# Side Effect Reduction of Prior and Processed Information on Survey Design

(Parts 1 & 2)

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 are likely generated from a different random process than the target one governing the surveys to be designed. The survey process, such as text classification, also makes the development of a survey difficult as it may vary from one human or machine to another. The impact of each error-prone set of information on the properties of the estimator can be significant. We are concerned with reducing the side effects of both the prior information and the processed information on the quality of the estimator of the parameter of interest during the 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 the data collection period, which involves the 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 as a random variable that has a joint distribution with some probability function. Then, at each time of 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 data collection period. In addition, the coefficient of reliability for a survey as both a whole set of processes and 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 areas, such as health, biometrics, industrial, commercial, finance, insurance, actuarial, and more, that require the estimation of quantities related to uncertain or imprecise information (i.e. $\psi$ ) to learn, model, and predict units (such as human and market) behaviours. Vague understanding of $\psi$ promotes approaches for the development of models to (1) explain known observations on prior information $\chi$ , (2) predict observations on the target information, and (3) relate prior information with the vague target 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 networks are inspired by the way biological neural networks in the human brain process information. It first learns a mapping $\boldsymbol{\psi}_k = f(\boldsymbol{v}_k; \boldsymbol{\lambda})$ from input $\boldsymbol{v}_k$ to output $\boldsymbol{\psi}_k$ given a sample of training examples $\wp = \{(\boldsymbol{\psi}_k, \boldsymbol{v}_k); k = 1,..., n\}$ of input-output pairs $(\boldsymbol{\psi}_k, \boldsymbol{v}_k)$ for unit $k$, where $\boldsymbol{\lambda}$ is a large vector of weights expressing the importance of the respective inputs to the output (Rosenblatt 1958, 1962) and $n$ is the number of training examples in the sample $\wp$. It then uses the uncovered patterns to predict unknown output using the best guess $\hat{\boldsymbol{\psi}}_k = f(\boldsymbol{v}_k; \hat{\boldsymbol{\lambda}})$, where $\hat{\boldsymbol{\lambda}}$ is the solution to an error minimization problem used to train the artificial neural network. Each training input $\boldsymbol{v}_k$ is a vector of numbers, representing units (possibly complex in nature) such as a person, an image, a sequence of characters or words, a video, etc. These are called features. 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 to (1) capture similarities within a set of labelled units represented by features, (2) to represent 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, fraud detection, and natural languages translation. 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, these questions arise: How can we collect 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 non-respondents, (3) the mode of data collection (e.g. in-person, by phone, by mail, or via the internet), and (4) the validation. In this study, we examine the general problem which includes the four desired properties stated in (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 between and 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 that fall within resource, quality, and timing constraints. Efficiency is an important issue because inefficient determination or allocation resources may lead to imprecise results and a misuse of these resources. To optimally determine the design parameters such as (1) the duration of the survey and (2) the amount of resources and their allocation between and within stages of the survey design, design pre-specification requires four steps:

1) **Specification Step.** Specification of: (1) the population of interest, (2) the parameter of interest, (3) the sampling frame and the sampling schemes, (4) the mode of data collection, (5) the nonresponse follow-up activities, (6) the validation activities, (7) the estimator to be used, (8) the precision function, (9) the cost function, and (10) the desired precision or the global cost.

2) **Prediction Step.** Obtaining prior information from the sampling frame, the administrative files, or the previous surveys is required to compute unknown quantities in formulas for both precision and cost functions.

3) **Optimization Step**. Optimization of some objective function – that involves both precision and cost functions.

4) **Decision Step**. Determination of the survey design parameter using the solution to the objective function.

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 $1$ 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 lack consistency between time periods. 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 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 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 target 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 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 predominantly 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 suggested to employ a more effective survey protocol for the sample cases (the least likely to respond) 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 detailed literature review 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 target random process, 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 a few observations from the expensive, 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 Squared Error (MSE) enhances information on (1) the survey results, (2) the comparisons between surveys, and (3) the contribution of the given survey as an addition to prior information. In Demnati (2018), we extended our work to cover the 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 based on its inputs. Automatic classification systems learn 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) the automatic 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 questions are 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 humans, random sampling, in combination with human validation, is 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 to the case where the error-prone prior information, say $^{(pri)}\boldsymbol{\chi}$, and the error-prone processed information, say $^{(pro)}\boldsymbol{\chi}$, 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 $\boldsymbol{\chi}$ can be carried out based on observations on $\boldsymbol{\psi}$, where $\boldsymbol{\chi}=(^{(pri)}\boldsymbol{\chi}^T, ^{(pro)}\boldsymbol{\chi}^T)^T$. We also assume that the error-prone information $\boldsymbol{\chi}$ has a potential bias $\boldsymbol{b}$ 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 bias. 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 $\boldsymbol{\chi}$, together with the joint probability function of $(\boldsymbol{\chi}^T, \boldsymbol{\psi}^T)^T$ 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 $\theta$ and an estimate $\hat{\theta}$ of it. Possible sources of error in surveys include sampling frame, sampling scheme, measurement, nonresponse, editing, imputation, disclosure-avoidance, etc. A criterion which is commonly used in judging the performance of an estimator $\hat{\theta}$ of a parameter $\theta$ is its MSE defined by $M(\hat{\theta}) = E\{(\hat{\theta}-\theta)^2\}$. Here, the parameter $\theta$ can be seen as the quantity that would be obtained under the ideal situation which consists of a census case with complete response and without any processing errors. 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 we can do is make this zero. Given that the remaining relative error or the relative missed information about $\theta$ 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 $\theta$ obtained after observing $\hat{\theta}$, i.e.,

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

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

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

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 to replace conditional variance, which depends on the joint distribution, with correlation as it can be calculated more easily. That being said, conditional independence is more meaningful and preferable than zero-correlation.

In an attempt to further discuss side effect reduction of both errors sources of information on the quality of the estimator of the parameter of interest during data collection period, our work below is organized as follows: in Section 2, we give a straightforward prescription of our wisdom design for designing and conducting a survey design; in Section 3, we study an example detailing the steps required for design pre-specification in the absence of processing errors; in Section 4, we revise the example design after 10 periods of data collection; in Section 5, some estimation methods for model parameters are discussed; in Section 6, classification model parameter is revised. Results of a simulation study are presented; and, in Section 7, response model parameter is revised.

## 2. The Wisdom Design

To revise the survey design in discrete intervals, we divided the continuous time of the entire data collection period into a sequence of continuous time periods: 1, 2, etc., and let $\mathrm{I}_{\min}$ denote the minimum length of the data collection period to obtain full responses. Suppose the survey limited length of duration of data collection is made up of $\mathrm{P}_{\max}$ phases, the $\mathrm{p}^{th}$ being of size $n_{\mathrm{p}}$ 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 $\mathrm{I}_{\max} < \mathrm{I}_{\min}$. Therefore, there would be an $N \times \mathrm{P}_{\max}$ rectangular array of phases of data collection, where $N$ is the size of the finite population. Because the design parameter $\boldsymbol{\Phi}$ is the solution to some objective function $\mathrm{O}(\boldsymbol{\Phi}; \boldsymbol{\psi}, \boldsymbol{\lambda})$ that requires the target information, we first considered the error-prone and error-free information as a random variable that has a joint distribution $f(\boldsymbol{\psi}, \boldsymbol{\chi}; \boldsymbol{\lambda})$ with vector parameter $\boldsymbol{\lambda}$. Then, at each time of survey data collection, after receiving the information that the target random process has taken specific values, we update the parameter $\boldsymbol{\lambda}$ of the joint probability distribution $f(\boldsymbol{\psi}, \boldsymbol{\chi}; \boldsymbol{\lambda})$ to revise the design specification $\boldsymbol{\Phi}$ in the course of its progress.

Our simplest prescription for designing and conducting a survey design is as follows:

**a. First specify the**

(1) the population of interest, (2) the parameter of interest, (3) the sampling frame and the sampling schemes, (4) the mode of data collection, (5) the nonresponse follow-up activities, (6) the validation activities, (7) the estimator to be used, (8) the precision function, (9) the cost function, and (10) the desired precision or the global cost.

**b. Then repeat continuously the following five steps until the end of data collection**

**b.1 Observation Step**

Obtain the next phase $\mathrm{p}$ of observation from respondents

- For design pre-specification: Obtain observations from the sampling frame, administrative files, or from previous surveys,
- This information is needed to compute unknown quantities in formulas for both the precision and cost functions.

**b.2 Validation Step.** Randomly validate some of the obtained values.

**b.3 Revision Step.** Update the vector parameter $\boldsymbol{\lambda}$ and the missing values of $\boldsymbol{\psi}$ using all available information i.e., (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 $\psi$ 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

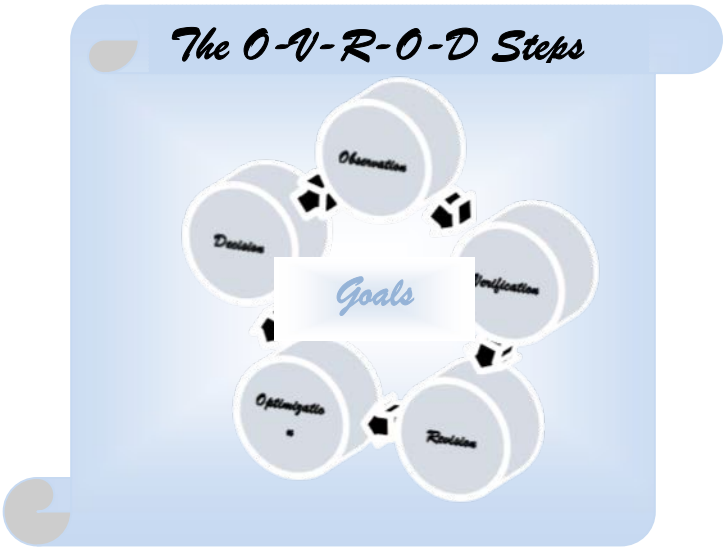
observed information until the end of phase $p$ of data collection, and $E_{\psi}$ denotes expectation with respect to the random process governing the component $\psi$. Note that $\psi_{p;k} = \psi_k$ when item $\psi_k$ is observed.

**b.4 Optimization Step.** Optimize the objective function $O(\boldsymbol{\Phi}; \boldsymbol{\psi}_p, \boldsymbol{\lambda}_p)$ – that involves both the precision and cost functions parameter, i.e., determine the optimal design parameter $\boldsymbol{\Phi}_p$ conditional on $\boldsymbol{\psi}_p$ and $\boldsymbol{\lambda}_p$ for possible future data collection periods.

**b.5 Decision Step.** Decide if the data collection should stop (i.e., $P = p$). If not (i.e. $P > p$), revise the specification of the design as necessary and repeat the five steps (b.1 to b.5) continuously after observing some realizations of the target process. Here $P$ denotes the number of periods of data collection.

**c. Finally, stop data collection and produce survey estimates.**

We refer to the above five steps as the Observation-Validation-Revision-Optimization-Decision (O-V-R-O-D) steps. The revision step incorporates learning and prediction, while the decision step incorporates actioning.




So, our approach is embedded in a continuous learning process that permits changes in methodology-process at any time of the 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. Therefore, this method does not use a fixed

design, although, an expected design is always pre-specified. Depending on the relationship between the error-prone information, the target information, and the stopping rules, only a few time periods may be sufficient to stop the data collection period.

## 3. Example of 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 the covariate $u$ on it; an address and contact 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 $\mathrm{I}_F$ 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 sample. Known values of survey design parameters such as $N$, $C_{\max}$, $\mathrm{I}_{\max}$, $\mathrm{I}_F$, and $D_F$ are provided in Table 1. The first task is to pre-specify the design to 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, (3) determine the resources and their allocation within stages of the survey design for each given period of data collection, and finally, (4) decide for the parameters of the design.

### 3.1 Specification Step

#### 3.1.1 Parameter of Interest

Estimate is wanted for specific subpopulation $\kappa$, called domain $\kappa$. 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 $\kappa$ of the units of interest or domain $\kappa$ be denoted as $P_\kappa$, and let $l_{\kappa;k} = 1_k(P_\kappa \mid P) = 1(k \in P_\kappa \mid k \in P)$ be the domain $\kappa$ membership indicator variable for unit $k$, where $1(condition \mid .)$ is the truth function, i.e., $1(condition \mid .) = 1$ if the *condition* is true and $1(condition \mid .) = 0$ if not. The domain total $Y_\kappa$ of a characteristic $y$ may be written as

$$Y_\kappa = \sum_k l_{\kappa;k} y_k \equiv \sum_k \dot{y}_{\kappa;k}, \tag{3.1}$$

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 $Y_\kappa$, obtained under the assumed ideal situation which consists of census case with complete response and without any processing error, acts as the "gold standard".

