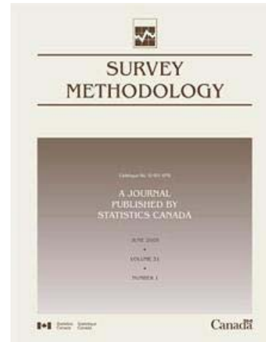
Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

Optimizing quality of response through adaptive survey designs



by Barry Schouten, Melania Calinescu and Annemieke Luiten

June 2013

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at infostats@statcan.gc.ca,

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

• Statistical Information Service	1-800-263-1136
• National telecommunications device for the hearing impaired	1-800-363-7629
• Fax line	1-877-287-4369

Depository Services Program

• Inquiries line	1-800-635-7943
• Fax line	1-800-565-7757

To access this product

This product, Catalogue no. 12-001-X, is available free in electronic format. To obtain a single issue, visit our website, www.statcan.gc.ca, and browse by “Key resource” > “Publications.”

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under “About us” > “The agency” > “Providing services to Canadians.”

Published by authority of the Minister responsible for Statistics Canada

Cette publication est aussi disponible en français.

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard symbols

The following symbols are used in Statistics Canada publications:

.	not available for any reference period
..	not available for a specific reference period
...	not applicable
0	true zero or a value rounded to zero
0^s	value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{p}	preliminary
^{r}	revised
x	suppressed to meet the confidentiality requirements of the *Statistics Act*
^{E}	use with caution
F	too unreliable to be published
*	significantly different from reference category ($p < 0.05$)

Survey Methodology, June 2013
Vol. 39, No. 1, pp. 29-58
Statistics Canada, Catalogue No. 12-001-X

Optimizing quality of response through adaptive survey designs

Barry Schouten, Melania Calinescu and Annemieke Luiten[1]

Abstract

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 quality given constraints on costs. Such designs are called adaptive survey designs. The basic ingredients of such designs are introduced, discussed and illustrated with various examples.

Key Words: Survey costs; Survey errors; 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, to reduce nonresponse selectivity or more, generally, to improve quality. Although a general terminology is lacking in the literature, such designs are usually 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, to illustrate their potential to improve efficiency of survey data collection, and to propagate their use in survey practice.

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 are observed during data collection. In other words, allocation of treatments is based on data that are 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

1. Barry Schouten, Statistics Netherlands and University Utrecht, PO Box 24500, 2490HA Den Haag, The Netherlands. E-mail: jg.schouten@cbs.nl; Melania Calinescu, VU University Amsterdam, Department of Mathematics, De Boelelaan 1081, 1081HV Amsterdam. E-mail: melania.calinescu@vu.nl; Annemieke Luiten, Statistics Netherlands, PO Box 4481, 6401CZ Heerlen, The Netherlands. E-mail: a.luiten@cbs.nl.

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), Murphy (2003) and Zajonc (2012).

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 distinction 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), Laflamme and Karaganis (2010), Phillips and Tabuchi (2009) and Peytchev, Riley, Rosen, Murphy and Lindblad (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 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 is that in previous design phases part of the sample has already responded. In this paper, it is assumed that historic data are available, that effective treatments are identified beforehand and that it is specified what linked data and paradata are going to be used to adapt the design.

What is new in this paper? We make three contributions. First, we set up a general mathematical framework for optimizing response quality given cost constraints. Second, we explicitly allocate different design features to different sample units within this framework. Third, we propose to optimize quality indicators for nonresponse error. The last two contributions are by themselves not completely new. Simple adaptive survey designs are already applied, *e.g.*, in the Dutch Labour Force Survey larger households are not interviewed by web or telephone and proxy reporting is only allowed by a member of the household core. Attempts to optimize survey design accounting for nonresponse error go at least as far back as Hartley and Monroe (1979). And there is a vast literature on optimizing timing and number of contact calls in interviewer surveys, *e.g.*, Kulka and Weeks (1988), Greenberg and Stokes (1990) and Kalsbeek, Botman, Massey and Liu (1994). What is new is the ensemble of all the pieces into a general mathematical framework that abstracts from single design features and that allows to apply general quality indicators. The main motivations for the advance of such a framework are the strong pressure on survey costs and the rise of web as a survey mode. Web has a strong quality-cost differential; it is cheap but has low response rates and has different measurement properties than interviewer modes. As such, web challenges the trade off between quality and costs. Although survey literature has devoted considerable attention to trade-offs in survey designs between the various surveys errors, *e.g.*, Lyberg, Biemer, Collins, de Leeuw, Dippo, Schwarz and Trewin (1997) and Dillman (2007), in survey practice there are still surprisingly few cases where differential design features are investigated and implemented. With this paper, we hope to provide a steppingstone for future research and discussion into adaptive survey designs.

In Section 2, we describe theory and concepts behind adaptive survey designs. In Section 3, we present an example based on virtual survey data, and, in Section 4 we discuss a simulation study based on real survey data. Finally, in Section 5 we end with 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 recruited members of a panel. Each unit will be assigned a strategy s from the set of candidate strategies $S = \{\varphi, s_1, s_2, \ldots, s_M\}$. 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 such as the type of contact mode and the type of survey mode, but the features may also describe different aspects 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).

