WORKING PAPER SERIES

WORKING PAPER NO 5, 2006

ESI

On the evaluation of the cost efficiency of nonresponse rate reduction efforts - some general considerations

by

Sara Tångdahl

http://www.oru.se/esi/wps

SE-701 82 Örebro
SWEDEN

ISSN 1403-0586

On the evaluation of the cost efficiency of nonresponse rate reduction efforts - some general considerations

Sara Tångdahl

August 11, 2006

Abstract

Virtually every survey today suffers from nonresponse to some extent. To counter this, survey administrators and researchers have a host of methods at their disposal, many of which are both expensive and time consuming. Reduction efforts, aiming at reducing the nonresponse rate, are an important part of the data collection process, but commonly also a substantial part of the available survey budget. We propose that the efficiency of the reduction efforts be evaluated in relation to the costs. In this paper we point in the direction of an evaluation procedure, using a measure of cost efficiency, that can be used in an "ideal" situation, where all relevant quantities are known. It can not be applied directly in practice, but will serve as a point of reference when practically feasible approaches are developed.

Contents

1 The problem

1.1 Introduction

Virtually every survey today suffers from nonresponse to some extent, and the nonresponse levels seem to be constantly increasing, see for example de Heer and de Leeuw (2002) and references therein. A common view is that the nonresponse rate is also an indicator of the size of the nonresponse error. Obviously, if the nonresponse rate is high, there is greater potential risk of large nonresponse error than if the nonresponse rate is low. However, there is no direct relation between the size of the nonresponse set and the size of the resultant error. This is shown in e.g. Tångdahl (2004). The main concern in the presence of nonresponse is the potential bias. If there was no risk of bias, we could simply draw a larger sample to compensate for the reduced number of observations.

To counter nonresponse, survey administrators and researchers have a host of methods at their disposal, many of which are both expensive and time consuming. One important category of methods is *reduction efforts*, i.e. all efforts that take place after the initiation of the data collection period and that aim at reducing the nonresponse rate. Methods for nonresponse rate reduction differ between data collection modes. In interviewer assisted surveys, by telephone or face-to-face, much of the success of call-backs or other efforts is up to the interviewers' ingenuity and skills. By contrast, mail surveys offer limited possibilities to vary strategies for follow-ups. Also, it is sometimes possible to probe into the reasons for nonresponse in interviewer assisted surveys, whereas these usually remain unknown in mail surveys.

When planning a survey, we must seek to allocate the limited resources so that they achieve the most error reduction, within the available budget. The efficiency of the reduction efforts must be measured by the nonresponse error reduction they result in (if any), but this reduction must also be weighed against the amount of resources that is required to achieve it. If the effect is small compared to the cost, we should either reallocate the budget to other survey activities or conduct the survey at a lower cost.

In this paper, we will discuss an approach to evaluate the reduction efforts in an "ideal" situation where all information required to calculate the expected costs and effect of each reduction effort is available to us. This allows us to formulate a procedure that we, though the conditions will never be met in practice, can use as a point of reference when developing procedures that can actually be applied in practice. It can also be used as a planning tool where the necessary calculations can be performed under reasonable assumptions about the true parameter values. The approach builds on the framework and results on estimator bias and variance given in Tångdahl (2004) and Tångdahl (2005), respectively. The bias and variance are there expressed as functions of the (unknown) response distributions.

The problem is formulated as one of evaluation, not optimization. One of the limitations of an optimization approach is that the optimality is valid only under the assumed model(s) and functions, and that only factors that are modeled explicitly are taken into account. The evaluation approach, on the other hand, is an important tool to guide survey decisions. The issue of whether we could improve the total quality by spending the money saved on reduction on alternative efforts to improve survey quality is not considered. The suggested approach should be used in planning for future surveys, but it is not intended to be used as a means to find the optimal break-off point in an ongoing survey.

In practice, we never have complete access to all quantities and population parameters required in the approach. Instead we have to settle for either of four typical cases. At best, we can have access to reliable estimates of the expected costs and effectiveness of the reduction efforts. In repeated surveys, one may be able to use data from past surveys to state close approximations to the true response distribution and required population quantities. Another, less powerful, possibility is to use available register values as proxies for the study variables. A third option, if resources are available, is to carry out an evaluation study to produce at least unbiased estimates of the population parameter or the corresponding sample quantity.

The least satisfactory situation is the case when we do not even have information as described above, since we in that case are not able to estimate the size of the nonresponse error. All we can do in this situation is to regard the estimates at the current cut-off point for final returns as the "best" estimate possible, the level with which to compare the estimates that would be produced if the data collection period was broken off earlier.

In these situations, it becomes important to reflect and incorporate the parameter uncertainty into the cost efficiency analysis and evaluation. This is discussed briefly in section 5 and will be dealt with in a subsequent paper.

Section 2 of this paper deals with alternative measures of quality and discusses aspects on cost modeling. Possible indicators of the cost efficiency are discussed in section 3, and in section 4 a strategy for evaluating the cost efficiency of the reduction strategy in the ideal case is proposed. Section 5 gives a discussion on how we could deal with the less than ideal case where we must estimate the relevant quantities required in the approach, and some concluding remarks are given.

1.2 General setup and definitions

Consider a finite study population $U = \{1, \ldots, k, \ldots, N\}$ of size N. The variable of interest is y, with the value y_k for the kth element. No assumptions are made about the form of the parameter to be estimated, but for simplicity, we will consider estimation of the total $t_{yU} = \sum_U y_k = \sum_{k \in U} y_k$. The sampling design is assumed to be single stage element sampling, but otherwise arbitrary.

We consider a mail survey where data collection is initiated by an attempted contact with all elements in the sample. This results in a number of response cases and nonresponse cases (e.g. non-contacts, refusals). At (usually) predetermined days, additional contact attempts are made with sample elements who have not yet responded or have not yet been categorized as definite nonresponse. After a sufficiently long time period, also usually at a predetermined time, say A, the data collection is terminated, giving a partitioning of the sample into a response set, $r^{(A)}$, and a nonresponse set, $s - r^{(A)}$. The choice of A is usually governed by the date when estimates must be published.

Remark 1 *In some cases, it is possible to extend the length of the data collection period by initiating it earlier. We will not consider this possibility, as we are mainly interested in the cost and effect of reduction activities that takes place during the period, regardless of its length.*

To evaluate estimator properties under such a data collection procedure, we must make some assumptions about the true, unknown, response mechanism. A response distribution is assumed to exist so that the generation of the response set is regarded as a second phase of random selection, although with unknown "selection" probabilities. This is a "quasi-randomization" approach, a term first introduced by Oh and Scheuren (1983).

The response probabilities are most likely affected by the general survey conditions and the specific survey setup, i.e. the *survey procedure*. In particular, we will consider the nonresponse rate reduction efforts as a source of influence on the response probabilities. Let the survey procedure $\mathfrak{P}^{(a)}$ represent the survey setup if we were to terminate data collection at time a. This specific procedure induces response distribution $RD^{(a)}$ with individual response probabilities $\theta_{k|s}^{(a)}$. We will assume that elements respond independently and that $\theta_{k|s}^{(a)} \geq \theta_{k|s}^{(a-1)}$ for all k. Throughout, quantities relating to time point a will be denoted with a superscript (a).