### 3.1.2 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 $\mathrm{I}_k$, 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 $h_{ki}(\boldsymbol{v}_{r;ki}, \boldsymbol{\lambda}_r)$, $h_{ki}$ 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 $\boldsymbol{v}_{r;ki}$ refers to both time-invariant and time-varying explanatory variables and $\boldsymbol{\lambda}_r$ is the unknown $q_r \times 1$ vector parameter to be estimated. For units with $t_k = i$, the probability of obtaining a response at time period $i$ could be expressed in terms of the hazard as

$$\Pr(t_k = i) = h_{ki} \prod_{j=1}^{i-1} (1 - h_{kj}) . \tag{3.2}$$

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

$$\Pr(t_k > i) = \prod_{j=1}^{i} (1 - h_{kj}) . \tag{3.3}$$

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

$$\xi_{1;k} = 1 - \prod_{i=1}^{1} (1 - h_{ki}) = \sum_{i=1}^{1} \Pr(t_k = i) . \tag{3.4}$$

Demnati (2017) also developed a discrete-time model for multiple kinds of results or events such as refusal, ineligibility, and mode of data collection, by extending the Bernoulli model to the multinomial model. Assume that there are $E$ specific results and the $(E+1)^{th}$ category of no response, where $E \ge 1$. Define a vector of result indicator variables as $r_{ki}^{(e)} = 1$ if outcome $e$ occurs

from unit $k$ at time period $i$, and $r_{ki}^{(e)} = 0$ if not, where $e \in \{1,...,E\}$, $E_r(r_{ki}^{(e)}) = h_{ki}^{(e)}$, $Var_r(r_{ki}^{(e)}) = h_{ki}^{(e)}(1-h_{ki}^{(e)})$, and for $e \neq e'$ $Cov_r(r_{ki}^{(e)}, r_{ki}^{(e')}) = -h_{ki}^{(e)}h_{ki}^{(e')}$. The combined discrete-time hazard is

$$h_{ki} = \sum_{e=1}^{E} h_{ki}^{(e)} = \Pr(t_k = i \mid t_k \geq i),$$

and the marginal probability of obtaining result $e$ after $\mathrm{I}$ time periods is given by

$$\xi_{1;k}^{(e)} = \sum_{i=1}^{\mathrm{I}} \Pr(r_{ki}^{(e)} = 1).$$

### 3.1.3 Modeling the Sample Selection Probabilities

We define the sample $\wp$ membership indicator variable as $a_k = 1_k(\wp \mid P)$; i.e., $a_k = 1$ if unit $k$ is selected in the sample $\wp$, and $a_k = 0$ if not. The conditional probability $\pi_k = \pi_k(\wp \mid P)$, that unit $k$ will be selected is constructed as

$$\log\{\pi_k/(1-\pi_k)\} = \boldsymbol{v}_{\wp;k}^T \boldsymbol{\Phi}_{\wp},$$

where $\boldsymbol{v}_{\wp;k} = (1, l_{\kappa;k})^T$ is the vector predictor, $\boldsymbol{\Phi}_{\wp} = (\Phi_{\wp;0}, \Phi_{\wp;1})^T$ is the unknown vector parameter to be determined, $\pi_k(\Omega \mid \Omega^*) = E\{1_k(\Omega \mid \Omega^*)\}$ is the set $\Omega$ inclusion probability for unit $k$ given $k \in \Omega^*$, and $E$ denotes expectation with respect to the inclusion mechanism.

### 3.1.4 Specification of the Follow-up Activity

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

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

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

### 3.1.5 Estimator to be used

Suppose that the response probabilities $\xi_{1;k}$ after $\mathrm{I}$ time periods of data collection are known for all population units. Then after $\mathrm{I}$ time periods of data collection and for general sampling design with known positive inclusion probabilities, $\pi_{1;k}$, an unbiased estimator of the domain total $Y_{\kappa}$ is given by

$$\breve{Y}_{\mathrm{I};\kappa} = \sum_k w_{\mathrm{I};k}\dot{y}_{\kappa;k} . \tag{3.5}$$

with

$$w_{\mathrm{I};\mathrm{k}} = d_{\mathrm{I};\mathrm{k}}(r_{\mathrm{I};\mathrm{k}} / \xi_{\mathrm{I};\mathrm{k}})$$

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 $\mathrm{I}$ time periods of data collection.

### 3.1.6 Derivation of the Variance Function

We may decompose the variance of $\breve{Y}_{\mathrm{I};\kappa}$ given by (3.5) as

$$Var(\breve{Y}_{\mathrm{I};\kappa}) = E_m E_\wp Var_r(\breve{Y}_{\mathrm{I};\kappa}) + E_m Var_\wp E_r(\breve{Y}_{\mathrm{I};\kappa}) + Var_m E_\wp E_r(\breve{Y}_{\mathrm{I};\kappa}) \equiv \mathrm{V}_r + \mathrm{V}_\wp + \mathrm{V}_m \equiv \mathrm{V}_{wope} \tag{3.6}$$

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 $\mathrm{V}_{wope}$ 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 (3.6) is given by

$$\mathrm{V}_r = \sum_k E_m(\dot{y}^2_{\kappa;k})(1-\xi_{\mathrm{I};k})/(\pi_{\mathrm{I};k}\xi_{\mathrm{I};k}) . \tag{3.7}$$

Under Poisson sampling, the second component $\mathrm{V}_\wp = E_m Var_\wp(\sum_k d_{\mathrm{I};k}\dot{y}_{\kappa;k})$ of (3.6) is given by

$$\mathrm{V}_\wp = \sum_k E_m(\dot{y}^2_{\kappa;k})(1-\pi_{\mathrm{I};k})/\pi_{\mathrm{I};k} . \tag{3.8}$$

Finally, under independent model mechanisms on $\dot{y}_k$, the last component of (3.6) is given by

$$\mathrm{V}_m = \sum_k Var_m(\dot{y}_{\kappa;k}) . \tag{3.9}$$

The sum of (3.7), (3.8), and (3.9) constitutes $\mathrm{V}_{wope} = \mathrm{V}_r + \mathrm{V}_\wp + \mathrm{V}_m$, the variance of $\breve{Y}_{\mathrm{I};\kappa}$ given by (3.5). 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 $v_{wope;0} = -\sum_k\{E(\dot{y}_{\kappa;k})\}^2$, and $v_{wope;k} = E(\dot{y}^2_{\kappa;k})$ .

### 3.1.7 Specification of the Cost Function

We may decompose the initial global cost over $\mathrm{I}$ time periods of data collection as

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

The fixed cost $C_{\mathrm{I}}$ 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 $C_f = \sum_k 1_k(\wp_1 \mid P)(1-r_{e_{f;k};k}^{(self)})l_{f;k}c_{f;k}$, where $r_{i;k}^{(self)}$ represents the response indicator under self-enumeration over $i$ time periods of data collection, $e_{f;k}$ 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 $C_{dc} = \sum_k 1_k(\wp_1 \mid P)\{r_{1;k}^{(M)}c_{dc;k}^{(M)} + r_{1;k}^{(W)}c_{dc;k}^{(W)}\}$, 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 $m \in \{M, W\}$.

### 3.1.8 Specification of the Objective Function

To create a design, we determine the number of time periods $\mathrm{I}$ of data collection (or equivalently the number of phases $\mathrm{P}$ with $\mathrm{I} = \sum_{\mathrm{p}=1}^{\mathrm{P}} n_{\mathrm{p}}$), the sample selection parameter $\boldsymbol{\Phi}_{\wp}$, and the follow-up model parameter $\boldsymbol{\Phi}_f$ by minimizing the variance, $\min_{\boldsymbol{\Phi}} Var(\breve{Y}_{1;\kappa})$, 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 $\boldsymbol{\Phi} = (\mathrm{I}, \boldsymbol{\Phi}_{\wp}^T, \boldsymbol{\Phi}_f^T)^T$, $\overline{C}_{wope} = C_{\mathrm{I}} + \overline{C}_{\wp} + \overline{C}_f + \overline{C}_{dc}$, $\overline{C}_{\wp} = \sum_k \pi_{1;k} c_{\wp;k}$, $\overline{C}_f = \sum_k \pi_{1;k}(1-\xi_{e_{f;k};k}^{(self)})\phi_{f;k}^{(1)} c_{f;k}$, and $\overline{C}_{dc} = \sum_k \pi_{1;k}\{\xi_{1;k}^{(M)}c_{dc;k}^{(M)} + \xi_{1;k}^{(W)}c_{dc;k}^{(W)}\}$. In this case and for any duration of data collection, a Lagrange multiplier can be used to find the constraint minimum of the variance. Therefore, in any duration of data collection, the objective function is given by

$$\mathrm{O}(\boldsymbol{\Phi}) = \sum_k v_{wo;k} / (\pi_{1;k}\xi_{1;k}) + \varsigma(\overline{C}_{wope} - C_{\max}), \tag{3.10}$$

where $\varsigma$ is the Lagrange multiplier. The optimization problem obtains a constrained minimum at the point where the estimating equations (EE) are set to zero, $\mathbf{o}(\boldsymbol{\Phi}) = \partial \mathrm{O}(\boldsymbol{\Phi}) / \partial \boldsymbol{\Phi} = \mathbf{0}$. Kokan (1963) discussed a 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 to 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 $\boldsymbol{\Phi} \equiv \boldsymbol{\Phi}(\boldsymbol{\psi})$, where $\boldsymbol{\psi}_k = (y_k, l_{\kappa;k}, h_{ki}^{(m)}(i = 1, ..., \mathrm{I}_{\max}), \boldsymbol{c}_k^T; \boldsymbol{\lambda})^T$, $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_y^T, \boldsymbol{\lambda}_\kappa^T, \boldsymbol{\lambda}_r^T, \boldsymbol{\lambda}_c^T)^T$, $\boldsymbol{\lambda}_y$ is the vector parameter associated with the model on $y$, $\boldsymbol{\lambda}_\kappa$ is the vector parameter associated with the domain $\kappa$ classification, $\boldsymbol{\lambda}_r$ is the vector parameter associated with the response model, and $\boldsymbol{\lambda}_c$ is the vector parameter associated with the cost model on $\boldsymbol{c}_k = (c_{\wp;k}, c_{f;k}, \boldsymbol{c}_{dc;k}^T)^T$. 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.

### 3.2 Prediction Step

It is clear from (3.10) 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 ${}^{(pri)}p_{\kappa;k}$ of being a member of the domain of interest $\kappa$ 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) an initial estimate of the vector cost ${}^{(pri)}\boldsymbol{c}_k$. To reduce the follow-up burden, we generated the follow-up entry time period $e_{f;k}$ from the uniform interval $[\mathrm{I}_F+1, \mathrm{I}_{\max}-D_F-2]$. So that the vector of available prior information for each unit $k$ in the frame is $\boldsymbol{\chi}_k=(u_k, {}^{(pri)}p_{\kappa;k}, {}^{(pri)}h_{ki}^{(m)}(i=1,...,\mathrm{I}_{\max}), e_{f;k}, {}^{(pri)}\boldsymbol{c}_k^T;\boldsymbol{\lambda}_\chi)^T$. Hence the estimator used for design pre-specification is

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

with ${}^{(1)}l_{\kappa;k}={}^{(pri)}l_{\kappa;k}$, ${}^{(1)}y_k=u_k$, ${}^{(1)}\xi_{1;k}={}^{(pri)}\xi_{1;k}$, $v_{wope;0}=-\sum_k u_k^{2\,(pri)}p_{\kappa;k}^2$, and $v_{wope;k}=u_k^{2\,(pri)}p_{\kappa;k}$ are the components of the variance under the assumptions that $u_k$ are constants, where $\xi_{1;k}=(1-\phi_{f;k})\xi_{1;k}^{(self)}+\phi_{f;k}\xi_{1;k}^{(self+f)}$, and $\xi_{1;k}^{(self+f)}$ is the probability of response under follow-up in addition to self-enumeration for unit $k$ during $\mathrm{I}$ 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**

| $N$ | $\mathrm{I}_{\max}$ | $\mathrm{I}_F$ | $C_{\max}$ | $c$ | $c_p$ | $c_f$ | $c_{dc}^{(M)}$ | $c_{dc}^{(I)}$ | $D_F$ |
|---|---|---|---|---|---|---|---|---|---|
| 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 of Data** | **Self-enumeration Only** | **With Follow-up** |
|---|---|---|

| Collection | 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 |

### 3.3 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 $\mathbf{\Phi}_{\wp}$ and $\mathbf{\Phi}_{f}$.

**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 | |
| | | | | | | | | | | $\Phi_{\wp;0}$ | $\Phi_{\wp;1}$ | $\Phi_{f;0}$ | $\Phi_{f;1}$ |
| 5 | 4174 | 0 | 413 | 8 | 5000 | 2 | 83 | 0 | 15 | 1.62 | .001 | -.27.46 | .47 |
| 10 | 3205 | 0 | 578 | 6 | 5000 | 4 | 64 | 0 | 32 | .58 | .000 | -21.22 | .33 |
| 15 | 2613 | 0 | 649 | 6 | 5000 | 6 | 52 | 0 | 42 | .09 | ..001 | -24.03 | .39 |
| 20 | 2235 | 0 | 683 | 6 | 5000 | 8 | 45 | 0 | 47 | -.21 | .000 | -148.5 | 4.12 |
| 25 | 1734 | 229 | 720 | 6 | 5000 | 10 | 35 | 11 | 44 | -.63 | .001 | -1.77 | .00 |
| 30 | 952 | 951 | 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 when the data collection period exceeds or equals 35 ($I \geq 35$).

### 3.4 Decision Step

It was decided to proceed with a data collection period of length 35 ( $I = 35$ ), $\boldsymbol{\Phi}_{\varphi} = (-1.525,.002)^T$ as the regression parameter for sampling, and $\boldsymbol{\Phi}_f = (11.87,4.28)^T$ as the regression parameter for nonresponse follow-up.

