Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Abstract

At the very outset of a pandemic, it is very important to be able to assess the spreading rate of the disease i.e., the rate of increase of infected people in a specific locality. Combating the pandemic situation critically depends on an early and correct prediction of, to what extent the disease may possibly grow within a short period of time. This paper attempts to estimate the spreading rate by counting the total number of infected persons at times. Adaptive clustering is especially suitable for forming clusters of infected persons distributed spatially in a locality and successive sampling is used to measure the growth in number of infected persons. We have formulated a 'chain ratio to regression type estimator of population total' in two occasions adaptive cluster successive sampling and studied the properties of the estimator. The efficacy of the proposed strategy is demonstrated through simulation technique as well as real life population which is followed by suitable recommendation.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

The recent experience of COVID-19 disaster (an extremely infectious disease caused by SARS-CoV-2 virus having common symptoms like fever, dry coughing and difficulty in breathing) has shown the world that procrastination and delay in action during the early outbreak of an epidemic...
may be lethal. At the very outset of the disease, it is critical to be able to measure the potential threat. Incorrect or under assessment may result in a catastrophe!

1.1. Strategy to control the spread

To control such infectious epidemic spread, the strategy of testing, contact tracing and isolating has been proved to be extremely effective. Contact tracing is a continual inverted tree traversal like process of identifying the people tested positive and in turn identifying those, who have come into close contact recently with any of the infected individuals.

The policy of quarantining may include
(a) Firstly, to estimate the total count of infected in a target area and to know the pockets where such people are abundant. This can be achieved by testing repeatedly after a regular interval (which is the whole of the present work).
(b) Estimating the number of individuals, an infected person has come in contact with (contact tracing). Irrespective of being symptomatic or asymptomatic, those should be quarantined.
(c) Assessing a priory, the probable number of infected persons in near future by executing step (a). This will enable us to develop medical and social infrastructural support to combat the disease before it is too late. Even many advanced countries have faced this problem in severe degree during the recent pandemic caused by COVID-19.

1.2. Proposed strategy to measure the growth rate

Since, initially the number of infected people will be less and most often they will be scattered non uniformly in the target area, the simple random sampling may fail to bring out the actual scenario (i.e., number of infected). Adaptive cluster sampling (henceforth referred to as ACS) is ideally suitable to measure the count of infected for such situation. Further, more we can identify the pockets of infection, more we shall be able to quarantine people to prevent further spread. According to Thompson (1990), ACS scheme is effective for epidemiological studies and contact tracing of highly contagious diseases such as Covid-19.

Works using adaptive cluster sampling may be found in Thompson (1990), Smith et al. (1995), Salehi and Seber (1997), Smith et al. (2003) and Dryver and Chao (2007) among others.

But, almost no attempt has been taken to utilize ACS in a situation where, the spatial distribution of infected persons is changing rapidly over time (a case of successive sampling). Add to this, the patchy nature of growth: abruptly scattered in the entire population. We have defined chain “ratio to regression” type estimator for population total using adaptive cluster successive sampling scheme. We have examined the performance of the proposed strategy through simulation study and have obtained encouraging results.

To pursue statistical analysis in such situation, we propose a new estimation strategy as follows:

Since, the entire process of spreading of disease is very much dynamic in nature; a successive sampling scheme is the one that naturally fits here to estimate the average as well as recent situation of spread. It is justified to sample the same population at different points of time to estimate the growth rate (rate of increase in the count of infected) precisely. Moreover, remembering the patchy nature of growth pointed out above, ACS should be a better choice in estimating the actual count. Convinced with the above observations and logic, therefore we have combined the strength of both successive sampling and adaptive clustering strategies to get the synergic effect as a whole.

2. The sampling methodology

It is mentioned and justified in introduction section why our proposed ACS strategy is combined with successive sampling model. We briefly describe both the ideas and their inter link.
2.1. Implementing ACS

It begins with representing the population of a specific area as a gridded square where, each grid presents a unit (i.e., household in this work). \( y_i \) denotes the value of study variable (i.e., the number of infected persons) in the \( i \)th unit. Then we randomly select some units and start forming cluster and thereby network from each of them. The guideline of forming such cluster and then network is called inclusion condition. Repeating such survey periodically, we get a very clear idea about the spatial distribution of infection i.e., the pockets where it is conspicuous and growing rapidly (hotspots). We have framed the following condition for the present work.