We thus define a sequence of response distributions, each corresponding to a specific survey procedure $\mathfrak{P}^{(a)}$, a representing possible points of time to terminate the data collection process, $a = 1, \ldots, A$, where A marks the point of time for termination of the data collection, i.e. the cut-off date for final returns, in the current survey setup.

We will use $\hat{t}_{yc}^{(a)}$ to denote an arbitrary estimator based on response set $r^{(a)}$, used to estimate the total t_{yU}, and $\hat{t}_{ys}$ will be the corresponding full response estimator. In standard survey practice, the term strategy is the combination of a sampling design and estimator. Here we introduce the term *survey strategy* to denote the combination of survey procedure, including the sampling design, and estimator. We will denote the survey strategy where the data collection is cut off at time a by $(\mathfrak{P}^{(a)}, \hat{t}_{yc}^{(a)})$.

2 Survey error and survey cost

When discussing survey quality, statisticians tend to talk about survey error, which can be regarded as the inverse of quality. The error in a survey statistic can be defined as the deviation between the survey result and the true value of the parameter one attempts to estimate. Seemingly, it is then a simple matter to choose the survey strategy that yields the smallest error. However, a comparison between different strategies with respect to survey quality must be done relative to the costs involved. The "best" survey strategy (the one with smallest possible total error) is not necessarily the most cost efficient. Neither should we choose the strategy that generates the lowest costs without regard to survey error. Instead, we should seek a survey strategy that is optimal in the sense that it balances error and cost. Tools to achieve this are error and cost modeling, discussed in this section.

2.1 Survey error and error modeling

It is a futile hope that all possible aspects of survey design and the resultant errors could be modeled and that a single mathematical model could be used to dictate the choice of an "optimal" survey strategy. Instead, we should regard error models as tools with which we can evaluate and compare different design alternatives, while taking into consideration features such as feasibility, time constraints and other practical aspects. More comprehensive discussions on the use of error models to guide survey decisions are given in e.g. Fellegi and Sunter (1974) and in Groves (1989).

The total survey error of an estimator consists of both variance and bias, from all error sources. The variance then includes all variable errors, not just sampling error, and the bias is the total effect of all systematic errors. In this paper, however, we will consider the special case where only two types of errors are present: *sampling error* and *nonresponse error*. Other types of nonsampling errors, such as measurement or coverage errors, will not be considered. The nonresponse error consists of both a systematic component, nonresponse bias, and components reflecting variability. The total variance of an estimator thus reflects both the sampling variance and the additional variability introduced by the generation of the response set. The nonresponse bias is discussed in Tångdahl (2004), and the components of the total variance in a survey with nonresponse and no other sampling errors present are discussed in Tångdahl (2005).

Remark 2 *Some commonly used estimators, such as regression or calibration estimators are only approximately unbiased, even under full response, due to their nonlinearity. In large samples, this "technical" bias will be small or negligible and is not considered here.*

2.2 Measures of survey error/quality

In a survey with nonresponse estimators may be greatly biased, so the focus of attention becomes that of accuracy rather than precision. The main purpose of nonresponse rate reduction efforts is generally to eliminate or at least reduce the nonresponse bias.

Let $L^{(a)} = L(\mathfrak{P}^{(a)}, \hat{t}_{yc}^{(a)}; t_{yU})$ be some suitably chosen measure of the error resulting from using the survey procedure $\mathfrak{P}^{(a)}$ and the estimator $\hat{t}_{yc}^{(a)}$ to estimate t_{yU}. The question is then what alternative error measures are feasible, and how to choose between the alternative measures in the present setting. For our purpose there is no obvious candidate, so our decision must be based on what properties of survey estimators that should be given priority.

Let $V(\hat{t}_{yc}^{(a)})$, $B(\hat{t}_{yc}^{(a)}) = E(\hat{t}_{yc}^{(a)}) - t_{yU}$ and $BR(\hat{t}_{yc}^{(a)}) = B(\hat{t}_{yc}^{(a)}) / \left[V(\hat{t}_{yc}^{(a)})\right]^{1/2}$ be the variance, bias and bias ratio of $\hat{t}_{yc}^{(a)}$, respectively. Several different measures of error have been proposed for different applications. Although mean square error is the most commonly used measure of error (or accuracy), other measures are possible, and in some applications perhaps more suited for evaluation of the quality.

The mean square error is

$$L^{(a)} = L(\mathfrak{P}^{(a)}, \hat{t}_{yc}^{(a)}; t_{yU}) = MSE(\hat{t}_{yc}^{(a)}) = V(\hat{t}_{yc}^{(a)}) + \left[B(\hat{t}_{yc}^{(a)})\right]^2 \qquad (1)$$

An alternative measure of the survey error is the squared bias ratio

$$L^{(a)} = L(\mathfrak{P}^{(a)}, \hat{t}_{yc}^{(a)}; t_{yU}) = \left[BR(\hat{t}_{yc}^{(a)})\right]^2 = \left(E(\hat{t}_{yc}^{(a)}) - t_{yU}\right)^2 / V(\hat{t}_{yc}^{(a)}) \qquad (2)$$

Variants of these are root MSE and the absolute bias ratio. Other measures, such as the expected absolute error

$$E|\hat{t}_{yc}^{(a)} - t_{yU}|$$

and the relative bias

$$B(\hat{t}_{yc}^{(a)})/t_{yU}$$

are feasible, but will not be considered here. The most commonly used measure in practice, however, is an estimate of the variance $V(\hat{t}_{yc}^{(a)})$, which may be grossly misleading if the nonresponse bias is not negligible.

We believe that in the presence of nonresponse, the risk of bias overwhelms any requirement of precision. This should be reflected in the measure of error since sampling variance in itself then no longer is a relevant indicator. Bias can severely distort the coverage rate of confidence intervals. Unless the number of observations becomes extremely low, we will have some control over the precision (sampling variance) through the sample size, making the variance a smaller problem. However, when we consider leaving out reduction efforts and cutting off the data collection period, the precision may become more important. There is no obvious choice of measure to show all aspects of the survey error, but the measures should at least take into account both estimator variance and estimator bias.

The Mean Square Error

The MSE is probably the most used measure of the error in a biased estimator (if there was no risk of bias, we could simply measure the error with the variance). It is a simple measure that takes both variance and (squared) bias into account. It is a squared error measure, so it is always positive and large deviations are given relatively more weight. However, when MSE is used as a measure of error in a survey with nonresponse, a relatively large nonresponse bias may be offset by a small variance. What we actually should do is minimize MSE, while also keeping the bias small. Also, in most cases in practice, the (nonresponse) variance will decrease over time just because the number of observations increases. So if MSE is used, any attempt to reduce the nonresponse rate would most likely lead to some improvement, even if the bias remains constant. A reduced variance may even "hide" an increased bias. Nevertheless, the MSE is a comprehensive and widely accepted and useful measure of error.

The Squared Bias Ratio

A small absolute bias ratio is required if confidence intervals are to be valid, since the bias distorts the coverage probability. Actually, the bias ratio is under certain conditions (see Särndal et al. (1992), p. 164-165) a direct measure of the coverage rate. The situation is complex, however, since the bias ratio may be high for an estimator with small variance although the relative bias itself is small. Such an estimator would then have a coverage rate way below the intended, while an estimator with the same relative bias but much larger variance can have a coverage rate close to the nominal level.