## 4. Revision after 10 Periods of Data Collection – Example Continuation

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. Table 6 displays the distribution of the cost. Table 7 represents respondents and non-respondents, 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 are errors in the prior information (e.g. domain classification and response behavior). The question is whether to decide between proceeding with the pre-specified design or whether the eventual efficiency of the estimator would be better enhanced by updating the design parameter. In Section 4.1, we optimize the objective function using the revised information; and in section 4.2, we update design parameters for the remaining time periods of data collection. Details on the revision of the prior information are given in sections 5,6, and 7.

**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) |

### 4.1 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 $\mathrm{I}$ of data collection (or equivalently the extra number of phases $\mathrm{P}$ with $10+\mathrm{I}=\sum_{\mathrm{p}=1}^{1+\mathrm{P}} n_{\mathrm{p}}$ and $n_1=10$), and the follow-up model parameter $\boldsymbol{\Phi}_f$ by minimizing the variance, $\min_{\boldsymbol{\Phi}_f} Var(\breve{Y}_{10+\mathrm{I};\kappa})$ for each remaining period of data collection, 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 $\boldsymbol{\Phi} = (\mathrm{I}, \boldsymbol{\Phi}_f^T)^T$, and $C_{10}$ is the total cost spent in the first phase, Here the estimator used for design revision is

$$^{(2)}\breve{Y}_{10+\mathrm{I};\kappa} = \sum_k d_{10+\mathrm{I};k} (^{(2)}r_{10+\mathrm{I};k} / ^{(2)}\xi_{10+\mathrm{I};k})^{(2)} I_{\kappa;k}\, ^{(2)}y_k ,$$

with $\xi_{10+\mathrm{I};k} = (1-\phi_{f;k})\xi_{10+\mathrm{I};k}^{(self)} + \phi_{f;k}\xi_{10+\mathrm{I};k}^{(self+f)}$, $\phi_{f;k} = 1 - \prod_{\mathrm{p}=1}^{2} \phi_{f;k}^{(\mathrm{p})}$, and $\phi_{f;k}^{(\mathrm{p})}$ is the conditional follow-up probability given that the unit was not followed-up. Note that $\phi_{f;k}^{(1)}$ is known, and $\phi_{f;k}^{(2)}$ is to be determined.

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 | | Extra Cost | | |
|---|---|---|---|---|---|---|---|
| | Follow-up | Respondents | CV | Extra Cost | Fixed | Follow-up | Data Collection |
| 5 | 0 | 203 | 3 | 740 | 300 | 0 | 440 |

Note that the expected coefficient of variation of .03 is less than the required one when the duration of data collection is 15(=10+5) time periods or more.

### 4.2 Decision Step

It was 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 time periods needed to complete the data collection.

## 5. Some Estimation Methods for Model Parameter

To proceed with the estimation, two elements are essential: (1) a set of known observations, and (2) a model that describes the distribution of the variable governing the observations. The purpose of the estimation methods is to find the parameter of the model that best explains the observed values. Two commonly used approaches to estimate model parameter from a random sample are the least squares estimation method and the maximum likelihood estimation method. In the least squares approach, we find the parameter of the model that yield the minimum sum of squared errors, while the purpose of the maximum likelihood approach is to find the parameter of the model that best explains the observed values in the sense of yielding the largest probability or likelihood of explaining the observed values. Thus, the least squares approach differs from the maximum likelihood approach mostly in terms of the criterion for estimating parameters: the former minimizes the sum of squared errors; the latter maximizes the probability of a model fitting the observations. A second difference is that in using maximum likelihood, one must make explicit assumption about the distribution of the variable. This precise distributional assumption is not necessary in least squares approach. In this section, the two methods of estimating the unknown parameter are discussed.

The problem of estimation reduces to find the parameter so that a criterion is to be maximized. The problem with finding the optimum of a function with respect to the parameter consists on: (1) differentiating the criterion with respect to the parameter, (2) setting this derivative to 0 to get the EE, and finally (3) solving the EE. The optimization process with respect to the model parameter is often solved using iterative methods. The iterative method continues until the convergence criteria is met, which is declared when the absolute of the estimates change is less than tolerance, while the maximum iteration number is respected. At convergence, the resulting estimate, usually denoted with a “hat”, “tilde”, etc., defines the estimator of the parameter. In this paper, we set the convergence criteria to 1E-8 and the maximum number of iteration to 500.

Solving the EE is usually difficult, if not impossible, in the cases of (1) non-linear forms such as multivariate normal distributions with explanatory variables, and artificial neural networks with multiple hidden layers, and (2) missing observations. In case of missing observations, the difficulty lies in: (1) the EE involve known observations, missing observations, and unknown model parameter, (2) the solution to the parameter requires the missing values and vice versa, and (3) the substitution of one EE into the other produces an unsolvable equation. The expectation-maximization (EM) algorithm used to find parameter estimate of models in case of missing observations is discussed in this section and applied in sections 6, and 7.

**5.1 Revision of a Design in the Course of its Progress**

When the duration of data collection period is taken into account, the likelihood function of the joint distribution under census data is defined for unit $k$ as

$$L_{\mathrm{I}_{\min};k}^{(\mathrm{I}_{\min})}(\boldsymbol{\lambda}) = f_{\mathrm{I}_{\min}}^{(\mathrm{I}_{\min})}(\boldsymbol{\psi}_k, \boldsymbol{\chi}_k), \tag{5.1}$$

where the subscript $\mathrm{I}$ in $f_{\mathrm{I}}^{(\mathrm{I}_{\min})}(\boldsymbol{\zeta})$ denotes that $\boldsymbol{\zeta}$ is observed during the interval $[0,\mathrm{I}]$. To simplify our notation we drop the superscript $\mathrm{I}_{\min}$, and write (5.1) as $L_{\mathrm{I}_{\min};k}(\boldsymbol{\lambda}) = f_{\mathrm{I}_{\min}}(\boldsymbol{\psi}_k, \boldsymbol{\chi}_k)$. The likelihood function under census case is $L(\boldsymbol{\lambda}) = \prod_k f_{\mathrm{I}_{\min}}(\boldsymbol{\psi}_k, \boldsymbol{\chi}_k)$, and the census parameter $\boldsymbol{\lambda}_{\mathrm{I}_{\min}}$ is defined as the solution to

$$\mathbf{S}_{\mathrm{I}_{\min}}(\boldsymbol{\lambda}) = \partial \log L(\boldsymbol{\lambda}) / \partial \boldsymbol{\lambda} = \textstyle\sum_k \partial \log f_{\mathrm{I}_{\min}}(\boldsymbol{\psi}_k, \boldsymbol{\chi}_k) / \partial \boldsymbol{\lambda} = \mathbf{0}. \tag{5.2}$$

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 $\mathrm{I}$ time periods of data collection, the joint observations on $\boldsymbol{\chi}$, and $\boldsymbol{\psi}$ are known during the $\mathrm{I}$ time periods, while only observations on $\boldsymbol{\chi}$ are known for the rest of the periods. Consequently, we decompose the likelihood of observed data for unit $k$ in two parts

$$L_{\mathrm{I};k}(\boldsymbol{\lambda}) = f_{\mathrm{I}_{mix}}(\boldsymbol{\chi}_k) f_{\mathrm{I}}(\boldsymbol{\psi}_k \mid \boldsymbol{\chi}_k), \tag{5.3}$$

in which case the log-likelihood is given by

$$\ell_{\mathrm{I};k}(\boldsymbol{\lambda}_{\min}) = \log f_{\mathrm{I}_{mix}}(\boldsymbol{\chi}_k) + \log f_{\mathrm{I}}(\boldsymbol{\psi}_k \mid \boldsymbol{\chi}_k), \tag{5.4}$$

where $f_{\mathrm{I}_{\min}}(\boldsymbol{\chi}_k) = \int f_{\mathrm{I}_{\min}}(\boldsymbol{\chi}_k, \boldsymbol{\psi}_k) d\boldsymbol{\psi}_k$, and $f_{\mathrm{I}}(\boldsymbol{\psi}_k \mid \boldsymbol{\chi}_k) = f_{\mathrm{I}}(\boldsymbol{\chi}_k, \boldsymbol{\psi}_k) / f_{\mathrm{I}}(\boldsymbol{\chi}_k)$. Note that $f_{\mathrm{I}_{\min}}(\boldsymbol{\chi}_k) f_{\mathrm{I}}(\boldsymbol{\psi}_k \mid \boldsymbol{\chi}_k) \to f_{\mathrm{I}_{\min}}(\boldsymbol{\psi}_k, \boldsymbol{\chi}_k)$ as $\mathrm{I} \to \mathrm{I}_{\min}$. In arriving to (5.3), we decomposed $\boldsymbol{\chi}$ in two parts: $\boldsymbol{\chi} = (\boldsymbol{\chi}_{[e_{dc;k},\mathrm{I}]}^T, \boldsymbol{\chi}_{[\mathrm{I}+1,\mathrm{I}_{\min}]}^T)^T$, where $\boldsymbol{\chi}_{[e_{dc;f},\mathrm{I}]}$ denotes observation during the interval $[e_{dc;k},\mathrm{I}]$, while $\boldsymbol{\chi}_{[\mathrm{I}+1,\mathrm{I}_{\min}]}$ denotes observation during the interval $[\mathrm{I}+1,\mathrm{I}_{\min}]$, and $e_{dc;k}$ denotes the entry time period for unit $k$ into data collection window. The joint distribution is given by

$$f(\boldsymbol{\chi}, \boldsymbol{\psi}) = f(\boldsymbol{\chi}_{[e_{dc;k},\mathrm{I}]}, \boldsymbol{\chi}_{[\mathrm{I}+1,\mathrm{I}_{\min}]}, \boldsymbol{\psi}) = f(\boldsymbol{\chi}_{[e_{dc;k},\mathrm{I}]}) f(\boldsymbol{\chi}_{[\mathrm{I}+1,\mathrm{I}_{\min}]} \mid \boldsymbol{\chi}_{[e_{dc;k},\mathrm{I}]}) f(\boldsymbol{\psi} \mid \boldsymbol{\chi}_{[e_{dc;k},\mathrm{I}]}) = f(\boldsymbol{\chi}) f(\boldsymbol{\psi} \mid \boldsymbol{\chi}_{[e_{dc;k},\mathrm{I}]}).$$

Taking the derivatives of (5.3) and adjusting for unequal probability of selection and response, we get the weighted EE

$$\breve{\mathbf{S}}_{\mathrm{I}}(\boldsymbol{\lambda}) = \textstyle\sum_k \mathbf{s}_{\mathrm{I}_{\min}}(\boldsymbol{\chi}_k; \boldsymbol{\lambda}) + \sum_k w_{\mathrm{I};k} \mathbf{s}_{\mathrm{I}}(\boldsymbol{\psi}_k; \boldsymbol{\lambda} \mid \boldsymbol{\chi}_k) = \mathbf{0}, \tag{5.5}$$

with

$$\mathbf{s}_{\mathrm{I}_{\min}}(\boldsymbol{\chi}_k; \boldsymbol{\lambda}) = \partial \log f_{\mathrm{I}_{\min}}(\boldsymbol{\chi}_k) / \partial \boldsymbol{\lambda}, \text{ and } \mathbf{s}_{\mathrm{I}}(\boldsymbol{\psi}_k; \boldsymbol{\lambda} \mid \boldsymbol{\chi}_k) = \partial \log f_{\mathrm{I}}(\boldsymbol{\psi}_k \mid \boldsymbol{\chi}_k) / \partial \boldsymbol{\lambda}, \tag{5.6}$$

where $r_{\mathrm{I};k}$ is the response indicator during the interval $[e_{dc;k},\mathrm{I}]$, and $\xi_{\mathrm{I};k} = E(r_{\mathrm{I};k})$ denotes the marginal probability of obtaining a response during $[e_{dc;k},\mathrm{I}]$. The EE given by (5.5) is unbiased for

the census EE given by (5.2), i.e., $E_r\{\breve{\mathbf{S}}_{\mathrm{I}}(\boldsymbol{\lambda})\} = \mathbf{S}_{\mathrm{I}_{\min}}(\boldsymbol{\lambda})$. The solution to (5.5) obtained by Newton-Raphson-type iterative method or the expectation maximization type algorithm gives the estimator $\breve{\boldsymbol{\lambda}}_{\mathrm{I}}$ of $\boldsymbol{\lambda}_{\mathrm{I}_{\min}}$.

