c=30
- $q^1 = q^2 = 0.5$; $q_L^1 = (0.2, 0.3, 0.5)'$, $q_L^2 = (1/3, 1/3, 1/3)'$
- $p_1 = 0.25$ $p_2 = 0.75$;

Table 3.2.1 and 3.2.2 provide the hypothetical response probabilities for the two subgroups when refusal conversion is applied and the cooperation and refusal probabilities for the first phase of data collection.

Table 3.2.1 - Response probabilities when refusal conversion is applied to young and old refusers given the assessment of propensity to respond.

	Young refuser					
	Good interviewer			*Less good interviewer*		
	Easy	*Medium*	*Difficult*	*Easy*	*Medium*	*Difficult*
r_{ij}^k	0.8	0.6	0.4	0.7	0.5	0.3

	Old refuser					
	Good interviewer			*Less good interviewer*		
	Easy	*Medium*	*Difficult*	*Easy*	*Medium*	*Difficult*
r_{ij}^k	0.9	0.7	0.5	0.8	0.6	0.4

Table 3.2.2 - Response and refusal probabilities in the first phase of data collection

	Young			*Old*		
	Easy	*Medium*	*Difficult*	*Easy*	*Medium*	*Difficult*
λ_i^k	0.5	0.6	0.7	0.2	0.3	0.4
μ_i^k	0.85	0.8	0.76	0.95	0.93	0.91

The optimal value of the R-indicator is 0.744. Table 3.2.3 shows the optimal values of the decision variables. It turns out that for all subpopulations and assessments, the decision variables are either 0 or 1, i.e. all re-assignments are non-probabilistic.

Table 3.2.3 – Optimal assignment of cases to interviewers

	Young			*Old*		
	Easy	*Medium*	*Difficult*	*Easy*	*Medium*	*Difficult*
Good	0	1	1	1	0	0
Less good	1	0	0	0	1	1

We investigate the changes in the optimal assignment if we would increase the number of interviewers. Figure 3.2.1 shows the evolution of the objective function and of the R-indicator for various numbers of available interviewers *M*.

For any interviewer number higher than 84 the optimal value cannot improve anymore, because both interviewer groups are sufficiently big in order to handle the entire sample. As a consequence, the cost constraint is not a true constraint anymore. The maximum value of the quality function is, however, different from zero even when the number of interviewers is unlimited. The lack of improvement after a certain level can be explained by the structure of the objective function that requires a balance between response rates over subpopulations. Although there are enough *good* interviewers to handle the entire number of refusals, the difference between the subpopulation response rates cannot be taken away. The re-assignment involves only the refuser follow-up and, hence, the impact is limited. If we would maximize the response rate then the allocation of interviewers will converge towards assigning only *good* interviewers to all cases.

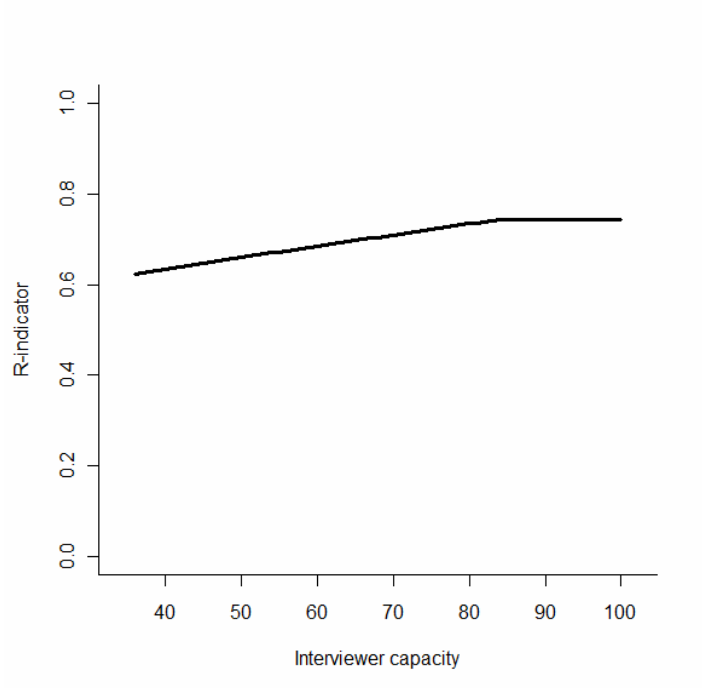

Figure 3.2.1 - Evolution of the R-indicator for different levels of the interviewer capacity