Changes in the bias ratio will be strongly influenced by changes in variance. The squared bias ratio will only decrease between times $a-1$ and a if the squared bias decreases by a larger amount than the variance, which, in turn, is likely to decrease just by the fact that more observations are available at time a. The reason for using the squared bias ratio instead of the bias ratio itself, is to avoid the difficulty of interpreting changes, since the bias ratio may be negative or shift from positive to negative or reversed.

There is a nonlinear relationship between the absolute bias ratio and the coverage probability of confidence intervals, so the effect of changes in the bias ratio will depend on its absolute value.

2.3 Survey costs and cost modeling

The total survey cost consists of 1) components that vary with sample and response set size, variable costs, and 2) components that do not, fixed costs. Both fixed and variable costs are made up of components that are mode specific, and to some extent also survey specific. An introduction to survey costs, mainly in interviewer assisted surveys, is given in Groves (1989), Ch.2.

An early reference where the costs of follow-ups are explicitly taken into account is Deming (1953). By defining a response model where response probabilities at successive calls are constant within population groups, he derives a callback strategy that will minimize mean squared error under the constraint of fixed total cost. An estimator is constructed as a weighted average of the means obtained at each call. The parameter of interest is the mean, and no auxiliary information is used in estimation. By contrast, we here consider standard estimators used in large scale statistics production, and allow the possibility of auxiliary information. Also, we do not impose the constraint of fixed total survey cost or fixed population groups with given response probabilities at each call.

The choice of cost function should reflect the purpose of the analysis and the specific features of the survey. In particular, the cost model should be parameterized in the same terms as the error model, in this case the response probabilities. In the present context, all costs depending on the sample size or number of respondents will be regarded as variable costs. The fixed costs cover operations such as planning and administration of the survey, construction and testing of the questionnaire, frame construction and sample selection. In telephone interview surveys or surveys with face-to-face interviewing, there are additional costs for training and administrating interviewers. The variable costs can be split up into components related to the sample size and the response set size at each reduction effort that is used in the survey. Also, additional costs depending on the response set size arise from handling received responses, mail opening, data registration/scanning and processing.

For our purpose, we need to define the total survey cost and its expected value for an approach where the data collection is terminated at an arbitrary point of time a, i.e. the cost of survey procedure $\mathfrak{P}^{(a)}$. An example of such a cost function in a specific survey setup, illustrating the different cost components as functions of the response distribution, is given in Appendix A.

Generally, when we compare two survey procedures $\mathfrak{P}^{(a-1)}$ and $\mathfrak{P}^{(a)}$, it is to be understood that all survey actions that make up the procedure $\mathfrak{P}^{(a-1)}$ are also part of the procedure $\mathfrak{P}^{(a)}$. This means that expected costs will never decrease over time. Let $C_T^{(a)}$ be the total cost associated with the survey procedure $\mathfrak{P}^{(a)}$ and let $E(C_T^{(a)})$ be its expected value under $RD^{(a)}$. Thus, we have

$$E(C_T^{(a)}) \geq E(C_T^{(a-1)}) \tag{3}$$

and also $C_T^{(a)} \geq C_T^{(a-1)}$ for all combinations $a-1$ and a.

Since the focus here lies on the reduction procedure, we need not model the fixed costs in detail, only as part of the total survey cost. In other applications it may be important to split the fixed costs into separate components. Numerically small components of the total cost and components that have no bearing on the current problem need not be separated.

Remark 3 *In postal surveys, there are some additional "semi"-fixed costs involved when reminders are sent, covering mainly work hours for administrating the mail-outs. They are fixed in the sense that they do not depend on the sample or response set sizes, but variable in the sense that they occur only due to the additional effort. These costs should also be taken into account when evaluating an effort and could be included as a variable cost. However, in most practical applications these costs are small relative to the variable costs and can hence be ignored.*

3 Cost efficiency

To evaluate the current strategy for nonresponse rate reduction, we will focus on both the error of nonresponse and the cost and effect of the reduction efforts. In particular, changes in nonresponse error and increase in cost induced by the reduction efforts used during the data collection period is of interest. To weigh the effect of a reduction effort against the cost, we seek a measure that incorporates both cost and error simultaneously and that reflects both the increased cost and the change in error that is induced by an additional survey effort. Assuming that we are in the ideal situation of knowing all population parameters, these measures can be calculated so the cost efficiency of each nonresponse rate reduction effort can basically be summarized in a single number, directly indicating which of the survey strategies is the most cost efficient. For a complete evaluation of the effect of the reduction efforts, we must combine the cost efficiency measure with other indicators. In section 4, a suggested approach on how such an evaluation of the reduction strategy could be made, using a cost efficiency measure in combination with other indicators, is presented. In this section, we discuss alternative measures of the cost efficiency that can be used.

The proposed measures are not absolute, but are evaluations of one survey strategy *relative to* another. The comparisons are made between *discrete* alternatives. Although a is basically continuous, so that time points $a-1$ and a can be chosen arbitrarily, we are generally only interested in comparing a limited number of time points during the data collection period.

3.1 Measures of cost efficiency

The concept of cost efficiency, or even of survey costs in general, is rarely treated in the survey literature in connection with nonsampling errors. One exception is Groves (1989), although his focus lies on error rather than cost. In the health economy and clinical trials literature, the term cost efficiency has grown increasingly popular over the past decades and occurs fairly often in the literature in that field. A general reference is Gold, Siegel, Russell, and Weinstein (1996).

The cost efficiency measures are a comparison between the survey strategies $(\mathfrak{P}^{(a-1)}, \hat{t}_{yc}^{(a-1)}; t_{yU})$ and $(\mathfrak{P}^{(a)}, \hat{t}_{yc}^{(a)}; t_{yU})$. The time points $a-1$ and a are chosen arbitrarily, but we are generally interested in the case when an action has been taken on the part of the survey organisation between time $a-1$ and a, so that we have strict inequality in (3).

There is no widely accepted general and formal definition of cost efficiency. Loosely, one can say that the survey procedure that generates the smallest error (or highest quality) relative to the cost of achieving it is the most cost efficient. In this, nothing is said about how the error should be defined or measured.

Two general measures of cost efficiency, so far leaving the error definition unspecified, are:

$$CE_1^{(a|a-1)} = CE_1\left[(\mathfrak{P}^{(a)}, \hat{t}_{yc}^{(a)})|(\mathfrak{P}^{(a-1)}, \hat{t}_{yc}^{(a-1)})\right] = \frac{L^{(a-1)}E(C_T^{(a-1)})}{L^{(a)}E(C_T^{(a)})} \tag{4}$$

and

$$CE_2^{(a|a-1)} = CE_2\left[(\mathfrak{P}^{(a)}, \hat{t}_{yc}^{(a)})|(\mathfrak{P}^{(a-1)}, \hat{t}_{yc}^{(a-1)})\right] = \frac{L^{(a)} - L^{(a-1)}}{E(C_T^{(a)}) - E(C_T^{(a-1)})} \quad (5)$$

where $CE^{(a|a-1)}$ is to be read as: the cost efficiency of the survey strategy $(\mathfrak{P}^{(a)}, \hat{t}_{yc}^{(a)})$ compared to that of $(\mathfrak{P}^{(a-1)}, \hat{t}_{yc}^{(a-1)})$.