Taylor linearization of $\breve{\mathbf{S}}_{\mathrm{I}}(\breve{\boldsymbol{\lambda}}_{\mathrm{I}})$ around $\boldsymbol{\lambda}^{(0)}$ gives the approximation

$$\mathbf{0} = \breve{\mathbf{S}}_{\mathrm{I}}(\breve{\boldsymbol{\lambda}}_{\mathrm{I}}) \approx \breve{\mathbf{S}}_{\mathrm{I}}(\boldsymbol{\lambda}^{(0)}) - \partial\breve{\mathbf{S}}_{\mathrm{I}}^{T}(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}\big|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^{(0)}} (\breve{\boldsymbol{\lambda}}_{\mathrm{I}} - \boldsymbol{\lambda}_{\mathrm{I}}),$$

or

$$\breve{\boldsymbol{\lambda}}_{\mathrm{I}} \approx \boldsymbol{\lambda}^{(0)} + \{\breve{\mathbf{J}}_{\mathrm{I}}(\boldsymbol{\lambda}^{(0)})\}^{-1}\breve{\mathbf{S}}_{\mathrm{I}}(\boldsymbol{\lambda}^{(0)}),$$

where $\breve{\mathbf{J}}_{\mathrm{I}}(\boldsymbol{\lambda}) = -\partial\breve{\mathbf{S}}_{\mathrm{I}}^{T}(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}$. Starting with a guessed value, $\boldsymbol{\lambda}^{(0)} = \boldsymbol{\lambda}_{\text{p-1}}$, then for $b = 1,2,...$ updates are made using Newton-Raphson method as follows

$$\boldsymbol{\lambda}^{(b)} = \boldsymbol{\lambda}^{(b-1)} + \{\breve{\mathbf{J}}_{\mathrm{I}}(\boldsymbol{\lambda}^{(b-1)})\}^{-1}\breve{\mathbf{S}}_{\mathrm{I}}(\boldsymbol{\lambda}^{(b-1)}).$$

### 5.2 The Expectation Maximization Algorithm

The Expectation Maximization (EM) algorithm introduced by Hartley (1958)—formalized and termed by Dempster et al. (1977) — has become a major tool for finding maximum likelihood estimate of $\boldsymbol{\lambda}$ 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 an 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 $f(\mathbf{d}_{o;}; \boldsymbol{\lambda})$. The objective is to estimate $\boldsymbol{\lambda}$ by $\hat{\boldsymbol{\lambda}} = \arg\max \ell(\mathbf{d}_{o;}; \boldsymbol{\lambda})$, where $\ell(\mathbf{d}_{o;}; \boldsymbol{\lambda}) = \log f(\mathbf{d}_{o;}; \boldsymbol{\lambda})$. Let $f(\mathbf{d}_{m;} \mid \mathbf{d}_{o;}; \boldsymbol{\lambda}) = f(\mathbf{d}; \boldsymbol{\lambda})/f(\mathbf{d}_{o;}; \boldsymbol{\lambda})$ 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 $\boldsymbol{\lambda}^{(e)}$, the E-step of the EM algorithm requires the calculation of a function of $\boldsymbol{\lambda}$, $Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}^{(e)})$, such that

$$Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}^{(e)}) = E\{\ell_c(\mathbf{d}; \boldsymbol{\lambda}) \mid \mathbf{d}_{o;}; \boldsymbol{\lambda}^{(e)}\} = \int \log f(\mathbf{d}; \boldsymbol{\lambda}) f(\mathbf{d}_m \mid \mathbf{d}_o; \boldsymbol{\lambda}^{(e)}) \partial\mathbf{d}_m, \tag{5.7}$$

where $\ell_c(\mathbf{d}; \boldsymbol{\lambda}) = \log f(\mathbf{d}; \boldsymbol{\lambda})$, $\boldsymbol{\lambda}$ is the parameter of interest, and $\boldsymbol{\lambda}^{(e)}$ 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 $\boldsymbol{\lambda}^{(e+1)}$, that maximizes $Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}^{(e)})$, i.e., $\hat{\boldsymbol{\lambda}}^{(e+1)} = \arg\max_{\boldsymbol{\lambda}} Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}^{(e)})$. If we iterate the E-step and M-step until convergence, under regularity conditions, the algorithm converges to the maximum likelihood estimate. Substituting $f(\mathbf{d}_m \mid \mathbf{d}_o; \boldsymbol{\lambda}^{(e)}) = f(\mathbf{d}; \boldsymbol{\lambda}^{(e)})/f(\mathbf{d}_o; \boldsymbol{\lambda}^{(e)})$ into (5.7), we get

$$Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}^{(e)}) = \{f(\mathbf{d}_o; \boldsymbol{\lambda}^{(e)})\}^{-1} \int \log f(\mathbf{d}; \boldsymbol{\lambda}) f(\mathbf{d}; \boldsymbol{\lambda}^{(e)}) \partial\mathbf{d}_m.$$

The M-step gives

$$\partial Q(\boldsymbol{\lambda}, \boldsymbol{\lambda}^{(e)}) / \partial \boldsymbol{\lambda} = \int \tau_k^{(e)} \{\partial \log f(\mathbf{d}; \boldsymbol{\lambda}) / \partial \boldsymbol{\lambda}\} \partial \mathbf{d}_m ,$$

where $\tau_k^{(e)} = f(\mathbf{d}; \boldsymbol{\lambda}^{(e)}) / f(\mathbf{d}_o; \boldsymbol{\lambda}^{(e)})$ .

As noticed by Anderson and Hinde (1988), the EM algorithm based on the complete data likelihood gives an iterative procedure that maximizes the marginal likelihood of $\boldsymbol{\chi}$ . In particular, we may write (5.6) as

$$\begin{aligned} \mathbf{s}_{\mathrm{I}_{\min}}(\boldsymbol{\lambda}; \boldsymbol{\chi}_k) &= \partial \log f_{\mathrm{I}_{\min}}(\boldsymbol{\chi}_k) / \partial \boldsymbol{\lambda} = \{f_{\mathrm{I}_{\min}}(\boldsymbol{\chi}_k)\}^{-1} \int f_{\mathrm{I}_{\min}}(\boldsymbol{\chi}_k, \boldsymbol{\psi}_k) \{\partial \log f_{\mathrm{I}_{\min}}(\boldsymbol{\chi}_k, \boldsymbol{\psi}_k) / \partial \boldsymbol{\lambda}\} d\boldsymbol{\psi} \\ &= \int \tau_{\mathrm{I}_{\min};k} \{\partial \log f_{\mathrm{I}_{\min}}(\boldsymbol{\chi}_k, \boldsymbol{\psi}_k) / \partial \boldsymbol{\lambda}\} d\boldsymbol{\psi}, \end{aligned}$$

and

$$\begin{aligned} \mathbf{s}_1(\boldsymbol{\lambda}; \boldsymbol{\psi}_k \mid \boldsymbol{\chi}_k) &= \partial \log f_1(\boldsymbol{\psi}_k \mid \boldsymbol{\chi}_k) / \partial \boldsymbol{\lambda} = \partial \log f_1(\boldsymbol{\psi}_k, \boldsymbol{\chi}_k) / \partial \boldsymbol{\lambda} - \partial \log f_1(\boldsymbol{\chi}_k) / \partial \boldsymbol{\lambda} \\ &= \partial \log f_1(\boldsymbol{\psi}_k, \boldsymbol{\chi}_k) / \partial \boldsymbol{\lambda} - \int \tau_{1;k} \{\partial \log f_1(\boldsymbol{\chi}_k, \boldsymbol{\psi}_k) / \partial \boldsymbol{\lambda}\} d\boldsymbol{\psi}, \end{aligned}$$

where $\tau_{1;k} = f_1(\boldsymbol{\chi}_k, \boldsymbol{\psi}_k) / f_1(\boldsymbol{\chi}_k)$ . We have assumed that the differential can be taken inside the integral sign.

## 6. Revision of the Classification Model Parameter

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 obtain, though it is expensive. Misclassification assumes 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.

### 6.1 Basic Aspects of the Bernoulli Distribution

The Bernoulli random variable $y$ is one with binary outcomes chosen from $\{0,1\}$ . Denote $p_i = \Pr(y = i)$ , $i = 0,1$ , with $\sum_{i=0}^{1} p_i = 1$ , then its probability density function is

$$f(y) = p_1^{y} p_0^{1-y} .$$

Next, consider the bivariate Bernoulli random vector $(y_1, y_2)$ , which takes values from $(0,0)$ , $(1,0)$ , $(0,1)$ , and $(1,1)$ . Denote $p_{ij} = \Pr(y_1 = i, y_j = j)$ , $i, j = 0,1$ , with $\sum_{i=0}^{1} \sum_{j=0}^{1} p_{ij} = 1$ , then the probability density function of $(y_1, y_2)$ can be written as

$$f(y_1, y_2) = p_{11}^{y_1 y_2} p_{10}^{y_1(1-y_2)} p_{01}^{(1-y_1)y_2} p_{00}^{(1-y_1)(1-y_2)} . \qquad (6.1)$$

The marginal distribution of $y_1$ in a bivariate Bernoulli vector $(y_1, y_2)$ is univariate Bernoulli with probability $(p_{10} + p_{11})$, and the conditional distribution of $y_1$ given $y_2$ is also univariate Bernoulli with density

$$f(y_1 \mid y_2) = f(y_1, y_2) / f(y_2).$$

When there are $N$ observations from a population with outcomes denoted as $\boldsymbol{y}_1 = (y_{1;1}, y_{2;1}), ..., \boldsymbol{y}_k = (y_{1;k}, y_{2;k}), ..., \boldsymbol{y}_N = (y_{1;N}, y_{2;N})$, the log likelihood of the complete data is

$$\ell_c(\boldsymbol{\lambda}; \boldsymbol{y}) = \log L_c(\boldsymbol{\lambda}; \boldsymbol{y}) = \sum_k \log f(y_{1;k}, y_{2;k}).$$

We now consider the E-step on the $(e+1)$ th iteration of the EM algorithm, where $\boldsymbol{\lambda}^{(e)}$ denote the value of $\boldsymbol{\lambda}$ after the $e^{th}$ EM iteration. Suppose now that values of the second variate $y_2$ is sometime missing, the current conditional expectation of the complete data log-likelihood is

$$Q(\boldsymbol{\lambda}; \boldsymbol{\lambda}^{(e)}) = E\{\ell_c(\boldsymbol{\lambda}; \boldsymbol{y}) \mid y_1, \boldsymbol{\lambda}^{(e)}\} = \sum_k \sum_{j=0}^{1} \tau_k^{(j|y_{1;k})(e)} \log f(y_{1;k}, j),$$

where $$\tau_k^{(j|y_{1;k})(e)} = \begin{cases} 1(j = y_{2;k}) & \text{if } y_{2;k} \text{ is observed} \\ \Pr(y_{2;k} \mid y_{1;k}, \boldsymbol{\lambda}^{(e)}) & \text{if not.} \end{cases}$$

### 6.2 Census Estimating Equation

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

$$\mathbf{P}_{\kappa;k} = \begin{pmatrix} 1 - p_{\kappa;k}^{(l|0)} & 1 - p_{\kappa;k}^{(l|1)} \\ p_{\kappa;k}^{(l|0)} & p_{\kappa;k}^{(l|1)} \end{pmatrix}.$$

The error-prone marginal probability ${}^{(er)}p_{\kappa;k}$ is given by

$${}^{(er)}p_{\kappa;k} = p_{\kappa;k}^{(l|1)} p_{\kappa;k} + p_{\kappa;k}^{(l|0)} (1 - p_{\kappa;k}),$$

where $p_{\kappa;k} = \Pr(l_{\kappa;k} = 1)$.

The census parameter is defined as the solution of

$$\mathbf{S}_{\mathrm{I}_{\min}}(\boldsymbol{\lambda}_\kappa) = \sum_k \partial \log f_{\mathrm{I}_{\min}}({}^{(er)}l_{\kappa;k}, l_{\kappa;k}) / \partial \boldsymbol{\lambda}_\kappa = \mathbf{0}, \tag{6.2}$$

where the joint distribution of $({}^{(er)}l_{\kappa;k}, l_{\kappa;k})$ is given by (6.1) with $(y_1, y_2) = ({}^{(er)}l_{\kappa;k}, l_{\kappa;k})$, $p_{11} = p_{\kappa;k} p_{\kappa;k}^{(l|1)}$, $p_{10} = p_{\kappa;k}(1 - p_{\kappa;k}^{(l|1)})$, $p_{01} = (1 - p_{\kappa;k}) p_{\kappa;k}^{(l|0)}$, and $p_{00} = (1 - p_{\kappa;k})(1 - p_{\kappa;k}^{(l|0)})$. Subsisting the expression of $f({}^{(er)}l_{\kappa;k}, l_{\kappa;k})$ into (6.2) and taking the derivative, we get

$$\mathbf{S}_{\mathrm{I}_{\min}}(\boldsymbol{\lambda}_\kappa) = \sum_k \mathbf{s}_{\mathrm{I}_{\min}}(\boldsymbol{\lambda}_\kappa; {}^{(er)}l_{\kappa;k}, l_{\kappa;k}) = \mathbf{0}\,, \tag{6.3}$$

with $$\mathbf{s}_{\mathrm{I}_{\min}}(\boldsymbol{\lambda}_\kappa; {}^{(er)}l_{\kappa;k}, l_{\kappa;k}) = \mathbf{s}_{\mathrm{I}_{\min}}(\boldsymbol{\lambda}_\kappa; l_{\kappa;k}) + \mathbf{s}_{\mathrm{I}_{\min}}(\boldsymbol{\lambda}_\kappa; l^{(1|1)}_{\kappa;k}) + \mathbf{s}(\boldsymbol{\lambda}_\kappa; l^{(1|0)}_{\kappa;k})\,,$$

where $$\mathbf{s}_{\mathrm{I}_{\min}}(\boldsymbol{\lambda}_\kappa; l_{\kappa;k}) = \dot{\boldsymbol{p}}_{\kappa;k}(l_{\kappa;k} - p_{\kappa;k})\{p_{\kappa;k}(1-p_{\kappa;k})\}^{-1}\,,$$

$$\mathbf{s}_{\mathrm{I}_{\min}}(\boldsymbol{\lambda}_\kappa; l^{(1|1)}_{\kappa;k}) = l_{\kappa;k}\left\{\dot{\boldsymbol{p}}^{(1|1)}_{\kappa;k}({}^{(er)}l_{\kappa;k} - p^{(1|1)}_{\kappa;k})\{p^{(1|1)}_{\kappa;k})(1-p^{(1|1)}_{\kappa;k})\}^{-1}\right\},$$

$$\mathbf{s}_{\mathrm{I}_{\min}}(\boldsymbol{\lambda}_\kappa; l^{(1|0)}_{\kappa;k}) = (1-l_{\kappa;k})\left\{\dot{\boldsymbol{p}}^{(1|0)}_{\kappa;k}({}^{(er)}l_{\kappa;k} - p^{(1|0)}_{\kappa;k})\{p^{(1|0)}_{\kappa;k})(1-p^{(1|0)}_{\kappa;k})\}^{-1}\right\},$$

$\dot{\boldsymbol{p}}_{\kappa;k} = \partial p_{\kappa;k}/\partial\boldsymbol{\lambda}_\kappa$, $\dot{\boldsymbol{p}}^{(1|1)}_{\kappa;k} = \partial p^{(1|1)}_{\kappa;k}/\partial\boldsymbol{\lambda}_\kappa$, and $\dot{\boldsymbol{p}}^{(1|0)}_{\kappa;k} = \partial p^{(1|0)}_{\kappa;k}/\partial\boldsymbol{\lambda}_\kappa$.