4. A simulation study based on real survey data

In this section a simple simulation study is presented where interviewer assignment is the design feature of interest. The example is taken from project RISQ, which is funded from the 7th EU Research Framework Programme. The website www.risq-project.eu gives background to the project as well as more detailed papers. The response probabilities used in the example are estimated from real survey data.

The Dutch Survey of Consumer Satisfaction (SCS) is a monthly telephone survey about the sentiments of households about their economic situation and expenditure. The survey provides insight into short-term economic development, and early indicators of differences in consumer trends. Each month 1500 households are sampled. The two most influential causes of nonresponse in the SCS are non-contact and refusal. Of the sample 95% is contacted, and of the contacted 71% of the households participate. The response rate is 67%.

One of the most important factors that affect participation is the interviewer. Interviewer's performance may vary greatly when it comes to obtaining response. In total 60 interviewers worked on the SCS during 2005. That means an interviewer had contact with 280 households on average. Interviewer participation rates ranged from 50% to 79%. The lowest rate of 50% was, however, exceptional as the one but lowest participation rate was 61%. The mean interviewer participation rate was 67%. Households were randomly assigned to interviewers in the CATI management system. Hence, it can be concluded that with respect to the interviewer the data are randomized. In the following the interviewer will be the design feature of interest. Hence, the survey strategy set S consists of sixty strategies, $S = \{I_1, I_2, \ldots, I_{60}\}$.

The available covariate vector X contains ethnicity, gender composition of the household core (male, female or mix), average age of the household core in 5-year classes, type of household, degree of urbanization of the neigborhood of residence and average value of houses in the neighborhood. No paradata is used so that the adaptive survey design is static. In the optimization the allocation probabilities $p(I_k \mid x)$ need to be chosen, i.e. it needs to be decided to which interviewers subpopulations are assigned.

The maximal bias

$$Q(p) = \frac{\hat{S}(\hat{\rho}_X)}{\hat{\bar{\rho}}_X}$$

is selected as quality function. For the SCS the maximal bias was 0.14.

To estimate the response propensities $\rho(I_k, x)$ for interviewers, two models are employed: a logistic regression for the contact propensities and a multilevel model for the participation propensities. The interviewers form the first level of the model and the households the second level. The multilevel model is used to separate individual participation propensities and interviewer participation propensities. By separating interviewer and individual, the interviewer effect can be isolated and interviewer assignment can be optimized and evaluated in a simulation study.

The logistic regression model for contact does not contain the interviewer as the interviewer is not responsible for the timing of call attempts.

For the interviewer effect it was first investigated whether it was sufficient to use a fixed slope multilevel model, i.e. the interviewer is added as a main effect only and there are no interactions with auxiliary variables. All covariates are selected for the multilevel model, but interactions with the interviewer were not significant at the 5% level.

The resulting model is

$$\log(\frac{\lambda_i}{1-\lambda_i}) = \alpha_0 + \alpha_X X_i \tag{16}$$

$$\log(\frac{\kappa_{ik}}{1-\kappa_{ik}}) = \beta_0 + \beta_X X_i + I_k \tag{17}$$

where λ_i is the contact propensity of household i, κ_{ik} is the participation propensity of household i given interviewer k was assigned to the household, X_i is the covariate vector containing the six auxiliary variables, α_0 and β_0 are constant terms or intercepts and α_X and β_X are the slope parameters. The response propensity now follows as $\rho(I_k, X_i) = \lambda_i \kappa_{ik}$.