In the literature many design features are suggested and evaluated, *e.g.*, Groves and Couper (1998) and Groves, Dillman, Eltinge and Little (2002). We refer to De Leeuw (2008) for a discussion of survey modes, to Dillman (2007) and De Leeuw, Callegaro, Hox, Korendijk and

Lensvelt-Mulders (2007) for an elaboration of advance letters and reminders, to Wagner (2008) for a discussion of contact protocol, to Barón, Breunig, Cobb-Clark, Gørgens and Sartbayeva (2009) for a review of incentives, to Kersten and Bethlehem (1984), Cobben (2009) and Lynn (2003) for research into condensed questionnaires, to Moore (1988) for a discussion of proxy reporting, and to Cobben (2009) for an example of interviewer assignment.

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, Luiten, Loosveldt, Beullens and Kleven (2010) and Schouten, Shlomo and Skinner (2011) for guidelines and examples on how to construct strategy sets.

With each population unit k a vector of covariates $X_k = (X_{k1}, X_{k2}, \ldots, X_{kp})^T$ 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 $\tilde{X}_k = (\tilde{X}_{k1}, \tilde{X}_{k2}, \ldots, \tilde{X}_{kq})^T$ may exist for unit k that reflects characteristics observed during data collection for 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 $\tilde{X}_k$ contains observations about the sampled person or household. Examples of X_k are gender, age, type of household or educational level. Examples of $\tilde{X}_k$ 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. $\tilde{X}_k$ 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 such as the interviewer that was assigned to the address.

The important distinction between X_k and $\tilde{X}_k$ is the level of availability. $\tilde{X}_k$ 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 $q(\tilde{x}, x)$ the joint distribution of X_k and $\tilde{X}_k$ in the sample. Furthermore, $q(\tilde{x} \mid x)$ denotes the conditional sample distribution. It is assumed that $q(x)$ and $q(\tilde{x}, x)$ 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 based on population characteristics available in registry and frame data 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 assigned to individual units because the choice of strategy depends on data that are 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 2.4 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 probability* 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 $\tilde{x}$ are observed. The following must hold

$$0 \leq p(s \mid x) \leq 1, \quad 0 \leq p(s \mid x, \tilde{x}) \leq 1 \tag{2.1}$$

$$\sum_s p(s \mid x) = 1, \quad \sum_s p(s \mid x, \tilde{x}) = 1, \tag{2.2}$$

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.*, $p = \{p(s_j \mid x, \tilde{x})\}_{1 \leq j \leq M, x, \tilde{x}}$ 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

$$\rho_X(x) = \sum_{s \in S} \sum_{\tilde{x}} q(\tilde{x} \mid x) p(s \mid x, \tilde{x}) \rho(s, x, \tilde{x}). \tag{2.3}$$

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 problem can now be formulated as

$$\max_p Q(p) \text{ given that } C(p) \leq C_{\max} \tag{2.4}$$

or as

$$\min_p C(p) \text{ given that } Q(p) \geq Q_{\min}, \tag{2.5}$$

where $C_{\max}$ represents the budget for a survey and $Q_{\min}$ minimum quality constraints. Problems (2.4) and (2.5) are called dual optimization problems, although the solutions to both problems may be different depending on the quality and cost constraints.

It is important to stress that the optimization of quality or costs is done only once, before survey data collection starts, and is not repeated during data collection. Hence, it is the strategy that is adapted to the population unit, and in case of a dynamic design to paradata about that unit, but it is not the optimization itself that is adapted. The optimization is based on historic survey data that includes the paradata that has become available in a survey. The joint density function $q(x, \tilde{x}),$ the response probabilities $\rho(x, \tilde{x}, s)$ and the cost function $c(x, \tilde{x}, s)$ are all estimated from historic survey data and are assumed to be given. Since in practice paradata becomes available only during data collection, the candidate strategies for units in the same stratum x are the same up to the moment the paradata $\tilde{x}$ becomes available. For instance, there may be the following four strategies: 1) two telephone call attempts and no follow-up, 2) two telephone call attempts and a follow-up with incentives, 3) three telephone call attempts and no follow-up, and 4) three telephone call attempts and a follow-up with incentives. The decision to make two or three call attempts is based on $x,$ while the follow up is decided upon using a telephone paradata observation $\tilde{x}.$ Thus, beforehand, it is estimated how many units will fall in stratum $(x, \tilde{x})$ and how many will receive a follow up, but only when $\tilde{x}$ is measured, the full strategy is known for individual units.

2.2 Quality objective functions

Adaptive survey designs, as discussed by the literature, typically focus on nonresponse error. In this section, we start with a general classification of quality functions, and then move to quality functions for nonresponse error. In general, a focus on nonresponse error is too narrow a view, especially, when the survey mode is one of the candidate design features in the adaptive survey design. Here, we do, however, not explicitly discuss other survey errors, but we return to this issue in the discussion. We refer to Calinescu, Schouten and Bhulai (2012) for an extension of adaptive survey designs to measurement error and Beaumont and Haziza (2011) for a discussion on adaptive survey designs and nonresponse variance.