### 6.3 Sample Estimating Equation

After observing I periods of data collection, an estimator of the census parameter based on only observed values from respondents is the solution to the following sample EE

$$\breve{\mathbf{S}}^{(r)}_{\mathrm{I}}(\boldsymbol{\lambda}_\kappa) = \sum_k w_{\mathrm{I};k}\mathbf{s}_{\mathrm{I}}(\boldsymbol{\lambda}_\kappa; {}^{(er)}l_{\kappa;k}, l_{\kappa;k}) = \mathbf{0}\,. \tag{6.4}$$

The solution to (6.4) is denoted by $\breve{\boldsymbol{\lambda}}^{(r)}_\kappa$, where the superscript $r$ stands for respondents.

In order to use all error prone values in addition to the error free observed values during I periods of data collection, we follow section (5.1) and decompose the joint distribution for unit $k$ in two parts

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

Taking the derivatives of the logarithm of the likelihood function $L(\boldsymbol{\lambda}_\kappa) = \prod_k f_{\mathrm{I}_{\min}}({}^{(er)}l_{\kappa;k}) f_{\mathrm{I}}(l_{\kappa;k}|{}^{(er)}l_{\kappa;k})$ and adjusting for the unequal selection probabilities and response probabilities, we get the sample EE

$$\begin{aligned}\breve{\mathbf{S}}^{(\varphi)}_{\mathrm{I}}(\boldsymbol{\lambda}_\kappa) &= \sum_k d_{\mathrm{I};k}\sum_{j=0}^{1}\tau^{(j|{}^{(er)}l_{\kappa;k})}_{\mathrm{I}_{\min};k}\mathbf{s}_{\mathrm{I}_{\min}}(\boldsymbol{\lambda}_\kappa; {}^{(er)}l_{\kappa;k}, j)\\ &+ \sum_k w_{\mathrm{I};k}\{\mathbf{s}_{\mathrm{I}}(\boldsymbol{\lambda}_\kappa; {}^{(er)}l_{\kappa;k}, l_{\kappa;k}) - \sum_{j=0}^{1}\tau^{(j|{}^{(er)}l_{\kappa;k})}_{\mathrm{I};k}\mathbf{s}_{\mathrm{I}}(\boldsymbol{\lambda}_\kappa; {}^{(er)}l_{\kappa;k}, j)\} = \mathbf{0},\end{aligned} \tag{6.5.a}$$

or

$$\begin{aligned}\breve{\mathbf{S}}^{(P)}_{\mathrm{I}}(\boldsymbol{\lambda}_\kappa) &= \sum_k \sum_{j=0}^{1}\tau^{(j|{}^{(er)}l_{\kappa;k})}_{\mathrm{I}_{\min};k}\mathbf{s}_{\mathrm{I}_{\min}}(\boldsymbol{\lambda}_\kappa; {}^{(er)}l_{\kappa;k}, j)\\ &+ \sum_k w_{\mathrm{I};k}\{\mathbf{s}_{\mathrm{I}}(\boldsymbol{\lambda}_\kappa; {}^{(er)}l_{\kappa;k}, l_{\kappa;k}) - \sum_{j=0}^{1}\tau^{(j|{}^{(er)}l_{\kappa;k})}_{\mathrm{I};k}\mathbf{s}_{\mathrm{I}}(\boldsymbol{\lambda}_\kappa; {}^{(er)}l_{\kappa;k}, j)\} = \mathbf{0}.\end{aligned} \tag{6.5.b}$$

The solution to (6.5) gives an estimator of the census parameter. Note that (6.5.a) uses error prone information from sampled units only, while (6.5.b) uses all error prone information from the frame. We may write (6.5) as

$$\breve{\mathbf{S}}^{(\varphi)}_{\mathrm{I}}(\boldsymbol{\lambda}_\kappa) = \sum_k w_{\mathrm{I};k}\mathbf{s}_{\mathrm{I}}(\boldsymbol{\lambda}_\kappa; {}^{(er)}l_{\kappa;k}, l_{\kappa;k}) + \sum_k (d_{\mathrm{I};k} - w_{\mathrm{I};k})\sum_{j=0}^{1}\tau^{(j|{}^{(er)}l_{\kappa;k})}_{\mathrm{I}_{\min};k}\mathbf{s}_{\mathrm{I}_{\min}}(\boldsymbol{\lambda}_\kappa; {}^{(er)}l_{\kappa;k}, j) = \mathbf{0}, \tag{6.6.a}$$

or

$$\breve{\mathbf{S}}^{(P)}_{\mathrm{I}}(\boldsymbol{\lambda}_\kappa) = \sum_k w_{\mathrm{I};k}\mathbf{s}_{\mathrm{I}}(\boldsymbol{\lambda}_\kappa; {}^{(er)}l_{\kappa;k}, l_{\kappa;k}) + \sum_k (1 - w_{\mathrm{I};k})\sum_{j=0}^{1}\tau^{(j|{}^{(er)}l_{\kappa;k})}_{\mathrm{I}_{\min};k}\mathbf{s}_{\mathrm{I}_{\min}}(\boldsymbol{\lambda}_\kappa; {}^{(er)}l_{\kappa;k}, j) = \mathbf{0}. \tag{6.6.b}$$

### 6.4 The EM Algorithm Estimator

Similarly, the EM algorithm can be applied to both the frame and the sample information. In both cases, the E step of the EM algorithm involves creating a set of "pseudo-data" in which the respondents are left intact and the non-respondents are fractionated into 2 partially complete pseudo-observations. The weight assigned to this pseudo-observation is the conditional probability that unit belongs to an associated domain given the observed data and prior estimation of the parameters. Once $\boldsymbol{\lambda}_{\kappa}^{(e)}$ has been obtained, estimates of the conditional probabilities can be formed for each $k$. The conditional probability that a non-respondent $k$ belongs to domain $j \in \{0,1\}$ is given by $\tau_k^{(j|^{(er)}l_{\kappa;k})(e)}$, where $\tau_k^{(j|^{(er)}l_{\kappa;k})(e)}$ is $\tau_k^{(j|^{(er)}l_{\kappa;k})}$ evaluated at $\boldsymbol{\lambda}_{\kappa}^{(e)}$. Estimation of $\boldsymbol{\lambda}_{\kappa}$ and $\tau_k^{(j|^{(er)}l_{\kappa;k})}$ are alternated repeatedly, where in their subsequent execution, the initial fit $\boldsymbol{\lambda}_{\kappa}^{(e)}$ is replaced by the current fit $\boldsymbol{\lambda}_{\kappa}^{(e+1)}$ for $\boldsymbol{\lambda}_{\kappa}$. The two versions of the current conditional expectation of the complete data log-likelihood are

$$\hat{Q}^{(\varphi)}(\boldsymbol{\lambda}_{\kappa};\boldsymbol{\lambda}_{\kappa}^{(e)}) = \sum_k d_{1;k} \sum_{j=0}^{1} \tau_k^{(j|^{(er)}l_{\kappa;k})(e)} \log f(\boldsymbol{\lambda}_{\kappa};{}^{(er)}l_{\kappa;k}, j), \tag{6.7.a}$$

and

$$Q^{(P)}(\boldsymbol{\lambda}_{\kappa};\boldsymbol{\lambda}_{\kappa}^{(e)}) = \sum_k \sum_{j=0}^{1} \tau_k^{(j|^{(er)}l_{\kappa;k})(e)} \log f(\boldsymbol{\lambda}_{\kappa};{}^{(er)}l_{\kappa;k}, j). \tag{6.7.b}$$

### 6.5 Simulation Study

We conducted a small simulation study to illustrate the performances of each estimator of the classification model parameter under (6.4), (6.5.a) and (6.7.a). We first used values $u_k$ and $e_{f;k}$ for each unit $k$ of the finite population of size $N = 2000$. Then, we generated the domain $\kappa$ membership indicator $l_{\kappa;k}$ from Bernoulli distribution with probability of membership $p_{\kappa;k} = \exp(\boldsymbol{v}_{\kappa;k}^T \boldsymbol{\lambda}_{\kappa}) / \{1 + \exp(\boldsymbol{v}_{\kappa;k}^T \boldsymbol{\lambda}_{\kappa})\}$ with $\boldsymbol{v}_{\kappa;k} = (1, u_k)^T$ and $\boldsymbol{\lambda}_{\kappa} = (1,.03)^T \equiv (\lambda_{\kappa}^{(1)}, \lambda_{\kappa}^{(2)})^T$. Then we generated the error-prone indicator ${}^{(er)}l_{\kappa;k}$ from the conditional Bernoulli distribution with probability of membership satisfying $logitp_k^{(1|l_{\kappa})} = l_{\kappa;} + (1 - l_{\kappa;}) p_k^{(1|0)}$ with $logitp_k^{(1|1)} = \boldsymbol{v}_{\kappa;k}^T \boldsymbol{\lambda}_{\kappa\kappa}^{(1|1)}$, $logitp_k^{(1|0)} = \boldsymbol{v}_{\kappa;k}^T \boldsymbol{\lambda}_{\kappa\kappa}^{(1|0)}$ $\boldsymbol{\lambda}_{\kappa\kappa}^{(1|1)} = (1,.01)^T \equiv (\lambda_{\kappa}^{(3)}, \lambda_{\kappa}^{(4)})^T$ and $\boldsymbol{\lambda}_{\kappa\kappa}^{(1|0)} = (-1,.05)^T \equiv (\lambda_{\kappa}^{(5)}, \lambda_{\kappa}^{(6)})^T$. Table 11 displays values of $\sum_k {}^{(er)}l_{\kappa;k} l_{\kappa;k}$, $\sum_k {}^{(er)}l_{\kappa;k}$ and $\sum_k l_{\kappa;k}$.

**Table 11: Values of $\sum_k {}^{(er)}l_{\kappa;k} l_{\kappa;k}$, $\sum_k {}^{(er)}l_{\kappa;k}$ and $\sum_k l_{\kappa;k}$**

| | **Error Free Classification $l_{\kappa;k}$** |
|---|---|

| **Error-prone** | | | 0 | 1 | **Total** |
|---|---|---|---|---|---|
| **Classification** | $^{(er)}l_{\kappa;k}$ | 0 | 177 | 396 | 573 |
| | | 1 | 177 | 1250 | 1427 |
| | **Total** | | 354 | 1646 | 2000 |

We maintained the population values $(u_k, e_{f;k}, {}^{(er)}l_{\kappa;k}, l_{\kappa;k})$ fix for $k = 1,...,N$, and we selected $A = 1000$ Poisson samples each with probability of selection $p_{\wp;k} = \exp(\boldsymbol{v}_{\wp;k}^T \boldsymbol{\Phi}_\wp)/\{1+\exp(\boldsymbol{v}_{\wp;k}^T \boldsymbol{\Phi}_\wp)\}$ from the population with $\boldsymbol{v}_{\wp;k} = (1, l_{\kappa;k})^T$ and $\boldsymbol{\Phi}_\wp = (-1.525, -.3)^T$. The process of generating respondents and nonrespondents from the simulated population is as follows. Self-enumeration response indicators for sampled unit $k$ are generated using $r_{ki} \sim M_3(1, h_{M;ki}, h_{I;ki}, h_{O;ki})$, with $h_{O;ki} = (1+\exp(\boldsymbol{v}_{rM;ki}^T \boldsymbol{\lambda}_r) + \exp(\boldsymbol{v}_{rI;ki}^T \boldsymbol{\lambda}_r))^{-1}$, $h_{m;ki} = h_{O;ki} \times \exp(\boldsymbol{v}_{rm;ki}^T \boldsymbol{\lambda}_r)$, and $\boldsymbol{v}_{rm;ki}^T \boldsymbol{\lambda}_r = \lambda_{r;0}^{(m)} + \lambda_{r;1}^{(m)} t + \lambda_{r;2}^{(m)} u_k$ for $m \in \{M, I\}$ and $i = 1,..., \mathrm{I}_{\max}$. For unit $k$ under follow-up data collection in addition to self-enumeration we used $\boldsymbol{v}_{rm;ki}^T \boldsymbol{\lambda}_r = \lambda_{r;0}^{(m)} + \lambda_{r;1}^{(m)} t + \lambda_{r;2}^{(m)} u_k + (i - e_{f;k} + 1)\lambda_{r;3}^{(m)}$ for $i = e_k,..., e_k + D_F$. Table 12 displays values of the response model parameter $\boldsymbol{\lambda}_r = (\boldsymbol{\lambda}_r^{(M)T}, \boldsymbol{\lambda}_r^{(I)T})^T$, where $\boldsymbol{\lambda}_r^{(M)} = (\lambda_{r;0}^{(M)}, \lambda_{r;1}^{(M)}, \lambda_{r;2}^{(M)}, \lambda_{r;3}^{(M)})^T$ and $\boldsymbol{\lambda}_r^{(I)} = (\lambda_{r;0}^{(I)}, \lambda_{r;1}^{(I)}, \lambda_{r;2}^{(I)}, \lambda_{r;3}^{(I)})^T$.