The measure CE_1 is proposed in Murthy (1967), with $L^{(a)} = MSE(\hat{t}_{yc}^{(a)})$, as a means to compare alternative strategies or estimators. It is the ratio of the amount of information per unit cost that the strategies supply, where the *information* is defined as the inverse of $L^{(a)}$.

CE_2 is a measure of cost efficiency frequently used in health economy applications for the purpose of resource allocation. This type of cost-effectiveness comparison is characterized as an *incremental* comparison between alternative intervention programs, where the incremental cost of an intervention is compared with the incremental effect. In the present application, the "interventions" are the additional reduction efforts.

Remark 4 *The cost efficiency measure CE_2 is actually defined in e.g. Gold et al. (1996) as the inverse of CE_2 as given by (5). In this application, however, there is a risk that the errors, or estimates of the errors, are the same at time $a-1$ and a, leading to division by zero. To avoid this, the inverse will be used instead.*

Based on the discussion in this section and in section 2.2, we can combine CE_1 and CE_2 with the suggested error measures into four alternative reasonable cost efficiency measures to consider.

$$CE_{1,MSE}^{(a|a-1)} = \frac{MSE(\hat{t}_{yc}^{(a-1)})E(C_T^{(a-1)})}{MSE(\hat{t}_{yc}^{(a)})E(C_T^{(a)})} \quad (6)$$

$$CE_{1,BR}^{(a|a-1)} = \frac{BR^2(\hat{t}_{yc}^{(a-1)})E(C_T^{(a-1)})}{BR^2(\hat{t}_{yc}^{(a)})E(C_T^{(a)})} \quad (7)$$

$$CE_{2,MSE}^{(a|a-1)} = \frac{MSE(\hat{t}_{yc}^{(a)}) - MSE(\hat{t}_{yc}^{(a-1)})}{E(C_T^{(a)}) - E(C_T^{(a-1)})} \quad (8)$$

$$CE_{2,BR}^{(a|a-1)} = \frac{BR^2(\hat{t}_{yc}^{(a)}) - BR^2(\hat{t}_{yc}^{(a-1)})}{E(C_T^{(a)}) - E(C_T^{(a-1)})} \quad (9)$$

It is not self-evident which measure of error and which cost efficiency measure that we should use. The error measures presented in section 2.2 both have their merits and their shortcomings, as have the suggested cost efficiency measures. Note that there is a possibility that the squared bias ratio is zero or nearly so, causing the ratio $CE_{1,BR}$ to be arbitrarily large. Regardless of the choice we make, no univariate measure could (or should) be used on its own to dictate survey decisions. This is formulated in Gold et al. (1996), although in a different context, as[1] :

> ... decisions in the real world are more complicated. Cost-effectiveness analysis provides valuable information about tradeoffs in the broad allocation of health resources, but other factors need to be considered as well [...] CEA is not a complete decision making process. The information it provides is, however, crucial to good decisions.

In the ideal case, the cost efficiency can be summarized in a single number, so decision rules are easily formulated. Of course, when the parameters must be estimated, there will be uncertainty about the true cost efficiency, so these decision rules must be adjusted.

The survey strategy $(\mathfrak{P}^{(a)}, \hat{t}_{yc}^{(a)})$ will be cost efficient relative to the alternative survey strategy $(\mathfrak{P}^{(a-1)}, \hat{t}_{yc}^{(a-1)})$ if

$$CE_1 > 1 \tag{10}$$

i.e. if it produces more information per unit cost. Another way to interpret this decision rule is that a relative change in information is considered cost-efficient only if the relative change in cost is the same or smaller. This is not an obvious choice of decision rule, since the utility of an increase in the amount of information is subjective and may vary between different surveys. Instead we could formulate a decision rule that reflects how a change in quality is valued, i.e. how much more we are willing to pay for a reduction (change) of the error. This could be done by replacing (10) with

$$CE_1 > \delta \tag{11}$$

where $\delta > 1$ as our decision rule.

[1] In this citation, CEA is short for cost-effectiveness analysis.

A reasonable decision rule when using CE_2 is that $(\mathfrak{P}^{(a)}, \hat{t}_{yc}^{(a)})$ is cost efficient relative to $(\mathfrak{P}^{(a-1)}, \hat{t}_{yc}^{(a-1)})$ if

$$CE_2 < \delta \tag{12}$$

where $\delta < 0$. In both (11) and (12), the constant δ is arbitrarily chosen, depending on how we value the change in quality.

Remark 5 *If the error measure can take on both positive and negative values, the cost efficiency measure CE_2 will become difficult to interpret, since we can not formulate an unambiguous decision rule. Such error measures should be avoided.*

In both measures, we compare two alternative strategies, rather than defining an absolute required level of e.g. information per unit cost. Thus, our conclusions will depend on *what* we compare, i.e. how we choose $a-1$ and a. Also, as the numerical examples in section 3.2 will show, the properties of the cost efficiency measures will depend largely on the choice of error measure.

3.2 Numerical comparisons

To illustrate the properties of the error and cost efficiency measures, a small numerical study is presented. The study is based on data from Statistics Sweden's 2003 survey *Gymnasieungdomars studieintresse*, on transition from upper secondary school to higher education.

The survey is carried out yearly among third-year students and data was collected by mail in October, November and December 2002 from a sample of 9023 students. The design is stratified simple random sampling and the frame population (students in the second-year student register from the previous year) is stratified by region, second-year study programme and gender. In the 2003 survey, a calibration estimator using known population totals, but no additional sample level information, was used. The auxiliary information is available from an updated student register and the calibration is done on the marginal frequencies of several categorical variables, among other the stratification variables, final mark in grade 9 and parents' level of education. The auxiliary variables are chosen because they seem to be correlated with either the important study variables or the response propensity. The total nonresponse rate was about 25 percent. The cost data in the examples is based on the actual fixed and variable costs generated by the separate mailings.

There are several study variables. Two of the most important ones are a) Intentions of third-year secondary school students with regard to pursuing studies at university level (*Yes/No/Not decided*), and b) The university programmes viewed by these students as the most attractive and interesting. For these variables, both totals and proportions are estimated. In this numerical illustration, only estimation of the proportion *Yes* and *No* on the first variable is studied.

There is a belief that the auxiliary information eliminates most of the bias associated with frame deficiencies and nonresponse. This may or may not be true, but to study the performance of the error and cost efficiency measures, two different scenarios concerning the bias, presented in the following, are considered. In the examples, the estimated standard errors from the survey are used, while the bias is estimated by the difference between the point estimate and the true population total, for different assumptions about the population total. The realized costs are used as estimates of expected costs at each time point. We use the estimates of the bias, standard errors and costs as if they are true values. The time points $a-1$ and a are chosen so that comparisons are made between successive efforts (as illustrated in table 1 on page 23). The estimates of the mean square error and the squared bias ratio are calculated each day during the data collection period.

Scenario A

We *assume* that the true proportions *Yes* and *No* are 44 percent and 31 percent, respectively. This gives a remaining bias at time A of 4 percent units for the proportion of students responding *Yes*, and -5 percent units for the proportion of students responding *No*, corresponding to relative biases of about 10 percent for the *Yes* variable, and about 20 percent for the *No* variable. The cost efficiency measures are plotted in figures 1 and 2 for the *Yes* variable, and in figures 3 and 4 for the *No* variable. The horizontal dotted lines in the figures mark the "standard" choice of δ.

