Side Effect Reduction of Prior and Processed Information on Survey Design

(First Part)

A. Demnati

Independent Researcher, Ottawa, Canada, Abdellatif_Demnati@msn.com

Abstract

It is difficult to design and conduct a survey because prior information on response rates and the like is likely generated from a different random process than the target one governing the surveys to be designed; and survey process, such as text classification, may vary from one human or machine to another. The impact of each error-prone sets of information on the properties of the estimator can be significant. We are concerned with reducing the side effect of both error-prone prior information and processed information on the quality of the estimator of the parameter of interest during survey data collection period. Nowadays, computer-assisted survey methods provide an instant variety of observations on the survey process and on the target random process governing the survey under consideration. These paradata, data, and quality measures enable the survey producer to make decisions regarding the need for methodology-process revision during survey data collection period, which involves consideration of both a model that represents how the target information relates to the error-prone information and the design that describes how the observations are obtained. We think of the error-prone and target information is a random process that has a joint distribution with some probability function. Then, at each phase of survey data collection, after receiving the information that the target random process has taken specific values, we update the joint probability distribution to revise the design specification in the course of the survey data collection period. Also, the coefficient of reliability for a survey as a whole set of processes as well as for a single process is further discussed.

Key Words: Multiple sources of information, Optimal resources allocation, Responsive design, Two-phase sampling, Unit classification, Wisdom design.

1. Introduction

There are a wide range of consumer market areas, such as health, biometrics, industrial, commercial, finance, insurance, actuarial and more that require the estimation of quantities related to uncertain or imprecise information ψ to learn, model, and predict human and market behaviours. Vague understanding of ψ promote approaches for the development of models to explain known observations on prior information χ, predict ones on ψ, and relates prior information with the target vague information after observing some of its realisations. It is hoped that the approach possesses some desirable properties, such as:

(1.1)

The fourth step incorporates prevention and correction. It is also desired that the approach can be used for tasks that are executed by humans. For instance, in the context of pattern classification, artificial neural network is inspired by the way biological neural networks in the human brain process information. It first learns a mapping $E(\boldsymbol{\psi}_k) = \mu(\boldsymbol{\nu}_k; \boldsymbol{\lambda}_{\psi})$ from input $\boldsymbol{\nu}_k$ to expected value of output $\boldsymbol{\psi}_k$ given a sample of training examples $\wp = \{(\boldsymbol{\psi}_k, \boldsymbol{\nu}_k); k = 1,...,n\}$ of input-output pairs $(\boldsymbol{\psi}_k, \boldsymbol{\nu}_k)$ for unit k, then uses the uncovered patterns to predict unknown output using the best guess $\hat{\boldsymbol{\psi}}_k = \mu(\boldsymbol{\nu}_k; \hat{\boldsymbol{\lambda}}_{\psi})$, where $E()$ denotes expectation with respect to the random process involved, n is the number of training examples in the sample $\wp$, $\boldsymbol{\lambda}_{\psi}$ is a (large) vector of weights expressing the importance of the respective inputs to the output (Rosenblatt 1958, 1962) and $\hat{\boldsymbol{\lambda}}_{\psi}$ is the solution to an error minimization problem used to train the artificial neural network. Each training input $\boldsymbol{\nu}_k$ is a vector of numbers, representing (possibly complex structured) units such as a person, an image, a sequence of characters or words, a video, etc. These are called features or covariates. The form of the output can in principle be anything, but most methods assume that $\boldsymbol{\psi}_k$ is a categorical variable. Artificial neural network is used universally for (1) capturing similarity within a set of labelled units represented by features; (2) high feature dimensionality; and, (3) when the relation between input and output information is vague or difficult to describe. Well known application includes text classification, email spam filtering, image classification, handwriting recognition, face recognition, and fraud detection. Artificial neural network achieves, to some extent, the first two desired properties stated by (1.1) based on a substantial sample of examples. It helps to realize the fact that, in order to train an artificial neural network, one needs a large sample that is random with no errors. Therefore, the question follows: How can we select observations on the target process in the absence of a training sample? How can we prevent and correct processing errors when collecting observations? It also helps to realize that there is a cost

associated with each stage of the process of obtaining the random sample such as: (1) the selection of units; (2) the optional follow-up in an attempt to receive a response from nonrespondents; and, (3) each mode of data collection (e.g. in Person, by phone, by mail, or via the internet). For this paper, we study the general problem which includes the four desired properties stated by (1.1) in the context of survey studies.

Survey or census studies start with a collection of distinct units of interest known as the population. There are multiple random variables attached to each unit, as each unit holds their own individual characteristics and aptitudes. Each particular study targets a small subset of these random variables. Measurements on some of these variables of interest are intended to be collected during the data collection stage from each selected unit and involve a questionnaire used to collect the data from the respondents. Meanwhile, measurements on the other set of these variables of interest are intended to be derived from one or more observed variables. These other variables are not directly included as items in the questionnaire. Both observed and derived measurements are used at the estimation stage to draw inferences about the parameter of interest associated with the given study.

At the planning stage of a survey, the question of determining resources and allocating them within different stages (such as sampling, nonresponse follow-up, data collection, validation) of the survey design is a difficult and critical one. Survey developers must justify resources to be used, and the survey managers should review the justification to ensure the survey produces results within resources, quality and timing constraints. Efficiency is an important issue because inefficient determination and/or allocation may lead to: 1) imprecise results; and, 2) a waste of time, resources, and money. To determine optimally, the design parameters such as (1) the duration of the survey; and, (2) the amount of resources and their allocation period, must be determined. Design pre-specification requires 3 steps:

1) **Specification step.** The specification of: a) the population of interest; b) the parameter of interest; c) the sampling frame and the sampling schemes; d) the nonresponse follow-up activities; e) the estimator to be used; f) the desired precision or the global cost; g) the cost function; and, h) the precision function;
2) **Prediction step.** Obtaining prior information, from the sampling frame, administrative files, or from previous surveys, needed to compute unknown quantities in formulas for both precision and cost functions; and,
3) **Optimization step**. Determining the survey design parameters through optimization of some objective function – that involves both precision and cost functions.

Suppose previous surveys suggest that the conditional probability of responding h (in a time period) for a unit in the population of interest is constant over time. When the conditional response probability h is constant over time, then the marginal response probability over I time periods is given by $\xi_1 = 1-(1-h)^1$. To reach a marginal probability of response close to 1 under constant conditional response probability, it will take around 17 time periods when $h = .5$, and over 100 time periods when $h = .1$. Collecting data over such a long period is time consuming, costly, and the results may vary from one time period to another. Because of this, the method of survey sampling when capturing information from (or estimating parameters with respect to) a population generated from such random processes is as follows: 1) selecting a random sample of units from the population; 2) increasing the level of efforts in terms of nonresponse follow-up activities to improve units cooperation; and, 3) monitoring the survey process to evaluate its quality and stability. Sampling is based on the idea that, within a certain margin of error, one can infer something about the parameter of interest from a small sample, as long as the sample was chosen at random. Efficient nonresponse follow-up requires information on the error-free target response mechanism governing the survey under consideration. It is difficult to pre-specify the design for certain surveys because prior information is more likely to be generated from a different random process than the one under consideration. A naive approach simplifies the problem under the assumption that resources should be big enough to have good estimates. However, a survey usually has a limited budget and time, and those in turn, in combination with the resource allocations used within the stages of the survey design based on prior information, determine its achievable quality. Nowadays, computer-assisted survey methods provide an instant variety of observations about the survey process and the target random process that can be used to revise survey design during its process. Although, previous survey designs are mostly done deterministically using prior information, there is a widespread need for adaptive or responsive design where the design is revised during the data collection period. The intent of such revision is to reduce errors attached to design pre-specification on prior information grounds. Groves and Heeringa (2006) introduced the concept of responsive design, formulated its objectives, and used paradata to guide mid-survey decisions affecting properties of the estimates. Peytchev *et al*. (2010) used paradata and other information to estimate the likelihood of any sample member becoming a non-respondent and suggest for those sample cases, least likely to respond, to employ a more effective survey protocol to gain unit cooperation. Schouten *et al*. (2013) considered adaptive design where each unit is assigned a follow-up treatment or strategy from a set of candidate strategies. A review of a substantial literature on adaptive and responsive designs is the paper by Tourangeau *et al*. (2016). In Demnati (2016), we formulated an optimization problem for designing a survey, and identified steps for its revision in the course of the data collection period. We considered the error-prone prior information and the error-free target information as a random variable with a joint distribution with some probability function. Then, we updated the joint

probability distribution after observing some of realizations of the error-free target random process at each phase of data collection, to revise the design specification in the course of the data collection period. The proposed approach makes full use of error-prone prior information while requiring only few observations from the expensive error-free target random process. A reliability coefficient for a survey as a whole set of processes, as well as for a single process, was also discussed. Such a coefficient when supplied with the Mean Square Error (MSE) enhances information on 1) the survey results; 2) the comparisons between surveys; and, 3) the contribution of the given survey as addition to prior information. In this paper, we extend our work to cover survey process.

A survey process such as data collection, measurement, text classification, or imputation, is the process by human or machine of taking provided responses and deriving them into a set of values that represent the targeted values of the complete survey variables of interest. Once obtained, the complete set of values is analyzed in the same way a set of complete observed responses can be. Here, automatic text classification also known as automatic text or document categorization is the task of automatically sorting a set of texts into predefined groups (or classes or categories) based on its inputs. Automatic classification system learns from previously classified texts the characteristics of one or more groups. Automatic classification means the automatic: 1) assignment of texts on the basis of their contents to a predefined set of groups which may not be predefined; and, 2) definition of each group. The advantages of automatic classifiers are obvious: 1) considerable savings in terms of both cost and expert manpower; and, 2) domain independence. A text is a sequence of characters or words, representing in the context of survey sampling the answer given in response to an open-ended question in a questionnaire. For example, open-ended question is used to classify units by industry code on the business register. This classification on the business register offers a convenient way for sampling and variance reduction, which is an example of partitioning a set of units into meaningful and useful groups. Even when the survey process is undertaken carefully, the process can be subjective, open to judgment and interpretation, and the results can vary from one human or machine to another. This means that the derived values cannot be determined with certainty, which in turn means that any survey process is fallible. It is thus customary for statistical agencies to both monitor survey process and collect data to evaluate its quality and stability. Although the accuracy of machines rivals that of human professionals, random sampling in combination with human validation still widespread for quality controls. The drawback of this approach is the cost of human power required for validation. Thus, survey process can be very tedious, cost consuming, and the challenge is to maintain a high degree of quality and stability of the survey process with a small validation sample.