**Table 12: Response Model Parameter**

| | **Mail** | **Internet** |
|---|---|---|
| $\lambda_{r;0}$ | -4 | -5 |
| $\lambda_{r;1}$ | .008 | .005 |
| $\lambda_{r;2}$ | .005 | .01 |
| $\lambda_{r;3}$ | 6 | 6 |

The probability of a follow-up is model as $p_{f;k} = \exp(\boldsymbol{v}_{f;k}^T \boldsymbol{\Phi}_f)/\{1+\exp(\boldsymbol{v}_{f;k}^T \boldsymbol{\Phi}_f)\}$ with $\boldsymbol{v}_{f;k} = (1, u_k)^T$ and $\boldsymbol{\Phi}_f = (-2, 2)^T$. Table 13 displays statistics on the realized samples.

**Table 13: Statistics on the Realized Samples**

| | **Average over samples** |
|---|---|
| **Sample Size** | 426 |
| **Pop. Size Estimate** | 1999 |
| **# Follow-ups** | 50 |
| **# Respondents** | 146 |
| **By Mail** | 107 |
| **By Internet** | 39 |

The vector parameter of interest is $\boldsymbol{\lambda}_\kappa$, the parameter associated with the model generating the domain membership indicators. Table 14 displays values of the classification model parameter $\boldsymbol{\lambda}_\kappa$.

Let $\hat{\theta}$ denote an estimator of the parameter of interest $\theta$. We calculated $\hat{\theta}$ from each repetition $a$ ($a = 1,..., A$), and its average $\bar{\hat{\theta}} = A^{-1} \sum_{a=1}^{A} \hat{\theta}_a$, where $\hat{\theta}_a$ is the value of $\hat{\theta}$ for the $a^{th}$ sample. The

simulated bias is calculated as $B(\hat{\theta}) = (\bar{\hat{\theta}} - \theta)$. We calculated $\bar{\hat{\theta}}$ and $B(\hat{\theta})$ for each component of the parameter of interest and those values are reported in Table 14. Table 14 clearly demonstrates that the bias is small for each response model parameter. The MSE of $\hat{\theta}$ is calculated as $M(\hat{\theta}) = A^{-1}\sum_{a=1}^{A}(\hat{\theta}_a - \theta)^2$. We calculated MSE ratios for each estimator $\hat{\theta}$ with the estimator solution to (5.4) and those values are reported in Table 14. Table 14 clearly indicates that all relative biases are small. Estimator using more prior information is more efficient than an estimator using only respondents. Our proposed estimator rivals the EM estimator under correct specification of the model generating the finite population; while it is more robust than the EM estimator under failure of the model' specification.

**Table 14: Estimator Average and Bias ( $\bar{\hat{\theta}}$ , $B(\hat{\theta})$ ); and MSE ratios $M(\hat{\theta})/M(\hat{\lambda}^{(o)})$**

| | | Parameter | | | | | |
|---|---|---|---|---|---|---|---|
| | | $\lambda_\kappa^{(1)}$ | $\lambda_\kappa^{(2)}$ | $\lambda_\kappa^{(3)}$ | $\lambda_\kappa^{(4)}$ | $\lambda_\kappa^{(5)}$ | $\lambda_\kappa^{(6)}$ |
| **EE** | **Statistics** | 1.0 | .03 | 1.00 | -.01 | -1.0 | .05 |
| (6.4) | $\bar{\hat{\theta}}$ | .95 | .03 | 1.04 | .006 | -1.02 | .09 |
| | $B(\hat{\theta})$ | -.05 | .007 | .04 | .016 | -.02 | .04 |
| | $M(\hat{\theta})/M(\hat{\lambda}^{(r)})$ | 1 | 1 | 1 | 1 | 1 | 1 |
| (6.5a) | $\bar{\hat{\theta}}$ | .95 | .03 | 1.00 | .007 | -1.01 | .09 |
| | $B(\hat{\theta})$ | -.04 | .006 | .004 | .02 | -.01 | .04 |
| | $M(\hat{\theta})/M(\hat{\lambda}^{(r)})$ | .04 | .02 | .004 | .02 | .01 | .03 |
| (6.7a) | $\bar{\hat{\theta}}$ | 1.00 | .04 | .91 | .009 | -1.0 | .08 |
| | $B(\hat{\theta})$ | .003 | .012 | -.08 | .02 | -.01 | .03 |
| | $M(\hat{\theta})/M(\hat{\lambda}^{(r)})$ | .003 | .01 | .08 | .01 | .07 | .06 |

Finally, Table 15 displays statistics on the number of repetitions to converge. It is clear from Table 16 that the EM algorithm requires much more iteration to converge than the proposed approach.

**Table 15: Statistics on the Number of Iterations**

| | **Proposed Method** | **EM Method** |
|---|---|---|
| **Average** | 10.13 | 27.29 |
| **Variance** | 16.37 | 92.37 |

## 7. Revision of the Response Model Parameter

Let ${}^{(er)}t_k$ represent the error-prone prior discrete random variable that indicates the time period $i$ when the response occurs for a randomly selected unit $k$ from the sample. After I time periods of

data collection, each unit $k$ is observed until the period ${}^{(er)}\mathrm{I}_k$, with ${}^{(er)}\mathrm{I}_k \le \mathrm{I}$. If the unit responses then ${}^{(er)}t_k = {}^{(er)}\mathrm{I}_k$, otherwise we only know that ${}^{(er)}t_k > \mathrm{I}$.

To specify a model for a joint distribution, it is often practice to parameterize the relationship between the two hazards in one of two ways. One way is that the joint distribution arises from the dependence of two survival times, which may result because a result of one affects the probability of responding of the other. A second way is that the joint distribution arises from shared observed or unobservable heterogeneity of the two observations. Hout, Duncan, and Sobel (1987) discuss these alternatives on the joint distribution of variables in multivariate contingency tables. The joint distribution of a discrete-time hazards, $\boldsymbol{t}_k = ({}^{(er)}t_k, t_k)^T$, could be specified in terms of the joint probability mass function

$$\Pr(\boldsymbol{t}_k) = \Pr({}^{(er)}t_k = i, t_k = j), \text{ for } i, j = 1,2,\ldots .$$

Shaked et al. (1995) defined the discrete bivariate conditional hazard rate function of $\boldsymbol{t}_k$ using the following five functions:

1) A first response occurs from the error-free process

$$h_{ki} = \Pr(t_k = i, {}^{(er)}t_k > i \mid t_k \ge i, {}^{(er)}t_k \ge i), \; i = 1,2,\ldots ,$$

2) A first response occurs from the error-prone process

$${}^{(er)}h_{ki} = \Pr({}^{(er)}t_k = i, t_k > i \mid t_k \ge i, {}^{(er)}t_k \ge i), \; i = 1,2,\ldots ,$$

3) Responses occur simultaneously from both processes

$${}^{(er)}h_{ki} = \Pr(t_k = i, {}^{(er)}t_k = i \mid t_k \ge i, {}^{(er)}t_k \ge i), \; i = 1,2,\ldots ,$$

4) A response occurs from the error-free process given that a response has been obtained from the error-prone process

$${}^{(t|er)}h_{i|j;k} = \Pr(t = i \mid t \ge i, {}^{(er)}t = j), \; i > j = 1,2,\ldots , \text{ and}$$

5) A response occurs from the error-prone process given that a response has been obtained from the error-free process

$${}^{(er|t)}h_{i|j;k} = \Pr({}^{(er)}t_k = i \mid {}^{(er)}t_k \ge i, t_k = j), \; i > j = 1,2,\ldots ,$$

provided the conditions in the above conditional probabilities have positive probabilities. Otherwise, we set these functions to be 1. The narrow sense bivariate geometric distribution (Esary and Marshall 1973), and results for the case of more than two discrete variables can be found in Shaked *et al.* (1995), who also give necessary and sufficient conditions on the five functions which ensure that they are hazard rate functions of some random vector $\boldsymbol{t}_k$.

### 7.1 EE for the Target Response Mechanism

We have

$$f(t_k) = \Pr(t_k = \mathrm{I}_k)^{\delta_k} \Pr(t_k > \mathrm{I}_k)^{1-\delta_k}, \tag{7.1}$$

where $\delta_k = 1$ if unit $k$ is uncensored (responds) under the true response mechanism and $\delta_k = 0$ if unit $k$ is censored. Substituting (3.2) and (3.3) into (7.1), yields

$$f(t_k) = \{h_{k\mathrm{I}_k} / (1 - h_{k\mathrm{I}_k})\}^{\delta_k} \prod_{i=1}^{\mathrm{I}_k} (1 - h_{ki}). \tag{7.2}$$

Expression (7.2) can be rewritten (Allison 1982) as

$$f(t_k) = \prod_{i=1}^{\mathrm{I}_k} \{h_{ki} / (1 - h_{ki})\}^{r_{ki}} \prod_{i=1}^{\mathrm{I}_k} (1 - h_{ki}),$$

where $r_{ki}$ is a sequence of the true response indicators defined for each unit $k$ whose values are defined as $r_{ki} = 1$ if the unit does respond in period $i$ and $r_{ki} = 0$ if the unit does not respond in period $i$. Taking the first derivatives of the logarithm of the objective function $L(\boldsymbol{\lambda}_r)L = \prod_k f(t_k)$ yields the census EE of the true response mechanism

$$^{(t)}\mathbf{S}(\boldsymbol{\lambda}_r) = \sum_k {}^{(t)}\mathbf{s}_k(\boldsymbol{\lambda}_r) = \mathbf{0}, \tag{7.3}$$

where $^{(t)}\mathbf{s}_k(\boldsymbol{\lambda}_r) = \partial \log f(t_k) / \partial \boldsymbol{\lambda}_r = \sum_{i=1}^{\mathrm{I}_k} \dot{\boldsymbol{h}}_{ki}(r_{ki} - h_{ki})\{h_{ki}(1 - h_{ki})\}^{-1}$, and $\dot{\boldsymbol{h}}_{ki} = \partial h_{ki} / \partial \boldsymbol{\lambda}_r$. For the logistic regression model $\log\{h_{ki} / (1 - h_{ki})\} = \boldsymbol{v}_{r;ki}^T \boldsymbol{\lambda}_r$, $\dot{\boldsymbol{h}}_{ki} = \boldsymbol{v}_{r;ki} h_{ki}(1 - h_{ki})$, $^{(t)}\mathbf{s}_k(\boldsymbol{\lambda}_r) = \sum_{i=1}^{\mathrm{I}_k} \boldsymbol{v}_{r;ki}(r_{ki} - h_{ki})$, and the matrix of second partial derivatives is

$$\frac{\partial^{(t)}\mathbf{S}^T(\boldsymbol{\lambda}_r)}{\partial \boldsymbol{\lambda}_r} = -\sum_k \sum_{i=1}^{\mathrm{I}_k} \boldsymbol{v}_{r;ki} h_{ki}(1 - h_{ki}) \boldsymbol{v}_{r;ki}^T \equiv -\mathbf{J}(\boldsymbol{\lambda}_r).$$

### 7 .2 Joint Distribution of the Error-prone and Error-free hazards

The conditional joint distribution of $(^{(er)}r_{ki}, r_{ki})$ is characterized by the matrix $\mathbf{H}_{ki}$ which depends on two probabilities: $h_{ki}^{(1|1)} = \Pr(^{(er)}r_{ki} = i | ^{(er)}t_k \geq i, r_{ki} = 1)$ and $h_{ki}^{(1|0)} = \Pr(^{(er)}r_{ki} = i | ^{(er)}t_k \geq i, r_{ki} = 0)$,

$$\mathbf{H}_{ki} = \begin{pmatrix} 1 - h_{ki}^{(1|0)} & 1 - h_{ki}^{(1|1)} \\ h_{ki}^{(1|0)} & h_{ki}^{(1|1)} \end{pmatrix},$$

where $^{(er)}r_{ki}$ is a sequence of the error-prone response indicators defined for each unit $k$ whose values are defined as $^{(er)}r_{ki} = 1$ if the unit does respond in period $i$ and $^{(er)}r_{ki} = 0$ if the unit does not respond in period $i$. Hence, the conditional joint distribution of $(^{(er)}r_{ki}, r_{ki})$ is given by (6.1) with $(y_1, y_2) = (^{(er)}r_{ki}, r_{ki})$, $p_{11} = h_{ki} h_{ki}^{(1|1)}$, $p_{10} = h_{ki}(1 - h_{ki}^{(1|1)})$, $p_{01} = (1 - h_{ki}) h_{ki}^{(1|0)}$, and $p_{00} = (1 - h_{ki})(1 - h_{ki}^{(1|0)})$.

