International Journal of Statistics and Applied Mathematics 2020; 5(4): 135-146

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2020; 5(4): 135-146
© 2020 Stats & Maths
www.mathsjournal.com
Received: 25-05-2020
Accepted: 27-06-2020

Mutanu Beth
Mathematics and Actuarial Science Department, South Eastern Kenya University, P.O. BOX 170 – 90200, Kitui, Kenya

Kahiri JM
Mathematics and Actuarial Science Department, Kenyatta University, P.O Box: 43844-00100, Nairobi, Kenya

Odongo LO
Mathematics and Actuarial Science Department, Kenyatta University, P.O Box: 43844-00100, Nairobi, Kenya

Corresponding Author:
Mutanu Beth
Mathematics and Actuarial Science Department, South Eastern Kenya University, P.O. BOX 170 – 90200, Kitui, Kenya

Effects of measurement and response errors on estimation based on centre sampling and two-phase centre sampling for elusive populations

Mutanu Beth, Kahiri JM and Odongo LO

DOI: https://doi.org/10.22271/maths.2020.v5.i4b.562

Abstract
Accurate survey data is important for planning and decision making. The presence of measurement and response errors in survey data can result in biased estimates as well as estimates with large standard errors, affecting the power of the statistical test to be carried out. It is important to investigate the effects of response errors when computing survey data so as to obtain reliable information for use by statisticians and policy makers. This paper investigates the effects of measurement and response errors in elusive surveys where the population has capture problems. The technique of centre sampling (CS) which entails collecting data on especially migrant populations is studied. In the technique, observations are made on those subjects who visit specified centres within the survey time period. This paper also investigates the technique of two-phase Centre Sampling (CS). A simulation study is carried out to establish the effect of response and measurement errors in centre sampling (CS) and two-phase Centre Sampling (CS). Using profiles at different levels of multiplicities, it was found that the errors affect the statistical estimates, even in small proportions to true values resulting in large estimation errors in both the mean and the variance. It is therefore recommended that both measurement and response errors be put into consideration when designing and carrying out a survey for more accurate results.

Keywords: Elusive populations**,** response errors, center sampling, two-phase sampling

1. Introduction

In elusive surveys, the population units are not only difficult to capture but response errors can be quite high especially if the study variable is an undesirable characteristic. Elusive populations with undesirable characteristics include populations such as homosexuals, drug addicts and illegal immigrants who are difficult to sample because they are rare in the general population and often not easy to identify. Such populations are generally best sampled and surveyed on specific locations or venues, commonly known as centers of sampling. The process is referred to as Centre Sampling (CS).

Centre sampling was developed in Italy to carry out surveys on immigrant population composed of regular people (those with a residence permit according to the law) and irregular people (those illegally residing in the country), (Blangiardo, 2004) [3]. Work on elusive surveys has been done by Mecatti (2002) [16] and Migliorati (2003) [18], amongst others. Fuller and Burmeister (1972) considered estimators for samples selected from two overlapping frames.

Most sampling theory methods are limited to sampling designs in which the selection of the samples is done before the survey, such that none of the decisions about the sampling depends on what is observed when gathering the data. In such typical surveys decisions such as how to sample during the experiment are arrived at and fixed in advance.

In repeated surveys, differences in response given by respondents may be caused by lack of understanding of the questions, difficult in determining a true value or due to lack of information required to answer the question correctly. The interviewer may contribute to the differences in responses by giving different interpretations which may lead to the interviewee not understanding the subject matter and the purpose of the survey (Burton and Blair, 1991) [6].

In studies where sampling is carried out instead of a census, two major categories of errors in survey estimates arise, namely; sampling errors and non-sampling errors.

Sampling errors occur when only a part of the population is studied to make inferences about a target population. Sampling error would therefore not exist if a census is conducted. Non sampling errors arise in the course of collecting data, through the measuring process, interviewing or during observations. The other source of non-sampling errors is non-response, where respondents fail to respond due to reasons associated with the study variable.

The centre sampling relies on the habit of the immigrants to congregate in particular places for social contacts, healthcare, religion or places such as churches, mosques, bus stations, markets, hospitals etc. Such centers which are finite in number are usually spread all over the geographical area of interest and overlap covering the entire population. Due to this, the number of centers is taken to be more than 2.

In order to handle the problem of non-response and response errors, a two-phase sampling is necessary since under the general conditions of a survey, the measurement for any individual varies over repeated trials of the survey measurement process.

The survey question and answer process is a four-step process that involves comprehension of the question, retrieval of the information from memory, assessment of the correspondence between the retrieved information and the requested information and communication (Sudman *et al*., 1996). The encoding of the information, a process outside the control of the survey interview, determines whether the information of interest is available for the respondent to retrieve from long term memory.

2. Estimation in Elusive surveys under non contaminated data

In this section consideration is made on estimation when data is assumed to be free of measurement and response errors under Centre sampling. Suppose the population is finite of size *N* and the study variable is μ . Let the population true (error free) values be $\{\mu_1,...,\mu_i,...,\mu_N\}$. Let there be *L* centres and define the indicator variable

$$\omega_{li} = \begin{cases} 1 \ \ if \ \ the \ \ i^{th} \ \ unit \ \ visits \ \ centre \ \ l, \ l = 1, 2, ..., L \\ 0 \ \ otherwise \end{cases}$$

Then unit *i* is characterized by the vector $\omega_r = [\omega_{r1},...,\omega_{rl},...,\omega_{rL}]$ which is referred to as the profile of that unit.

Each unit which has one of the $2^L - 1$ possible ordered collections of *L*-tuples of the digits zero and one has unique profile. The null *L*-tuple is not a profile. Denote by $\omega_r = [\omega_{r1},...,\omega_{rl},...,\omega_{rL}]$ the r^{th} profile, where $\boldsymbol{\omega_i} = (\omega_{i1}, \omega_{i2}, .. \omega_{iL})$ is one of the $\boldsymbol{\omega_r}$, $2^L - 1$ possible profiles.

Profiles are very important practical tools since they can be assumed observable by asking the sampled unit the question; “Which other centers among the *L* considered and besides the one where you were sampled do you attend?”.

According to Mecatti and Migliorati (2003) [17], the population mean is given by

$$\bar{\mu} = \frac{1}{N} \sum_{r=1}^{2^L-1} \sum_{q=1}^{N_{l\omega_r}} \mu_{lrq}$$

Where μ_{lrq} is the qth unit with profile r in centre l

If $\boldsymbol{m_r} = \sum_{\boldsymbol{l}=1}^{\boldsymbol{L}} \boldsymbol{\omega_{lr}}$ denotes the number of centers attended by units with profile $\boldsymbol{\omega}_r$ i.e. indicating the multiplicity of the r^{th} profile (Birnbaum and Sirken, 1965; Thompson, 2002) [2, 22], then using profiles and multiplicity, $\bar{\mu}$ can be expressed as

$$\bar{\mu} = \frac{1}{N} \sum_{l=1}^{L} \sum_{r=1}^{2^L-1} \frac{1}{m_r} \sum_{q=1}^{N_{l\omega_r}} \mu_{lrq}$$

and

$$\tilde{\mu}_l = \frac{1}{N_l} \sum_{r=1}^{2^L-1} \frac{1}{m_r} \sum_{q=1}^{N_{l\omega_r}} \mu_{lrq} \qquad (1)$$

is the mean of μ in the l^{th} center adjusted for multiplicity.