Models (16) and (17) were fitted to the SCS data set. Next, the estimated interviewer effect I_k was used to smooth the participation propensities. This was done using the following steps:

1. Remove the interviewer component I_k from the estimated propensities.
2. Sort the remaining interviewer-corrected components in ascending order, from persistent nonrespondents to persistent respondents.
3. Sort the interviewer components I_k in descending order, from the very good interviewers to the least good interviewers
4. Match the interviewers to the households, i.e. match good interviewers to persistent nonrespondents and vice versa.
5. Add both components to form new response propensities.

The fifth step assures that interviewers have the same workload when the sample units are artificially re-assigned, and that the response rate must be unchanged. As a consequence, costs are also the same after re-assignment. However, since strong interviewers are assigned to difficult addresses and weak interviewers to easy addresses, the individual interviewer response rates change and differences between interviewers are levelled out.

Table 4.1 contains the response rate and maximal bias before and after re-assignment of interviewers. The re-assignment leads to a smaller maximal bias and, hence, higher perceived quality for the same costs. The predicted response rate after re-assignment is slightly higher than in the SCS. The slight increase is probably caused by small interactions between the interviewer and the covariates.

Table 4.1: The response rate and maximal bias of the SCS and the SCS after re-assignment of interviewer.

	$\overline{\rho}_X$	$Q(p)$
SCS	67%	0.14
Re-assignment	68%	0.07

5. A pilot study

In this section we summarize the results of a fieldwork pilot into an adaptive survey design for the Survey on Consumer Satisfaction (SCS) conducted in 2009. For a detailed account of the design and analysis of the pilot we refer to Luiten and Wetzels (2009) and Fosen et al (2010).

SCS fieldwork is conducted in the first ten workdays of each month. Because the SCS is conducted monthly, a wealth of information is available about contact and cooperation characteristics of former sample units. This accumulated knowledge was the first of the sources used to develop the adaptive design for the experimental group.

The pilot was conducted in two consecutive months, alongside the regular SCS, during the same 10 day fieldwork period, with a similar sampling method, a similar sample size and the same interviewers. The SCS served as a control group.

It was the objective of the pilot to maximize the R-indicator while restricting the response rate to be at least as high as for the regular design and the costs to be smaller or equal to the regular budget.

The design features that were allowed to be adapted to the sample units were the survey mode, the contact protocol and the interviewer assignment. Apart from telephone also web and paper were added as candidate survey modes. Web and paper questionnaires were available from a different pilot study. Web and paper were optional as the first mode in a sequential mixed-mode design. In all cases, telephone was used, either as the first mode or as the follow-up mode. The contact protocol allowed for different timing and number of calls, and allowed for prioritizing cases. Based on their SCS work in 2008 and the first half of 2009, interviewers were classified in three categories, according to the cooperation rates achieved. A top quartile of the best interviewers (mean cooperation rate in 2008-2009: 82%), a middle group of the second and third quartile (cooperation rate 74%) and a third group in the lowest quartile (66%).

The representativeness of the response was assessed relative to demographic and socio-economic auxiliary variables: ethnic group, gender, age, type of household, degree of urbanization of area of residence, percentage of nonwestern nonnatives at postal code area and monthly income.

In the adaptive survey design two data collection phases were distinguished; makingcontact and obtaining participation. It was anticipated that the two phases involve different processes and have a different impact on the overall R-indicator. Evidently, the contact protocol mostly affects the first phase of getting contact while the interviewer assignment affects only the second phase of obtaining participation. Consequently, the various subpopulations induced by the selected auxiliary variables were potentially treated differently in both phases.

Tables 5.1 shows the resulting contact, cooperation and response rates for the pilot and the regular SCS. Table 5.2 contains the R-indicators for both samples. The cooperation rates and R-indicators are relative to contact. The pilot was successful in maintaining response rates at traditional values while improving the representativeness of the response with respect to the selected auxiliary variables. The costs of the pilot sample were lower than for the regular SCS.

Table 5.1: Contact, cooperation and response rates.

	Pilot	*SCS*
Contacted	93.9%	93.5%
Cooperation	67.9%	67.2%
Response	63.8%	62.8%

Table 5.2: R-indicators for contact, cooperation and response[2].