After $\mathrm{I}$ time periods of data collection, we may decompose the joint distribution of $\boldsymbol{t}_k$ as

$$f(^{(er)}t_k, t_k) = f(^{(er)}t_k \mid t_k) f(t_k),$$

where $f(t_k)$ is given by (7.1). Taking the first derivatives of the logarithm of the likelihood function $L(\boldsymbol{\lambda}_r) = \prod_k f(^{(pri)}t_k, t_k)$ yields the census EE of the true response mechanism

$$\mathbf{S}(\boldsymbol{\lambda}_r) = \sum_k \mathbf{s}_k(\boldsymbol{\lambda}_r; ^{(er)}t_k, t_k) = \mathbf{0}, \tag{7.4}$$

where $\mathbf{s}_k(\boldsymbol{\lambda}_r;{}^{(er)}t_k,t_k)=\partial\log f({}^{(er)}t_k,t_k)/\partial\boldsymbol{\lambda}_r={}^{(t)}\mathbf{S}(\boldsymbol{\lambda}_r)+{}^{(er|t)}\mathbf{S}(\boldsymbol{\lambda}_r)$, ${}^{(t)}\mathbf{S}(\boldsymbol{\lambda}_r)$ is given by (7.3), ${}^{(er|t)}\mathbf{S}(\boldsymbol{\lambda}_r)=\sum_k \mathbf{s}_k(\boldsymbol{\lambda}_r;{}^{(er)}t_k\,|\,t_k)$ and $\mathbf{s}_k(\boldsymbol{\lambda}_r;{}^{(er)}t_k\,|\,t_k)=\partial\log f({}^{(er)}t_k\,|\,t_k)/\partial\boldsymbol{\lambda}_r$. The solution to (7.4) defines the census parameter.

### 7.3 Sample Estimating Equation

Taking the derivatives of the logarithm of the objective function $L(\boldsymbol{\lambda}_r)=\prod_k f_{\mathrm{I}_{\min}}({}^{(er)}t_k)f_{\mathrm{I}}(t_k|{}^{(er)}t_k)$ and adjusting for the unequal selection probabilities, we get the sample EE

$$\hat{\mathbf{S}}_{\mathrm{I}}(\boldsymbol{\lambda}_r)=\sum_k \mathbf{s}_{\mathrm{I}_{\min}}(\boldsymbol{\lambda}_r;{}^{(er)}t_k)+\sum_k d_{\mathrm{I};k}\{\mathbf{s}_{\mathrm{I}}(\boldsymbol{\lambda}_r;{}^{(er)}t_k,t_k)-\mathbf{s}_{\mathrm{I}}(\boldsymbol{\lambda}_r;{}^{(er)}t_k)\}=\mathbf{0}\,, \tag{7.5}$$

where $\mathbf{s}(\boldsymbol{\lambda}_r;{}^{(er)}t_k)=\partial\log f({}^{(er)}t_k)/\partial\boldsymbol{\lambda}_r$.

Now, taking the derivatives of the logarithm of $\prod_k f({}^{(er)}t_k)$ yields the census EE associated with the error-prone response mechanism

$${}^{(er)}\mathbf{S}(\boldsymbol{\lambda}_r)=\sum_k \mathbf{s}_k(\boldsymbol{\lambda}_r;{}^{(er)}t_k)=\mathbf{0}\,, \tag{7.6}$$

with $\mathbf{s}_k(\boldsymbol{\lambda}_r;{}^{(er)}t_k)=\partial\log f({}^{(er)}r_{ki})/\partial\boldsymbol{\lambda}_r={}^{(er)}\dot{\boldsymbol{h}}_{ki}({}^{(er)}r_{ki}-{}^{(er)}h_{ki})/\{{}^{(er)}h_{ki})(1-{}^{(er)}h_{ki})\}$,

where ${}^{(er)}h_{ki}=h_{ki}^{(1|1)}h_{ki}+h_{ki}^{(1|0)}(1-h_{ki})$ and $\dot{\boldsymbol{h}}_{ki}=\partial h_{ki}/\partial\boldsymbol{\lambda}_r$. We use the EM algorithm to derive the solution of (7.6). The E-step on the $(e+1)^{th}$ iteration of the EM algorithm,

$$\begin{aligned}Q(\boldsymbol{\lambda}_r;\boldsymbol{\lambda}_r^{(e)})&=E\{\textstyle\sum_k \delta_k \sum_{i=1}^{\mathrm{I}} l_{k;i}f({}^{(er)}t_k,t_k)|{}^{(er)}t_k,\boldsymbol{\lambda}_r^{(e)}+\sum_k(1-\delta_k)(t_k>\mathrm{I})f({}^{(er)}t_k,t_k)|{}^{(er)}t_k,\boldsymbol{\lambda}_r^{(e)}\}\\&=\textstyle\sum_k \delta_k \sum_{i=1}^{\mathrm{I}}\Pr(t_k=i|{}^{(er)}t_k,\boldsymbol{\lambda}_r^{(e)})\log f({}^{(er)}t_k,t_k=i)\\&+\textstyle\sum_k(1-\delta_k)\Pr(t_k>\mathrm{I}|{}^{(er)}t_k,\boldsymbol{\lambda}_r^{(e)})\log f({}^{(er)}t_k,t_k>\mathrm{I}),\end{aligned}$$

where $\boldsymbol{\lambda}_r^{(e)}$ denote the value of $\boldsymbol{\lambda}_r$ after the $e^{th}$ EM iteration. It remains to derive the conditional Probabilities of ${}^{(er)}t_k\,|\,t_k$ and its derivatives.

### 7 .3 Conditional Probabilities of ${}^{(er)}t_k\,|\,t_k$

The conditional probability of obtaining a response under the error-prone response mechanism at time period $i$ could be expressed in terms of the hazard as

**Case 1.1.** $\mathrm{I}_k\geq{}^{(er)}\mathrm{I}_k$ & $r_{k({}^{(er)}\mathrm{I}_k)}=1$

For unit with ${}^{(er)}t_k={}^{(er)}\mathrm{I}_k$, the probability of obtaining a response at time period $i$ could be expressed in terms of the hazard as

$$\Pr({}^{(er)}t_k={}^{(er)}\mathrm{I}_k\,|\,\mathrm{I}_k\geq{}^{(er)}\mathrm{I}_k,r_{k({}^{(er)}\mathrm{I}_k)}=1)=\textstyle\prod_{i=1}^{({}^{(er)}\mathrm{I}_k-1)}(1-h_{ki}^{(1|0)})h_{k({}^{(er)}\mathrm{I}_k)}^{(1|1)}\,. \tag{7.7.a}$$

For units with ${}^{(er)}t_k>{}^{(er)}\mathrm{I}_k$, the probability of obtaining a response can be expressed as

$$\Pr({}^{(er)}t_k > {}^{(er)}\mathrm{I}_k \mid \mathrm{I}_k \geq {}^{(er)}\mathrm{I}_k, r_{k^{(er)}\mathrm{I}_k} = 0) = \prod_{i=1}^{{}^{(er)}\mathrm{I}_k - 1}(1 - h_{ki}^{(1|0)})(1 - h_{k^{(er)}\mathrm{I}_k}^{(1|1)}). \quad (7.7.b)$$

**Case 1.2.** $\mathrm{I}_k \geq {}^{(er)}\mathrm{I}_k$ & $r_{k^{(er)}\mathrm{I}_k} = 0$

For unit with ${}^{(er)}t_k = {}^{(er)}\mathrm{I}_k$, the probability of obtaining a response at time period $i$ could be expressed in terms of the hazard as

$$\Pr({}^{(er)}t_k = {}^{(er)}\mathrm{I}_k \mid \mathrm{I}_k \geq {}^{(er)}\mathrm{I}_k, r_{k^{(er)}\mathrm{I}_k} = 0) = \prod_{i=1}^{{}^{(er)}\mathrm{I}_k - 1}(1 - h_{ki}^{(1|0)}) h_{k^{(er)}\mathrm{I}_k}^{(1|0)}. \quad (7.8.a)$$

For units with ${}^{(er)}t_k > {}^{(er)}\mathrm{I}_k$, the probability of obtaining a response can be expressed as

$$\Pr({}^{(er)}t_k > {}^{(er)}\mathrm{I}_k \mid \mathrm{I}_k \geq {}^{(er)}\mathrm{I}_k, r_{k^{(er)}\mathrm{I}_k} = 0) = \prod_{i=1}^{{}^{(er)}\mathrm{I}_k}(1 - h_{ki}^{(1|0)}). \quad (7.8.b)$$

**Case 2.1.** $\mathrm{I}_k < {}^{(er)}\mathrm{I}_k$ & $r_{k\mathrm{I}_k} = 1$

For unit with ${}^{(er)}t_k = {}^{(er)}\mathrm{I}_k$, the probability of obtaining a response at time period $i$ could be expressed in terms of the hazard as

$$\begin{aligned}\Pr({}^{(er)}t_k = {}^{(er)}\mathrm{I}_k \mid \mathrm{I}_k < {}^{(er)}\mathrm{I}_k, r_{k\mathrm{I}_k} = 1) = {} & \prod_{i=1}^{\mathrm{I}_k - 1}(1 - h_{ki}^{(1|0)}) \\ & \times (1 - h_{k\mathrm{I}_k}^{(1|1)}) \\ & \times \prod_{i=\mathrm{I}_k+1}^{{}^{(er)}\mathrm{I}_k - 1}\{h_{ki}(1 - h_{ki}^{(1|1)}) + (1 - h_{ki})(1 - h_{ki}^{(1|0)})\} \\ & \times \{h_{k^{(er)}\mathrm{I}_k} h_{k^{(er)}\mathrm{I}_k}^{(1|1)} + (1 - h_{k^{(er)}\mathrm{I}_k}) h_{k^{(er)}\mathrm{I}_k}^{(1|0)}\}.\end{aligned} \quad (7.9.a)$$

For units with ${}^{(er)}t_k > {}^{(er)}\mathrm{I}_k$, the probability of obtaining a response can be expressed as

$$\begin{aligned}\Pr({}^{(er)}t_k > {}^{(er)}\mathrm{I}_k \mid \mathrm{I}_k < {}^{(er)}\mathrm{I}_k, r_{k\mathrm{I}_k} = 1) = {} & \prod_{i=1}^{\mathrm{I}_k - 1}(1 - h_{ki}^{(1|0)}) \\ & \times (1 - h_{k\mathrm{I}_k}^{(1|1)}) \\ & \times \prod_{i=\mathrm{I}_k+1}^{{}^{(er)}\mathrm{I}_k}\{h_{ki}(1 - h_{ki}^{(1|1)}) + (1 - h_{ki})(1 - h_{ki}^{(1|0)})\}.\end{aligned} \quad (7.9.b)$$

**Case 2.2.** $\mathrm{I}_k < {}^{(er)}\mathrm{I}_k$ & $r_{k\mathrm{I}_k} = 0$

For unit with ${}^{(er)}t_k = {}^{(er)}\mathrm{I}_k$, the probability of obtaining a response at time period $i$ could be expressed in terms of the hazard as

$$\begin{aligned}\Pr({}^{(er)}t_k = {}^{(er)}\mathrm{I}_k \mid \mathrm{I}_k < {}^{(er)}\mathrm{I}_k, r_{k\mathrm{I}_k} = 0) = {} & \prod_{i=1}^{\mathrm{I}_k}(1 - h_{ki}^{(1|0)}) \\ & \times \prod_{i=\mathrm{I}_k+1}^{{}^{(er)}\mathrm{I}_k - 1}\{h_{ki}(1 - h_{ki}^{(1|1)}) + (1 - h_{ki})(1 - h_{ki}^{(1|0)})\} \\ & \times \{h_{k^{(er)}\mathrm{I}_k} h_{k^{(er)}\mathrm{I}_k}^{(1|1)} + (1 - h_{k^{(er)}\mathrm{I}_k}) h_{k^{(er)}\mathrm{I}_k}^{(1|0)}\}.\end{aligned} \quad (7.10.a)$$

For units with ${}^{(er)}t_k > {}^{(er)}\mathrm{I}_k$, the probability of obtaining a response can be expressed as

$$\begin{aligned}\Pr({}^{(er)}t_k > {}^{(er)}\mathrm{I}_k \mid \mathrm{I}_k < {}^{(er)}\mathrm{I}_k, r_{k\mathrm{I}_k} = 0) = {} & \prod_{i=1}^{\mathrm{I}_k}(1 - h_{ki}^{(1|0)}) \\ & \times \prod_{i=\mathrm{I}_k+1}^{{}^{(er)}\mathrm{I}_k}\{h_{ki}(1 - h_{ki}^{(1|1)}) + (1 - h_{ki})(1 - h_{ki}^{(1|0)})\}.\end{aligned} \quad (7.10.b)$$

### 7.4 Derivatives of the Conditional Probability of ${}^{(er)}t_k \mid t_k$

We have

$$f({}^{(er)}t_k \mid t_k) = \Pr({}^{(er)}t_k = {}^{(er)}\mathrm{I}_k \mid t_k)^{{}^{(er)}\delta_k} \Pr({}^{(er)}t_k > {}^{(er)}\mathrm{I}_k \mid t_k)^{1-{}^{(er)}\delta_k}, \tag{7.11}$$

where ${}^{(er)}\delta_k = 1$ if unit $k$ is uncensored (responds) under the error-prone response mechanism and ${}^{(er)}\delta_k = 0$ if unit $k$ is censored.

Consider first the form given by

$$g(y) = \{q/(1-q)\}^{y}(1-q).$$

Taking the first derivatives of the logarithm of the objective function $g(y)$ yields

$$\partial \log g(y)/\partial \boldsymbol{\lambda} = \dot{\boldsymbol{q}}(y-q)/\{q(1-q)\},$$

where $\dot{\boldsymbol{q}} = \partial q/\partial \boldsymbol{\lambda}$.