Design pre-specification as well as survey process are special cases of measurement error which refers here to the case where the error-prone prior information, say $\boldsymbol{\chi}$, and the error-prone processed information, say $\aleph$, are not necessarily identical to the error-free (or target) information, say $\boldsymbol{\psi}$, of the process underlying the population of interest. We assume that the assessment of error in $\boldsymbol{\chi}$ and the assessment of error in $\aleph$ can be carried out based on observations on $\boldsymbol{\psi}$. We also assume that the error-prone prior information $\boldsymbol{\chi}$ and the error-prone processed information $\aleph$ have a potential bias $\boldsymbol{b}_{\boldsymbol{\chi}}$ and $\boldsymbol{b}_{\aleph}$ respectively when used to estimate $\boldsymbol{\psi}$ and that the error-free information $\boldsymbol{\psi}$ has no error. Thus, the assessment of errors allows quantification of such biases. Under two random processes, we are interested in the error-free random variable $\boldsymbol{\psi}$, knowing its probability function, the probability function of another random variable $(\boldsymbol{\chi}^T, \aleph^T)^T$, together with the joint probability function of $(\boldsymbol{\chi}^T, \aleph^T, \boldsymbol{\psi}^T)^T$ with vector parameter denoted by $\boldsymbol{\lambda}$. It is assumed that the sampling frame has no coverage bias. It is also assumed that values of the error-prone prior information are available for all units in the population, while values of the error-free variable are unknown but observable.

Once an estimate of $\boldsymbol{\lambda}$, of a realization of $\boldsymbol{\psi}$, or of the parameter of interest is obtained, the question follows; what is the reliability of this estimate? In a general sense, reliability of an estimate refers to the degree to which the estimate is free from error and therefore truly measures the parameter that it is intended to measure. When reliability measures are available at all various stages of the survey process, they can serve as performance measures. Such measures enable the survey manager to make decisions regarding the need for methodology-process modification. As there is no general reliability measure that would capture all information on the impact of each stage of the survey design on the ultimate estimate, the survey manager tends to combine various measures to get a broader effect and interactions between different factors of the survey process. A key step in defining reliability was the introduction of an error criterion that measures, in a probabilistic sense, the error between the desired parameter θ and an estimate $\hat{\theta}$ of it. Possible sources of error in surveys include sampling frame, sampling scheme, nonresponse, measurement, editing, imputation, disclosure-avoidance, etc. A criterion which is commonly used in judging the performance of an estimator $\hat{\theta}$ of a parameter θ is its MSE defined by $M(\hat{\theta}) = E\{(\hat{\theta} - \theta)^2\}$. Here, the parameter θ can be seen as the quantity that would be obtained under the ideal situation which consists of census case with complete response and without any processing error. We can also interpret the MSE formula via the MSE decomposition. For any random variable z, we have $E(z^2) = E\{[z - E(z)]^2\} + \{E(z)\}^2$. Applying this to $z = \hat{\theta} - \theta$ we get

$$E\{(\hat{\theta} - \theta)^2\} = E\{[(\hat{\theta} - \theta) - E(\hat{\theta} - \theta)]^2\} + \{E(\hat{\theta} - \theta)\}^2 . \quad (1.2)$$

The first term of (1.2) is the variance of $\hat{\theta}-\theta$. It is the error of the estimator due to the random processes involved. The second term of (1.2) is the square of the bias of $\hat{\theta}$, the best one can do is make this zero. Given that the remaining relative error or the relative missed information about θ based on the knowledge of $\hat{\theta}$ is given by $Var(\theta \mid \hat{\theta})/Var(\theta)$, Demnati(2016) defined the coefficient of reliability as the proportion of knowledge or the proportion of attained information about θ obtained after observing $\hat{\theta}$, i.e.,

$$\mathrm{K}\{\theta;\hat{\theta}\} = 1 - \frac{Var(\theta \mid \hat{\theta})}{Var(\theta)}. \tag{1.3}$$

If $Var(\theta \mid \hat{\theta}) = Var(\theta)$ then $\mathrm{K}\{\theta;\hat{\theta}\} = 0$, and if $Var(\theta \mid \hat{\theta}) = 0$ then $\mathrm{K}\{\theta;\hat{\theta}\} = 1$; so that $0 \le \mathrm{K}\{\theta;\hat{\theta}\} \le 1$. Under the normality assumption, the coefficient of reliability (1.3) reduces to the square of the correlation coefficient

$$\mathrm{K}_N\{\theta;\hat{\theta}\} = \left(\frac{Cov(\theta,\hat{\theta})}{\sigma_\theta \sigma_{\hat{\theta}}}\right)^2 \equiv \rho^2_{\theta\hat{\theta}}. \tag{1.4}$$

Tenenbein (1970) introduced the square of the correlation coefficient given by (1.4) as a measure of reliability between the error-prone and error-free classification variables to measure the strength of the relationship between the true and fallible classifications; i.e. it measures how well the true classification can be predicted from the fallible classification on a given unit. Expression (1.4) gives a convenient way to compute the coefficient of reliability: It is reasonable in practice to replace conditional variance, which depends on the joint distribution, with correlation which can be calculated more easily. That being said, conditional independence is more meaningful and preferable than zero-correlation.

In an attempt to discuss side effect reduction of both errors sources of information on the quality of the estimator of the parameter of interest during survey data collection period, Sections 2 and 3 consider the case of no processing errors; while sections 4 and 5 consider the case with the inclusion of processing errors. In detail our work below is organized as follows: in Section 2, we study steps required for design pre-specification when designing a survey in the absence of processing errors; in Section 3, we update design pre-specification after observing some realizations of the target process; in Section 4, we study the extra steps required for design pre-specification in the presence of processing errors; and, in Section 5, we update design pre-specification after observing some realization of the target process.

2. Design Pre-specification under No Processing Errors

There will be a survey to be conducted on a population of size N, where the sampling frame had covariate u on it; an address and telephone information. The initial request for response is by Web

or mail. The main interest is to estimate a domain total of the variable of interest y, where the sampled units are to be classified to the domain of interest based on their response to an open-ended question in the questionnaire. The upper limit on the coefficient of variation of the estimator is set to 0.05. The budget is constrained to a global cost of $C_{\max}$, while the maximum duration of the survey data collection is constrained to $\mathrm{I}_{\max}$ time periods. After I_F time periods of data collection under self-enumeration, there will be an optional follow-up for those who had not responded. The duration of a follow-up is D_F periods of time. Poisson sampling is to be used for the selection of the main sample. Known values of survey design parameters such as N, $C_{\max}$, $\mathrm{I}_{\max}$, I_F, and D_F are provided in Table 1. The first task is to pre-specify the design that better enhance the quality of the estimator while respecting the survey design constraints. In particular, we have to: 1) derive steps required for design pre-specification; 2) present briefly available prior information; and finally, 3) determine the resources and their allocation within stages of the survey design.

We divide the continuous time of the entire survey process period into a sequence of continuous time periods: 1, 2, and so on, and let $\mathrm{I}_{\min}$ denote the minimum length of data collection period to obtain full responses on the target information. Suppose that the survey limited length of duration of data collection is made up of $\mathrm{I}_{\max}$ time periods; or equivalently $\mathrm{P}_{\max}$ phases. The p^{th} phase being of size $n_{\mathrm{p}}(\geq 1)$ time periods, so that the limited duration of data collection is made up of $\mathrm{I}_{\max} = \sum_{\mathrm{p}=1}^{\mathrm{P}_{\max}} n_{\mathrm{p}}$ time periods, with $\mathrm{I}_{\max} < \mathrm{I}_{\min}$. We then shall be dealing with $N \times \mathrm{P}_{\max}$ rectangular array of phases of data collection.

2.1 Specification Step

2.1.1 Parameter of Interest

Estimate is wanted for specific subpopulation κ, called domain κ. The methodology behind estimating parameters for domains, based on observation of randomly selected units, is well described by survey literature. See for example Cochran (1977); or, Särndal *et al*. (1992). Let the specific subpopulation κ of units of interest or domain κ be denoted as P_κ, and let $l_{\kappa;k} = 1_k(P_\kappa \mid P)$ be the domain κ membership indicator variable for unit k, where $1_k(\Omega \mid \Omega^*) = 1(k \in \Omega \mid k \in \Omega^*)$ is the set Ω membership indicator variable for unit k given that unit k belongs to the set Ω^*, $1(condition)$ is the truth function, i.e., $1(condition) = 1$ if the *condition* is true, $1(condition) = 0$ if not. The domain total Y_κ of a characteristic y may be written as

$$Y_\kappa = \sum_k l_{\kappa;k} y_k \equiv \sum_k \dot{y}_{\kappa;k}, \tag{2.1}$$

where $\sum_k$ denotes sum overall population units, $\boldsymbol{y} = (y_1,...,y_N)^T$ is the vector of values of the characteristic of interest y, and $\dot{y}_{\kappa;k} = l_{\kappa;k} y_k$. The parameter Y_κ, obtained under the assumed ideal situation which consists of census case with complete response and without any processing error, plays the role of a "gold standard".