	Pilot	*SCS*
Contacted	0.87	0.83
Cooperation	0.89	0.87
Response	0.85	0.77

6. Discussion

This paper describes survey designs in which different population units receive different treatment or survey strategies. Differences between population units are reflected by covariates from either linked data from registrations or paradata. Survey strategies are defined as different specifications of survey design features. Such designs are termed adaptive survey designs as they adapt or tailor data collection to the population of interest. Adaptive survey designs resemble responsive survey designs; the main difference regards the presence of an initial design phase in

[2] R-indicators have a precision that depends on the sample size. In table 5.2 the difference in the R-indicator for response is significant at the 5% level. The differences for contact and cooperation are not significant at the 5% level.

responsive designs that is used to learn how effective and costly various design features are.

Basic ingredients of adaptive survey designs are survey strategies, population covariates, response propensities, cost functions and strategy allocation probabilities. Adaptive survey designs attempt to optimize response quality by assigning different strategies to different population units. The strategy allocation probabilities represent the decision variables in the optimization.

Response quality and survey costs are known to be complex and hard to quantify. However, they determine the optimization problem. A number of quality functions is presented that have come up in the recent literature. Furthermore, the optimization depends strongly on the quality of anticipated response propensities and costs. Both must either be estimated in an initial phase of data collection like is done in responsive designs or based on historic data on the same survey or similar surveys. Given the complexity of quality – costs trade-offs and the robustness of estimates of response and costs, it is imperative that adaptive survey designs are kept simple. In other words the set of candidate survey strategies and the vector of covariates should be manageable.

Adaptive and responsive survey designs are beginning to emerge in survey practice. Although tailoring of design features is done for several years, it has rarely been done in a systematic, mathematical way by optimizing quality given costs or vice versa. The terms adaptive and responsive designs have been suggested only very recently. As a result research into such designs is still young.

There are still a lot of open questions and the field of adaptive survey designs is in its infancy. Future research into these designs should address and investigate the role of quality functions, cost functions and the estimation of response propensities based on either historical survey data or randomized contrasts measured in initial design phases or pilot studies.

Acknowledgements: The authors thank James Wagner, Mick Couper, Fannie Cobben and Mariëtte Vosmer for their useful comments on the draft version of this paper.

References

Andridge, R., Little, R. (2010), Proxy pattern-mixture analysis for survey nonresponse, Paper presented at 21st Internation workshop on Household Survey Nonresponse, August 30 - September 2, Nürnberg, Germany.

Barón, J.D., Breunig, R.V., Cobb-Clark, D., Gørgens, T., Sartbayeva, A. (2009), Does the effect of incentive payments on survey response rates differ by income support history, Journal of Official Statistics, 25, 483 – 507.

Cobben, F. (2009), Nonresponse in sample surveys. Methods for analysis and adjustment, Ph.D. Thesis, University of Amsterdam.

De Leeuw, E.D. (2008), Choosing the Method of Data Collection. In: De Leeuw, E.D., Hox, J.J. & Dillman, D.A. (eds.), *International Handbook of Survey Methodology*, Larence Erlbaum Associates, New York, USA, pp. 113-135.

De Leeuw, E., Callegaro, M, Hox, J., Korendijk, E., & Lensvelt-Mulders, G. (2007), The influence of advance letters on response in telephone surveys, Public Opinion Quarterly, 71, 413-443.

Dillman, D. A. (2007), Mail and internet surveys: the tailored design method, Wiley, Hoboken, New Jersey, USA.

Fosen, J., Kleven, Ø, Lagerstrøm, B., Luiten, A., Wetzels, W. (2010), Indicators and data collection control. Reports from the pilots in Statistics Norway and Statistics Netherlands, RISQ deliverable 7, www.risq-project.eu

Greenberg, B.S, Stokes, S.L. (1990), Developing an optimal call scheduling strategy for a telephone survey, Journal of Official Statistics, 6, 421 – 435.

Groves, R.M. (1989), Survey errors and survey costs, Wiley, New York, USA.