**Case 1.1.** $\mathrm{I}_k \geq {}^{(er)}\mathrm{I}_k$ & $r_{k^{(er)}\mathrm{I}_k} = 1$

Substituting (7.10) into (7.11) yields

$$f({}^{(er)}t_k \mid t_k) = \{h^{(1|1)}_{k^{(er)}\mathrm{I}_k}/(1-h^{(1|1)}_{k^{(er)}\mathrm{I}_k})\}^{{}^{(er)}\delta_k} \prod_{i=1}^{{}^{(er)}\mathrm{I}_k - 1}(1-h^{(1|0)}_{ki})(1-h^{(1|1)}_{k^{(er)}\mathrm{I}_k}). \tag{7.12.1.a}$$

Expression (7.12.1.a) can be rewritten as

$$\begin{aligned} f({}^{(er)}t_k \mid t_k) = {} & \prod_{i=1}^{{}^{(er)}\mathrm{I}_k - 1}\{h^{(1|0)}_{ki}/(1-h^{(1|0)}_{ki})\}^{{}^{(er)}r_{ki}}\{h^{(1|1)}_{k^{(er)}\mathrm{I}_k}/(1-h^{(1|1)}_{k^{(er)}\mathrm{I}_k})\}^{{}^{(er)}r_{k^{(er)}\mathrm{I}_k}} \\ & \times \prod_{i=1}^{{}^{(er)}I_k - 1}(1-h^{(1|0)}_{ki})(1-h^{(1|1)}_{k^{(pri)}\mathrm{I}_k}). \end{aligned} \tag{7.12.1.b}$$

Taking the first derivatives of the logarithm of (7.12.1.b) yields

$$\begin{aligned} \mathbf{s}_k(\boldsymbol{\lambda}_r;{}^{(er)}t_k \mid t_k) = \partial \log f({}^{(er)}t_k \mid t_k) = {} & \textstyle\sum_{i=1}^{{}^{(er)}\mathrm{I}_k - 1} \dot{\boldsymbol{h}}^{(1|0)}_{ki}({}^{(er)}r_{ki} - h^{(1|0)}_{ki})/\{h^{(1|0)}_{ki}(1-h^{(1|0)}_{ki})\} \\ & + \dot{\boldsymbol{h}}^{(1|1)}_{k^{(er)}\mathrm{I}}({}^{(er)}r_{k^{(er)}\mathrm{I}} - h^{(1|1)}_{kr_{k^{(er)}\mathrm{I}_k}})/\{h^{(1|1)}_{kr_{k^{(er)}\mathrm{I}_k}}(1-h^{(1|1)}_{kr_{k^{(er)}\mathrm{I}_k}})\}. \end{aligned} \tag{7.12.1.c}$$

For the logistic regression models

$$\mathbf{s}_k(\boldsymbol{\lambda}_r;{}^{(er)}t_k \mid t_k) = \textstyle\sum_{i=1}^{{}^{(er)}\mathrm{I}_k - 1} \boldsymbol{v}^{(1|0)}_{ki}({}^{(er)}r_{ki} - h^{(1|0)}_{ki}) + \boldsymbol{v}^{(1|1)}_{k^{(er)}\mathrm{I}_k}({}^{(er)}r_{k^{(er)}\mathrm{I}_k} - h^{(1|1)}_{k^{(er)}\mathrm{I}_k}).$$

**Case 1.2.** $\mathrm{I}_k \geq {}^{(er)}\mathrm{I}_k$ & $r_{k^{(er)}\mathrm{I}_k} = 0$

Substituting (7.10) into (7.11) yields

$$f({}^{(er)}t_k \mid t_k) = \{h^{(1|0)}_{k^{(er)}\mathrm{I}_k}/(1-h^{(1|0)}_{k^{(er)}\mathrm{I}_k})\}^{{}^{(er)}\delta_k} \prod_{i=1}^{{}^{(er)}\mathrm{I}_k}(1-h^{(1|0)}_{ki}). \tag{7.12.2.a}$$

Expression (7.12.2.a) can be rewritten as

$$f({}^{(er)}t_k \mid t_k) = \prod_{i=1}^{{}^{(er)}\mathrm{I}_k}\{h^{(1|0)}_{ki}/(1-h^{(1|0)}_{ki})\}^{{}^{(er)}r_{ki}} \prod_{i=1}^{{}^{(er)}\mathrm{I}_k}(1-h^{(1|0)}_{ki}). \tag{7.12.2.b}$$

Taking the first derivatives of the logarithm of (7.12.2.b) yields

$$\mathbf{s}_k(\boldsymbol{\lambda}_r;{}^{(er)}t_k \mid t_k) = \partial \log f({}^{(er)}t_k \mid t_k) = \textstyle\sum_{i=1}^{{}^{(er)}\mathrm{I}_k} \dot{\boldsymbol{h}}^{(1|0)}_{ki}({}^{(er)}r_{ki} - h^{(1|0)}_{ki})/\{h^{(1|0)}_{ki}(1-h^{(1|0)}_{ki})\}. \tag{7.12.2.c}$$

For the logistic regression models

$$\mathbf{s}_k(\boldsymbol{\lambda}_r;{}^{(er)}t_k \mid t_k) = \textstyle\sum_{i=1}^{{}^{(er)}\mathrm{I}_k} \boldsymbol{v}^{(1|0)}_{ki}({}^{(er)}r_{ki} - h^{(1|0)}_{ki}).$$

**Case 2.1.** $\mathrm{I}_k < {}^{(er)}\mathrm{I}_k$ & $r_{\mathrm{I}_k} = 1$

Substituting (7.10) into (7.11) yields

$$f({}^{(er)}t_k \mid t_k) = \{{}^{(er)}h_{k({}^{(pri)}\mathrm{I}_k)} / (1-{}^{(er)}h_{k({}^{(pri)}\mathrm{I}_k)})\}^{{}^{(er)}\delta_k} \times \prod_{i=1}^{\mathrm{I}_k-1}(1-h_{ki}^{(1|0)}) \times (1-h_{k\mathrm{I}_k}^{(1|1)}) \times \prod_{i=\mathrm{I}_k+1}^{{}^{(er)}\mathrm{I}_k}(1-{}^{(er)}h_{ki}). \quad (7.12.3.a)$$

Expression (7.12.3.a) can be rewritten as

$$f({}^{(er)}t_k \mid t_k) = \prod_{i=1}^{\mathrm{I}_k-1}\{h_{ki}^{(1|0)} / (1-h_{ki}^{(1|0)})\}^{{}^{(er)}r_{ki}} \times \{h_{k\mathrm{I}_k}^{(1|1)} / (1-h_{k\mathrm{I}_k}^{(1|1)})\}^{{}^{(pri)}r_{k\mathrm{I}_k}} \times \prod_{i=\mathrm{I}_k+1}^{{}^{(er)}\mathrm{I}_k}\{{}^{(er)}h_{ki} / (1-{}^{(er)}h_{ki})\}^{{}^{(er)}r_{ki}} \times \prod_{i=1}^{\mathrm{I}_k-1}(1-h_{ki}^{(1|0)}) \times (1-h_{k\mathrm{I}_k}^{(1|1)}) \times \prod_{i=\mathrm{I}_k+1}^{{}^{(er)}\mathrm{I}_k}(1-{}^{(er)}h_{ki}). \quad (7.12.3.b)$$

Taking the first derivatives of the logarithm of (7.12.3.b) yields

$$\begin{aligned}\mathbf{s}_k(\boldsymbol{\lambda}_r;{}^{(er)}t_k \mid t_k) = \partial \log f({}^{(er)}t_k \mid t_k) &= \textstyle\sum_{i=1}^{\mathrm{I}_k-1} \dot{\boldsymbol{h}}_{ki}^{(1|0)}({}^{(er)}r_{ki} - h_{ki}^{(1|0)}) / \{h_{ki}^{(1|0)}(1-h_{ki}^{(1|0)})\} \\ &+ \dot{\boldsymbol{h}}_{k\mathrm{I}_k}^{(1|1)}({}^{(er)}r_{k\mathrm{I}_k} - h_{k\mathrm{I}_k}^{(1|1)}) / \{h_{k\mathrm{I}_k}^{(1|1)}(1-h_{k\mathrm{I}_k}^{(1|1)})\} \\ &+ \textstyle\sum_{i=\mathrm{I}_k+1}^{{}^{(pri)}\mathrm{I}_k}\{{}^{(er)}\dot{\boldsymbol{h}}_{ki}({}^{(er)}r_{ki} - {}^{(er)}h_{ki}) / \{{}^{(er)}h_{ki}(1-{}^{(er)}h_{ki})\}.\end{aligned} \quad (7.12.3.c)$$

**Case 2.2.** ${}^{(er)}\mathrm{I}_k < \mathrm{I}_k$ & $r_{k\mathrm{I}_k} = 0$

Substituting (7.10) into (7.11) yields

$$f({}^{(er)}t_k \mid t_k) = \{{}^{(er)}h_{k({}^{(pri)}\mathrm{I}_k)} / (1-{}^{(er)}h_{k({}^{(er)}\mathrm{I}_k)})\}^{{}^{(er)}\delta_k} \times \prod_{i=1}^{\mathrm{I}_k}(1-h_{ki}^{(1|0)}) \times \prod_{i=\mathrm{I}_k+1}^{{}^{(er)}\mathrm{I}_k}(1-{}^{(er)}h_{ki}). \quad (7.12.4.a)$$

Expression (7.12.4.a) can be rewritten as

$$f({}^{(er)}t_k \mid t_k) = \prod_{i=1}^{\mathrm{I}_k}\{h_{ki}^{(1|0)} / (1-h_{ki}^{(1|0)})\}^{{}^{(er)}r_{ki}} \times \prod_{i=\mathrm{I}_k+1}^{{}^{(er)}\mathrm{I}_k}\{{}^{(er)}h_{ki} / (1-{}^{(er)}h_{ki}\}^{{}^{(er)}r_{ki}} \times \prod_{i=1}^{\mathrm{I}_k}(1-h_{ki}^{(1|0)}) \times \prod_{i=\mathrm{I}_k+1}^{{}^{(er)}\mathrm{I}_k}(1-{}^{(er)}h_{ki}). \quad (7.12.4.b)$$

Taking the first derivatives of the logarithm of (7.12.4.b) yields

$$\begin{aligned}\mathbf{s}_k(\boldsymbol{\lambda}_r;{}^{(er)}t_k \mid t_k) &= \partial \log f({}^{(er)}t_k \mid t_k) \\ &= \textstyle\sum_{i=1}^{\mathrm{I}_k} \dot{\boldsymbol{h}}_{ki}^{(1|0)}({}^{(er)}r_{ki} - h_{ki}^{(1|0)}) / \{h_{ki}^{(1|0)}(1-h_{ki}^{(1|0)})\} \\ &+ \textstyle\sum_{i=\mathrm{I}_k+1}^{{}^{(er)}\mathrm{I}_k}{}^{(pri)}\dot{\boldsymbol{h}}_{ki}({}^{(er)}r_{ki} - {}^{(er)}h_{ki}) / \{{}^{(er)}h_{ki}(1-{}^{(er)}h_{ki})\}.\end{aligned} \quad (7.12.4.c)$$

## Concluding Remarks

We formulated an optimization problem for designing a survey and we identified steps for its revision during the survey data collection period. We considered the error-prone prior, the error-prone processed, and the 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 realizations of the error-free random process at each phase of the 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 target

information. 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**

Allison, P. D. 1982. Discrete-time Methods for the Analysis of Event Histories. S. Leinhardt (ed.). *Sociological Methodology*. San Francisco: Jossey-Bass. 61–98.

Anderson, D. A. and J. P. Hinde. 1988. Random Effects in Generalized Linear Models and The EM Algorithm. *Communications in Statistics - Theory Methods*, 17. 3847-3856.

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 $2 \times 2$ Tables. *Biometrics*, 10, pp. 478-483.

Cauchy, M. A. 1847. Méthode générale pour la résolution des systèmes d'équations simultanées. Comptes Rendus Hebd. Séances Acad. Sci. Paris, 25, 536-538.

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 JSM Proceedings, Survey Research Methods Section. Alexandria, VA : American Statistical Association. 258-271.

Demnati, A. 2017. Extending Hansen and Hurwitz's Approach for Nonresponse in Sample Survey. In JSM Proceedings, Survey Research Methods Section. Alexandria, VA : American Statistical Association. 2438-2459.

Demnati, A. 2018. Side Effect Reduction of Prior and Processed Information on Survey Design. In JSM Proceedings, Survey Research Methods Section. Alexandria, VA : American Statistical Association. 1194-1218.

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.

Esary , J . D . A. W. Marshall. 1973. Multivariate Geometric Distributions Generated by a Cumulative Damage Processes. Naval Postgraduate School Report. Available at: *http://calhoun.nps.edu/handle/10945/30003*

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.

Hout, M., O. D. Duncan, and M. E. Sobel. 1987. Association and Heterogeneity: Structural Models of Similarities and Differences. *Sociological Methodology*, 17, 145-184.

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 Scheme for Estimating from Misclassified Multinomial Data with Applications to Sampling Inspection. *Technometrics*, 14, 187-202.

Shaked, M., J.G. Shanthikumar, and J.B. Valdez-Torres. 1995. Discrete Hazard Rate Functions. *Computers & Operations Research*. 22, 391-402.