2.2.1 Covariate-based and item-based quality functions

When quality is optimized according to (2.4), 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 (2.5), then quality may be multi-dimensional (but cost functions should be one-dimensional).

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 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. Covariate-based quality functions can only be related to the nonresponse bias of the covariates that are included.

2.2.2 Optimizing quality of response

We, first, describe briefly a number of quality functions that have appeared in recent literature. Next, we discuss the choice of a quality function.

The most well-known covariate-based quality function for nonresponse 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. The response rate is represented as the mean response propensity

Response rate: $$Q(p) = \bar{\rho} = \sum\nolimits_{x,\tilde{x},s} q(x,\tilde{x}) p(s \mid x,\tilde{x}) \rho(x,\tilde{x},s) \qquad (2.6)$$

Schouten, Cobben and Bethlehem (2009) propose two covariate-based quality functions, the R-indicator and a measure they call the maximal or worst-case nonresponse bias. The label of the second indicator is misleading as it is only an estimator of the maximal bias of the unadjusted mean of respondents, not the true maximal bias. A better label is the coefficient of variation of response propensities, which we will use here. The measures can be written as

R-indicator: $$Q(p) = R(\rho_Z) = 1 - 2S(\rho_Z) \qquad (2.7)$$

Coefficient of variation: $$Q(p) = CV(\rho_Z) = \frac{S(\rho_Z)}{\bar{\rho}}, \qquad (2.8)$$

where the representativeness may be evaluated with respect to linked data only, $Z = X$, or with respect to a vector containing both linked data and paradata, $Z = (X, \tilde{X})^T$. The standard deviations of the response propensities, $S(\rho_X)$ and $S(\rho_{X,\tilde{X}})$, can be written in terms of the strategy allocation probabilities as

$$S(\rho_X) = \sqrt{\sum\nolimits_x q(x) \left(\sum\nolimits_{\tilde{x},s} q(\tilde{x} \middle| x) p(s \middle| x,\tilde{x}) \hat{\rho}(x,\tilde{x},s) - \bar{\rho} \right)^2} \qquad (2.9a)$$

$$S(\rho_{X,\tilde{X}}) = \sqrt{\sum\nolimits_{x,\tilde{x}} q(x,\tilde{x}) \left(\sum\nolimits_s p(s \middle| x,\tilde{x}) \hat{\rho}(x,\tilde{x},s) - \bar{\rho} \right)^2}. \qquad (2.9b)$$

Särndal and Lundström (2010) and Särndal (2011a and b) propose indicators that are very similar in definition and nature to (2.7) and (2.8). These indicators were derived from the perspective of calibration and so-called balanced response, and may be used as alternatives to (2.7) or (2.8).

An example of an item-based quality function for nonresponse is presented by Groves and Heeringa (2006). For a specific target variable Y, Groves and Heeringa (2006) suggest the nonresponse bias of the unadjusted mean of respondents

Estimated nonresponse bias: $$Q(p) = \frac{\operatorname{cov}(Y, \rho_X)}{\bar{\rho}}, \qquad (2.10)$$

with $\operatorname{cov}(Y, \rho_X)$ the response covariance between the target variable and the response propensities given covariates X. It can be written as

$$\operatorname{cov}(Y, \rho_X) = \frac{\sum_x q(x)\rho_X(x)(\rho_X(x) - \bar{\rho}_X)(y(x) - \bar{y}_R)}{\bar{\rho}}, \qquad (2.11)$$

with $\rho_X(x)$ as in (2.3), $y(x)$ the mean value of Y for $X = x$ and $\bar{y}_R$ the expected mean of respondents. Again, (2.11) can be extended to include paradata $\tilde{X}$.

All quality functions in this section are defined as population parameters. In practice, they need to be estimated from survey data. The true $\rho_X(x)$ need to be replaced by estimators $\hat{\rho}_X(x)$, based on some form of regression, and the summations over the population will be replaced by design weighted summations over the sample. We return to the estimation of propensities in Section 2.4.

Now, how to choose a quality function? All quality functions mentioned here attempt to measure the impact of nonresponse beyond that of a mere reduction in sample size. They do this based on covariates from linked data and paradata. The rationale behind optimizing these quality functions is that stronger traces of nonresponse error on these covariates may imply larger nonresponse error on other variables as well; the quality functions are viewed as process quality indicators rather than product quality indicators. Although appealing, this conjecture clearly needs empirical support. The choice of a quality function for nonresponse should be based on the set of key survey variables, the population parameters of interest and the estimators that are going to be employed. The response rate and R-indicator do not aim at a specific population parameter or estimator. The coefficient of variation focuses on population means, but it is not specific to any survey variable or nonresponse adjustment, while the estimated nonresponse bias originates from the same perspective, but it is applied to a single survey variable. If a survey carries multiple key survey variables, then an item-based quality function for nonresponse is to be avoided, as it may lead to conflicting optimization problems. If a survey has a single key variable, then it is effective to either use an item-based quality function or to restrict to the most relevant covariates only in covariate-based quality functions. If a survey has multiple uses, then, in our view, it is too restrictive to focus on a specific population parameter and estimator, and we favor the R-indicator to any other quality function. If it can be expected that users will focus on population means or totals, then the coefficient of variation is to be preferred, in our opinion, as it does not assume a specific adjustment method.