By defining $\alpha_l = \frac{N_l}{N}$ as the weight of the l^{th} center with respect to the population, then

$$\bar{\mu} = \sum_{l=1}^{L} \alpha_l \tilde{\mu}_l \qquad (2)$$

If an unbiased estimator $\hat{\bar{\mu}}$ unbiased for $\bar{\mu}$ under a general sampling design is given, then

$$\hat{\bar{\mu}} = \sum_{l=1}^{L} \alpha_l \hat{\bar{\mu}}_l \tag{3}$$

is an unbiased estimator for $\bar{\mu}$ and, by assuming independence of samples among centres, it has variance,

$$\text{var}\left(\hat{\bar{\mu}}\right) = \sum_{l=1}^{L} \alpha_l^2 \,\text{var}\left(\hat{\bar{u}}_l\right).$$

2.1. Extension of Centre Sampling to Two phase Centre Sampling (CS) design

The two phase Centre Sampling (CS) can be applied to cut on budgetary allocations or in cases whereby a number of centres are fairly large in size. It can also be applied when centres are expected to be fairly similar the characteristic μ observed such that it might be impractical to exhaustively enumerate the selected centres. Under two phase CS design, *n* out of L centres are selected at the first phase with inclusion with respect to probabilities $\pi_i, \left(l = 1, \ldots, 2^L - 1\right)$. The n_j units are drawn from the l^{th} selected centre in the second phase, under sampling design not necessarily conforming to the first phase.

Using profiles $\pi_{rq,j}$ to represent the second stage inclusion probability of the q^{th} unit with profile ω_r attending the l^{th} centre selected at first stage, the following estimator;

$$\hat{\tilde{\mu}}_l = \frac{1}{N_l} \sum_{r=1}^{2^L-1} \frac{1}{m_r} \sum_{s=1}^{n_{l\omega_r}} \frac{\mu_{lrs}}{\pi_{lrs}} \tag{4}$$

is unbiased for $\tilde{\mu}_l$ under the second phase sampling design (Mecatti and Migliorati, 2003) [17].

Thus

$$\hat{\bar{\mu}} = \sum_{l=1}^{n} \alpha_l \frac{\hat{\tilde{\mu}}_l}{\pi_l} \tag{5}$$

is an unbiased estimator for μ according to standard theory for two phase centre sampling (Hedayat and Sinha,1991) [12]. By using sample totals instead of sample means the estimator (4) agrees with Maiti, Pal and Sinha (1993) [14] estimator for the population size *N* under two-dimensional sampling. Using standard results from two phase centre sampling, the estimator (4) has variance

$$\text{var}\left(\hat{\bar{\mu}}\right) = \sum_{l=1}^{L} \alpha_l^2 \tilde{\mu}_l^2 \left(\frac{1}{\pi_l} - 1\right) + \sum_{l=1}^{L} \sum_{\substack{l'=1 \\ l' \neq 1}}^{L} \alpha_l \alpha_{l'} \tilde{\mu}_l \tilde{\mu}_{l'} \left(\frac{\pi_{ll'}}{\pi_l \pi_{l'}} - 1\right) + \sum_{l=1}^{L} \frac{\alpha_l^2}{\pi_l} \text{var}\left(\hat{\tilde{\mu}}_l\right) \tag{6}$$

with $\pi_{ll'}$ being the joint inclusion probability of the pair of centers *l* and l' at the first phase selection and $\text{var}\left(\tilde{\mu}_l\right)$ conform with the general form (5).

Assuming $\pi_{ll'} > 0, \left(l \neq l' = 1, \ldots, L\right)$, according to Mecatti and Migliorati (2003) [17] +, an unbiased estimator of the variance (6) is given by

$$\hat{v}\left(\hat{\bar{\mu}}\right) = \sum_{l=1}^{n} \frac{\alpha_l \bar{y}_l^2}{\pi_l} \left(\frac{1}{\pi_l} - 1\right) + \sum_{l=1}^{n} \sum_{\substack{l'=1 \\ l' \neq 1}}^{n} \frac{\alpha_l \alpha_{l'} \tilde{\mu}_l \tilde{\mu}_{l'}}{\pi_{ll'}} \left(\frac{\pi_{ll'}}{\pi_l \pi_{l'}} - 1\right) + \sum_{l=1}^{n} \frac{\alpha_l^2}{\pi_l} \hat{v}\left(\hat{\tilde{\mu}}_l\right) \tag{7}$$

3. Estimation under Centre Sampling (CS) for elusive population when there are measurement and response errors

Estimation is now considered in the presence of measurement and response errors. It is assumed that there exists a true value for each population unit which we denote by μ_{lrq} for unit q with profile r in centre l in the population of size *N*.

In the process of making observations, a response error occurs for individual q with profile r in centre l such that rather than observing μ_{lrq}, we observe y_{lrq} where y_{lrq} is modeled by

$$y_{lrq} = \mu_{lrq} + e_{lrq}, \tag{8}$$

e_{lrq} being the measurement error. Note that the value μ_{lrq} does not change for unit q if it appears in other centres.

Let the sample size selected from centre l be n_l. Let n_{lr} be number of units selected having profile r. The error term e_{lrq} is assumed to have

$$E\left(e_{lrq}\right)=0 \text{ and } V\left(e_{lrq}\right)=\sigma_{lrq}^2$$
$$\operatorname{cov}\left(e_{lrq},e_{lrq'}\right)=\sigma_{lrrqq'}$$
$$\operatorname{cov}\left(e_{lrq},e_{lr'q'}\right)=\sigma_{lrr'qq'}$$
$$\operatorname{cov}\left(e_{lrq},e_{l'r'q'}\right)=\sigma_{ll'rr'qq'}$$

Define $\hat{\theta}_l$ as the estimator of $\tilde{\mu}_l$ given by

$$\hat{\theta}_l=\frac{1}{N_l}\sum_{r=1}^{2^L-1}\frac{1}{m_r}\sum_{q\in s}\frac{y_{lrq}}{\pi_{lrq}} \tag{9}$$

Where π_{lrq} is the 1st order inclusion probability of unit q with profile r in centre l and $\hat{\theta}_l$ is the estimate of $\tilde{\mu}_l$

The expectation of $\hat{\theta}_l$ is considered as a two-phase process where in the 1st phase a unit is selected into the sample and in the 2nd phase, an error randomly occurs from the population of possible errors. The expectation is therefore

$$E\left(\hat{\theta}_l\right)=E_pE_m\left(\hat{\theta}_l\right)$$
$$E_m\left(\hat{\theta}_l\right)=E_m\left(\frac{1}{N_l}\sum_{r=1}^{2^L-1}\frac{1}{m_r}\sum_{q\in s}\frac{y_{lrq}}{\pi_{lrq}}\right)$$
$$E_m\left(\hat{\theta}_l\right)=\frac{1}{N_l}\sum_{r=1}^{2^L-1}\frac{1}{m_r}\sum_{q\in s}\frac{\left(\mu_{lrq}+b_{lrq}\right)}{\pi_{lrq}}=\frac{1}{N_l}\sum_{r=1}^{2^L-1}\frac{1}{m_r}\sum_{q\in s}\frac{\mu_{lrq}^*}{\pi_{lrq}} \tag{10}$$