2.1.2 Response Mechanism

We now give a brief account of the Demnati modeling approach of the response indicators as discrete-time hazard. See for example Demnati (2017). Let t represent the discrete random variable that indicates the time period i when the response occurs for a randomly selected unit from the sample. We assume that every unit in the sample lives through each successive discrete time period until the unit responds or is censored by the end of data collection. Then each unit k is observed until some period I_k, with $\mathrm{I}_k \le \mathrm{I}_{\max}$. Observation of the unit could be discontinued for two reasons: 1) the unit response; or, 2) the survey data collection period ends. In the first case, $t_k = \mathrm{I}_k$. In the second case, we only know that $t_k > \mathrm{I}_{\max}$. Units with $t_k > \mathrm{I}_{\max}$ are right-censored – when they respond is unknown. Because response occurrence is intrinsically conditional, we characterized t by its conditional probability function – the distribution of the probability that a response will occur in each time period given that it has not already occurred in a previous time period – known as the discrete-time hazard function. Discrete-time hazard $h_{ki}(\boldsymbol{x}_{ki}, \boldsymbol{\beta})$, h_{ki} for short, is defined as the conditional probability that unit k will respond in time period i, given that the unit did not respond prior to i:

$$h_{ki} = \Pr(t_k = i \mid t_k \ge i),$$

where $\boldsymbol{x}_{ki}$ refers to both time-invariant and time-varying explanatory variables and $\boldsymbol{\beta}$ is the unknown $q_r \times 1$ vector parameter to be estimated. For unit with $t_k = i$, the probability of obtaining a response at time period i could be expressed in terms of the hazard as

$$\Pr(t_k = i) = h_{ki} \prod_{j=1}^{i-1} (1 - h_{kj}).$$

For units with $t_k > i$, the probability of obtaining a response can be expressed as

$$\Pr(t_k > i) = \prod_{j=1}^{i} (1 - h_{kj}).$$

The marginal probability of obtaining a response after $\mathrm{I}_{\max}$ time periods is given by

$$\xi_k = 1 - \prod_{i=1}^{\mathrm{I}_{\max}} (1 - h_{ki}) = \sum_{i=1}^{\mathrm{I}_{\max}} \Pr(t_k = i).$$

2.1.3 Estimator To Be Used

Suppose that the response probabilities $\xi_{\mathrm{I};k}$ after I time periods of data collection are known for all population units. Then for general sampling design with known positive inclusion probabilities, $\pi_{\mathrm{I};k}$, an unbiased estimator of the domain total Y_κ given by (2.1) after I time periods of data collection is given by

$$\breve{Y}_{\mathrm{I};\kappa} = \sum_k d_{\mathrm{I};k}(r_{\mathrm{I};k} / \xi_{\mathrm{I};k})\dot{y}_{\kappa;k}, \tag{2.2}$$

where $d_{\mathrm{I};k} = d_k(\wp_{\mathrm{I}} \mid P) = 1_k(\wp_{\mathrm{I}} \mid P)/\pi_{\mathrm{I};k}$ are the design weights associated with the random sample $\wp_{\mathrm{I}}$ obtained after I time periods of data collection, $\pi_{\mathrm{I};k} = \pi_k(\wp_{\mathrm{I}} \mid P)$, $\pi_k(\Omega \mid \Omega^*) = E_\Omega\{1_k(\Omega \mid k \in \Omega^*)\}$ is the set Ω inclusion probability for unit k given $k \in \Omega^*$, and E_Ω denotes expectation with respect to the inclusion mechanism.

2.1.4 Derivation of the Variance Function

We may decompose the variance of $\breve{Y}_{\mathrm{I};\kappa}$ given by (2.2) as

$$Var(\breve{Y}_{\mathrm{I};\kappa}) = E_m E_\wp Var_r(\breve{Y}_{\mathrm{I};\kappa}) + E_m Var_\wp E_r(\breve{Y}_{\mathrm{I};\kappa}) + Var_m E_\wp E_r(\breve{Y}_{\mathrm{I};\kappa}) \equiv \mathrm{V}_r + \mathrm{V}_\wp + \mathrm{V}_m \equiv \mathrm{V}_{wope} \tag{2.3}$$

where $Var_\wp$, Var_r and Var_m denote variance with respect to the sampling design, the response mechanism and the model on $\dot{y}$ respectively, and the subscript "$wope$" in V_{wope} stands for "without processing error". Under independent mechanism on $r_{\mathrm{I};k}$, the first component $\mathrm{V}_r = E_m E_\wp Var_r\{\sum_k d_{\mathrm{I};k}(r_{\mathrm{I};k} / \xi_{\mathrm{I};k})\dot{y}_{\kappa;k}\}$ of (2.3) is given by

$$\mathrm{V}_r = \sum_k E_m(\dot{y}^2_{\kappa;k})(1-\xi_{\mathrm{I};k})/(\pi_{\mathrm{I};k}\xi_{\mathrm{I};k}). \tag{2.4}$$

Under Poisson sampling, the second component $\mathrm{V}_\wp = E_m Var_\wp(\sum_k d_{\mathrm{I};k}\dot{y}_{\kappa;k})$ of (2.3) is given by

$$\mathrm{V}_\wp = \sum_k E_m(\dot{y}^2_{\kappa;k})(1-\pi_{\mathrm{I};k})/\pi_{\mathrm{I};k}. \tag{2.5}$$

Finally, under independent model mechanisms on $\dot{y}_k$, the last component of (2.3) is given by

$$\mathrm{V}_m = \sum_k Var_m(\dot{y}_{\kappa;k}). \tag{2.6}$$

The sum of (2.4), (2.5), and (2.6) constitutes $\mathrm{V}_{wope} = \mathrm{V}_r + \mathrm{V}_\wp + \mathrm{V}_m$, the variance of $\breve{Y}_{\mathrm{I};\kappa}$ given by (2.2). It follows that, we can express $Var(\breve{Y}_{\mathrm{I};\kappa})$ as

$$Var(\breve{Y}_{\mathrm{I};\kappa}) = v_{wope;0} + \sum_h v_{wope;k}/(\pi_{\mathrm{I};k}\xi_{\mathrm{I};k}),$$

where $v_{wope;0} = -\sum_k\{E_m(\dot{y}_{\kappa;k})\}^2$, and $v_{wope;k} = E(\dot{y}^2_{\kappa;k})$.

2.1.5 Specification of the Follow-up Activity

We define the nonresponse follow-up indicator variables as $l_{f;k}^{(\mathrm{p})}=1$ if unit k is assigned to the follow-up activity at phase p, and $l_{f;k}^{(\mathrm{p})}=0$ if not, where $l_{f;k}^{(\mathrm{p})}$ are realizations of independent distributed variables according to a Bernoulli distribution, $B(\phi_{f;k}^{(\mathrm{p})})$, $\phi_{f;k}^{(\mathrm{p})}$ is the probability of a follow-up, and the subscript "f" stands for "follow-up". The follow-up probability is constructed as

$$\log\{\phi_{f;k}^{(\mathrm{p})}/(1-\phi_{f;k}^{(\mathrm{p})})\}=\boldsymbol{v}_{f;k}^{(\mathrm{p})T}\boldsymbol{\lambda}_{f}^{(\mathrm{p})},$$

where $\boldsymbol{v}_{f;k}^{(\mathrm{p})}=(1,\dot{y}_{\kappa;k})^{T}$ is the vector predictor and $\boldsymbol{\lambda}_{f}^{(\mathrm{p})}=(\lambda_{f;0}^{(\mathrm{p})},\lambda_{f;1}^{(\mathrm{p})})^{T}$ is the unknown vector parameter to be determined.

2.1.6 Specification of the Initial Cost Function

We may decompose the initial global cost over I time periods of data collection as

$$C_{wope}=C_{\mathrm{I}}+C_{\wp}+C_{f}+C_{dc}.$$

The fixed cost C_{I} is given by $C_{\mathrm{I}}=c\times\mathrm{I}$, where c is a fixed cost per time period. The sampling component $C_{\wp}$ is given by $C_{\wp}=\sum_{k}1_{k}(\wp_{\mathrm{I}}\mid P)c_{\wp;k}$, where $c_{\wp;k}$ is the sampling cost for unit k. The follow-up component C_{f} is given by $C_{f}=\sum_{k}1_{k}(\wp_{\mathrm{I}}\mid P)(1-r_{e_k;k}^{(self)})l_{f;k}^{(1)}c_{f;k}$, where $r_{i;k}^{(self)}$ represents the response indicator under self-enumeration over i time periods of data collection, e_k is the follow-up entry time period for unit k, and $c_{f;k}$ is the follow-up cost for unit k. The data collection cost C_{dc} is given by $C_{dc}=\sum_{k}1_{k}(\wp_{\mathrm{I}}\mid P)\{r_{\mathrm{I};k}^{(M)}c_{dc;k}^{(M)}+r_{\mathrm{I};k}^{(W)}c_{dc;k}^{(W)}\}$, where the superscripts "M" and "W" stand for "Mail" and "Web" respectively, and $c_{dc;k}^{(m)}$ is the data collection cost associated with mode $m\in\{M,W\}$.

2.1.7 Modeling the Sample Selection Probabilities

The conditional probability that unit k will be selected in phase p, given that the unit was not selected prior to p is constructed as

$$\log\{\pi_{\wp;k}^{(\mathrm{p})}/(1-\pi_{\wp;k}^{(\mathrm{p})})\}=\boldsymbol{v}_{\wp;k}^{(\mathrm{p})T}\boldsymbol{\lambda}_{\wp}^{(\mathrm{p})},$$

where $\boldsymbol{v}_{\wp;k}^{(\mathrm{p})}=(1,l_{k})^{T}$ is the vector predictor and $\boldsymbol{\lambda}_{\wp}^{(\mathrm{p})}=(\lambda_{\wp;0}^{(\mathrm{p})},\lambda_{\wp;1}^{(\mathrm{p})})^{T}$ is the unknown vector parameter to be determined.