However, even more important than the choice of the indicator is the set of linked data and paradata that are input to the indicator. If a survey has one or only a few key variables, then the selected linked data and paradata can and should relate to those variables. If, however, a survey has a wide range of survey variables, then one must restrict necessarily to auxiliary variables that generally distinguish persons or households.

So far, we have ignored the impact of nonresponse on precision, while requirements for the precision of the survey estimates may be given explicitly. There are two options to include precision in the optimization. First, one may add an additional constraint. The straightforward choice would be a constraint on the minimum number of respondents, possibly for a number of population subgroups; thereby avoiding to specify the population parameters and estimators. Second, the nonresponse variance may be included in the quality function itself, *i.e.*, one would consider indicators for the mean square error. This option is proposed and elaborated by

Beaumont and Haziza (2011). Under the second option, one again has to consider the set of key survey variables, the population parameter and the estimation strategy, as precision is specific to an estimator for some population parameter for a single survey variable.

2.3 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 traveling costs play a dominant role.

If a large number of design features is optional, then the cost functions have complex forms with many overhead 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.

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. The cost function $C(p)$ is the sum of two components

$$C(p) = C_F + C_V(p), \tag{2.12}$$

of which only the second, the variable component, depends on the allocation probabilities.

In general, with a survey strategy $s,$ costs $c(\tilde{x}, x, s)$ are associated with population units from group $(\tilde{x}, x).$ 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 $c(\tilde{x}, x, s)$ is a relative cost function as it describes only the contribution of the strategy to the variable cost component $C_V(p)$

$$C_V(p) = \sum_{\tilde{x},x,s} q(\tilde{x}, x)p(s \mid x, \tilde{x})c(\tilde{x}, x, s). \tag{2.13}$$

Three remarks are in order. First, the derivation of fixed and variable cost components is complicated when a survey organization runs many surveys in parallel. On the one hand, the interaction between surveys makes it hard to separate costs per survey, especially when strategies are tailored. 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. As a consequence, the cost function becomes a vector and the constraint a vector of constraints. 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. It is important to monitor data collection closely and to build in indicators for strategies.

2.4 Estimating response probabilities

Next to cost parameters and quality 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, as 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.

Literature on household surveys gives an extensive list of models for response that include design features. 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, and may even be biased themselves when they are based on similar, but different surveys. In the optimization, this uncertainty can be accounted for by allowing response propensities to be random variables rather than fixed quantities. The randomness demands for sensitivity analyses and evaluations of the robustness of the optimization that provide insight into the variation of quality and costs when the survey is conducted multiple times (under the same circumstances).

2.5 The optimization problem

One may take two approaches to the optimization of (2.3) and (2.4): a trial-and-error approach or a mathematical optimization. In this paper, we concentrate on a mathematical framework and optimization, but one may be more modest and introduce adaptive survey designs gradually through pilot studies and field tests.

Quality functions (2.6), (2.7), (2.8) and (2.10) all are functions of the strategy allocation probabilities p. The response rate is a linear function of the allocation probabilities, which makes it relatively easy to optimize using standard optimization software (*e.g.*, the *linprog* package in R or any other software that can address linear programming problems). Still, as far as we know, due to the high dimensionality of p there is no closed form solution to (2.4) even for linear problems. In general, however, the quality functions are nonlinear, nonconvex functions with respect to the allocation probabilities, and cannot be optimized without numerical or Monte Carlo methods. The complexity of the problem grows quickly as a function of the number of candidate strategies and the number of subgroups based on linked data and paradata.

Current statistical softwares contain procedures or packages that can handle nonlinear optimization problems, like *nlm* or *nlminb* in R or *proc optmodel* in SAS. However, nonlinear nonconvex problems may require long computational times or may converge to local optima. For this reason, specialized optimization softwares such as Xpress, Baron or AMPL are recommended.

In the examples of Section 3 and 4, we perform a number of optimizations. The optimization problem of Section 3 is relatively simple; the quality objective function is the R-indicator which is evaluated against two population subgroups. For two subgroups the optimization can be rewritten as a linear programming problem. For the example of Section 4, we were able to construct an algorithm that converges to the optimal solution in a small number of steps. All optimizations were programmed in R and the code is available upon request.

3 A dynamic adaptive survey design: Re-assigning interviewers in a follow-up survey

In this section, we provide an example of a dynamic adaptive design: the re-assignment of interviewers based on observations of the propensity to cooperate. The example is based on hypothetical response propensities and cost functions. 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 it is not available for sample units who were not contacted during the first phase. It provides a judgement on the propensity that the sample unit 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 sample units are divided into respondents, refusers and noncontacts. 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 (2.7) as the quality objective function. We split the sample using $X = (\text{age})$ 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 maximized.