Where $\mu_{lrq}^*=E_m(y_{lrq})=\mu_{lrq}+b_{lrq}$

The expectation over sampling is

$$E_{pm}\left(\hat{\theta}_l\right)=E_p\left[\frac{1}{N_l}\sum_{r=1}^{2^L-1}\frac{1}{m_r}\sum_{q\in s}\frac{\mu_{lrq}^*}{\pi_{lrq}}\right]$$
$$=E_p\left[\frac{1}{N_l}\sum_{r=1}^{2^L-1}\frac{1}{m_r}\sum_{q=1}^{N_{\omega_r}}\frac{\mu_{lrq}^*I_{lrq}}{\pi_{lrq}}\right]$$
$$=\frac{1}{N_l}\sum_{r=1}^{2^L-1}\frac{1}{m_r}\sum_{q=1}^{N_{\omega_r}}\mu_{lrq}^*=\theta_l^* \tag{11}$$

$$E_{pm}\left(\hat{\mu}\right)=\frac{1}{N_l}\sum_{r=1}^{2^L-1}\frac{1}{m_r}\sum_{l=1}^{N_{lwr}}\frac{\mu_{lrq}^*E_p\left(I_{lrq}\right)}{\pi_{lrq}}$$
$$=\frac{1}{N_l}\sum_{r=1}^{2^L-1}\frac{1}{m_r}\sum_{l=1}^{N_{lwr}}\mu_{lrq}^* =\text{E}(\hat{\theta}_l) \tag{12}$$

We can therefore write

$$E_{pm}\left(\hat{\mu}\right)=E_{pm}\left(\sum_{L=1}^{L}\alpha_l\hat{\theta}_l\right)=\sum_{L=1}^{L}\alpha_l\theta_l^*=\theta^*$$

Where $\theta^*=E_{pm}(\hat{\theta})$ and $\theta_l^*=E_{pm}(\hat{\theta}_l)$

The bias due to measurement errors is

$$Bias\left(\hat{\theta}\right)=E_{pm}\left(\hat{\theta}\right)-\mu=\theta^*-\mu$$

The variance of the estimate is given by

$$V\left(\hat{\theta}_l\right)=V_pE_m\left(\hat{\theta}_l\right)+E_pV_m\left(\hat{\theta}\right)$$

$$V_p\left[E_m\left(\hat{\theta}_l\right)\right]=V_p\left[\frac{1}{N_l}\sum_{r=1}^{2^L-1}\frac{1}{m_r}\sum_{q\in s}\frac{\mu_{lrq}^*}{\pi_{lrq}}\right]$$

$$=\frac{1}{N_l^2}\left\{\sum_{r=1}^{2^L-1}\frac{1}{m_r^2}\left[\sum_{q=1}^{N_{\omega_r}}\frac{\mu_{lrq}^*}{\pi_{lrq}^2}V\left(I_{lrq}\right)+\sum\sum_{q\neq q'}\frac{\mu_{lrq}^*\mu_{lrq'}^*}{\pi_{lrq}\pi_{lrq'}}\operatorname{cov}\left(I_{lrq},I_{lrq'}\right)\right]+\right.$$
$$\left.\sum\sum_{r\neq r'}\frac{1}{m_rm_{r'}}\sum_{q=1}^{N_{\omega_r}}\sum_{q'=1}^{N_{\omega_r}}\frac{\mu_{lrq}^*\mu_{lrq'}^*}{\pi_{lrq}\pi_{lr'q'}}\operatorname{cov}\left(I_{lrq},I_{lrq'}\right)\right\}$$

$$=\frac{1}{N_l^2}\left\{\sum_{r=1}^{2^L-1}\frac{1}{m_r^2}\left[\sum_{q=1}^{N_{\omega_r}}\frac{\mu_{lrq}^*}{\pi_{lrq}^2}V\left(I_{lrq}\right)+\sum\sum_{q\neq q'}\frac{\mu_{lrq}^*\mu_{lrq'}^*}{\pi_{lrq}\pi_{lrq'}}\pi_{lrqq'}\right]+\right.$$
$$\left.\sum\sum_{r\neq r'}\frac{1}{m_rm_{r'}}\sum_{q=1}^{N_{\omega_r}}\sum_{q'=1}^{N_{\omega_r}}\frac{\mu_{lrq}^*\mu_{lrq'}^*}{\pi_{lrq}\pi_{lr'q'}}\pi_{lrr'qq'}\right\} \quad (13)$$

We then obtain

$$E_pV_m\left(\hat{\theta}_l\right)$$

$$V_m\left(\hat{\theta}_l\right)=V_m\left[\frac{1}{N_l}\sum_{r=1}^{2^L-1}\frac{1}{m_r}\sum_{q\in s}\frac{y_{lrq}}{\pi_{lrq}}\right]$$

$$=\frac{1}{N_l^2}\left[\sum_{l=1}^{L}\frac{1}{m_r^2}\left(\sum_{q\in s}\frac{V\left(e_{lrq}\right)}{\pi_{lrq}}+\sum\sum_{q\neq q'}\frac{\operatorname{cov}\left(e_{lrq},e_{lrq'}\right)}{\pi_{lrq},\pi_{lrq'}}\right)\right]$$

$$=\frac{1}{N_l^2}\sum_{r=1}^{2^L-1}\frac{1}{m_r^2}\left[\sum_{q\in s}\frac{\sigma_{lrq}^2}{\pi_{lrq}^2}+\sum\sum_{q\neq q'}\frac{\sigma_{lrqq'}^2}{\pi^2_{lrqq'}}\right]+\frac{1}{N_l^2}\sum\sum_{r\neq r'}\frac{1}{m_rm_{r'}}\sum\sum_{q\neq q'}\sigma_{lrr'qq'}$$

Then

$$E_p\left(V_m\right)=E_p\left[\frac{1}{N_l^2}\sum_{r=1}^{2^L-1}\frac{1}{m_r^2}\sum_{q=1}^{N_{\omega_r}}\frac{\sigma_{lrq}^2}{\pi_{lrq}^2}I_{lrq}+\sum\sum_{q\neq q'}^{N_{\omega_r}}\frac{\sigma_{lrrqq'}}{\pi_{lrq}\pi_{lrq}}I_{lrq}I_{lrq'}\right]$$

$$=\frac{1}{N_l^2}\sum_{r=1}^{2^L-1}\frac{1}{m_r^2}\left[\sum_{q=1}^{N_{\omega_r}}\frac{\sigma_{lrq}^2\pi_{lrq}}{\pi_{lrq}^2}+\sum\sum_{q\neq q'}^{N_{\omega_r}}\frac{\sigma_{lrrqq'}\pi_{lrrqq'}}{\pi_{lrq}\pi_{lrq'}}\right]+\frac{1}{N_l^2}\sum\sum_{r\neq r'}\frac{1}{m_rm_{r'}}\sum\sum_{q\neq q'}\frac{\sigma_{lrr'qq'}\pi_{lrr'qq'}}{\pi_{lrq}\pi_{r'q'}} \quad (14)$$

Finally,

$$V\left(\hat{\theta}\right)=\sum\alpha_l^2V\left(\hat{\theta}_l\right)$$, assuming independence between centres to get,

$$V\left(\hat{\theta}\right)=\sum_{l=1}^{L}\frac{N_l^2}{N^2}\frac{1}{N_l^2}\left\{\sum_{r=1}^{2^L-1}\frac{1}{m_r^2}\left[\mu_{lrq}^{*2}\left(\frac{1}{\pi_{lrq}}-1\right)+\sum\sum_{q\neq q'}\mu_{lrq}^{*}\mu_{lrq'}^{*}\left(\frac{\pi_{lrrqq'}}{\pi_{lrq}\pi_{lrq'}}-1\right)\right]\right.$$

$$+\sum\sum_{r\neq r'}\frac{1}{m_r m_{r'}}\sum\sum_{q\neq q'}\frac{\mu_{lrq}^{*}\mu_{lrq'}^{*}}{\pi_{lrq}\pi_{lrq'}}\left(\pi_{lrqrq'}-\pi_{lrq}\pi_{lrq'}\right)$$

$$\left.+\sum_{l=1}^{L}\frac{N_l^2}{N^2}\frac{1}{N_l^2}\sum_{r=1}^{2^L-1}\frac{1}{m_r^2}\left[\sum\frac{\sigma_{lrq}^2\pi_{lrq}}{\pi_{lrq}^2}+\sum\sum\sigma_{lrqq'}\frac{\pi_{lrrqq'}}{\pi_{lrq}\pi_{lrq'}}\right]\right\} \quad (15)$$