2.1.1. Inclusion condition

The ACS is started by choosing any \( n \) units (i.e., households) randomly where at least one infected people is found. Let \( y_i \) denote the number of infected found in \( i \)th unit. Then we choose randomly any 4 new households in the neighborhood to each of the selected ones and test for disease. The members tested positive are added to the count. We stop the process when all four new adjacent households are free from the disease i.e., when we reach edge elements. Thus against each initial unit \((i, y_i)\) we build up a network \( \psi_i \) of size \( m_i \) (\( i = 1, 2, \ldots, n \)). The corresponding total of \( y \) value is expressed in (2). One may note that concept of contract tracing is not used for the purpose.

2.2. The successive sampling model

In situations, where the value of study variable in population is changing in time, successive sampling is the best way to measure the average and current value of it. The problem of sampling on two successive occasions with a part retained from earlier sampling units was first introduced by Jessen (1942). He used the entire information collected during past surveys to make current estimation more accurate. After Jessen successive sampling was supplemented by Yates (1949), Patterson (1950), Tikkiwal (1951), Rao and Graham (1964) and many others. In this paper, we denote the second (or current) and the first occasion study variables (i.e., total number of infected persons in the target area) by \( y \) and \( x \) respectively. \( z \) is the auxiliary variable denoting the number of deaths due to the disease recorded from that area. We want to estimate total number of infected persons i.e., \( \tau_y = \sum_{i=1}^{N} y_i \) at second (current) occasion. It is a common practice in successive sampling that information on previous occasion (i.e., the total number of infected persons \( \tau_x = \sum_{i=1}^{N} x_i \) at first occasion) is used as auxiliary information at second (current) occasion.

2.3. Combined approach

Finally, we unite the above two strategies to derive a single estimation process. We consider a specific locality for example, the wards under a metro corporation. Let there be a total of \( N \) households (treated as units in the paper) within the area. We randomly choose \( n \) households to start the first occasion ACS. In the first occasion, a sampling process through testing continues for 7 days say. Let \( x \) be the study variable (first occasion) denoting the number of infected people. Let \((x_1, x_2, x_3, \ldots, x_n)\) be the initial counts of \( x \) in randomly chosen \( n \) units. An adaptive sampling is done and let the \( i \)th unit is grown up to a cluster. A network, henceforth denoted by \( \psi_i \), \((i = 1, 2, \ldots, n)\) is defined as the cluster thus formed minus the edge units. \( |\psi_i| = m_i \) is the size of \( i \)th network. Therefore, in the first occasion the total number of initial units drawn is \( n \) and the sum total of units after ACS is

\[
m = \sum_{i=1}^{n} m_i.
\]

Also, the estimate of total number of infected people i.e., \( \tau_x = \sum_{i=1}^{N} x_i \) after first occasion successive sampling following the version Hansen and Hurwitz estimator (1943) used by Thompson (1990) in
where \( (w_x)_i \) is the average of the variable of interest in the network that includes unit \( i \) of the initial sample, i.e.; \( (w_x)_i = \frac{1}{m_i} \sum_{j \in \psi_i} x_j \); \( (w_x)_i \) = average size of ith network.

Also, \( f_i \) is the number of units in the initial sample which fall in the network \( \psi_i \). \( f_i = 0 \) if no unit in the initial sample falls in the network \( \psi_i \). Thus, \( f_i \) (selected from the \( m_i \) units in the network) follows a hyper geometric distribution \( (N, m_i, n) \) with \( E(f_i) = \frac{nm_i}{N} \).