Let n be the sample size of the survey. The population proportions of the two subpopulations, *young* and *old*, are denoted by $q(1)$ and $q(2)$. We let $q(\tilde{x} \mid x)$ be the conditional probability that a sample unit from age subpopulation x is of type $\tilde{x}$, where $\tilde{x} \in \{\text{easy, medium, difficult}\}$. Furthermore, let $\lambda(x, \tilde{x})$ be the probability that a sample unit of type $\tilde{x}$ from age subpopulation x is a refusal. If a person is not a refuser, then $\mu(x, \tilde{x})$ 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 $p_s M$ represents the number of interviewers with skill $s \in S = \{\text{good, less good}\}$, $0 \leq p_s \leq 1$ and $p_{\text{good}} + p_{\text{less good}} = 1$. The set S forms the set of strategies, *i.e.*, we want to assign each refuser to either a good or a less good interviewer. We assume that each interviewer can handle at most c refusal cases in the second phase of the survey. The probability that a refusal of type $\tilde{x}$ from subpopulation x will respond if contacted by an interviewer of skill s is denoted by $\rho(s, x, \tilde{x})$ and it is again assumed to be known from previous surveys.

Let $\{p(s \mid x, \tilde{x})\}_{x, \tilde{x}}$ be the set of decision variables, where $p(s \mid x, \tilde{x})$ represents the probability that a sample unit of type $\tilde{x}$ will be assigned to an interviewer of skill s given that he/she belongs to subpopulation x. In other words, we allow for a random assignment of sample units to the two interviewer groups.

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

$$n\sum_{x,\tilde{x}} q(x)q(\tilde{x}\,|\,x)\,p(s\,|\,x,\tilde{x})\,\lambda(x,\tilde{x}) \le Mp_s c, \quad \forall s \in S.$$

In other words, the total number of refusers that can be assigned to interviewers of skill s is restrained to the maximum possible workload for that skill group.

The response propensity for a unit from subpopulation x can now be derived as

$$\sum_{\tilde{x}} q(\tilde{x}\,|\,x)\Big[(1-\lambda(x,\tilde{x}))\mu(x,\tilde{x}) + \lambda(x,\tilde{x})\sum\nolimits_s p(s\,|\,x,\tilde{x})\rho(s,x,\tilde{x})\Big],$$

and form the input to the R-indicator.

Now, consider the following input data for the example: a sample size of $n = 2{,}000$, a total of 80 interviewers, $M = 80$, a maximal workload of 30 cases per interviewer, $c = 30$, an age distribution equal to $q(1) = q(2) = 0.5$, conditional distributions of refusal type $q(\tilde{x}\,|\,1) = (0.2, 0.3, 0.5)'$ and $q(\tilde{x}\,|\,2) = (1/3, 1/3, 1/3)'$ and 25% of the interviewers are classified as good, $p_1 = 0.25 = 1 - p_2$.

Tables 3.1 and 3.2 give the hypothetical response probabilities $\rho(s,x,\tilde{x})$ for the two subgroups when refusal conversion is applied, as well as the cooperation probabilities $\mu(x,\tilde{x})$ and refusal probabilities $\lambda(x,\tilde{x})$.

We optimize the R-indicator with respect to the two age groups. For two strata, it can be shown that the R-indicator is maximal when the absolute distance between the two strata response propensities is minimal. The optimal value of the R-indicator turns out to be 0.827. Table 3.3 shows the optimal values of the decision variables; all but one of the decision variables $p(s\,|\,x,\tilde{x})$ are either 0 or 1, *i.e.*, the re-assignments are mostly non-probabilistic. The exception is the subpopulation of young persons with medium response propensity assessment.

Table 3.1
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**
$\rho(s, 1, \tilde{x})$	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**
$\rho(s, 2, \tilde{x})$	0.9	0.7	0.5	0.8	0.6	0.4

Table 3.2
Refusal and cooperation probabilities in the first phase of data collection

	Young			Old		
	Easy	**Medium**	**Difficult**	**Easy**	**Medium**	**Difficult**
$\lambda(x, \tilde{x})$	0.5	0.6	0.7	0.2	0.3	0.4
$\mu(x, \tilde{x})$	0.85	0.8	0.76	0.95	0.93	0.91

Table 3.3
Optimal assignment of cases to interviewers

	Young			Old		
	Easy	**Medium**	**Difficult**	**Easy**	**Medium**	**Difficult**
Good	1	0.83	1	0	0	0
Less good	0	0.17	0	1	1	1

It is useful to compare the optimal allocation to a random allocation of interviewers in order to see how much is gained. If we would randomly assign the refusals to the interviewers, then the value of the R-indicator equals 0.749. The optimal assignment, thus, leads to a considerable increase in the R-indicator. The response rates are, respectively, 72.0% and 70.1% for the optimal and the random assignment.

If we increase the number of interviewers, while fixing the maximal number of cases per interviewer as well as the other parameters, then for any interviewer number higher than $M = 84$ the R-indicator does not improve. Both interviewer groups are sufficiently big to handle the entire sample and the cost constraint is no real constraint anymore. The R-indicator for $M = 84$ is equal to 0.830 and the response rate is 72.1%. If we would maximize the response rate rather than the R-indicator, then the allocation of interviewers will converge towards assigning only *good* interviewers to all cases.