3.1 Simple Random Sampling

Under simple random sampling without replacement,

$$\pi_{lrq}=\frac{n_l}{N_l},\quad \pi_{lrrqq}=\frac{n_l\left(n_l-1\right)}{N_l\left(N_l-1\right)}$$

$$E\left(I_{lrq}\right)=\frac{n_l}{N_l}=\pi_{lrq}=\frac{n_l}{N_l}$$

$$V\left(I_{lrq}\right)=\pi_{lrq}-\pi_{lrq}^2=\frac{n_l}{N_l}\left(1-\frac{n_l}{N_l}\right)$$

$$C_{OV}\left(I_{lrq},I_{lrq'}\right)=C_{OV}\left(I_{lrq},I_{lr'q'}\right)=\pi_{lrqq'}-\pi_{lrq}\pi_{lrq'}=\frac{n_l(n_l-N_l)}{N_l^2(N_l-1)}$$

This gives

$$V\left(\hat{\theta}\right)=\sum_{l=1}^{L}\frac{\alpha_l^2\left(N_l-n_l\right)}{n_lN_l\left(N_l-1\right)}\left[\sum_{r=1}^{2^L-1}\sum_{q=1}^{N_{\omega_r}}\frac{\mu_{lrq}^{*}}{m_r^2}-\frac{1}{N_l}\left(\sum_{r=1}^{2^L-1}\frac{\mu^{*}{}_{lrq}}{m_r}\right)^2\right]$$

$$+\sum_{l=1}^{L}\frac{\alpha_l^2}{n_lN}\sum_{r=1}^{2^L-1}\frac{1}{m_r^2}\left[\sum_{q=1}^{N_{\omega_r}}\sigma_{lrq}^2+\frac{n_l-1}{N_l-1}\sum\sum\sigma_{lrqq'}\right] \quad (16)$$

An estimator of the variance is

$$\hat{V}\left(\hat{\theta}\right)=\sum_{l=1}^{L}\frac{\alpha_l^2}{n_l^2\left(n_l-1\right)}\left[n_l\sum_{r=1}^{2^L-1}\sum_{s=1}^{n_{l\omega_r}}\frac{y_{lrs}^2}{m_r^2}-\left(\sum_{r=1}^{2^L-1}\sum_{s=1}^{n_{l\omega_r}}\frac{y_{lrs}}{m_r}\right)^2\right] \quad (17)$$

Then

$$V\left(\hat{\theta}_l\right)=\frac{1}{N^2}\left\{\sum_{r=1}^{2^L-1}\frac{1}{m_r}\left[\sum_{q=1}^{N_{lwr}}\mu_{lrq}^2\left(\frac{N_l}{n_l}-1\right)+\sum_{q\neq q'}\sum\mu_{lrq}\mu_{lrq'}\left(\frac{n_i-N_l}{n_l\left(N_l-1\right)}\right)\right.\right.$$

$$\left.+\sum\sum_{r\neq r'}\frac{1}{m_rm_{r'}}\sum_{q\neq q'}^{N_{lwr}}\sum^{N_{l\omega_{r'}}}\mu_{lrq}\mu_{lr'q'}\frac{\left(n_{l'}-N_{l'}\right)}{n_{l'}\left(N_{l'}-1\right)}\right\} \quad (18)$$

The variance of the estimator can be written as

$$V_{pm}\left(\hat{\theta}\right)=V_pE_m\left(\hat{\theta}\right)+E_pV_m\left(\hat{\theta}\right)$$

Where,

$$V_pE_m\left(\hat{\theta}\right)=\sum_L \alpha_l^2\frac{1}{N_l^2}\left\{\sum_{r=1}^{2^L-1}\frac{1}{m_r^2}\left[\sum_{q=1}^{N_{lwr}}\frac{\mu_{lrq}^{*2}\left(1-\frac{n_l}{N_l}\right)}{\frac{n_l}{N_l}}+\sum_{q\neq q'}^{N_{lwr}}\sum\frac{\mu_{lrq}^{*}\mu_{lrq'}^{*}}{\frac{n_l}{N_l}\frac{n_l}{N_l}}\left(\frac{n_l\left(n_l-1\right)}{N_l\left(N_l-1\right)}-\frac{n_l}{N_l}\frac{n_l}{N_l}\right)\right]\right.$$

$$\left.+\sum_{r\neq r'}\sum\frac{1}{m_rm_{r'}}\sum_{q=1}^{N_{lwr}}\sum_{q'=1}^{N_{lwr'}}\frac{\mu_{lrq}^{*}\mu_{lr'q'}^{*}}{\frac{n_l}{N_l}\frac{n_l}{N_l}}\left(\frac{n_l\left(n_l-1\right)}{N_l\left(N_l-1\right)}-\frac{n_l}{N_l}\frac{n_l}{N_l}\right)\right\}$$

$$=\sum_L \alpha_l^2\frac{1}{N_l^2}\left\{\sum_{r=1}^{2^L-1}\frac{1}{m_r^2}\left[\left(\frac{N_l}{n_l}-1\right)\sum_{q=1}^{N_{lwr}}\mu_{lrq}^{*2}+\frac{(n_l-N_l)}{n_l(N_l-1)}\sum_{q\neq q'}^{N_{lwr}}\sum\mu_{lrq}^{*}\mu_{lrq'}^{*}\right]\right.$$

$$\left.+\sum_{r\neq r'}\sum\frac{1}{m_rm_{r'}}\sum_{q=1}^{N_{lwr}}\sum_{q'=1}^{N_{lwr'}}\mu_{lrq}^{*}\mu_{lr'q'}^{*}\frac{(n_{l'}-N_{l'})}{n_{l'}(N_{l'}-1)}\right\} \qquad (19)$$

And

$$E_pV_m\left(\hat{\theta}\right)=\sum_L \alpha_l^2\frac{1}{N_l^2}\left\{\sum_{r=1}^{2^L-1}\frac{1}{m_r^2}\left[\sum_{q=1}^{N_{lwr}}\frac{\sigma_{lrq}^2}{\frac{n_l}{N_l}}+\sum_{q\neq q'}^{N_{lwr}}\sum\frac{\sigma_{lrqq'}}{\frac{n_l}{N_l}\frac{n_l}{N_l}}\frac{n_l\left(n_l-1\right)}{N_l\left(N_l-1\right)}\right]\right.$$

$$\left.+\sum_{r\neq r'}\sum\frac{1}{m_rm_{r'}}\sum_{q=1}^{N_{lrq}}\sum_{q'=1}^{N_{lr'q'}}\frac{\sigma_{lrr'qq'}}{\frac{n_l}{N_l}\frac{n_l}{N_l}}\frac{n_l\left(n_l-1\right)}{N_l\left(N_l-1\right)}\right\}$$

$$=\frac{1}{N_l^2}\left\{\sum_{r=1}^{2^L-1}\frac{1}{m_r^2}\left[\frac{N_l}{n_l}\sum_{q=1}^{N_{lwr}}\sigma_{lrq}^2+\frac{(n_l-N_l)}{n_l(N_l-1)}\sum_{q\neq q'}^{N_{lwr}}\sum\sigma_{lrqq'}\right]+\sum_{r\neq r'}\sum\frac{1}{m_rm_{r'}}\sum_{q=1}^{N_{lrq}}\sum_{q'=1}^{N_{lr'q'}}\sigma_{lrr'qq'}\frac{(n_{l'}-N_{l'})}{n_{l'}(N_{l'}-1)}\right\} \qquad (20)$$

4. Estimation under two phase Centre Sampling (CS) for elusive population when there are measurement and response errors

In two phase centre sampling a sample say S' is selected at the first phase (phase 1) and some auxiliary variable X is observed. In phase 2, a subsample is selected from the phase 1 sample.