2.1.8 Specification of the Initial Objective Function

To create a design, we determine the number of time periods I of data collection (or equivalently the number of phases P with $\mathrm{I}=\sum_{\mathrm{p}=1}^{\mathrm{P}} n_{\mathrm{p}}$), the initial samples selection parameter $\boldsymbol{\lambda}_{\varphi}^{(1)}$, and the initial follow-up model parameter $\boldsymbol{\lambda}_{f}^{(1)}$ by minimizing the variance, $\min_{\boldsymbol{a}} Var(\breve{Y}_{\mathrm{I};\kappa})$, subject to constraint on the expected cost, $\overline{C}_{wope} \le C_{\max}$, and constraint on the duration $1 \le \mathrm{I} \le \mathrm{I}_{\max}$, where $\boldsymbol{a}=(\mathrm{I},\boldsymbol{\lambda}_{\varphi}^{(1)T},\boldsymbol{\lambda}_{f}^{(1)T})^{T}$, $\overline{C}_{wope}=C_{\mathrm{I}}+\overline{C}_{\varphi}+\overline{C}_{f}+\overline{C}_{dc}$, $\overline{C}_{\varphi}=\sum_{k}\pi_{\mathrm{I};k}c_{\varphi;k}$, $\overline{C}_{f}=\sum_{k}\pi_{\mathrm{I};k}(1-\xi_{e_k;k}^{(self)})\phi_{f;k}^{(1)}c_{f;k}$, $\overline{C}_{dc}=\sum_{k}\pi_{\mathrm{I};k}\{\xi_{\mathrm{I};k}^{(M)}c_{dc;k}^{(M)}+\xi_{\mathrm{I};k}^{(W)}c_{dc;k}^{(W)}\}$, and $\pi_{\mathrm{I};k}=\pi_{\varphi}^{(1)}$. In this case a Lagrange multiplier can be use to find the constraint minimum of the variance. So then, the objective function is given by

$$G(\boldsymbol{a})=\sum_{k} v_{wo;k}/(\pi_{\mathrm{I};k}\xi_{\mathrm{I};k})+\varsigma(\overline{C}_{wope}-C_{\max}), \tag{2.7}$$

where ς is the Lagrange multiplier. The optimization problem obtains a constrained minimum at the point where the estimating equations (EE) are set to zero, $g(\boldsymbol{a})=\partial G(\boldsymbol{a})/\partial\boldsymbol{a}=\mathbf{0}$. Kokan (1963) discussed similar allocation problem extensively under stratified simple random sampling and showed how it can be adapted to cover many common sample allocations. We have used the concept of EE to define a set of simultaneous equations involving both the data and the unknown parameter which are to be solved in order define the estimate of the parameter. This concept of EE is more general than the concept of estimating functions having zero mean for the k^{th} component at the true parameter which includes the log-likelihood estimating functions as well as least square estimating functions.

We do not have an explicit solution, but nonlinear programming can be used to get a constraint minimum $\boldsymbol{a}\equiv\boldsymbol{a}(\boldsymbol{\psi}_k)$, where $\boldsymbol{\psi}_k=(y_k,l_{\kappa;k},\xi_{i;k}^{(m)},e_k,\boldsymbol{c}_k^{T};\boldsymbol{\lambda}_{\psi})^{T}$, $\boldsymbol{\lambda}_{\psi}=(\boldsymbol{\lambda}_{y}^{T},\boldsymbol{\lambda}_{l}^{T},\boldsymbol{\lambda}_{r}^{T},\boldsymbol{\lambda}_{c}^{T})^{T}$, $\boldsymbol{\lambda}_{y}$ is the vector parameter associated with the model on y, $\boldsymbol{\lambda}_{l}$ is the vector parameter associated with the classification model on l_{κ}, $\boldsymbol{\lambda}_{r}$ is the vector parameter associated with the response model, and $\boldsymbol{\lambda}_{c}$ is the vector parameter associated with the cost model on $\boldsymbol{c}_k=(c_{\varphi;k},c_{f;k},\boldsymbol{c}_{dc;k}^{T})^{T}$. The first two components y_k and $l_{\kappa;k}$ of $\boldsymbol{\psi}_k$ are referred as data in adaptive design literature, while the rest of the vector $\boldsymbol{\psi}_k$ is referred as paradata.

2.2 Prediction Step

It is clear from (2.7) that the optimization problem cannot be performed since $\boldsymbol{\psi}_k$ are unknown. From the sampling frame, the variable u is used to approximate y. From the modeling of previous surveys, it was possible to assign to each unit k in the sampling frame: 1) an initial estimated probability ${}^{(pri)}p_k$ of being a member of the domain of interest κ given the covariate; 2) an initial estimated conditional probability ${}^{(pri)}h_{ki}^{(m)}$ of responding by mode m and by time period i; and, 3)

a follow-up entry time period ${}^{(pri)}e_k$ generated from the uniform interval $[\mathrm{I}_1+1, \mathrm{I}_{\max}-D_F-2]$. So that the vector of available prior information for each unit k in the frame is $\boldsymbol{\chi}_k=(u_k, {}^{(pri)}p_k, {}^{(pri)}h_{ki}^{(m)}, {}^{(pri)}e_k, {}^{(pri)}\boldsymbol{c}_k^T; \boldsymbol{\lambda}_{\chi})^T$, $i=1,...,\mathrm{I}_{\max}$. Hence the estimator used for design pre-specification is

$${}^{(1)}\breve{Y}_{\mathrm{I};\kappa}=\sum_k d_{\mathrm{I};k}({}^{(1)}r_{\mathrm{I};k}/{}^{(1)}\xi_{\mathrm{I};k}){}^{(1)}l_{\kappa;k}{}^{(1)}y_k,$$

with ${}^{(1)}l_{\kappa;k}={}^{(pri)}l_{\kappa;k}$, ${}^{(1)}y_k=u_k$, ${}^{(1)}\xi_{\mathrm{I};k}={}^{(pri)}\xi_{\mathrm{I};k}$, $v_{wope;0}=-\sum_k u_k^{2\,(pri)}p_k^2$, and $v_{wope;k}=u_k^{2\,(pri)}p_k$ are the components of the variance under the assumptions that u_k are constants, where $\xi_{\mathrm{I};k}=(1-\phi_{f;k})\xi_{\mathrm{I};k}^{(self)}+\phi_{f;k}\xi_{\mathrm{I};k}^{(self+f)}$, $\phi_{f;k}=\phi_{f;k}^{(1)}$, and $\xi_{\mathrm{I};k}^{(self+f)}$ is the probability of response under follow-up in addition to self-enumeration for unit k during I time periods of data collection. Table 2 gives the initial estimate of the size of the domain of interest and its total of u, while Table 3 displays the response rates for different durations of data collection under the prior response model parameter.

Table 1. Prior Information on the Survey

N	$\mathrm{I}_{\max}$	I_F	$C_{\max}$	c	c_p	c_f	$c_{dc}^{(M)}$	$c_{dc}^{(I)}$	D_F
5000	40	3	5000	20	1	3	2	1	3

Table 2: Prior Information on the Domain of Interest

Domain Size	**Domain Total**
2 495	46 238

Table 3: % of Expected Number of Respondents based on Prior Information

Duration	**Self-enumeration Only**			**With Follow-up**		
	Mail	**Internet**	**Both**	**Mail**	**Internet**	**Both**
5	5	5	10	6	5	11
10	9	9	18	20	12	32
15	12	12	25	32	18	50
20	15	15	31	42	23	65
25	18	18	35	51	26	77
30	20	20	40	59	30	89
35	22	21	43	66	32	98
40	24	23	47	67	33	100

2.3 Optimization Step

Using the error-prone prior information as input to the optimization problem, Table 4 displays the values of the design parameters: the expected sample size, the expected number of follow-ups, the expected number of respondents, and the expected coefficient of variation in percentage. Table 4

also displays the expected ratios in percentage for the fix cost, the sampling cost, the follow-up cost and the data collection cost. Finally for more information, Table 4 displays estimates of regression parameters $\boldsymbol{\lambda}_{\varphi}^{(1)}$ and $\boldsymbol{\lambda}_{f}^{(1)}$.

Table 4: Resources Allocation Based on Prior Information

Duration	Expected					% Cost Ratio				Regression Parameter Estimates			
	Sample Size	#Follow-up	# Respondents	CV	Total Cost	Fixed	Sampling	Follow-up	Data Collection	Sampling		Follow-up	
										$\lambda_{\varphi;0}^{(1)}$	$\lambda_{\varphi;1}^{(1)}$	$\lambda_{f;0}^{(1)}$	$\lambda_{f;1}^{(1)}$
5	4174		413	8	5000	2	83	0	15	1.62	.001	-.27.46	.47
10	3205		578	6	5000	4	64	0	32	.58	.000	-21.22	.33
15	2613		649	6	5000	6	52	0	42	.09	..001	-24.03	.39
20	2235		683	6	5000	8	45	0	47	-.21	.000	-148.5	4.12
25	1734		720	6	5000	10	35	11	44	-.63	.001	-1.77	.00
30	952		834	6	5000	12	19	41	28	-1.43	.005	523	5.56
35	894	894	877	5	5000	14	18	39	29	-1.52	.000	11.87	4.28
40	868	867	867	5	5000	16	17	38	29	-1.56	.013	7.89	19.12

Note: The required coefficient of variation (cv) of .05 is reached only for $I \geq 10$.

3. Adaptive Design

It was decided to proceed with $I = 35$, $\boldsymbol{\lambda}_{\varphi}^{(1)} = (-1.525,.002)^T$ as the regression parameter for sampling, and $\boldsymbol{\lambda}_{f}^{(1)} = (11.87,.4.28)^T$ as the regression parameter for nonresponse follow-up. We computed descriptive statistics on the observed data and on the predicted data based on the prior information, after observing realisations over 10 time periods of data collection. Table 5 displays the realized sample size, the number of respondents, and the number of follow-ups. While Table 6 displays the distribution of the cost. Respondents and nonrespondents are represented by Table 7, while Table 8 displays the classifications of the respondents. These tables show that the first phase, composed of 10 time periods, goes better than predicted in the selected sample. Using only 147 follow-ups instead of the predicted number of 165, the number of respondents improved from the predicted 284 to the observed 389 (Table 5). This improvement in the number of respondents, increased the data collection cost from the predicted cost of 452 to the cost spent of 682 (Table 6). Tables 7 and 8 show clearly that there is error in prior information (e.g. domain classification and response behavior). The question is now to decide whether proceeding with the pre-specified design is a good idea, or whether the eventual efficiency of the estimator would be better enhanced by updating the design parameter. We first give in section 3.1 a brief description of the Demnati (2016)' adaptive approach to revise the survey design during its progress. In Section 3.2 we update

the classification model to illustrate the revision process, and in section 3.3 we revise the design parameter using the updated information for the remaining time periods of data collection.