4 A static adaptive survey design: Assigning telephone interviewers

In this section, a simulation study is presented where telephone interviewer assignment is the design feature of interest. The response probabilities used in the example are estimated from real telephone 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 1,500 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, with respect to the interviewer the data are randomized (or interpenetrated). In the following, the interviewer will be the design feature of interest. The survey strategy set S consists of sixty strategies, $S = \{s_1, s_2, \ldots, s_{60}\}$.

From the available auxiliary variables a vector X was selected containing 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. Especially age, average house value and type of household relate to key statistics deduced from the SCS. No paradata were available in this study. Therefore, the adaptive survey design is static. In the optimization the allocation probabilities $p(s_k \mid x)$ need to be chosen, *i.e.*, it needs to be decided to which interviewers subpopulations based on X are assigned $\left(\text{such that } \sum_k p(s_k \mid x) = 1\right)$.

The coefficient of variation of the response propensities ρ_X defined by (2.8) is selected as the target quality function. To estimate the response propensities $\rho(s_k, x)$ for interviewers, a multilevel model is used with the identity link function, *i.e.*, a linear regression with two levels. The interviewers form the first level of the model and the households the second level. The multilevel model is used to separate individual response propensities and interviewer response propensities. The rationale is that by separating interviewer and individual, the interviewer effect can be isolated and interviewer assignment can be optimized. We chose a linear model as it allows for easy optimization. Since the propensities are never close to 0 or 1, the linear model produces almost the same estimates as a logit or probit model.

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 pre-selected covariates gave significant contributions to the multilevel model, but none of the interactions with the interviewer were significant at the 5% level. For this reason, we restrained ourselves to the following main effect model

$$\rho(s_k, x_i) = \beta_0 + \beta_X x_i + \beta_k \tag{4.1}$$

where x_i is the covariate vector of household $i, 1 \leq i \leq n, n$ the sample size, β_k is the (fixed) interviewer effect for interviewer k, β_0 is the constant term or intercept and β_X is the slope parameter. We let $\rho(x_i) = \sum_k p(s_k \mid x_i) \rho(s_k, x_i)$ denote the response propensity of sample unit i.

Model (4.1) was fitted to the SCS data set. Next, the estimated interviewer effect β_k was used to optimize the coefficient of variation, subject to two cost constraints: both the total interview time and the individual number of calls for each interviewer must be the same as in the original design. Since the telephone management system handles the calls, the interview time is the dominant component in the costs. If we fix the total interview time, then we constrain costs to be the same as for the regular SCS. Since interviewers can handle only a certain amount of calls, we must also fix the number of calls they are allocated to. The first constraint implies that we fix the response rate, as the total interview time is the multiple of the average individual interview time and the number of respondents. The SCS questionnaire is simple and does not contain any nested sets of survey items. As a result the individual interview time shows hardly any variation over population subgroups. The second constraint is equal to

$$\sum_i p(s_k \mid x_i) = n_k, \tag{4.2}$$

where n_k is the pre-specified number of calls for interviewer k and $\sum_k n_k = n$.

We optimize the coefficient of variation by distributing the β_k's to the households. Due to the additive nature of the model, it is easy to show that any permutation of the interviewers to the cases leads to the same average response propensity and, hence, to the same interview time and costs. The average response propensity is

$$\bar{\rho} = \frac{1}{n} \sum_i \rho(x_i) = \frac{1}{n} \sum_{i,k} p(s_k \mid x_i)(\beta_0 + \beta_X x_i + \beta_k) = \beta_0 + \frac{1}{n} \sum_i \beta_X x_i + \frac{1}{n} \sum_k n_k \beta_k,$$

which does not depend on the set of allocation probabilities $p(s_k \mid x)$. As a consequence, optimizing the coefficient of variation amounts to optimizing the variance of the response propensities $S^2(\rho_X)$.

If we restrict ourselves to 0-1 decision variables, *i.e.*, $p(s_k \mid x) \in \{0,1\}, \forall x, k,$ then it is relatively easy to show that the optimal allocation corresponds to linking the best interviewers to the most difficult sample units and vice versa. In other words, the sample units are sorted by putting the individual response propensities without the interviewer effect, $\beta_0 + \beta_X x_i,$ in an increasing order, and the interviewers are sorted in a decreasing order based on their interviewer effect, β_k. If two sample units i and j are allocated to two different interviewers, say k and l, and $\beta_X x_i < \beta_X x_j, \beta_k < \beta_l$ and $p(s_k \mid x_i) = p(s_l \mid x_j) = 1,$ then it is optimal to switch the two interviewers, *i.e.*, $p(s_l \mid x_i) = p(s_k \mid x_j) = 1.$ This can be shown as follows. The difference in variance $S^2(\rho_X)$ is proportional to

$$\begin{aligned} \Delta S^2(\rho_X) &= (\beta_0 + \beta_X x_i + \beta_k - \bar{\rho})^2 \\ &\quad +(\beta_0 + \beta_X x_j + \beta_l - \bar{\rho})^2 - (\beta_0 + \beta_X x_i + \beta_l - \bar{\rho})^2 - (\beta_0 + \beta_X x_j + \beta_l - \bar{\rho})^2 \\ &= 2(\beta_l - \beta_k)(\beta_0 + \beta_X x_j - \bar{\rho}) - 2(\beta_l - \beta_k)(\beta_0 + \beta_X x_i - \bar{\rho}) \\ &= 2(\beta_l - \beta_k)(\beta_X x_j - \beta_X x_i) > 0. \end{aligned} \tag{4.3}$$