Let as usual π_{lrq} be the inclusion probability of unit q after the two phases. We can write

$\pi_{lrq} =$ P(unit q is selected after both phases)

= P(unit q is selected at phase 1) ×P(unit q is selected at phase 2 given that it was selected at phase 1)

$= \pi_{lrq1} \times \pi_{lrq2}$

Or simply

$\pi_{lrq} = \pi_{lrq1}.\pi_{lrq2}$

The estimate of μ is therefore

$$\hat{\theta}=\sum_{l=1}^{L}\alpha_l\hat{\theta}_l$$

Where $\hat{\theta}_l = \frac{1}{N_l}\sum_{r=1}^{2^L-1}\frac{1}{m_r}\sum_{q\in s}\frac{y_{lrq}}{\pi_{lrq}}$

The expectation now becomes

$$E_{p_1p_2m}(\hat{\theta}_l) = E_{p_1p_2}\frac{1}{N_l}\sum_{r}^{2^L-1}\frac{1}{m_r}\sum_{q\in s}\frac{\mu^*_{lrq}}{\pi_{lrq}}$$

$$E_{p_1}E_{p_2}\left[\frac{1}{N_l}\sum_{r}^{2^L-1}\frac{1}{m_r}\sum_{q\in s'}\frac{\mu^*_{lrq}I_{lrqs}}{\pi_{lrq}}\right] = E_{p_1}\left[\frac{1}{N_l}\sum_{r=1}^{2^L-1}\frac{1}{m_r}\sum_{q\in s'}\frac{\mu^*_{lrq}}{\pi_{lrq_1}}\right]$$

$$= E_{p_1}\left[\frac{1}{N_l}\sum_{r=1}^{2^L-1}\frac{1}{m_r}\sum_{q=1}^{N_{lwr}}\frac{\mu^*_{lrq}}{\pi_{lrq_1}}I_{lrqs'}\right] \text{ since } E_{p_2}(I_{lrqs}) = \pi_{lrq_2} \quad E_{p_1}(I_{lrqs'}) = \pi_{lrq_1}$$

$$= \frac{1}{N_l}\sum_{r=1}^{2^L-1}\frac{1}{m_r}\sum_{q=1}^{N_{lwr}}\mu^*_{lrq} = \theta^* \tag{21}$$

The variance under the two-phase sampling and measurement model is given by

$$V\left(\hat{\theta}_l\right) = V_{P_1P_2m}\hat{\theta}_l = E_{p_1}E_{p_2}E_m\left(\hat{\theta}_l - \hat{\theta}\right)^2$$

$$= V_{p_1}E_{p_2}E_m\left(\hat{\theta}_l\right) + E_{p_1}V_{p_2}E_m\left(\hat{\theta}_l\right) + E_{p_1}E_{p_2}V_m\left(\hat{\theta}_l\right)$$

The three components are obtained as follows,

$$V_{p_1}E_{p_2}E_m\left(\hat{\theta}\right) = V_{p_1}E_{p_2}E_m\left(\frac{1}{\mathrm{N_l}}\sum_{r=1}^{2^L-1}\frac{1}{\mathrm{m_r}}\sum\frac{\mathrm{y_{lrq}}}{\pi_{\mathrm{lrq}}}\right)$$

$$= \mathrm{V}_{p_1}\mathrm{E}_{p_2}\left(\frac{1}{N_l}\sum_{r}^{2^L-1}\frac{1}{m_r}\sum_{q\in s}\frac{u^*_{lrq}}{\pi_{lrq}}\right)$$

$$= V_{p_1}E_{p_2}\left(\frac{1}{N_l}\sum_{r}^{2^L-1}\frac{1}{m_r}\sum_{q\in s}\frac{u^*_{lrq}I_{lrqs}}{\pi_{lrq}}\right)$$

$$= V_{p_1}\left(\frac{1}{N_l}\sum_{r}^{2^L-1}\frac{1}{m_r}\sum_{q\in s'}\frac{u^*_{lrq}}{\pi_{lrq_1}}\right) = V_{p_1}\left(\frac{1}{N_l}\sum_{r}^{2^L-1}\frac{1}{m_r}\sum_{q=1}^{N_{l\omega_r}}\frac{u^*_{lrq}I_{lrqs'}}{\pi_{lrq_1}}\right)$$

$$= \frac{1}{N_l^2}\{\sum_{r}^{2^L-1}\frac{1}{m_r}\left[\sum_{q=1}^{N_{\omega r}}\frac{u^{*2}_{lrq}\pi_{lrq_1}\left(1-\pi_{lrq_1}\right)}{\pi^2_{lrq_1}} + \sum\sum_{q\neq q_1}^{N_{l\omega_r}}\frac{u^*_{lrq}u^*_{lrq_1}}{\pi_{lrq_1}\pi_{lrq_1'}}\left(\pi_{lrqq_1} - \pi_{lrq_1}\pi_{lrq'_1}\right)\right]$$

$$+\sum\sum_{r\neq r'}\frac{1}{m_rm_{r'}}\sum_{q=1}^{N_{l\omega_r}}\sum_{q'=1}^{N_{l\omega_{r'}}}\frac{u^*_{lrq}u^*_{lrq'}}{\pi_{lrq_1}\pi_{lrq_1'}}\left(\pi_{lr_1'qq'} - \pi_{lrq_1}\pi_{lr'q'_1}\right)\} \tag{22}$$

Then

$$E_{p_1}V_{p_2}E_m\left(\hat{\theta}\right) = E_{p_1}V_{p_2}\left[\frac{1}{N_l}\sum_{r}^{2^L-1}\frac{1}{m_r}\sum_{q\in s}\frac{u^*_{lrq}}{\pi_{lrq}}\right]$$

$$= E_{p_1}V_{p_2}\left[\frac{1}{N_l}\sum_{r}^{2^L-1}\frac{1}{m_r}\sum_{q\in s'}\frac{u^*_{lrq}I_{lrq_2}}{\pi_{lrq}}\right]$$

$$= E_{p_1} \frac{1}{N_l^2} \{ \sum_{r=1}^{2^L-1} \frac{1}{m_r^2} \left[\sum_{q \in s'} \frac{u_{lrq}^{*2} \pi_{lrq_2} \left(1-\pi_{lrq_2}\right)}{\pi_{lrq}^2} + \sum \sum_{q \neq q'}^{S'} \frac{u_{lrq}^{*} u_{lrq'}^{*}}{\pi_{lrq}\pi_{lrq'}} \left(\pi_{lrqq'_2} - \pi_{lrq_2}\pi_{lrq'_2}\right) \right]$$

$$+ \sum\sum_{r \neq r'} \frac{1}{m_r m_{r'}} \sum_{q=1}\sum_{q'=1}^{S'} \frac{u_{lrq}^{*} u_{lr'q'}^{*}}{\pi_{lrq}\pi_{lr'q'}} \left(\pi_{lrr'qq'_2} - \pi_{lrq_2}\pi_{lr'q'_2}\right) \}$$