Table 5 : Observed Counts After 10 Time Periods of Data Collection

Sample Size		878
Observed Counts	**# Respondents**	389
	# Follow-up	147
Predicted Counts (Based on Prior Information	**# Respondents**	284
	# Follow-up	165

Table 6 : Observed Costs After 10 Time Periods of Data Collection

Observed Costs	**Sampling**	878
	Follow-up	441
	Data Collection	682
	M	586
	W	96
	Fixed	200
	Total	2201
Predicted Costs (Based on Prior Information	**Sampling**	878
	Follow-up	495
	Data Collection	452
	M	336
	W	116
	Fixed	200
	Total	2025

Table 7 : Counts of Respondents and Nonrespondents after 10 Time Periods of Data Collection

		Observed Information		**Total**
		Respondents	**Non Respondents**	
Prior Information	**Respondents**	196	88	284
	Nonrespondents	193	401	594
Total		389	489	878

Table 8 : Respondents Classification after 10 Time Periods of Data Collection

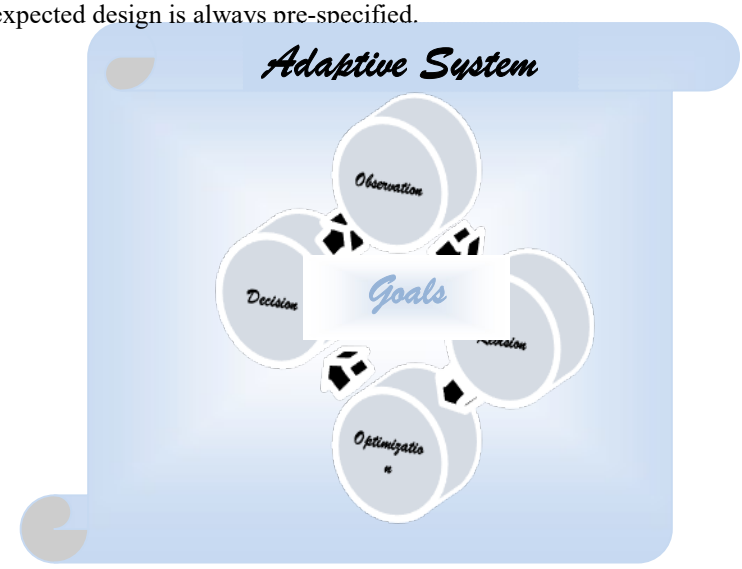
		Observed Information		**Total**
		In Domain	**Outside Domain**	
Prior Information	**In Domain**	110	63	173
	Outside Domain	106	110	216
Total		216	173	389

Table 9 : Domain Estimation after 10 Time Periods of Data Collection		
	Estimation based on	
	Prior Values	Observed Values (%estimated cv)
Domain Size	2 127	2 476 (7.38)
Domain Total	38 789	142 089 (7.47)

3.1 Demnati Adaptive Method

We now give a brief account of the Demnati (2016) approach to revise the survey design during its progress. The method begins with a pre-specified design based on prior information then design revision consists on: 1) accumulating observations on the target process governing the survey under consideration from the sampled respondent; 2) updating information used for design specification; 3) determining the design parameter for the remaining time periods of data collection using the updated information; and, 4) revising, if necessary, specification of the design. After completing the fourth step, the stopping rules are consulted to see if the data collection should stop. If not, decide if the design parameter should change; and repeat the four steps continuously. The above four steps are termed as the Observation-Revision-Optimization-Decision (O-R-O-D) steps. The second step incorporates learning and prediction, while the fourth step incorporates actioning.

While most existing literature on adaptive designs faces nonresponse using a traditional approaches organized in 2 or 3 phases during which the design is extant (Groves and Heeringa 2006), our approach is embedded in a continuing learning process that permits changes in methodology-process at any time of data collection period as a result of an increase in acquiring information and facts, while relating phases and stages of the design to each other. Such changes are guided by the primary survey objective. The method does not, therefore, use a fixed design, although an expected design is always pre-specified.



- **Observation Step:** Obtain next phase p of observations on the error-free process.

- **Revision Step:**
  - **Estimation/Imputation**: 1) Update $\boldsymbol{\lambda}_{\mathrm{p}-1}$ to get $\boldsymbol{\lambda}_{\mathrm{p}}$ using $\mathbf{d}_{\mathrm{o;p}}$; and, 2) Impute missing values of each component ψ of $\boldsymbol{\psi}$ to get $\psi_{\mathrm{p};k} = E_{\psi}(\psi_k \mid \mathbf{d}_{\mathrm{o;p}}, \boldsymbol{\lambda}_{\mathrm{p}})$, where $\mathbf{d}_{\mathrm{o;p}}$ denotes all observed information until the end of phase p of data collection, and E_{ψ} denotes expectation with respect to the random process governing the component ψ. Note that $\psi_{\mathrm{p};k} = \psi_k$ when item ψ_k is observed.
- **Optimization Step:**
  - Determine the optimal design parameter $\boldsymbol{a} = (\mathrm{P}, \boldsymbol{\lambda}_{\wp}^{(\mathrm{p})T}, \boldsymbol{\lambda}_{f}^{(\mathrm{p})T})^T$ conditional on $\boldsymbol{\psi}_{\mathrm{p}}$ and $\boldsymbol{\lambda}_{\mathrm{p}}$. The solution is denoted by $\boldsymbol{a}_{\mathrm{p}} = (\mathrm{P}_{\mathrm{p}}, \boldsymbol{\lambda}_{\wp}^{(\mathrm{p})T}, \boldsymbol{\lambda}_{f}^{(\mathrm{p})T})^T$.
- **Decision Step:**
  - Decide if the data collection should stop (i.e., $\mathrm{p} = \mathrm{P}_{\mathrm{p}}$), if not, decide if the methodology-process should change, and then repeat the four steps continuously after observing some realizations of the target process.

.

Depending on the relationship between the error-prone prior information, the target information; and, the stopping rules, only a few time periods may be sufficient to stop the data collection period.

3.2 Revision of the Classification Model Parameter

We now give an example of the revision process in the Revision Step. The effects of misclassification in categorical data on estimators have been discussed for some time by Bross (1954) and others. Tenenbein (1970, 1972) proposed two-phase sampling to protect against error, assuming that error-free classification is possible to obtained, though it is expensive. Misclassification assumed that two measuring devices are available to classify units into one of numerous mutually exclusive groups. The first device is a cheaper procedure which tends to misclassify units; the second device is an expensive procedure which classifies units correctly.

The domain-classification for unit k is characterized by the matrix $\mathbf{P}_k$ who depends on two conditional probabilities: $p_k^{(1|1)} = \Pr({}^{(pri)}l_{\kappa;k} = 1 \mid l_{\kappa;k} = 1)$ which consists of the probability of classifying the domain of interest given that unit belongs truly to the domain of interest, and $p_k^{(1|0)} = \Pr({}^{(pri)}l_{\kappa;k} = 1 \mid l_{\kappa;k} = 0)$ which consists of the probability of classifying the domain of interest given that unit do not belongs truly to the domain of interest. Hence

$$\mathbf{P}_k = \begin{pmatrix} 1-p_k^{(1|0)} & 1-p_k^{(1|1)} \\ p_k^{(1|0)} & p_k^{(1|1)} \end{pmatrix}.$$

The marginal distributions of ${}^{(pri)}p_k$ is given by

$${}^{(pri)}p_k = p_k^{(1|1)}p_k + p_k^{(1|0)}(1-p_k).$$

The census parameter $\boldsymbol{\lambda}_{\mathrm{I}_{\min}}$ is defined as the solution of

$$\mathbf{S}_{\mathrm{I}_{\min}}(\boldsymbol{\lambda}) = \sum_k \partial \log f_{\mathrm{I}_{\min}}({}^{(pri)}l_{\kappa;k}, l_{\kappa;k})/\partial\boldsymbol{\lambda} = \mathbf{0}, \tag{3.1}$$

where the subscript I in $f_{\mathrm{I}}(\boldsymbol{\zeta})$ denotes that $\boldsymbol{\zeta}$ is observed during the interval $[0,\mathrm{I}]$. Hence when the data collection period is considered, the census case means that $\mathrm{I} = \mathrm{I}_{\min}$ and $d_k(\wp) = 1_k(\wp) = \pi_k = 1$. The solution to (3.1) obtained by Newton-Raphson-type iterative method or the Expectation Maximization (EM) algorithm gives the census parameter $\boldsymbol{\lambda}_{\mathrm{I}_{\min}}$ associated with $\boldsymbol{\lambda}$. The census parameter $\boldsymbol{\lambda}_{\mathrm{I}_{\min}}$, obtained under the assumed ideal situation which consists of census case with complete response and without any processing error, plays the role of a "gold standard".

After observing I time periods of data collection, Demnati (2016) decomposed the joint distribution for unit k in two parts

$$f_{\mathrm{I}}({}^{(pri)}l_{\kappa;k}, l_{\kappa;k}) = f_{\mathrm{I}_{\min}}({}^{(pri)}l_{\kappa;k}) f_{\mathrm{I}}(l_{\kappa;k}|{}^{(pri)}l_{\kappa;k}). \tag{3.2}$$

Note that $f_{\mathrm{I}_{\min}}({}^{(pri)}l_{\kappa;k}) f_{\mathrm{I}}(l_{\kappa;k}|{}^{(pri)}l_{\kappa;k}) \to f_{\mathrm{I}_{\min}}(l_{\kappa;k}, {}^{(pri)}l_{\kappa;k})$ as $\mathrm{I} \to \mathrm{I}_{\min}$. We may write $f_{\mathrm{I}}(l_{\kappa;k}|{}^{(pri)}l_{\kappa;k})$ as $f_{\mathrm{I}}(l_{\kappa;k}|{}^{(pri)}l_{\kappa;k}) = f_{\mathrm{I}}(l_{\kappa;k}, {}^{(pri)}l_{\kappa;k})/f_{\mathrm{I}}({}^{(pri)}l_{\kappa;k})$. The log-likelihood for unit k is given by

$$\ell_{\mathrm{I};k}(\boldsymbol{\lambda}) = \log f_{\mathrm{I}_{\min}}({}^{(pri)}l_{\kappa;k}) + \log f_{\mathrm{I}}({}^{(pri)}l_{\kappa;k}\,|\,l_{\kappa;k}), \tag{3.3}$$

where

$$\log f_{\mathrm{I}_{\min}}({}^{(pri)}l_{\kappa;k}) = {}^{(pri)}l_{\kappa;k}\log({}^{(pri)}p_k) + (1-{}^{(pri)}l_{\kappa;k})\log(1-{}^{(pri)}p_k),$$

$$\begin{aligned}\log f_{\mathrm{I}}(l_{\kappa;k}|{}^{(pri)}l_{\kappa;k}) = {}&{}^{(pri)}l_{\kappa;k}\{l_{\kappa;k}\log(\tau_k^{(1|1)}) + (1-l_{\kappa;k})\log(1-\tau_k^{(1|1)})\} \\ &+ (1-{}^{(pri)}l_{\kappa;k})\{l_{\kappa;k}\log(\tau_k^{(1|0)}) + (1-l_{\kappa;k})\log(1-\tau_k^{(1|0)})\},\end{aligned}$$