Under this scenario, the measures $CE_{1,MSE}$ and $CE_{2,MSE}$ both indicate for the *Yes* variable that the last reminder is not cost efficient at all. Looking at $CE_{2,MSE}$, the conclusion is that the first reminder is more cost efficient than the second, while $CE_{1,MSE}$ is almost equal for the two. The measures CE_{1,BR^2} and CE_{2,BR^2} indicate that none of the efforts are cost efficient, the

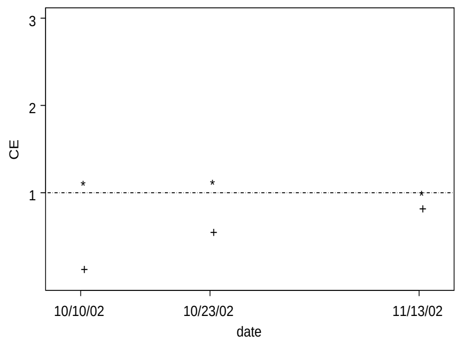

Figure 1: proportion Yes, * = $CE_{1,MSE}$, + = CE_{1,BR^2}

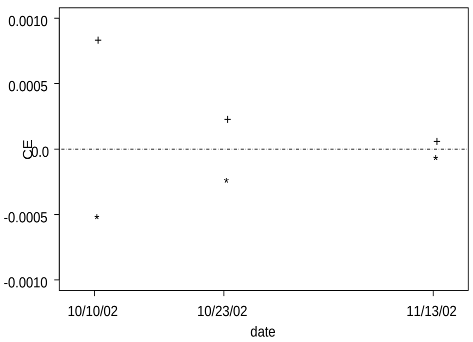

Figure 2: proportion Yes, * = $CE_{2,MSE}$, + = CE_{2,BR^2}

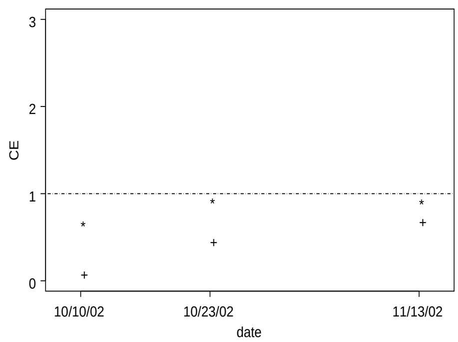

Figure 3: proportion No, * = $CE_{1,MSE}$, + = CE_{1,BR^2}

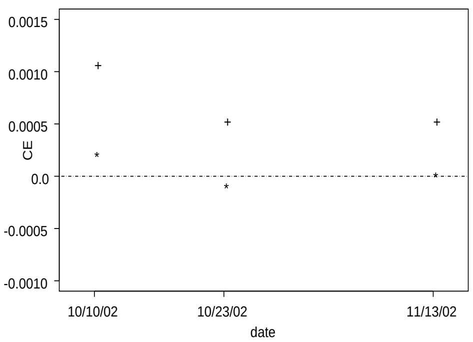

Figure 4: proportion No, * = $CE_{2,MSE}$, + = CE_{2,BR^2}

first least so. This pattern is nearly the opposite compared with the MSE based cost efficiency measures. For the *No* variable, the patterns for the MSE- and BR^2-based measures are the same, only the magnitudes differ. They show that none of the efforts are cost efficient, with the exception of the second reminder according to $CE_{2,MSE}$. What causes these results to appear? An analysis of the error measures, shown in figures 5, 6, 7 and 8 respectively, gives the answer. The vertical dotted lines in the figures mark the dates when the reminders are mailed out, i.e. when reduction efforts are made.

For the *Yes* variable, the MSE drops rapidly in the beginning of the data collection period, so we get a large positive effect (increase in quality) from

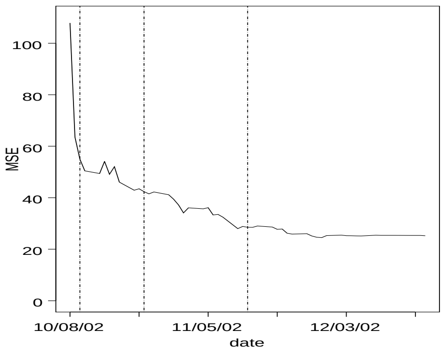

Figure 5: Mean square error, proportion Yes

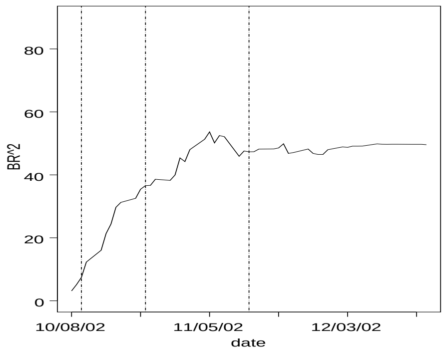

Figure 6: Squared bias ratio, proportion Yes

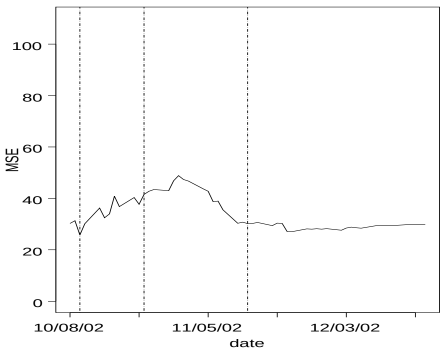

Figure 7: Mean square error, proportion No

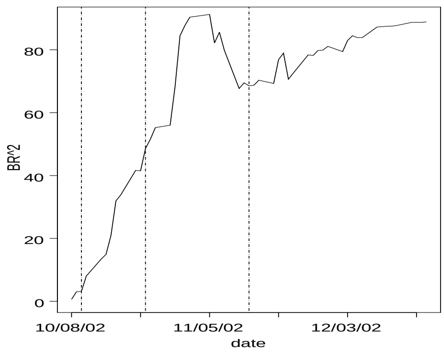

Figure 8: Squared bias ratio, proportion No

the first effort. The MSE continues to drop, but at a slower rate, until it finally stabilizes, giving no effect at all from the last effort. The squared bias ratio for the *Yes* variable is increasing during most of the data collection period, but stabilizes at the end. This is due to the fact that the standard error decreases much more rapidly than the bias (not shown). Thus, the error, as measured by the squared bias ratio is *increasing* over time.

The *No* variable shows an increase in both the squared bias ratio and the MSE initially, but at the middle of the data collection period they both drop. The increase in the squared bias ratio is dramatic initially. The MSE eventually stabilizes, while the squared bias ratio starts to increase again after the sudden drop.

Scenario B:

Under this scenario, it is *assumed* that there is essentially no bias remaining at time A, for either of the variables. The cost efficiency measures are plotted in figures 9 and 10 for the *Yes* variable, and in figures 11 and 12 for the *No* variable. The horizontal dotted lines in the figures mark the "standard" choice of δ.