$$= E_{p_1} \frac{1}{N_l^2} \{ \sum_{r=1}^{2^L-1} \frac{1}{m_r^2} \left[\sum_{q=1}^{N_{l\omega_r}} \frac{u_{lrq}^{*2}}{\pi_{lrq}^2} \pi_{lrq_2} \left(1-\pi_{lrq_2}\right) I_{rq_1} + \sum \sum_{q \neq q'}^{N_{l\omega_r}} \frac{u_{lrq}^{*} u_{lrq'}^{*}}{\pi_{lrq}\pi_{lrq'}} (\pi_{lrqq'_2} - \pi_{lrq_2}\pi_{lrq'_2}) I_{lrr'qq's'} \right]$$

$$+ \sum\sum_{r \neq r'} \frac{1}{m_r m_{r'}} \sum_{q=1}^{N_{l\omega_r}} \sum_{q' \neq 1}^{N_{l\omega_{r'}}} \frac{u_{lrq}^{*} u_{lr'q'}^{*}}{\pi_{lrq}\pi_{lrq'}} \left(\pi_{lrr'qq'_2} - \pi_{lrq_2}\pi_{lr'q'_2}\right) \mathrm{I}_{lrr'qq's'} \}$$

$$= \frac{1}{N_l^2} \{ \sum_{r=1}^{2^L-1} \frac{1}{m_r^2} \left[\sum_{q=1}^{N_{l\omega_r}} \frac{u_{lrq}^{*2}}{\pi_{lrq}} \left(1-\pi_{lrq_2}\right) + \sum \sum_{q \neq q'}^{N_{l\omega_r}} \frac{u_{lrq}^{*} u_{lrq'}^{*}}{\pi_{lrq}\pi_{lrq'}} \pi_{lrqq'_1} (\pi_{lrqq'_2} - \pi_{lrq_2}\pi_{lrq'_2}) \right]$$

$$+ \sum\sum_{r \neq r'} \frac{1}{m_r m_{r'}} \sum_{q=1}^{N_{l\omega_r}} \sum_{q' \neq 1}^{N_{l\omega_{r'}}} \frac{u_{lrq}^{*} u_{lr'q'}^{*}}{\pi_{lrq}\pi_{lr'q'}} \pi_{lrr'qq'_1} \left(\pi_{lrr'qq'_2} - \pi_{lrq_2}\pi_{lr'q'_2}\right) \} \tag{23}$$

Then

$$E_{p_1} E_{p_2} V_m = E_{p_1} E_{p_2} V_m \left(\frac{1}{N_l} \sum_{1}^{2^L-1} \frac{1}{m_r} \sum_{q \in s} \frac{y_{lrq}}{\pi_{lrq}} \right)$$

$$= E_{p_1} E_{p_2} \frac{1}{N_l^2} \left\{ \sum_{r=1}^{2^L-1} \frac{1}{m_r^2} \left[\sum_{q \in s} \frac{\sigma_{lrq}^2}{\pi_{lrq}^2} + \sum \sum_{q \neq q'}^{S} \frac{\sigma_{lrqq'}}{\pi_{lrq}\pi_{lrq'}} \right] + \sum\sum_{r \neq r'} \frac{1}{m_r m_{r'}} \sum_{q=1}^{S} \sum_{q'=1}^{S} \frac{\sigma_{lrr'qq'}}{\pi_{lrq}\pi_{lr'q'}} \right\}$$

$$E_{p_1} E_{p_2} V_m = E_{p_1} E_{p_2} \frac{1}{N_l^2} \left\{ \sum_{r=1}^{2^L-1} \frac{1}{m_r^2} \left[\sum_{q \in s'} \frac{\sigma_{lrq}^2}{\pi_{lrq}^2} I_{lrqs} + \sum \sum_{q \neq q'}^{S'} \frac{\sigma_{lrqq'} I_{lrr'qq's}}{\pi_{lrq}\pi_{lrq'}} \right] + \sum\sum_{r \neq r'} \frac{1}{m_r m_{r'}} \sum \sum_{q \neq q'}^{S'} \frac{\sigma_{lrr'qq'} I_{lrr'qq's}}{\pi_{lrq}\pi_{lr'q'}} \right\}$$

$$= E_{p_1} \frac{1}{N_l^2} \left\{ \sum_{r=1}^{2^L-1} \frac{1}{m_r^2} \left[\sum_{q \in s'} \frac{\sigma_{lrq}^2}{\pi_{lrq1}^2 \pi_{lrq2}} + \sum \sum_{q \neq q'}^{S'} \frac{\sigma_{lrqq'}^2 \pi_{lrqq'_2}}{\pi_{lrq}\pi_{lrq'}} \right] + \sum\sum_{r \neq r'} \frac{1}{m_r m_{r'}} \sum_{q=1}^{S'} \sum_{q' \neq 1} \frac{\sigma_{lrr'qq'} \pi_{lrr'qq'_2}}{\pi_{lrq}\pi_{lr'q'}} \right\}$$

$$= E_{p_1} \frac{1}{N_l^2} \left\{ \sum_{r=1}^{2^L-1} \frac{1}{m_r^2} \left[\sum_{q=1}^{N_{l\omega_r}} \frac{\sigma_{lrq}^2}{\pi_{lrq_1}^2 \pi_{lrq2}} I_{lrqs'} + \sum \sum_{q \neq q'}^{N_{l\omega_r}} \frac{\sigma_{lrqq'}}{\pi_{lrq}\pi_{lrq'}} \pi_{lrr'qq'_2} I_{lrr'qq's'} \right] + \sum\sum_{r \neq r'} \frac{\sigma_{lrr'qq'}}{\pi_{lrq}\pi_{lr'q'}} \pi_{lrr'qq'_2} I_{lrr'qq's'} \right\}$$

$$= \frac{1}{N_l^2} \left\{ \sum_{r=1}^{2^L-1} \frac{1}{m_r^2} \left[\sum_{q=1}^{N_{l\omega_r}} \frac{\sigma_{lrq}^2}{\pi_{lrq1}\pi_{lrq2}} + \sum\sum \frac{\sigma_{lrqq'} \pi_{lrr'qq'}}{\pi_{lrq}\pi_{lrq'}} \right] + \sum\sum_{r \neq r'} \frac{1}{m_r m_{r'}} \sum\sum \frac{\sigma_{lrr'qq'} \pi_{lrr'qq'}}{\pi_{lrq}\pi_{lr'q'}} \right\}$$

$$= \frac{1}{N_l^2} \left\{ \sum_{r=1}^{2^L-1} \frac{1}{m_r^2} \left[\sum_{q=1}^{N_{l\omega_r}} \frac{\sigma_{lrq}^2}{\pi_{lrq1}\pi_{lrq2}} + \sum\sum \frac{\sigma_{lrqq'} \pi_{lrr'qq'}}{\pi_{lrq}\pi_{lrq'}} \right] + \sum\sum_{r \neq r'} \frac{1}{m_r m_{r'}} \sum\sum \frac{\sigma_{lrr'qq'} \pi_{lrr'qq'}}{\pi_{lrq}\pi_{lr'q'}} \right\}$$

$$= \frac{1}{N_l^2} \left\{ \sum_{r=1}^{2^L-1} \frac{1}{m_r^2} \left[\sum_{q=1}^{N_{l\omega_r}} \frac{\sigma_{lrq}^2}{\pi_{lrq}} + \sum\sum_{q \neq q'} \frac{\sigma_{lrqq'} \pi_{lrr'qq'}}{\pi_{lrq}\pi_{lrq'}} \right] + \sum\sum_{r \neq r'} \frac{1}{m_r m_{r'}} \sum\sum \frac{\sigma_{lrr'qq'} \pi_{lrr'qq'}}{\pi_{lrq}\pi_{lr'q'}} \right\} \tag{24}$$