From (4.3), we can conclude that there is a decrease in variance, and, hence, in the coefficient of variation, if we swap the two interviewers for cases i and j. From this argument, it follows easily that the optimal solution is as suggested. In a similar fashion, but requiring more algebra, it can be shown that the optimal solution for probabilistic allocations, $p(s_k \mid x) \in [0,1],$ is the same.

The first two rows of table 4.1 contain the average response propensity and the coefficient of variation before and after re-assignment of interviewers. The coefficient of variation dropped from 0.117 to 0.035. In order to get an idea of the significance of the change in the quality function, we computed bootstrap standard errors. For each bootstrap, the re-assignment of interviewers was performed. The errors are given in table 4.1.

Table 4.1
The average response propensity and coefficient of variation of the regular SCS, the SCS after re-assignment of interviewers without and with adjustment for interview time. Bootstrap standard errors are given within brackets.

SCS	Adjustment for interview time?	$\bar{\rho}$	$Q(p)$
Regular	-	70.8%	0.117 (0.005)
Re-assignment	No	70.8%	0.035 (0.003)
Re-assignment	Yes	70.8%	0.034 (0.003)

The reader may have noticed that fixed numbers of interviewer cases do not imply fixed numbers of interviews per interviewer. In fact, by rearranging the interviewers, the good interviewers will do fewer interviews as they get the harder cases, while the less good interviewers do more interviews. As a result, the good interviewers will work smaller numbers of hours than they would do in the regular SCS and the less good interviewers will work more. This would be an undesirable side effect, which can, however, be adjusted relatively easy. Starting from the optimal solution, and sorting again the sample units based on their individual response propensities without the interviewer effect, we can shift neighbouring cases from less good interviewers to better interviewers. This is done in such a way that the total interview time per interviewer does not exceed that of the regular SCS. One can again prove that this procedure leads to a new optimal solution where the constraint on the fixed number of cases in (4.2) is replaced by the constraint on the fixed number of interviews

$$\sum_i p(s_k \mid x_i)\,\rho(s_k, x_i) = r_k, \tag{4.4}$$

where r_k is the pre-specified number of interviews. Table 4.1 presents the coefficient of variation for the optimal solution given (4.4). The response rate remains fixed, and the coefficient of variation is marginally smaller.

In 2009, the SCS survey has been used as an instrument to test a static adaptive survey design. We refer to Luiten and Wetzels (2009) and Luiten and Schouten (2013) for details. Interviewer assignment was one of the main design features that were adapted. Other design features were the survey mode and the contact protocol. Apart from telephone, also web was selected as a potential survey mode. Sample units with low estimated contact probabilities were assigned to more intensive contact protocols and were prioritized. Based on historical SCS data, contact and response probabilities were estimated. The pilot succeeded in significantly improving the coefficient of variation, while fixing the response rate and budget.

In this section, we presented a simulation study where good telephone interviewers get more difficult cases. This may in practice lead to annoyance among these interviewers. When implementing such a design, one should carefully instruct interviewers beforehand. In the 2009 SCS pilot, this did not lead to any negative comments from interviewers. In face-to-face surveys, a re-assignment of interviewers cannot be done so easily as travel costs may change drastically. Still, within densely populated inteviewer regions, re-assignment may be an option.

5 Discussion

This paper describes survey designs in which different population units receive different treatments 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. Basic ingredients of adaptive survey designs are survey strategies, population covariates, response propensities, cost and quality 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.

We believe this paper contributes to the literature in three ways: it presents a general framework, it explicitly opts to choose from a set of strategies in making a quality-cost trade off, and it focuses on indicators for nonresponse error. The last two components can be found in the survey literature; it is the generalization to multiple design features and nonresponse error that is new. In its most modest form, adaptive survey designs are a stratified allocation of survey strategies over different population subgroups. In its most ambitious form, adaptive survey designs are extensions of sampling designs to multiple strategies and with a focus on nonresponse error. However, even in its most modest form, adaptive survey designs may include survey modes, incentives, reminders, length of fieldwork in face-to-face surveys, interviewer assignment and type of reporting.

Adaptive survey designs lend themselves best to settings where surveys are run repeatedly for a longer time period. In such settings, the historic information is strongest. The designs also lend themselves to survey institutes that conduct many surveys that are relatively similar in topics and budget. New and one-time only surveys ask for modesty and caution. However, this would also be true for single strategy designs. Adaptive survey designs may account for the lack of strong historic data by allowing for uncertainty in response propensities and other parameters, and by introducing a learning period or initial design phase.