A second drive of ACS is taken after an incubation period (say 14 days) is over and is treated as the second (current) occasion sampling in successive model. At this time we retain \( n'(< n) \) number of units from the first occasion randomly and let the initial values of \( y \) are \( (y_1, y_2, y_3, \ldots, y_{n'}) \) this time. Thus in the second occasion the total number of initial units chosen is \( n' \) and the sum total of that after ACS is

\[
m' = \sum_{i=1}^{n'} m'_i \text{ similarly.}
\]

The corresponding estimate of total of \( y \) value i.e., \( \tau_y = \sum_{i=1}^{n'} y_i \) is obtained as

\[
\hat{\tau}_{yHH} = \frac{N}{n'} \sum_{i=1}^{n'} y_i \frac{s_i}{m'_i} = \frac{N}{n'} \sum_{i=1}^{n'} \frac{1}{m'_i} \sum_{j \in \psi_i} y_j = \frac{N}{n'} \sum_{i=1}^{n'} (w'_y)_i
\]

Moreover, \( u \) initial units are chosen randomly from \( N \) total units of original population afresh at the same time for ACS. Let the initial values of \( y \) be \( (y_1, y_2, y_3, \ldots, y_u') \) for fresh sample. Let ith unit is grown to the size \( m''_i \), \( (i = 1, 2, 3, \ldots, u) \)

Thus in the second occasion fresh sampling, the total number of initial units chosen is \( u \) and the sum total of that after ACS is

\[
u' = \sum_{i=1}^{u} m''_i
\]

The corresponding estimate of total count of \( y \) value \( \tau_y \) is

\[
\hat{\tau}_{yHH} = \frac{N}{u} \sum_{i=1}^{u} y_i \frac{s_i}{m''_i} = \frac{N}{u} \sum_{i=1}^{u} \frac{1}{m''_i} \sum_{j \in \psi_i} y_j = \frac{N}{u} \sum_{i=1}^{u} (w''_y)_i
\]

Remembering the condition to be generally accomplished in successive sampling, we have: \( n = n' + u \) i.e., the sum of retained units (\( n' \)) and fresh units (\( u \)) must be equal to initial \( n \) units.

To elucidate the entire sampling strategy we provide below a labeled diagram of the procedure described in 2.3.1.
2.3.1. Diagrammatic presentation of sampling method

3. Formulation of proposed estimator

3.1. Notations used

The expression of the unbiased Hansen Hurwitz estimators \( \hat{\tau}_{xHH} \) and \( \hat{\tau}_{yHH} \) of total count \( \tau_x \) and \( \tau_y \) in the adaptive sense are already defined in Sampling Methodology Section 2.3.

In addition to that we shall use the following notations and expressions throughout the text:

The corresponding variances of \( \hat{\tau}_x \) and \( \hat{\tau}_y \) are:

\[
V(\hat{\tau}_{xHH}) = \frac{N(N - n)}{(N - 1)} \sum_{i=1}^{N} \left( \frac{(w_{x(i)})_i - \hat{\tau}_x}{N} \right)^2, \quad V(\hat{\tau}_{yHH}) = \frac{N(N - n')}{(N - 1)} \sum_{i=1}^{N} \left( \frac{(w_{y(i)})_i - \hat{\tau}_y}{N} \right)^2
\]

the population and sample correlations are:

\[
\rho_{xy} = \frac{S_{yHH}S_{xHH}}{S_{yHH}^2} = \sqrt{\sum_{i=1}^{N} \left( \frac{(w_{y(i)})_i - \hat{\tau}_y}{N} \right)^2 \sum_{i=1}^{N} \left( \frac{(w_{x(i)})_i - \hat{\tau}_x}{N} \right)^2}
\]

and

\[
\gamma_{xy} = \frac{S_{yHH}S_{xHH}}{S_{yHH}^2} = \frac{\sum_{i=1}^{N} \left( \frac{(w_{y(i)})_i - \hat{\tau}_{yHH}}{N} \right) \left( \frac{(w_{x(i)})_i - \hat{\tau}_{xHH}}{N} \right)}{\sqrt{\sum_{i=1}^{N} \left( \frac{(w_{y(i)})_i - \hat{\tau}_{yHH}}{N} \right)^2 \sum_{i=1}^{N} \left( \frac{(w_{x(i)})_i - \hat{\tau}_{xHH}}{N} \right)^2}}
\]

respectively.
3.2. Structure of proposed estimator

In the literature of statistical estimation, regression estimator provides lowest MSE and hence, is considered to be the most reliable one. To get a maximum gain of successive model utilizing auxiliary variable information, we wish to deploy a chain ratio to regression estimator. Kiregyera (1980, 1984) pioneered in using such structure in estimating mean in two phase simple random sampling. We derive the structure of proposed estimator adhering to successive scheme at outer layer and using ACS strategy at inner layer as follows:

Let $T_{n'}$ be the current occasion estimator of total count ($y$). We consider a structure as:

$$T_{n'} = y_{n'}^* + \alpha(x_{nm}^* - x_{n'm}^*)$$

value of $\alpha$ is determined by minimizing $T_{n'}$

(7)

where $x_{nm}^*$ and $x_{n'm}^*$ are ratio estimators based on auxiliary variable $z$ defined as:

$$x_{nm}^* = \frac{x_{nm}}{z_{nm}} \tau_z$$

and $x_{n'm}^* = \hat{\tau}_{yHH}$

Further, we can apply ACS model to estimate $x_{nm}^*$ and $y_{n'm}^*$ using the technique adopted earlier by Thompson (1990). Hence $x_{nm}^* = \hat{\tau}_{yHH}$ and $y_{n'm}^* = \hat{\tau}_{yHH}$. We also derive the structure of $y_{n'm}^*$, $x_{n'm}^*$.

Let $T_{uu'}$ be the regression estimator for fresh sampling in second (current) occasion. Then

$$T_{uu'} = y_{uu'}^* + \beta(\hat{\tau}_z - z_{uu'})$$

(8)

where $\hat{\tau}_{yHH} = \frac{N}{u} \sum_{i=1}^u (w_i^*)$ used to find $y_{uu'}^*$ for fresh occasion and $\beta$ is determined from the condition of minimum value of $T_{uu'}$.

Thus, we propose the final estimator $T_{PM-SA}$ as a linear convex combination of $T_{n'm'}$ and $T_{uu'}$ as

$$T_{PM-SA} = \varphi T_{n'm'} + (1 - \varphi) T_{uu'}$$

(9)

$0 < \varphi < 1$ where $\varphi$ is a real constant.

4. Forming the MSE expression of $T_{n'm'}$ and $T_{uu'}$

The efficacy of an estimator is established by ensuring its variance (or MSE in case of biased estimator) is smaller than earlier ones under similar conditions. The MSE of our proposed estimator $T_{PM-SA}$ up to the second order of approximations are derived under large sample approximations using the following transformations:

$$y_{n'm'} = (1 + e_0)\tau_y, \quad x_{n'm'}^* = (1 + e_1)\tau_x, \quad z_{uu'} = (1 + e_2)\tau_x, \quad x_{nm} = (1 + e_3)\tau_x, \quad y_{uu'}^* = (1 + e_4)\tau_y,$$

$$z_{uu'} = (1 + e_5)\tau_z$$

(10)

such that $|e_i| < 1$ and $E(e_i) = 0$ for $i = 1, 2, ..., 5$.

Further, we have the expected values of the above parameters as

$$E(e_0^2) = \lambda_1 c_1^2, \quad E(e_1^2) = \lambda_1 c_1^2, \quad E(e_2^2) = \lambda_2 c_2^2, \quad E(e_3^2) = \lambda_2 c_2^2,$$

$$E(e_0 e_1) = \lambda_1 c_1 c_2, \quad E(e_0 e_2) = \lambda_2 c_1 c_3, \quad E(e_3 e_2) = \lambda_2 c_3 c_3,$$

$$E(e_1 e_2) = \lambda_2 c_1 c_3, \quad E(e_1 e_3) = \lambda_2 c_1 c_3, \quad E(e_2 e_3) = \lambda_2 c_1 c_3,$$

where $\lambda_1 = \frac{N(N-n)}{n}, \lambda_2 = \frac{N(N-n)}{n}; n' < n < N$, $\lambda_3 = \frac{N(N-u)}{u}; u < n' < n$

Also, $x_{nm}^* = \frac{\tau_z}{x_{4nm}} $$

We have expanded the term binomially and ignored higher orders of $e$'s.