4.1. Simple random sampling in both phases

$$V_{p_1} E_{p_2} E_m\left(\hat{\theta}\right) = \frac{1}{N_l^2} \{ \sum_{r=1}^{2^L-1} \frac{1}{m_r^2} \left[\sum_{q=1}^{N_{l\omega_r}} u_{lrq}^{*} \frac{\frac{n_l'}{N_l}\left(1-\frac{n_l'}{N_L}\right)}{\frac{n_l'^2}{N_l^2}} + \sum_{q \neq q'}^{N_{l\omega_r}} \sum \frac{u_{lrq}^{*} u_{lrq'}^{*}}{\frac{n_l'}{N_l} \cdot \frac{n_l'}{N_l}} \left(\frac{n_l'\left(n_l'-1\right)}{N_l\left(N_l-1\right)} - \frac{n_l'}{N_l} \cdot \frac{n_l'}{N_l} \right) \right]$$

$$+\sum_{r\neq r'}\sum\frac{1}{m_r m_r}N_l\sum_{q=1}^{N_{l\omega_r}}\sum_{q'=1}^{N_{l\omega_{r'}}}\frac{u^*_{lrq}u^*_{lr'q}}{\frac{n'_l}{N_l}\cdot\frac{n'_l}{N_l}}\left(\frac{n'_l\left(n'_l-1\right)}{N_l\left(N_l-1\right)}-\frac{n'_l}{N_l}\cdot\frac{n'_l}{N_l}\right)\}$$

$$=\frac{1}{N_l^2}\left\{\sum_{r=1}^{2^L-1}\frac{1}{m_r^2}\left[\frac{N_l-n'_l}{n'_l}\sum_{q=1}^{N_{l\omega_r}}u^*_{lrq}+\frac{\left(n'_l-N_l\right)}{n'_l\left(N_l-1\right)}\sum_{q\neq q'}^{N_{l\omega_r}}\sum u^*_{lrq}u^*_{lrq'}\right]+\frac{\left(n'_l-N_l\right)}{n'_l\left(N_l-1\right)}\sum_{r\neq r'}\frac{1}{m_r m_{r'}}\sum_{q=1}^{N_{l\omega_r}}\sum_{q'=1}^{N_{l\omega_{r'}}}u^*_{lrq}u^*_{lrq_1}\right\} \quad (25)$$

Secondly

$$E_{p_1}V_{p_2}E_m\left(\hat{\theta}\right)=\frac{1}{N_l^2}\{\sum_{r=1}^{2^L-1}\frac{1}{m_r^2}\left[\sum_{q=1}^{N_{l\omega_r}}\frac{u^{*2}_{lrq}\left(1-\frac{n_l}{n'_l}\right)}{\frac{n_l}{n'_l}}+\sum_{q\neq q'}^{N_{l\omega_r}}\sum\frac{u^*_{lrq}u^*_{lrq'}}{\frac{n_l}{n'_l}\cdot\frac{n_l}{n'_l}}\frac{n_l\left(n_l-1\right)}{n'_l\left(n'_l-1\right)}\left(\frac{n_l\left(n_l-1\right)}{n'_l\left(n'_l-1\right)}-\frac{n_l}{n'_l}\cdot\frac{n_l}{n'_l}\right)\right]$$

$$+\sum_{r\neq r'}\sum\frac{1}{m_r m_{r'}}\sum_{q=1}^{N_{l\omega_r}}\sum_{q'=1}^{N_{l\omega_{r'}}}\frac{u^*_{lrq}u^*_{lrq'}}{\frac{n_l}{n'_l}\cdot\frac{n_l}{n'_l}}\frac{n_l\left(n_l-1\right)}{n'_l\left(n'_l-1\right)}\left(\frac{n_l\left(n_l-1\right)}{n'_l\left(n'_l-1\right)}-\frac{n_l}{n'_l}\cdot\frac{n_l}{n'_l}\right)$$

$$=\frac{1}{N_l^2}\left\{\sum_{r=1}^{2^L-1}\frac{1}{m_r^2}\left[\frac{n'_l-n_l}{n_l}\sum_{q=1}^{N_{l\omega_r}}u^{*2}_{lrq}+\frac{\left(n_l-1\right)\left(n_l-n'_l\right)}{n'_l\left(n'_l-1\right)^2}\sum_{q\neq q'}^{N_{l\omega_r}}\sum u^*_{lrq}u^*_{lrq'}\right]+\frac{\left(n_l-1\right)\left(n_l-n'_l\right)}{n'_l\left(n'_l-1\right)^2}\sum_{r\neq r'}\sum\frac{1}{m_r m_{r'}}\sum_{q=1}^{N_{l\omega_r}}\sum_{q'=1}^{N_{l\omega_{r'}}}u^*_{lrq}u^*_{lrq_1}\right\} \quad (26)$$

Similarly

$$E_{p_1}E_{p_2}V_m=\frac{1}{N_l^2}\left\{\sum_{r=1}^{2^L-1}\frac{1}{m_r^2}\left[\frac{N_l n'_l}{n'_l n_l}\sum_{q=1}^{N_{l\omega_r}}\sigma^2_{lrq}+\sum_{q\neq q'}^{N_{l\omega_r}}\sigma_{lrqq'}\frac{\frac{n_l\left(n_l-1\right)}{n'_l\left(n'_l-1\right)}}{\frac{n_l}{n'_l}\cdot\frac{n_l}{n'_l}}\right]+\sum_{r\neq r'}\sum\frac{1}{m_r m_{r'}}\sum_{q=1}^{N_{l\omega_r}}\sum_{q'=1}^{N_{l\omega_{r'}}}\sigma_{lrr'qq'}\frac{\frac{n_l\left(n_l-1\right)}{n'_l\left(n'_l-1\right)}}{\frac{n_l}{n'_l}\cdot\frac{n_l}{n'_l}}\right\}$$

$$=\frac{1}{N_l^2}\left\{\sum_{r=1}^{2^L-1}\frac{1}{m_r^2}\left[\frac{N_l}{n_l}\sum_{q=1}^{N_{l\omega_r}}\sigma^2_{lrq}+\frac{n'_l\left(n_l-1\right)}{n_l\left(n'_l-1\right)}\sum_{q\neq q'}^{N_{l\omega_r}}\sum\sigma_{lrqq'}\right]+\frac{n'_l\left(n_l-1\right)}{n_l\left(n'_l-1\right)}\sum_{r\neq r'}\sum\frac{1}{m_r m_{r'}}\sum_{q=1}^{N_{l\omega_r}}\sum_{q'=1}^{N_{l\omega}}\sigma_{lrr'qq'}\right\} \quad (27)$$

5. A numerical example

Considering an example with four centers. This implies that there are 2^L-1=15 possible profiles with multiplicities as shown in Table 1.

Table 1: Profiles and multiplicities of the 4 centers

Centers	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	0	0	0	1	1	1	0	0	0	1	1	1	0	1
2	0	1	0	0	1	0	0	1	1	0	1	1	0	1	1
3	0	0	1	0	0	1	0	1	0	1	1	0	1	1	1
4	0	0	0	1	0	0	1	0	1	1	0	1	1	1	1
Multiplicities	1	1	1	1	2	2	2	2	2	2	3	3	3	3	4

A normal population of size N=200 was randomly simulated with mean=60 and variance=10. The population values were then assigned to the profiles randomly. The results are as shown below,

Estimate $\hat{\mu}=59.9700$ and Estimate $Var(\hat{\mu})=9.4061$

A sample of size n=60 is then randomly selected from the N=200 simulated values to obtain $\mu_i,\ i=1,2,...,\mathrm{n}$. A population with measurement and response errors is then generated as $y_i,\ i=1,2,...,\mathrm{n}$, where $y_i=\mu_i+\mathrm{e}_i,\ i=1,2,...,\mathrm{n}$ and $\mathrm{e}_i\sim N(k\mu_i,p\mu_i)$, where k and p are positive constants. The population with measurement and response errors is randomly assigned to the profiles for different k and p. The sample sizes for the centres are, $n_1=32,\ \mathrm{n}_2=34,\ n_3=35 \text{ and } n_4=32$. The results are obtained as shown in Table 2.