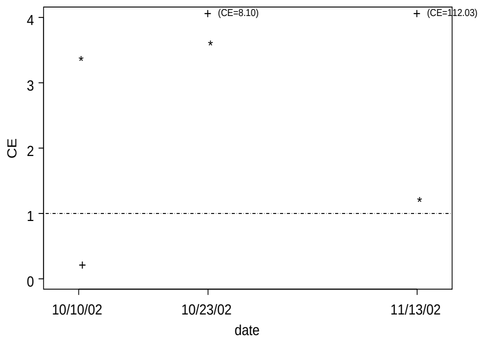

Figure 9: proportion Yes, $* = CE_{1,MSE}$, $+ = CE_{1,BR^2}$

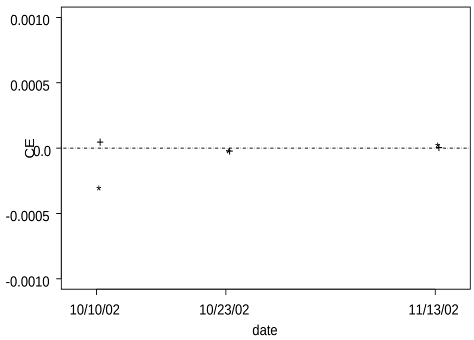

Figure 10: proportion Yes, $* = CE_{2,MSE}$, $+ = CE_{2,BR^2}$

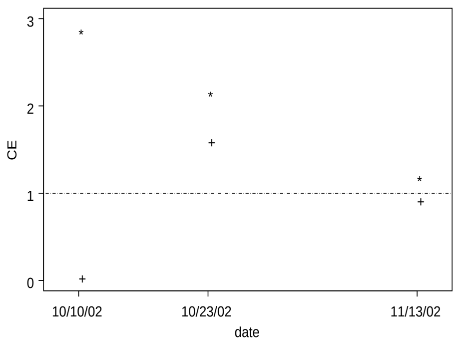

Figure 11: proportion No, $* = CE_{1,MSE}$, $+ = CE_{1,BR^2}$

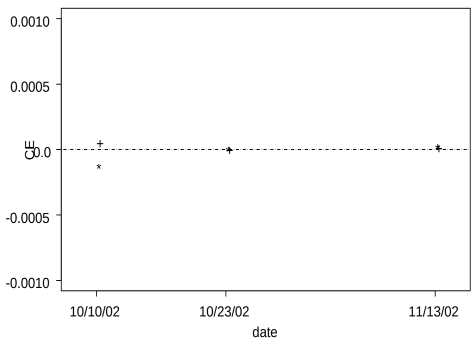

Figure 12: proportion No, $* = CE_{2,MSE}$, $+ = CE_{2,BR^2}$

The measure $CE_{1,MSE}$ indicates that the first effort is cost efficient, both for the *Yes*-variable and for the *No*-variable. $CE_{2,MSE}$ shows a marginal cost efficiency for the first effort, but none at all for the last two, while CE_{2,BR^2} indicates that none of the efforts are cost efficient. This holds for both variables. The measure CE_{1,BR^2} breaks down for the *Yes* variable since the squared bias ratio is very near zero at the end of the data collection period.

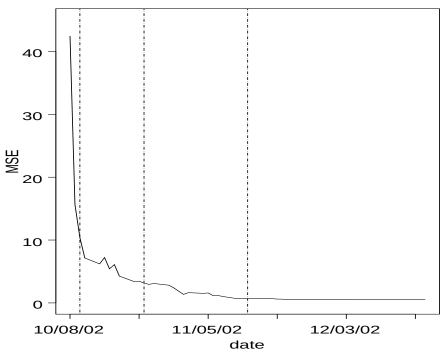

Figure 13: Mean square error, proportion Yes

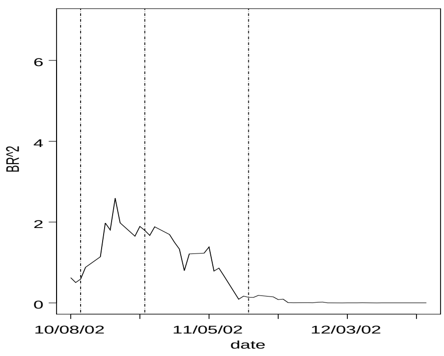

Figure 14: Squared bias ratio, proportion Yes

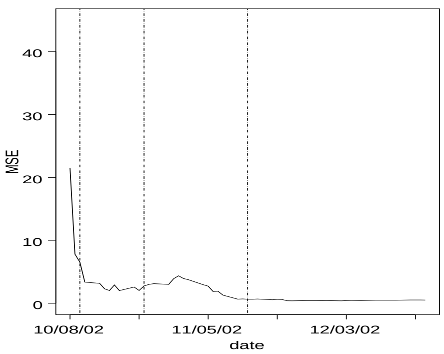

Figure 15: Mean square error, proportion No

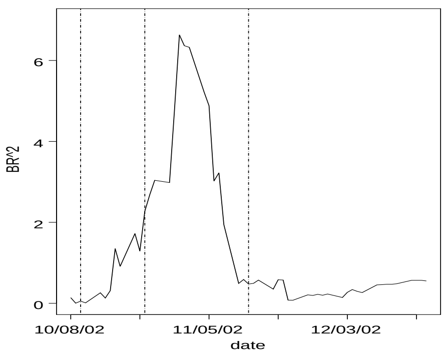

Figure 16: Squared bias ratio, proportion No

In figures 13, 14, 15 and 16 respectively, we look more closely at what causes this. The horizontal dotted lines in the figures mark the “standard” choice of δ.

For the *Yes* variable, the MSE drops (rapidly in the beginning) while the squared bias ratio increases. For the *No* variable, the MSE decreases initially, rises in the middle of the data collection period, but eventually stabilizes. The squared bias ratio increases dramatically initially, but drops off to stabilize at a low level.

This simple numerical example shows the impact of the error measure on the cost efficiency measures and the conclusions. It also shows that the conclusions about the cost efficiency, of the reduction efforts may be quite different for different study variables. As discussed in section 2.2, it also seems that the squared bias ratio may, under unfortunate circumstances, indicate that the error increases, despite the fact that both the absolute bias and the standard error decrease, as discussed in section 2.2.

4 An approach to evaluate the reduction efforts

4.1 Some opening remarks

The measures proposed in section 3.1 can be used as rough indicators of the cost efficiency of the reduction efforts, mainly follow-ups or callbacks. In a full evaluation of the follow-up strategy, additional factors must be considered; the cost efficiency measure alone does not tell the full story.

This section deals with an approach where the reduction efforts are evaluated with the use of one of the cost efficiency measures suggested in section 3.1, combined with other indicators. As mentioned previously, the suggested approach is an evaluation procedure where the cost and effect of every effort is evaluated, given the reduction strategy as a whole. The problem – and the proposed solution in the ideal case – is formulated in descriptive terms and the conclusions drawn about whether to exclude an effort or not should only be applied in future surveys, not in an ongoing survey.

In a situation without nonresponse, it is good practice to formulate precision requirements prior to survey planning and budget allocation. These requirements should be based on and reflect the precision needs of the major users of the survey results. However, when nonresponse can be expected, formulation of the precision requirements is a more difficult task. We should then formulate quality requirements involving both variance and bias. The requirement as regards the nonresponse bias should be that no bias is acceptable, but in practice it is not realistic to believe that we can eliminate the bias completely. This raises the question of how much we are willing to spend on error reduction. The monetary value of a reduced error is subjective and may vary between surveys, depending on the intended use of the statistics and the survey manager's assessment of user needs for accuracy.