In our view, the focus on nonresponse error is an important part of the framework. In this paper, we aim at representativeness of response. This aim comes from our conviction that nonresponse is always not-missing-at-random. We see larger deviations from missing-completely-at-random mechanisms for relevant auxiliary variables as indications of stronger not-missing-at-random nonresponse on survey variables given these auxiliary variables. Theoretically, this does not have to be true. Consider a simple binary yes-no survey question and 50% nonresponse. The extreme cases arise when all nonrespondents would say either yes or no. They can do so regardless of the choice of auxiliary variables and, hence, the maximal nonresponse bias on this question is the same for whatever choice of auxiliary variables. Hence, research should provide empirical support for the focus on indirect measures for nonresponse error.

Future research into adaptive and responsive designs is also needed for other questions. Research should extend designs to multiple survey errors and should investigate the robustness of designs for misspecification of models for response propensities. Until now, adaptive and responsive survey designs have focused on the nonresponse error and ignored the response or measurement error. It is well known, however, that some survey design features, *e.g.*, survey mode or interviewers, may have a strong impact on the response error and, consequently, on the total survey error. Adaptive survey designs should, therefore, account for measurement error as well, when it can be expected that design features have a strong differential impact on response error. Optimization accounting for multiple errors represents an important area of future research.

Adaptive survey designs should in all cases be modest in the number of strategies employed in order to avoid an overly complex survey process and optimization on propensities and cost functions that are subject to uncertainty. Nevertheless, a structured way of looking is always to be preferred; adaptive designs provide such a framework and accommodate a structured search for enhanced survey designs.

Acknowledgements

The authors thank James Wagner, Mick Couper, Fannie Cobben and Mariëtte Vosmer for their useful comments on the draft version of this paper. The authors also thank the associate editor and referees for their useful remarks which have greatly improved the paper.

References

Barón, J.D., Breunig, R.V., Cobb-Clark, D., Gørgens, T. and 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.

Beaumont, J.-F., and Haziza, D. (2011). A theoretical framework for adaptive collection designs. Paper presented at 5th International Total Survey Error Workshop, June 21-23, Quebec, Canada.

Calinescu, M., Schouten, B. and Bhulai, S. (2012). Adaptive survey designs that minimize nonresponse and measurement risk, Discussion paper, Statistics Netherlands. Available at through www.cbs.nl.

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 *International Handbook of Survey Methodology*, (Eds., E.D. De Leeuw, J.J. Hox and D.A. Dillman), Larence Erlbaum Associates, New York, U.S.A., 113-135.

De Leeuw, E., Callegaro, M., Hox, J., Korendijk, E. and 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, U.S.A.

Greenberg, B.S., and 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, U.S.A.

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. and Little, R. (2002). Survey Nonresponse, New York: Wiley Series in Probability and Statistics.

Groves, R.M., and 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.

Hartley, H.O., and Monroe, H. (1979). Interviewer assignments which minimize the effects of nonsampling errors. *Proceedings of the Survey Research Methods Section*, American Statistical Association, 216-220.

Heyd, J.M., and 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. and 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., and 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., and 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., and Schouten, B. (2013). Adaptive fieldwork design to increase representative household survey response. *Journal of the Royal Statistical Society*, Series A, 176, 1, 169-190.

Luiten, A., and 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., de Leeuw, E.D., Dippo, C., Schwarz, N. and Trewin, D. (1997). Survey measurement and process quality, Wiley Series in Probability and Statistics, Wiley, Hoboken, New Jersey, U.S.A.

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

Mohl, C., and 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, U.S.A.

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. and Lindblad, M. (2010). Reduction of Nonresponse Bias in Surveys through Case Prioritization, Survey Research Methods, 4, 1, 21-29.

Phillips, O., and Tabuchi, T. (2009). Responsive design for the survey of labour and income dynamics. *Proceedings: Symposium 2009, Longitudinal Surveys: from design to analysis*, Statistics Canada.

Särndal, C.-E. (2011a). Three factors to signal non-response bias with applications to categorical auxiliary variables. *International Statistical Review*, 79, 2, 233-254.

Särndal, C.-E. (2011b). The 2010 Morris Hansen Lecture: Dealing with survey nonresponse in data collection, in estimation. *Journal of Official Statistics*, 27, 1, 1-21.

Särndal, C.-E., and Lundström, S. (2010). Design for estimation: Identifying auxiliary vectors to reduce nonresponse bias. *Survey Methodology*, 36, 2, 131-144.

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

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

Schouten, J.G., Shlomo, N. and Skinner, C. (2011). Indicators for monitoring and improving representativeness of response. *Journal of Official Statistics*, 27, 2, 231-253.

Wagner, J. (2008). Adaptive survey design to reduce nonresponse bias, Ph.D. thesis, University of Michigan, U.S.A.

Zajonc, T. (2012). Bayesian inference for dynamic treatment regimes: Mobility, equity and efficiency in student tracking. *Journal of the American Statistical Association*, 107, 80-92.