$\tau_k^{(1|1)} = p_k^{(1|1)}p_k/{}^{(pri)}p_k$ is the conditional probability that unit k belongs really to the domain of interest given that the unit is classified into the domain of interest, and $\tau_k^{(1|0)} = (1-p_k^{(1|1)})p_k/(1-{}^{(pri)}p_k)$ is the conditional probability that unit k belongs really to the domain of interest given that the unit is not classified into the domain of interest. Taking the derivatives of (3.3) and adjusting for unequal probability of selection and for the response mechanism, we get the weighted EE

$$\hat{\mathbf{S}}_{\mathrm{I}}(\boldsymbol{\lambda}; l_{\kappa;k}, {}^{(pri)}l_{\kappa;k}) = \sum_k \{\mathbf{s}_{\mathrm{I}_{\min};k}(\boldsymbol{\lambda}; {}^{(pri)}l_{\kappa;k}) + d_k(\wp)(r_{\mathrm{I};k}/\xi_{\mathrm{I};k})\mathbf{s}_{\mathrm{I};k}(\boldsymbol{\lambda}; l_{\kappa;k}|{}^{(pri)}l_{\kappa;k})\} = \mathbf{0}, \tag{3.4}$$

where $\mathbf{s}_{\mathrm{I}_{inax};k}(\boldsymbol{\lambda}; {}^{(pri)}l_{\kappa;k}) = \partial \log f_{\mathrm{I}_{\min}}({}^{(pri)}l_{\kappa;k})/\partial\boldsymbol{\lambda}$ and $\mathbf{s}_{\mathrm{I};k}(\boldsymbol{\lambda}; l_{\kappa;k}|{}^{(pri)}l_{\kappa;k}) = \partial \log f_{\mathrm{I}}(l_{\kappa;k}|{}^{(pri)}l_{\kappa;k})/\partial\boldsymbol{\lambda}$.

Note that $E\{\hat{\mathbf{S}}_{\mathrm{I}}(\boldsymbol{\lambda}; l_{\kappa;k}, {}^{(pri)}l_{\kappa;k})\} = \mathbf{S}_{\mathrm{I}_{\min}}(\boldsymbol{\lambda})$ the censuss EE given by (3.1). The estimator $\hat{\boldsymbol{\lambda}}_{\mathrm{I}}$ of $\boldsymbol{\lambda}$ is obtained using the following update step.

Update Step of $\boldsymbol{\lambda}$: Starting with a guessed value, $\boldsymbol{\lambda}^{(0)} = \boldsymbol{\lambda}_{\text{I-1}}$, then for $b = 1,2,...$ updates are made using

$$\boldsymbol{\lambda}^{(b)} = \boldsymbol{\lambda}^{(b-1)} + \{\hat{\mathbf{J}}_{\text{I}}(\boldsymbol{\lambda}^{(b-1)})\}^{-1}\hat{\mathbf{S}}_{\text{I}}(\boldsymbol{\lambda}^{(b-1)}; l_{\kappa;k}, {}^{(pri)}l_{\kappa;k}),$$

where $\hat{\mathbf{J}}_{\text{I}}(\boldsymbol{\lambda}) = -\partial\hat{\mathbf{S}}_{\text{I}}^{T}(\boldsymbol{\lambda}; l_{\kappa;k}, {}^{(pri)}l_{\kappa;k})/\partial\boldsymbol{\lambda}$.

One may use the EM algorithm to derive the maximum likelihood estimate of $\boldsymbol{\lambda}$. The EM algorithm introduced by Hartley (1958)—formalized and termed by Dempster et al. (1977) — has become a major tool for finding maximum likelihood estimates in situations considered practically intractable such as missing data. Let $\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_N$ be independent identically distributed random variables from a distribution indexed by an unknown parameter $\boldsymbol{\lambda}$. For each unit k, we divide the vector $\mathbf{d}_k$ into an observed and a unobserved (or missing) parts: $\mathbf{d}_k = (\mathbf{d}_{o;k}^T, \mathbf{d}_{m;k}^T)^T$. This notation does not imply that always the same dimension of the vector is not observed. Any dimension could not be observed. The observed data $\mathbf{d}_o$ are supposed to be generated from the density $g(\mathbf{d}_{o;}; \boldsymbol{\lambda})$. The objective is to estimate $\boldsymbol{\lambda}$ by $\hat{\boldsymbol{\lambda}} = \arg\max \ell(\mathbf{d}_{o;}; \boldsymbol{\lambda})$, where $\ell(\mathbf{d}_{o;}; \boldsymbol{\lambda}) = \log g(\mathbf{d}_{o;}; \boldsymbol{\lambda})$. Let $k(\mathbf{d}_{m;} \mid \mathbf{d}_{o;}; \boldsymbol{\lambda}) = f(\mathbf{d}; \boldsymbol{\lambda})/g(\mathbf{d}_{o;}; \boldsymbol{\lambda})$ the conditional density of the unobserved part $\mathbf{d}_m$ given the observed part $\mathbf{d}_o$. Using some initial value for $\boldsymbol{\lambda}$, say $\boldsymbol{\lambda}^{(b)}$, the E-step of the EM algorithm requires the calculation of a function of $\boldsymbol{\lambda}$, $Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}^{(b)})$, such that $Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}^{(b)}) = E\{\ell_c(\mathbf{d}; \boldsymbol{\lambda}) \mid \mathbf{d}_{o;}; \boldsymbol{\lambda}^{(b)}\} = \int \log f(\mathbf{d}; \boldsymbol{\lambda}) k(\mathbf{d}_m \mid \mathbf{d}_o; \boldsymbol{\lambda}^{(b)}) \partial\mathbf{d}_m$, where $\boldsymbol{\lambda}$ is the parameter of interest, and $\boldsymbol{\lambda}^{(b)}$ is the value of $\boldsymbol{\lambda}$ in the previous iteration. Then, the M step of the EM algorithm intent to choose the value of $\boldsymbol{\lambda}$, say $\boldsymbol{\lambda}^{(b+1)}$, that maximizes $Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}^{(b)})$, i.e., $\hat{\boldsymbol{\lambda}}^{(b+1)} = \arg\max Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}^{(b)})$. If we iterate the E-step and M-step until convergence, under regularity conditions, the algorithm converges to the maximum likelihood estimate.

We use instead in this paper a variant of the EM algorithm called the Stochastic Expectation Maximization (SEM) algorithm. The idea underlying SEM algorithm (Broniatowski *et al*. 1983, Celeux and Diebolt 1985) is to replace the derivation and maximization of $Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}^{(b)})$ by the generation $\mathbf{d}_m^{(b)}$ values for the unobserved part $\mathbf{d}_m$ from $k(\mathbf{d}_{m;} \mid \mathbf{d}_{o;}; \boldsymbol{\lambda}^{(b)})$ and then update $\boldsymbol{\lambda}$ from the complete data $\mathbf{d}^{(b)} = (\mathbf{d}_1^{(b)}, ..., \mathbf{d}_N^{(b)})^T$ with $\mathbf{d}_k^{(b)} = (\mathbf{d}_{o;k}^T, \mathbf{d}_{m;k}^{(b)T})^T \quad k = 1, ..., N$. The stochastic step of the SEM relies on the random imputation by generating observations from their conditional density of the unobserved part given the observed part and the current value of the parameter. Hence, the SEM algorithm takes the following form. From an arbitrary starting value $\boldsymbol{\lambda}^{(0)}$ a sequence ($\boldsymbol{\lambda}^{(b)}, b = 1,2,..$) is formed by going through the stochastic E-step or SE Step and a M Step. SE Step: Given a value of $\boldsymbol{\lambda}^{(b)}$, simulate values $\mathbf{d}_k^{(b)}$ from the conditional distribution of $\mathbf{d}_k$ given $\mathbf{d}_{o;k}$, i.e.,

draw $\mathbf{d}_{m;k}^{(b)} \sim k(\mathbf{d}_{m;k} \mid \mathbf{d}_{o;k})$. M Step: Maximize the resulting complete data log-likelihood, $\sum_k \log f(\mathbf{d}_k^{(b)}; \boldsymbol{\lambda})$ with $\mathbf{d}_k^{(b)} = (\mathbf{d}_{o;k}^T, \mathbf{d}_{m;k}^{(b)T})^T$, and let the maximum be the next value $\boldsymbol{\lambda}^{(b+1)}$.

Let $\mathbf{d}_k = ({}^{(pri)}l_{\kappa;k}, l_{\kappa;k})^T$ denotes a bivariate discrete variable with mean $\boldsymbol{\mu}_k = ({}^{(pri)}p_k, p_k)^T$ and covariance $\boldsymbol{\Sigma}_k$. We have $E(l_{\kappa;k} \mid {}^{(pri)}l_{\kappa;k}) = {}^{(pri)}l_{\kappa;k}\tau_k^{(1|1)} + (1 - {}^{(pri)}l_{\kappa;k})(1 - \tau_k^{(1|0)})$.

Given $\mathbf{d}_k$, the joint distribution is given by

$$f({}^{(pri)}l_{\kappa;k}, l_{\kappa;k}) = \{p_k^{(1|1)} p_k\}^{{}^{(pri)}l_{\kappa;k} l_{\kappa;k}} \times \{p_k^{(1|0)}(1-p_k)\}^{{}^{(pri)}l_{\kappa;k}(1-l_{\kappa;k})} \times \{(1-p_k^{(1|1)})p_k\}^{(1-{}^{(pri)}l_{\kappa;k})l_{\kappa;k}} \times \{(1-p_k^{(1|0)})(1-p_k)\}^{(1-{}^{(pri)}l_{\kappa;k})(1-l_{\kappa;k})}$$

.