Other factors must also be considered, since quality in a survey refers to all aspects of the product which are of relevance for how well it meets users' needs and expectations. A quality concept used by e.g. Statistics Sweden in quality declarations has the following main components: (1) *Contents*, (2) *Accuracy*, (3) *Timeliness*, (4) *Coherence and comparability*, (5) *Availability and clarity*. The statistician's assessment of the survey strategy must be formed with regard to all of these aspects. The balance between these sources of error is discussed in e.g. Biemer and Lyberg (2003) and in Groves (1989).

4.2 Suggested approach

To facilitate understanding of the basic ideas, we will introduce an example of a specific survey procedure. Without any loss of generality, we can describe the general approach in terms of this procedure. The survey procedure in general and the time intervals between two successive steps in the setup may vary between surveys. The procedure is chosen as a fairly standard setup of a postal survey conducted by a large scale statistical agency. The survey in the numerical examples in section 3.2 follows this setup. Consider the following: All preventive efforts are regarded as fixed, i.e. we assume that they have been decided upon prior to data collection. So, among many other activities, the questionnaire and the sampling frame have been constructed, the letter of introduction has been written and the sample has been selected according to the preferred design. We will also assume that precision requirements have been formulated, and that levels of δ have been agreed upon.

The current survey setup from this point on consists of the following stages:

1. Data collection is initiated at time Q by the mailing of the questionnaire and a letter of introduction to all elements in the sample.

2. A specified number of days later, at time TR, a thank you/reminder card is sent to all elements in the sample.

3. At time $R1$, usually within one or two weeks from TR, a reminder, including a new questionnaire, is sent to those elements in the sample that have not yet returned the questionnaire.

4. A second reminder, including a questionnaire, is sent to those elements that belong to the nonresponse at time $R2$.

5. A third and last reminder, including a questionnaire, is sent at time $R3$ to those that still have not responded to the survey request.

6. At time A, the data collection period is terminated. The sample elements who have not yet returned the questionnaire are classified as ultimate nonresponse.

In the following, let T' denote the time point just before T. In a postal survey, it is natural to let T' be the day before e.g. a reminder is mailed out.

At statistical agencies carrying out a large number of surveys, the timing between mailings is usually determined according to some standard setup. The wording and layout of the advance letter that accompany the questionnaire also usually follows some given standard.

Since the effect of the different efforts can not always be separated, the data collection procedure must be evaluated as a whole. The effect of previous efforts may influence the apparent effect of subsequent efforts, but not the other way around. [2] The evaluation should therefore be performed stepwise, starting, in the ideal case considered here, with the last effort. If an effort, based on the procedure described below, is not judged as cost efficient, we move on to the next to last effort, and so on until all efforts have been evaluated. Of course, other aspects than merely the error as given by just one error measure must be considered as well. From the numerical comparisons in section 3.2, we see that it is important to base conclusions on the evaluation of *all* efforts, to look for interaction effects and to weigh in other aspects of the survey strategy, quality and feasibility.

Remark 6 *In some surveys, it might be suspected that the response quality is not as good for late respondents, i.e. that late responses to a larger extent suffer from measurement errors. We do not address this issue here.*

Assuming that we are in the ideal situation where all quantities in the cost efficiency measures can be calculated, we could perform the following procedure for each effort that is to be evaluated. In the survey setup described here, we recommend choosing time points $a-1$ and a as indicated in table 1, to give "pairwise" evaluation of the efforts. In telephone interview surveys, it is more natural to define the time points in terms of the number of callbacks instead of days between efforts. One should be aware that the choice of $a-1$

[2] We will disregard the possibility that people respond in order to avoid being subjected to future callbacks or reminders.

Evaluation of	$a-1$	a
R3	R3'	A
R2	R2'	R3'
R1	R1'	R2'
TR	TR'	R1'

Table 1: Choice of time points $a-1$ and a

and a may impact the conclusions. A complementary way to choose time points is to fix a at A and choose $a-1$ as $R3'$, $R2'$ and so on. The conclusions about the cost efficiency then applies to the combination of reduction efforts used between $a-1$ and a. The combination of efforts may in fact be cost efficient, although none of the efforts evaluated separately are themselves cost efficient.

For every choice of $a-1$ and a, the following steps are taken:

Step 1 The chosen cost efficiency measure is calculated. The decision rule is that the strategy $(\mathfrak{P}^{(a)}, \hat{t}_{yc}^{(a)})$ is cost efficient, relative to $(\mathfrak{P}^{(a-1)}, \hat{t}_{yc}^{(a-1)})$, if

$$CE_1^{(a|a-1)} > \delta,\ \delta > 0 \qquad \text{or} \qquad CE_2^{(a|a-1)} < \delta,\ \delta < 0$$

depending on which measure is chosen. The constant δ is chosen to reflect the importance of a quality improvement.

Step 2 The variance of the point estimator at time $a-1$ is calculated, to determine if the precision requirements are met and $\mathfrak{P}^{(a-1)}$ possible.

Step 3 The magnitude of the absolute and relative bias is calculated. Even if one or more efforts are not found to be cost efficient and thus could be excluded, there may still be a large bias present. If so, the complete survey procedure must be reappraised.

Step 1 indicates whether the reduction effort at time a, or, rather, if the procedure $\mathfrak{P}^{(a)}$, is more cost efficient than $\mathfrak{P}^{(a-1)}$, while steps 2 and 3 provide additional information to help guide the decision making process. Taken together, these indicators provide a basis for making a decision on whether to exclude an effort or not, and also an evaluation of the survey procedure as a whole.

What conclusions shall be drawn? As has been pointed out previously, the statistician must use his experience and discrimination in the evaluation and in drawing conclusions based on the results from steps 1-3 above. Here we can only give some general comments. These are valid even if the cost efficiency, the variance and the bias must be estimated, but then the decision rules must be adjusted to reflect this uncertainty.

If step 1 should indicate that the effort being evaluated *is* in fact cost efficient, then steps 2 and 3 become less important. However, even if the effort at time a reduces the error, we should still check for remaining bias and that the precision requirements are met. If there seems to be a substantial bias remaining even at time a, we should look for ways to eliminate it. Should the variance be too large, we could consider taking a bigger sample in future surveys. If it is the nonresponse that causes the variance to be too large, we should consider alternative ways to handle the nonresponse. Both the bias and the variance can also be reduced by using a better estimator or better auxiliary information in the current estimator.

When step 1 indicates that the effort is in fact *not* cost efficient, the obvious consequence should be that we exclude this particular effort from the data collection procedure and possibly terminate the data collection period earlier. But before doing that, it is essential that steps 2 and 3 have been taken, as the recommended course of action depends on what they indicate. If both the bias and the variance at time $a-1$ are "small enough" (what this means must be determined from case to case), the effort can be excluded without negative effects on estimator properties. If the bias is sufficiently small at time $a-1$ and the variance is too large at time $a-1$ but not at time a, we must look for ways to reduce the variance, other than nonresponse rate reduction. Unless we succeed in reducing the variance, we should probably not terminate data collection at time $a-1$.

We may find that the bias at time $a-1$ is large and that we do not succeed in reducing it by reducing the nonresponse rate. Then it does not matter what the variance is. We should then consider reworking the whole survey procedure, so that the bias can be eliminated or at least reduced.