Under above transformations (10) the estimators in (7) and (8) take the forms:

$$T_{n'm'} = y_{n'm'}^* + \alpha(x_{nm}^* - x_{n'm}^*) = (1 + e_0)\tau_y + \alpha((1 + e_2 - e_3)\tau_x - (1 + e_1)\tau_x)$$

and

$$T_{uu'} = y_{uu'}^* + \beta(\tau_z - z_{uu'}) = (1 + e_4)\tau_y + \beta(\tau_z - (1 + e_5)\tau_z)$$
\(T_{n'm'}\) and \(T_{uw'}\) are both regression to ratio type estimators. Therefore, both the estimators are biased. So we take MSE of each of them as follows.

\[
M(T_{n'm'}) = E(T_{n'm'} - \tau y)^2 = E[y_{n'm'}^* + \alpha(x_{nm}^* - x_{n'm'}^*)]^2 \\
= E[(1 + e_0)\tau y - \tau y + \alpha((1 + e_2 - e_3)\tau x - (1 + e_1)\tau x)]^2 \\
= \tau_y^2 E(e_0^2) + \alpha^2 E(\tau x (e_2 - e_3 - e_1))^2 + 2\tau_y \tau x \alpha(\lambda_2 c_y c_x \rho_{yx} - \lambda_2 c_y c_z \rho_{yz} - \lambda_1 c_y c_x \rho_{yx}) \\
\]

on simplification and substitution of the values of \(E(e_0^2), E(e_2^2), E(e_3^2)\) and \(E(e_0 e_1), E(e_0 e_2), E(e_2 e_3), E(e_2 e_1)\) and \(E(e_2 e_1)\) we obtain:

\[
M(T_{n'm'}) = \tau_y^2 \lambda_1 c_y^2 + \alpha^2 \tau_y^2 [\lambda_1 c_x^2 - \lambda_2 c_y^2 + \lambda_2 c_z^2] + 2\tau_y \tau x \alpha[\lambda_2 c_y c_x \rho_{yx} - \lambda_2 c_y c_z \rho_{yz} - \lambda_1 c_y c_x \rho_{yx}] \\
\tag{11}
\]

Similarly for fresh sample we have,

\[
M(T_{uw'}) = E[(1 + e_0)\tau y + \beta \tau z - (1 + e_5)\tau z - \tau y]^2 \\
= E[\tau y e_4 - \beta \tau z e_3 \tau = \tau_y^2 E(e_4^2) + \beta^2 \tau_z^2 E(e_3^2) - 2\beta \tau y \tau z E(e_4 e_3) \\
= \tau_y^2 \lambda_3 c_y^2 + \beta^2 \tau_z^2 c_z^2 - 2\beta \tau y \tau z \lambda_3 c_y \rho_{yz} \\
\tag{12}
\]

It may be noted from the constructions of the estimators \(T_{n'm'}\) and \(T_{uw'}\) from Eqs. (7) and (8) that, they have real constants \(\alpha\) and \(\beta\) whose values are to be found by minimizing their respective MSE. This is done by the theory of maxima–minima in calculus.

Thus, we proceed to minimize first the MSEs of \(T_{uw'}\) and \(T_{n'm'}\) given in (11) and (12) as under:

\[
\frac{\partial M(T_{n'm'})}{\partial \alpha} = 0 \Rightarrow 2\alpha \tau_y^2 [\lambda_1 c_y^2 - \lambda_2 c_y^2 + \lambda_2 c_z^2] + 2\tau_y \tau x [\lambda_2 c_y c_x \rho_{yx} - \lambda_2 c_y c_z \rho_{yz} - \lambda_1 c_y c_x \rho_{yx}] = 0 \\
\]

Hence, \(\alpha = - \frac{\tau_y [\lambda_2 c_y c_x \rho_{yx} - \lambda_2 c_y c_z \rho_{yz} - \lambda_1 c_y c_x \rho_{yx}]}{\tau x [\lambda_1 c_y^2 - \lambda_2 c_y^2 + \lambda_2 c_z^2]} \)