Given $\mathbf{d}_k, k \in \wp$, one could estimate the parameter using the following EE:

$$\hat{\mathbf{U}}(\boldsymbol{\lambda}) = \sum_k d_k(\wp)\{\mathbf{u}(l_{\kappa;k}; \boldsymbol{\lambda}) + \mathbf{u}(l_{\kappa;k}^{(1|1)}; \boldsymbol{\lambda}) + \mathbf{u}(l_{\kappa;k}^{(1|0)}; \boldsymbol{\lambda})\} = \mathbf{0},$$

where $\mathbf{u}(l_{\kappa;k}; \boldsymbol{\lambda}) = \sum_k \dot{\boldsymbol{p}}_k (l_{\kappa;k} - p_k)\{p_k(1-p_k)\}^{-1}$,

$$\mathbf{u}(l_k^{(1|1)}; \boldsymbol{\lambda}) = \sum_{\kappa;k} l_{\kappa;k} \left\{\dot{\boldsymbol{p}}_k^{(1|1)} ({}^{(pri)}l_{\kappa;k} - p_k^{(1|1)})\{p_k^{(1|1)})(1-p_k^{(1|1)})\}^{-1}\right\},$$

$$\mathbf{u}(l_{\kappa;k}^{(1|0)}; \boldsymbol{\lambda}) = \sum_k (1-l_{\kappa;k}) \left\{\dot{\boldsymbol{p}}_k^{(1|0)} ({}^{(pri)}l_{\kappa;k} - p_k^{(1|0)})\{p_k^{(1|0)})(1-p_k^{(1|0)})\}^{-1}\right\},$$

$\dot{\boldsymbol{p}}_k = \partial p_k / \partial \boldsymbol{\lambda}$, $\dot{\boldsymbol{p}}_k^{(1|1)} = \partial p_k^{(1|1)} / \partial \boldsymbol{\lambda}$, and $\dot{\boldsymbol{p}}_k^{(1|0)} = \partial p_k^{(1|0)} / \partial \boldsymbol{\lambda}$.

3.3 Optimization Step

So we use the phase 1 (10 time-periods of observations): 1) updated domain classification model; 2) updated response model; and, 3) updated values of the variable of interest; to determine the extra number of time periods I of data collection (or equivalently the extra number of phases P with $10 + \mathrm{I} = \sum_{\mathrm{p}=1}^{1+\mathrm{P}} n_{\mathrm{p}}$ and $n_1 = 10$), and the revised follow-up model parameter $\boldsymbol{\lambda}_f^{(2)}$ by minimizing the variance, $\min_{\boldsymbol{a}} Var(\breve{Y}_{10+\mathrm{I};\kappa})$, subject to constraint on the expected cost, $\overline{C}_{wope} \le C_{\max} - C_{10}$, and constraint on the duration $0 \le \mathrm{I} \le \mathrm{I}_{\max} - 10$, where $\boldsymbol{a} = (\mathrm{I}, \boldsymbol{\lambda}_f^{(2)T})^T$, and C_{10} is the total cost spent in the first phase, Here the estimator used for design revision is

$${}^{(2)}\breve{Y}_{\mathrm{I};\kappa} = \sum_k d_{\mathrm{I};k} ({}^{(2)}r_{\mathrm{I};k} / {}^{(2)}\xi_{\mathrm{I};k}) {}^{(2)}l_{\kappa;k}\, {}^{(2)}y_k,$$

with $\xi_{\mathrm{I};k} = (1-\phi_{f;k})\xi_{\mathrm{I};k}^{(self)} + \phi_{f;k}\xi_{\mathrm{I};k}^{(self+f)}$, and $\phi_{f;k} = 1 - \prod_{\mathrm{p}=1}^{2} \phi_{\wp;k}^{(\mathrm{p})}$.

Using the revised information as input to the optimization problem, Table 10 displays the revised values of the design parameters: the expected duration to reach the required cv, the expected number of follow-ups, the expected number of respondents, and the expected coefficient of

variation in percentage. Table 10 also displays the expected fix cost, the expected follow-up cost and the expected data collection cost.

Table 10: Extra Resources Allocation

Extra Duration	Expected Number of Extra		Expected		Cost		
	Follow-up	Respondents	CV	Cost	Fixed	Follow-up	Data Collection
5	0	203	3	740	300	0	440

Note that the required of .05 is expected to be reached after 15(=10+5) time periods.

3.4 Decision Step

It was then decided to proceed without follow-up for a maximum of 5 time periods of data collection. Instead, after a time period of data collection is complete, the new observations are included and the follow-up decision is revised. This would be conducted on a continual basis for however many number of time periods needed until data collection is complete.

4. Design Pre-specification under Processing Errors

We now consider the case where the classification process is fallible. A Poisson subsampling is used to validate the processed classification, as well as the prior classification.

4.1 Estimator

A naïve estimator of the domain total Y_κ is given by

$$\breve{Y}_{\mathrm{I};\kappa}^{(N)} = \sum_k d_{\mathrm{I};k}(r_{\mathrm{I};k}/\xi_{\mathrm{I};k})^{(pro)}l_{\kappa;k}y_k \equiv \sum_k d_{\mathrm{I};k}(r_{\mathrm{I};k}/\xi_{\mathrm{I};k})\aleph_{\kappa;k} \tag{4.1}$$

where $^{(pro)}l_{\kappa;k}$ is the error-prone processed classification indicator for unit k, $\aleph_{\kappa;k}={}^{(pro)}l_{\kappa;k}y_k$, and the superscript "*pro*" stands for the error-prone "processed" information. The conditional bias induced by the naïve estimator $\breve{Y}_{\mathrm{I};\kappa}^{(N)}$ is given by $B_{\mathrm{I};\kappa}^{(N)} = \sum_k E(\aleph_{\kappa;k} - \dot{y}_{\kappa;k})$. In order to remove the potential processing bias, we subsample from respondents, obtain the true classification for subsampled units, and estimate the bias by

$$\breve{B}_{\mathrm{I};\kappa}^{(N)} = \sum_k d_{\mathrm{I};k}(r_{\mathrm{I};k}\,|\,\xi_{\mathrm{I};k})d_{\mathrm{v}|\mathrm{I};k}^{(2|\mathrm{I})}(\aleph_{\kappa;k} - \dot{y}_{\kappa;k}),$$

where $d_{\mathrm{v}|\mathrm{I};k}^{(2|\mathrm{I})}$ are the design weights associated with the validation subsample $\wp_{\mathrm{v}|\mathrm{I}}$. Finally we adjust $\breve{Y}_{\mathrm{I};\kappa}^{(N)}$ to get a bias-adjusted estimator

$$\breve{Y}_{\mathrm{I};\kappa}^{(ad)} = \breve{Y}_{\mathrm{I};\kappa}^{(N)} - \breve{B}_{\mathrm{I};\kappa}^{(N)}. \tag{4.2}$$

If $E_{\mathrm{v};\wp}(d_{\mathrm{v}|\mathrm{I};k}^{(2|\mathrm{I})})=1$, then $\breve{Y}_{\mathrm{I};\kappa}^{(ad)}$ is unbiased estimator for Y_{κ}, where $E_{\mathrm{v};\wp}$ denotes conditional expectation with respect to the subsampling design.

4.2 Derivation of the Variance Function

Assume that I time periods of survey process are completed, with $1<\mathrm{I}\le \mathrm{I}_{\max}$, and consider variance derivation of the estimator given by (4.2). We may first decompose the variance of $\breve{Y}_{\mathrm{I};\kappa}^{(ad)}$ as

$$Var(\breve{Y}_{\mathrm{I};\kappa}^{(ad)}) = EVar_{\mathrm{v};\wp}(\breve{Y}_{\mathrm{I};\kappa}^{(ad)}) + VarE_{\mathrm{v};\wp}(\breve{Y}_{\mathrm{I};\kappa}^{(ad)}) \equiv \mathrm{V}_{\mathrm{v};\wp} + \mathrm{V}_{wope} \tag{4.3}$$

where $Var_{\mathrm{v};\wp}$ denotes conditional variance with respect to the subsampling design.

For Poisson subsampling with constant subsampling probabilities $\pi_{\mathrm{I};k}^{(2|\mathrm{I})}$, the first component $\mathrm{V}_{\mathrm{v};\wp} = EVar_{\mathrm{v};\wp}\{\sum_k d_{\mathrm{I};k}(r_{\mathrm{I};k}/\xi_{\mathrm{I};k})d_{\mathrm{v}|\mathrm{I};k}^{(2|\mathrm{I})}(\aleph_{\kappa;k}-\dot{y}_{\kappa;k})\}$ is given by

$$\mathrm{V}_{\mathrm{v};\wp} = \sum_k M(\aleph_{\kappa;k}-\dot{y}_{\kappa;k})(1-\pi_{\mathrm{I};k}^{(2|\mathrm{I})})/(\pi_{\mathrm{I};k}\xi_{\mathrm{I};k}\pi_{\mathrm{I};k}^{(2|\mathrm{I})}), \tag{4.4}$$

where $M(\aleph_{\kappa;k}-\dot{y}_{\kappa;k}) = E(\aleph_{\kappa;k}-\dot{y}_{\kappa;k})^2$. In the absence of processing error $\aleph_{\kappa;k}=\dot{y}_{\kappa;k}$ and $\mathrm{V}_{\mathrm{v};\wp}=0$. The sum of (2.3) and (4.4) constitutes $Var(\breve{Y}_{\mathrm{I};k}^{(ad)}) = \mathrm{V}_{\mathrm{v};\wp}+\mathrm{V}_r+\mathrm{V}_{\wp}+\mathrm{V}_m$, the variance of $\breve{Y}_{\mathrm{I};k}^{(ad)}$ given by (4.2). It follows that, we can express $Var(\breve{Y}_{\mathrm{I};k}^{(ad)})$ as

$$Var(\breve{Y}_{\mathrm{I};k}^{(ad)}) = v_0 + \sum_h v_k/(\pi_{\mathrm{I};k}\xi_{\mathrm{I};k}) + \sum_k v_{\mathrm{v};k}/(\pi_{\mathrm{vI};k}\xi_{\mathrm{I};k}),$$

where $v_0 = -\sum_k\{E_m(\dot{y}_{\kappa;k})\}^2$, $v_k = E(\dot{y}_{\kappa;k}{}^2{}_{\kappa;k}) - M(\aleph_{\kappa;k}-\dot{y}_{\kappa;k})$, $v_{\mathrm{v};k} = M(\aleph_{\kappa;k}-\dot{y}_{\kappa;k})$, and $\pi_{\mathrm{vI};k} = \pi_{\mathrm{I};k}\pi_{\mathrm{vI};k}^{(2|\mathrm{I})}$.