Remark 7 *It may be argued that if the bias is too large at time A, it is needless to evaluate the reduction efforts, since we know the data collection is not working satisfactorily. However, it may be that the reduction efforts* do *have an effect on the error, but that this is not enough. If the current efforts are cost efficient, it is possible to conclude that we should* add *efforts* *to the existing data collection procedure, instead of replacing it entirely.*

Note that there are some difficulties and unresolved issues in the approach, even if we assume that all relevant parameters and population characteristics are known. Firstly, it can be difficult to specify δ. This is not the statistician's task, but should be done in consultation with the users and/or client. Secondly, the cost efficiency measures can be difficult to interpret since the results may depend on what time points $a-1$ and a are chosen. It is thus important that care is taken when the time points are chosen, and that all reduction efforts are evaluated before definite conclusions are drawn, so that a comprehensive assessment of the data collection can be made.

5 Dealing with uncertainty

The approach described in this paper requires that we are in an ideal situation where all relevant quantities are known or at least possible to calculate. This is of course not the case in practice, as was indicated in the introduction. What we have done here is to point in the direction of a possible "ideal" procedure. It must be stressed that the approach only serves as a tool with which we can evaluate the current survey setup and, in particular, the nonresponse rate reduction procedure. As with any tool, if used as intended, it can be a great help but if misused and used without discrimination, important aspects of the survey procedure and estimator quality can be missed, leading to the wrong conclusions.

What remains to be solved is how to apply this procedure in practice, when relevant quantities such as the bias, the total variance and the expected costs must be estimated. We must also find ways to communicate the results from the cost efficiency analysis and evaluation, and the associated uncertainty, to the users or clients.

In practice, we can find ourselves in one of the following typical situations:

i. Reliable estimates of the required quantities, such as the estimator bias and variance at different time points, as well as expected costs, are available.

ii. Auxiliary variables are available, allowing for an analysis where these variables are used as proxies for the study variable. However, the response probabilities are unknown, so the analysis must be based on (realistic) assumptions about the response distribution.

iii. An evaluation study is carried out in order to learn the values of the population parameters or the corresponding sample values, or at least unbiased estimates of them. These can then be used to produce better estimates of the nonresponse bias and related quantities.

iv. No additional information is available, other than data on costs and response sets. It is not possible to perform an evaluation according to one of the above situations, but we can calculate the standard point estimators and variance estimators for any choice of a.

If an analysis according to one of the above situations is not possible, we can regard the estimates at the current cut-off date for final returns as the best estimate possible. This would then define the level with which we compare the other estimates.

The implementation of the suggested evaluation approach depends on our ability to estimate a number of population unknowns. Unless special efforts are made, our evaluation must rely heavily on unverifiable model assumptions. In each of the above situations it is the statistician's task to produce such estimates, perhaps under a variety of reasonable assumptions, to asses the possible range of values for the true values. In addition, this may provide some insight about the sensitivity of our conclusions about the cost efficiency to the model assumptions and uncertainties.

Also, since we build the cost efficiency measures on estimated quantities, we must consider the inferential aspects. Our CE measures will only be estimates of the true ones. In each of the cases described above, we must work out how to deal with the uncertainty, related to the specific situation. Attempts along these lines will be made in a subsequent paper.

References

Biemer, P. P. and L. E. Lyberg (2003). *Introduction to Survey Quality.* New York: Wiley.

de Heer, W. and E. de Leeuw (2002). Trends in household survey nonresponse: A longitudinal and international comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. A. Little (Eds.), *Survey Nonresponse*, pp. 41–54. Wiley.

Deming, W. E. (1953). On a probability mechanism to attain an economic balance between the resultant error of response and the bias of nonresponse. *Journal of the American Statistical Association 48*, 66–77.

Fellegi, I. P. and A. B. Sunter (1974). Balance between different sources of survey errors – some Canadian experiences. *Sankhya, Series C 36*, 119–142.

Gold, M. R., J. S. Siegel, L. B. Russell, and M. C. Weinstein (1996). *Cost-Effectiveness in Health and Medicine.* Oxford: Oxford University Press.

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

Murthy, M. (1967). *Sampling theory and Methods.* Calcutta: Statistical Publishing Society.

Oh, H. and F. Scheuren (1983). Weighting adjustment for unit nonresponse. In W. Madow, I. Olkin, and D. Rubin (Eds.), *Incomplete data in sample surveys*, Volume 2, pp. 143–184. Academic Press.

Särndal, C. E., B. Swensson, and J. Wretman (1992). *Model Assisted Survey Sampling.* New York: Springer-Verlag.

Tångdahl, S. (2004). Nonresponse bias for some common estimators and its change over time in the data collection process. ESI Working Paper Series 2004:13, Örebro University.

Tångdahl, S. (2005). The variance of some common estimators and its components under nonresponse. ESI Working Paper Series 2005:9, Örebro University.

A An example of a specific cost function

For this follow-up procedure, a reasonable cost function for the procedure $\mathfrak{P}^{(a)}$ is given by

$$C_T^{(a)} = C_F + C_V^{(a)}$$

where $C_V^{(a)}$, the variable cost for the complete data collection up to time a, is a sum of cost components due to each reduction measure. We identify the per element costs

- c_Q, cost of the initial mailing of the questionnaire, including paper, printing, questionnaire and return envelopes, enclosing, addressing and postage.
- c_{TR}, cost of a thank you/reminder card, including paper, printing, addressing and postage.
- c_{R1}, c_{R2}, c_{R3}, cost of reminder 1, 2 and 3 respectively, including paper, printing, addressing, postage and return envelope
- c_P, cost of processing received responses, including registration of incoming responses, data registration or scanning and editing.

In the present situation, it is reasonable to assume that the per element costs are fixed and thus independent of the sample size, although this is an oversimplification. For the domain of applicability of the cost model, however, the assumption should be sufficiently realistic.

The total variable cost for the current survey procedure $\mathfrak{P}^{(A)}$ is

$$\begin{aligned} C_V^{(A)} =& C_Q + C_{TR} + C_{R1} + C_{R2} + C_{R3} + C_P^{(A)} \\ =& nc_Q + nc_{TR} + (n - m^{(R1')})c_{R1} + (n - m^{(R2')})c_{R2} \\ &+ (n - m^{(R3')})c_{R3} + m^{(A)}c_P \end{aligned}$$

Depending on the choice of cut-off point for final returns, some of the costs may be avoided. The variable cost for the procedure $\mathfrak{P}^{(a)}$ can, in this specific example, be expressed

$$C_V^{(a)} = C_Q + C_{TR} + \sum_{j \le a} C^{(j)} + C_P^{(a)} \text{ for } j = R1, R2, R3$$

Since $m^{(j)} = \sum_s R^{(j)}_{k|s}$, the number of respondents, is a random quantity under the assumed true response distribution, $C_V^{(a)}$ becomes a random variable with expected value

$$
\begin{aligned}
E_p E_{RD^{(a)}} \left(C_V^{(a)} \right) =& E_p \left(n(c_Q + c_{TR}) + \sum_{j \le a} (n - \textstyle\sum_s \theta^{(j')}_{k|s}) c_j + \textstyle\sum_s \theta^{(a)}_{k|s} c_P \right) \\
=& n(c_Q + c_{TR}) + E_p \left(\sum_{j \le a} (n - \textstyle\sum_s \theta^{(j')}_{k|s}) c_j + \textstyle\sum_s \theta^{(a)}_{k|s} c_P \right)
\end{aligned}
$$