Groves, R.M. and Couper, M.P. (1998), Nonresponse in Household Interview Surveys, Wiley series in probability and statistics, Survey methodology section

Groves, R.M., Dillman, D., Eltinge, J., Little, R. (2002), Survey Nonresponse, New York: Wiley Series in Probability and Statistics

Groves, R.M., Heeringa, S.G. (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.

Heyd, J.M., Carlin, B.P. (1999), Adaptive design improvements in the continual reassessment method for Phase I studies, Statistics in Medicine, 18, 1307 – 1321.

Kalsbeek, W.D., Botman, S.L., Massey, J.T., Liu, P.W. (1994), Cost-efficiency and the number of allowable call attempts in the National Health Interview Survey, Journal of Official Statistics, 10, 133 – 152.

Kersten, H.M.P., and Bethlehem, J.G. (1984), Exploring and reducing the non-response bias by asking the basic question, BPA 627-84-M1, CBS, Voorburg

Kulka, R.A., Weeks, M.F. (1988), Towards the development of optimal calling protocols for telephone surveys: a conditional probabilities approach, Journal of Official Statistics, 4, 319 – 322.

Laflamme, F., Karaganis, M. (2010), Implementation of responsive collection design for CATI surveys at Statistics Canada, Paper presented at Q2010, 3 – 6 May, Helsinki, Finland.

Luiten, A., Wetzels, W. (2009), Indicators and data collection control. Work plan and preliminary findings pilot Statistics Netherlands, RISQ deliverable 8.2, www.risq-project.eu.

Lyberg, L.E., Biemer, P., Collins, M., Leeuw, E.D. de, Dippo, C., Schwarz, N., Trewin, D. (1997), Survey measurement and process quality, Wiley Series in Probability and Statistics, Wiley, Hoboken, New Jersye, USA.

Lynn, P. (2003), PEDAKSI: Methodology for Collecting Data about Survey Non-Respondents, Quality and Quantity, Vol. 37, No. 3, pp. 239 – 261.

Mohl, C., Laflamme, F (2007), Research and responsive design options for survey data collection at Statistics Canada, Proceedings of ASA Joint Statistical Meeting, Section 293, July 29 – August 2, Salt Lake City, USA.

Moore, J. (1988), Self/Proxy response status and survey response quality. A review of literature, Journal of Official Statistics, 4, 2, 155 – 172.

Murphy, S.A. (2003), Optimal dynamic treatment regimes, Journal of the Royal Statistical Society: Series B, 65, 331 – 355.

Peytchev, A., Riley, S., Rosen, J., Murphy, J., & Lindblad, M. (2010), Reduction of Nonresponse Bias in Surveys through Case Prioritization, Survey Research Methods, 4(1), 21–29.

Rao, R.S., Glickman, M.E., Glynn, R.J. (2008), Stopping rules for surveys with multiple waves of nonrespondent follow up, Statistics in Medicine, 27, 2196 – 2213.

Särndal, C.E. and Lundström, S. (2008), Assessing auxiliary vectors for control of nonresponse bias in the calibration estimator, Journal of Official Statistics, 24, 167-191.

Schouten, B., Cobben, F., Bethlehem, J. (2009), Indicators for the representativeness of survey response, Survey Methodology, 35 (1), 101 – 113.

Schouten, B., Luiten, A., Loosveldt, G., Beullens, K., Kleven, Ø. (2010), Monitoring and changing data collection through R-indicators and partial R-indicators, RISQ Working paper, available at www.risq-project.eu.

Tabuchi, T., Laflamme, F., Philips, O., Karaganis, M., Villeneuve, A. (2009), Responsive design for the survey of labour and income dynamics, Proceedings of the Statistics Canada Symposium 2009, Longitudinal Surveys: from design to analysis, Statistics Canada.

Wagner, J. (2008), Adaptive survey design to reduce nonresponse bias, PhD thesis, University of Michigan, USA.

Wagner, J., Raghunathan, T. (2007), Bayesian approaches to sequential selection of survey design protocols, Proceedings of ASA Joint Statistical Meeting, Section 501, July 29 – August 2, Salt Lake City, USA.