Table 2: Mean and variance when there are measurement and response errors

P	k	$\hat{\mu}$	Estimate ($\hat{\theta}$)	$Var(\hat{\mu})$	Estimate $Var(\hat{\theta})$
0	0	60.0500		9.4600	
0.01	0.05		60.2250		9.4850
	0.1		61.8800		10.0846
	0.2		63.6800		10.8073
	0.5		69.1000		13.3584
0.05	0.05		60.9850		9.7543
	0.1		61.9050		10.0950
	0.2		63.7250		10.8231
	0.5		69.1250		13.3794

In Table 2 the mean and variance increase as the value of k and p increases. This implies that the more the measurement and response errors the more the mean and variance are overestimated. We deduce that for an increasing error (k) on the mean level, the measurement and response errors consequently increase.

5.1. Two phase Centre Sampling

A sample of size n=60 is then randomly selected from the N=200 simulated values as the first phase units to obtain $\mu_i,\ i=1,2,...,\text{n}$. A population with measurement and response errors is then generated as $y_i,\ i=1,2,...,\text{n}$, where $y_i=\mu_i+\text{e}_i,\ i=1,2,...,\text{n}$ and $\text{e}_i \sim N(k\mu_i, p\mu_i)$, where k and p are positive constants. A sub sample of size n=30 is then selected from the sample of 60 to obtain second phase units. The second phase units with measurement and response errors is randomly assigned to the profiles for different k and p. The sample sizes for the centres are, $n_1=16,\ \text{n}_2=16,\ n_3=17$ and $n_4=16$, The results are obtained as shown in Table 3.

Table 3: Mean and variance when there are measurement and response errors

P	*k*	$\hat{\mu}$	Estimate ($\hat{\theta}$)	$Var(\hat{\mu})$	Estimate $Var(\hat{\theta})$
0	0	64.3396		15.6731	
0.01	0.05		60.6629		12.7545
	0.1		65.2555		15.4595
	0.2		70.2731		18.7168
	0.5		85.4404		30.2640
0.05	0.05		63.0509		14.1860
	0.1		65.6978		15.8042
	0.2		70.6936		19.0541
	0.5		85.7666		30.6431

In Table 3 as the values of p and k increases, the mean and variance are overestimated. Compared to the results in Table 2, both the means and variance estimates are higher in Table 3, that represents the second phase sampling. Although two phase sampling is a strategy to reduce cost, when measurement and response errors are present the estimates are compromised. The presence of measurement and response errors cannot be ruled out in any survey.

6. Conclusion

Carrying out of any survey involves enormous amounts of finances, a lot of time and meticulous planning and coordination of the process. The data obtained is used for important issues such as preparation of budgets, development of strategic plans as well as designing intervention measures for certain undesirable trends in the society. It is therefore imperative that the sources of errors in surveys are identified so as to reduce their effects on the survey data obtained. This paper has shown that the presence of measurement and response errors in elusive surveys greatly affects the results by overestimating and underestimating the estimates. This will seek to improve efficiency on planning, forecasting as well policy formulation processes.

7. References

1. Battistin, E. Errors in Survey Reports of Consumption Expenditures, working paper (03/07), Institute for Fiscal Studies, London, 2003.
2. Birnbaum ZW, Sirken MG. *Design of sample surveys to estimate the prevalence of rare diseases: Three unbiased estimates,* Vital and Health Statistics. Ser. 2, 11, Government Printing Office, Washington, 1965.
3. Blangiardo GC, Migliorati S, Terzera L. Center sampling: from applicative issues to methodologocal aspects. Atti della XLII Riunione Scientifica della SIS, volume Sessioni plenarie e Sessioni specializzate, 2004, 377-388.
4. Bollinger CR. Measurement Error in the Current Population Survey: A Nonparametric Look, Journal of Labor Economics, 1998; 16(3):576-594.

5. Bound J, Krueger AB. The Extent of Measurement Error in Longitudinal Earning Data: Do Two Wrongs Make a Right? Journal of Labor Economics, 1991; *9*(1):1-24.
6. Burton S, Blair E. Task conditions, response formation processes, and response accuracy for behavioral frequency questions in surveys. Public Opinion Quarterly. 1991; 55(1):50-79.
7. Chase DR, Harada M. Response Error in Self-Reported Recreation Participation, Journal of Leisure Research, 1984; 16(4):322-329.
8. Deming EW. On errors in survey. American Sociological Review, 1944; 9(4):359-369.
9. Ecklcr AR, Hurwitz WN. Response variance and biases in censuses and surveys. Bulletin of the International Statistical Institute, 1958; 36(2):12-35.
10. Hansen M, Hurwitz W, Bershad MA. Measurements Errors in Census and Surveys, Bullettin of the International Statistical Institute, 1961 38(2).
11. Hanson RH, Marks ES. Influence of the interviewer on the accuracy of survey results. Journal of the American Statistical Association. 1958; 53(283):635-655.
12. Hedayat AS, Sinha BK. Design and inference in finite population sampling. Wiley, 1991.
13. Kish L. Survey sampling. New York: Wiley, 1965
14. Maiti, Pal M, Sinha K, Bikas. Estimating unknown dimensions of a binary matrix with application to estimation of the size of a mobile population, In Probability and Statistics, S. K. Basu and K. Sinha, Bimal. Editors. Narosa Publishing House, New Delhi, 1993, 220-233.
15. Mathiowetz NA, Ouncan GJ. Out of work, out of mind: Response error in retrospective reports of unemployment. Journal of Business and Economic Statistics. 1988; 6(2):221-229.
16. Mecatti F. Centre sampling and dual frame surveys: comparisons among estimators for the total, Technical Report, Department of Statistics, University of Milan-Bicocca, Italy. Submitted, 2002.
17. Mecatti F, Migliorati S. Center Sampling: A Strategy for Elusive Population Surveys. Department of Statistics, University of Milan-Bicocca, Italy. Submitted, 2003.
18. Migliorati S. New developments in centre sampling, Technical Report, Department of Statistics, University of Milan-Bicocca, Italy. Submitted, 2003
19. Neter J, Waksberg J. A study of response errors in expenditures data from household interviews. Journal of the American Statistical Association, 1964; 59(305):18-55.
20. Pratesi M, Rocco E. Centre Sampling for Estimating Elusive Population Size. Department of Statistics, University of Milan-Bicocca, Italy, 2002; 62(4):745-757.
21. Schwarz NE, Sudman SE. Answering questions: Methodology for determining cognitive and communicative processes in survey research. Jossey-Bass, 1996.
22. Thompson SK, Collins LM. Adaptive sampling in research on risk-related behaviors. Drug and Alcohol Dependence, 2002; 68:57-67.
23. Wyatt RD, Weeks BA, Hamilton PB, Burmeister HR. Severe oral lesions in chickens caused by ingestion of dietary fusariotoxin T-2. Appl. Environ. Microbiol. 1972; 24(2):251-257.