then

\[
M(T_{n'm'}) = \tau_y^2 \lambda_1 c_y^2 + \tau_y^2 \frac{[\lambda_2 c_y c_x \rho_{yx} - \lambda_2 c_y c_z \rho_{yz}]^2}{\lambda_1 c_y^2 - \lambda_2 c_y^2 + \lambda_2 c_z^2} - \frac{2\tau_y^2 [\lambda_1 c_y^2 - \lambda_2 c_y^2 + \lambda_2 c_z^2] - \lambda_2 c_y c_x \rho_{yx} - \lambda_2 c_y c_z \rho_{yz}]^2}{\lambda_1 c_y^2 - \lambda_2 c_y^2 + \lambda_2 c_z^2} \\
= \tau_y^2 c_y \left\{ \lambda_1 - \frac{[\lambda_2 c_y c_x \rho_{yx} - \lambda_2 c_y c_z \rho_{yz}]^2}{\lambda_1 c_y^2 - \lambda_2 c_y^2 + \lambda_2 c_z^2} \right\} \\
\tag{13}
\]

Similarly, \(\frac{\partial M(T_{uw'})}{\partial \beta} = 0 \Rightarrow \beta = \frac{\tau_y c_y \rho_{yz}}{\tau_2 c_z} \)

Hence, \(M(T_{uw'}) = \tau_y^2 \lambda_3 c_y^2 - \tau_y^2 \lambda_3 c_y^2 \rho_{yz} = \tau_y^2 \lambda_3 c_y^2 (1 - \rho_{yz}^2) \)

**Remark 4.1.** Since the samples of current and fresh occasion are independent, therefore the estimators \(T_{n'm'}\) and \(T_{uw'}\) are uncorrelated and their covariance term is zero. We also note that fresh sampling is totally independent of first occasion sampling. Hence, their correlation is also zero.

**Remark 4.2.** It is to be mentioned that the values of \(\alpha\) and \(\beta\) depend on unknown population parameters as \(\lambda_1, \lambda_2, \lambda_3, c_y, c_y^2, c_z^2, \tau_y, \tau_x, \rho_{yx}\) and \(\rho_{yz}\). These unknown population parameters may be estimated from their respective sample estimates or from past data or may be guessed from experience gathered over time. Such problems are also considered by Reddy (1978), Tracy et al. (1996).
5. Obtaining the expression of final estimator $T_{PM\_SA}$

Therefore, we obtain the MSE of the final estimator $T_{PM\_SA}$ as follows:

$$M(T_{PM\_SA}) = E[T_{PM\_SA} - \tau_y]^2 = E[\varphi T_{n'm'} + (1 - \varphi)T_{uu'} - \tau_y]^2, \quad 0 < \varphi < 1$$

$$= E[\varphi(T_{n'm'} - \tau_y) + (1 - \varphi)(T_{uu'} - \tau_y)]^2 = \varphi^2 M(T_{n'm'}) + (1 - \varphi)^2 M(T_{uu'})$$

The MSE of combined estimator $T_{PM\_SA}$ is minimized with respect to the real parameter $\varphi$ following optimum replacement strategy.

The minimum expression of $M(T_{PM\_SA})$ with respect to $\varphi$ is:

$$M(T_{PM\_SA}) = M(T_{n'm'}) M(T_{uu'})$$

$$= \tau_y^2 c_y^2 \left\{ \lambda_1 [\lambda_1 c_y^2 - \lambda_2 c_x^2 + \lambda_2 c_z^2] - [(\lambda_2 - \lambda_1) c_x \rho_{yx} - \lambda_2 c_z \rho_{yz}]^2 \right\} \left\{ \lambda_3 (1 - \rho_{yz}^2) \right\}$$

$$+ \lambda_y^2 c_x^2 \left\{ \lambda_1 [\lambda_1 c_y^2 - \lambda_2 c_x^2 + \lambda_2 c_z^2] - [(\lambda_2 - \lambda_1) c_x \rho_{yx} - \lambda_2 c_z \rho_{yz}]^2 \right\} + \lambda_y^2 c_z^2 \left\{ \lambda_3 (1 - \rho_{yz}^2) \right\}$$

(15)

6. Efficiency comparison between proposed estimator and the conventional ones

We have considered the version of Hansen–Hurwitz estimator used by Thompson (1990) i.e., $\hat{\tau}_y^{HH}$ for comparing with the proposed one. The Hansen–Hurwitz estimator for total count in adaptive sense defined by:

$$\hat{\tau}_y^{HH} = \frac{N}{n} \sum_{i=1}^{n} \left[ \frac{1}{m_i} \sum_{j \in \psi_i} y_j \right] = \frac{N}{n} \sum_{i=1}^{n} w(y)_i$$

where $w(y)_i = \frac{1}{m_i} \sum_{j \in \psi_i} y_j$ = average size of ith network.