4.3 Specification of the Cost Function

We may decompose the global cost as

$$C = C_{wope} + C_{\mathrm{v}}.$$

The subsampling and validation component C_{v} is given by $C_{\mathrm{v}} = \sum_k 1_k(\wp_{\mathrm{I}}|P)r_{\mathrm{I};k}1_k(\wp_{\mathrm{vI}}|\wp_{\mathrm{I}}^{(r)})c_{\mathrm{v};k}$, where $\wp_{\mathrm{I}}^{(r)}$ is the sample of respondents at time period I, and $c_{\mathrm{v;k}}$ is the subsampling-validation cost for unit k.

4.4 Modeling the Subsample Selection Probabilities

The conditional probability that unit k will be selected in phase p, given that the unit was not selected prior to p is constructed as

$$\log\{\pi_{\wp;k}^{(2|\mathrm{I})(\mathrm{p})}/(1-\pi_{\wp;k}^{(2|\mathrm{I})(\mathrm{p})})\} = \boldsymbol{\nu}_{\wp;k}^{(2|\mathrm{I})(\mathrm{p})T}\boldsymbol{\lambda}_{\wp}^{(2|\mathrm{I})(\mathrm{p})},$$

where $\boldsymbol{v}_{\wp;k}^{(2|1)(\mathrm{p})} = (1, l_{dc;k}^{(W)})^T$ is the vector predictor, $l_{dc;k}^{(W)} = 1$ if the mode of data collection is Web, and $l_{dc;k}^{(W)} = 0$ if not, and $\boldsymbol{\lambda}_{\wp}^{(2|1)(\mathrm{p})} = (\lambda_{\wp;0}^{(2|1)(\mathrm{p})}, \lambda_{\wp;1}^{(2|1)(\mathrm{p})})^T$ is the unknown vector parameter to be determined.

4.5 Specification of the Objective Function

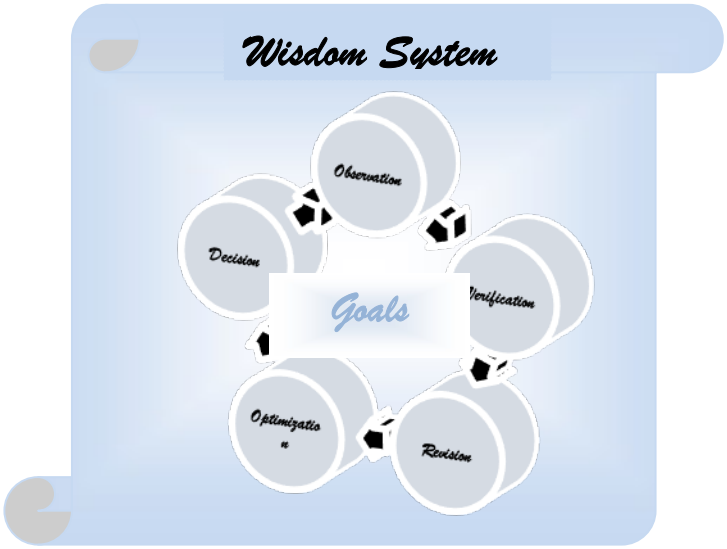
To create a design, we determine the number of phases P of data collection (or equivalently the number of time periods $\mathrm{I}_{\mathrm{P}} = \sum_{\mathrm{p}=1}^{\mathrm{P}} n_{\mathrm{p}}$), the samples selection parameter $(\boldsymbol{\lambda}_{\wp}^T, \boldsymbol{\lambda}_{\wp}^{(1|2)T})^T$, and the follow-up model parameter $\boldsymbol{\lambda}_f$ by minimizing the variance, $\min_{\boldsymbol{a}} Var(\breve{Y}_{\mathrm{I};\kappa}^{(ad)})$, subject to constraint on the expected cost: $\overline{C} \le C_{\max}$, and constraint on the duration $1 \le \mathrm{P} \le \mathrm{P}_{\max}$, where the expected cost is given by $\overline{C} = \overline{C}_{wope} + \overline{C}_{\mathrm{v}}$, where $\overline{C}_{\mathrm{v}} = \sum_k \pi_{\mathrm{I};k} \xi_{\mathrm{I};k} \pi_{\mathrm{vI};k}^{(2|1)} c_{\mathrm{v};k}$.

6. Wisdom Design

We now extend our approach to cover the validation task: 1) accumulate observations from the sampled units on the processed information and the target information governing the survey under consideration; 2) validate the processed outputs for the subsample units; 3) update information used for design specification; and, 4) revise, if necessary, specification of the design for the remaining time periods of survey process. After completing the fifth step, the stopping rules are consulted to see if the survey process should stop. If not, the five steps are repeated continuously to detect any discrepancies between observed and expected information. We refer to the above five steps as the Observation-Validation-Revision-Optimization (O-V-R-O) steps.

- **Observation Step:** Obtain next phase p of observations on the error-prone process.
- **Validation Step:** Obtain next phase p of observations on the error-free process.
- **Revision Step:**
- **Estimation/ Imputation**: 1) Update $\boldsymbol{\lambda}_{\mathrm{p}-1}$ to get $\boldsymbol{\lambda}_{\mathrm{p}}$ using $\mathbf{d}_{\mathrm{o;p}}$; and, 2) Impute missing values of each component ψ of $\boldsymbol{\psi}$ to get $\psi_{\mathrm{p};k} = E_{\psi}(\psi_k \mid \mathbf{d}_{\mathrm{o;p}}, \boldsymbol{\lambda}_{\mathrm{p}})$,
- **Optimization Step:**
  - Determine the optimal design parameter $\boldsymbol{a} = (P, \boldsymbol{\lambda}_{\wp}^T, \boldsymbol{\lambda}_{\wp}^{(2|1)T}, \boldsymbol{\lambda}_f^T)^T$ conditional on $\boldsymbol{\psi}_{\mathrm{p}}$ and $\boldsymbol{\lambda}_{\mathrm{p}}$. The solution is denoted by $\boldsymbol{a}_{\mathrm{p}} = (P_{\mathrm{p}}, \boldsymbol{\lambda}_{\wp;\mathrm{p}}^T, \boldsymbol{\lambda}_{\wp;\mathrm{p}}^{(2|1)T}, \boldsymbol{\lambda}_{f;\mathrm{p}}^T)^T$.
- **Decision Step:**

- Decide if the data collection should stop (i.e., $p = P_p$), if not, decide if the methodology-process should change, and then repeat the five steps continuously after observing some realizations of the processed/target information.



Concluding Remarks

We formulated an optimization problem for designing a survey, and we identified steps for its revision during the survey process period. We considered the error-prone prior, error-prone processed and error-free information as a random variable with a joint distribution with some probability function. Then, we updated the joint probability distribution after observing some of realizations of the error-free random process at each phase of survey process to revise the survey design specification. The proposed approach makes full use of both error-prone sets of information while requiring only few observations from the error-free and expensive random process. Since revision of a design indicates when a design is nearly "optimal", and how the error-free information varies from the error-prone prior and processed information, the revision of the design has an important role to play in survey quality and cost.

References

Broniatowski, M., G. Celeux, and J. Diebolt. 1983. Reconnaissance de Mélanges de Densités par un Algorithme d'Apprentissage Probabiliste. Data Analysis and Informatics, Diday E. et al, eds., 359-374 Amsterdam, North Holand.

Bross, I. 1954. Misclassification in 2×2 Tables. *Biometrics*, 10, pp. 478-483.

Celeux, G. and J. Diebolt. 1985. The SEM Algorithm: a Probabilistic Teacher Algorithm Derived from the EM Algorithm for the Mixture Problem, *Computational Statistics Quarterly*, 2, 73-82.

Cochran, W. 1977. *Sampling Techniques*. New York: John Wiley & Sons, Inc.

Demnati, A. 2016. Responsive Design – Side Effect Reduction of Prior Information on Survey Design. In Joint Statistical Meeting of the American Statistical Association Proceedings, July 30 – August 4, Chicago, Illinois, USA

Demnati, A. 2017. Extending Hansen and Hurwitz's Approach for Non Response in Sample Survey. In the Joint Statistical Meeting of the American Statistical Association Proceedings, July 29 – August 3, Baltimore, Maryland, USA.

Dempster, A. P., N.M. Laird and D.B. Rubin. 1977. Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society B.* 39, pp. 1–38.

Groves, R.M. and S.G. Heeringa. 2006. Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs. *Journal of the Royal Statistical Society*, Series A, 169, 439-457.

Hartley, H. O. 1958. Maximum likelihood estimation from incomplete data. *Biometrics*. 4, pp. 174–194.

Kokan, A. R. 1963. Optimum Allocation in Multivariate Surveys. Journal of the Royal Statistical Society, Series A., 139, 80-95.

Peytchev, A., S. Riley, J. Rosen, J. Murphy, and M. Lindblad. 2010. Reduction of Nonresponse Bias in Surveys Through Case Prioritization. *Survey Research Methods*, 4, 21-29.

Rosenblatt, F. 1958. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review* 65, 386-408.

Rosenblatt, F. 1962. *Principles of Neurodynamics*. Spartan Books, Washington D.C.

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

Schouten, B., M. Calinescu, and A. Luiten. 2013. Optimizing Quality of Response Through Adaptive Survey Designs. *Survey Methodology*, 39, 29-58.

Tourangeau, R., J.M. Brick, S. Lohr, and J. Li. 2016. Adaptive and Responsive Survey Design: A Review and Assessment. *Journal of the Royal Statistical Society*, Series A.

Tenenbein A. 1970. A Double Sampling Scheme for Estimating from Binomial Data with Misclassifications. *Journal of the American Statistical Association*, 65, 1350-1361.

Tenenbein A. 1972. A Double Sampling Sc heme for Estimating from Misclassified Multinomial Data with Applications to Sampling Inspection. *Technometrics*, 14, 187-202.