The variance of the estimator $\hat{\tau}_y^{HH}$ is obtained as:

$$V(\hat{\tau}_y^{HH}) = \frac{N(N - n)}{(N - 1)} \sum_{i=1}^{N} \left[ (w(y)_i) - \frac{\tau_y}{N} \right]^2, \quad \tau_y = \sum_{i=1}^{N} y_i$$

(16)

We have calculated the percentage relative efficiency (PRE) factor of our estimator against the one described in (16). The expression of PRE of $T_{PM\_SA}$ with respect to $\hat{\tau}_y^{HH}$ is given by:

Therefore, we have

$$\text{PRE} = \frac{V(\hat{\tau}_y^{HH})}{M(T_{PM\_SA})} \times 100$$

(17)

7. Numerical illustrations

The proposed sampling methodology is, on the first hand tested through simulation study and then we applied it on some real life situation. Details of the outcome is demonstrated below.

7.1. Numerical illustrations through simulation study

Populations at different phases of spread have been prepared with the help of the statistical software R. The following algorithm (similar to the work of Thompson, 1990) is used here and is implemented in R software.
Table 1
Containing first occasion data (400 units considered in a gridted area).

|   | 0 | 1 | 2 | 3 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 2 | 2 | 2 | 1 | 2 | 1 | 0 | 1 | 2 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 2 |
| 1 | 1 | 1 | 0 | 1 | 2 | 1 | 1 | 1 | 2 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 2 | 2 | 0 | 0 |
| 1 | 2 | 3 | 0 | 1 | 0 | 0 | 2 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 2 | 2 | 0 | 0 |
| 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 2 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| 1 | 2 | 1 | 0 | 0 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 |
| 1 | 0 | 0 | 0 | 2 | 1 | 2 | 1 | 3 | 2 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 1 | 2 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 1 | 1 | 0 | 3 | 0 | 1 | 1 | 1 | 2 | 1 | 1 |
| 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 3 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 2 | 0 | 2 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 2 | 0 | 2 | 0 | 0 | 0 | 1 | 1 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 1 | 0 | 1 | 0 | 2 | 1 | 2 | 0 | 0 |
| 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 0 | 0 | 1 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| 0 | 2 | 1 | 0 | 2 | 1 | 2 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 3 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 |
| 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 3 | 2 |

7.1.1. Algorithm for generating population and other parameters

Population of size 400 is generated keeping in mind that number of zero infected units will be more initially, which will gradually decrease as the epidemic falls out. We have generated six population sets (having six different pair of $\rho_{YX}$ and $\rho_{YZ}$ values) as shown in Table 7 and each population set consists of four datasets for first, current and fresh occasion study variable (with different values of mean and SD) and one of aux. var. (z). For a chosen pair of $\rho_{YX}$ and $\rho_{YZ}$ we execute the following steps.

**Step 1.** We generate a mixture of two random normal populations of $\mu = 1$, $\sigma = 1$ and $\mu = 0$, $\sigma = 1$ with a proportion 1:3 (to get greater number of zeros initially) subject to a total size of 400. The negatives are replaced by zero. This creates the initial population.

**Step 2.** Randomly choose 8 ($= n$) grid positions (i.e., units) and then choose 8 clusters of different sizes to create 1st occasion sample. Calculate the mean and SD of each.

**Step 3.** Retain randomly 5 grid positions from 1st occasion (step 2).

**Step 4.** Again generate a mixture of two random normal populations of $\mu = 2$, $\sigma = 1$ and $\mu = 0$, $\sigma = 2$ (to reduce the number of zeros) with a proportion 1:3 subject to total size 400. The negatives are replaced by zero. This generates the population before second occasion. Now, create 5 new clusters from the retained 5 grid positions from 1st occasion.

**Step 5.** Finally, generate a mixture of random normal populations of $\mu = 2.5$, $\sigma = 1$ and $\mu = 0$, $\sigma = 1$ with a proportion 1:2 (to further reduce the zeros in population) subject to total size 400. The negatives are replaced by zero. Choose 3 ($= u$) grid positions and generate new three clusters from these grid positions. Compute the mean and SD. This generates the population before fresh sample.

**Step 6.** After generating x and y data, we now similarly generate population for z (death records) for the chosen value of $\rho_{YZ}$ using a standard normal distribution with $\mu = 0$ and $\sigma = 0.4$. Computed the mean and SD.

**Step 7.** The entire process from step 1 through step 6 is repeated six times for six different pair values of $\rho_{YX}$ and $\rho_{YZ}$ and outcomes are recorded.

For the three population tables of different occasions are furnished, in Table 4 we show one example data on number of death records (z).

We calculate the fixed and variable parameters and are shown in Table 5.

**Note 7:** Only one set of population data (for $\rho_{YX} = 0.085$ and $\rho_{YZ} = 0.8$) and parameter details (Tables 1–5) with network size in different occasions are shown for $n = 8$ and $n' = 5$ for the convenience of reader. According to the step 7 of above algorithm, we have generated 6 different
Table 2
Second occasion data (400 units considered in a gridded area).

| 1 | 4 | 3 | 2 | 0 | 1 | 2 | 0 | 0 | 0 | 2 | 0 | 2 | 0 | 0 | 0 | 0 | 3 | 2 | 2 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 1 | 0 | 0 | 2 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 5 | 1 | 1 | 3 |
| 2 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 2 | 1 | 3 | 2 | 0 | 0 |
| 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 2 | 0 | 1 | 0 | 1 | 2 | 2 | 1 | 4 | 2 | 1 | 0 | 0 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 1 | 2 | 2 | 2 | 1 | 3 | 1 | 2 | 1 | 3 | 1 | 0 | 2 | 1 | 0 | 0 | 0 | 0 |
| 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 1 | 1 | 3 | 3 | 2 | 3 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 2 | 0 | 1 | 2 | 1 | 3 | 1 | 2 | 2 |
| 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 2 | 3 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 |
| 1 | 0 | 1 | 1 | 0 | 3 | 3 | 1 | 0 | 2 | 2 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 1 | 2 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 2 | 3 | 2 | 4 |
| 2 | 3 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 2 | 0 | 0 | 1 | 0 | 1 | 0 | 1 |
| 2 | 0 | 2 | 4 | 3 | 1 | 3 | 0 | 1 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 3 | 4 | 2 | 1 | 3 | 1 | 1 | 1 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 4 | 2 | 2 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 0 | 0 | 0 | 3 | 0 | 3 | 3 | 1 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 3 | 1 | 1 | 1 | 2 | 2 | 0 | 1 | 0 | 0 | 0 |
| 0 | 1 | 2 | 0 | 1 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 3 | 0 | 3 | 2 | 1 | 1 | 3 | 1 |

Table 3
Fresh occasion data (400 units considered in a gridded area of interest).

| 2 | 3 | 2 | 3 | 3 | 3 | 3 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 3 | 2 | 3 | 3 | 3 | 3 | 2 | 4 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 1 | 2 | 3 | 4 | 1 | 2 | 2 | 1 | 1 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 2 |
| 1 | 2 | 2 | 4 | 3 | 3 | 3 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 2 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 | 2 | 4 | 1 | 3 | 2 | 4 | 3 | 3 | 3 | 0 | 0 | 1 | 1 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 4 | 1 | 2 | 4 | 2 | 2 | 1 | 2 | 2 | 3 |
| 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 1 | 3 | 1 | 2 | 3 | 4 | 3 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 2 | 3 | 4 | 3 | 0 | 2 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 2 | 0 | 0 | 2 |
| 2 | 3 | 3 | 3 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| 0 | 0 | 2 | 1 | 2 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 2 | 1 | 2 | 1 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 4 | 3 | 3 | 2 | 2 | 2 | 3 | 0 | 0 | 0 | 2 |
| 0 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 2 | 4 | 1 | 3 | 3 | 0 | 0 |
| 1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 3 | 2 | 3 | 4 | 2 | 0 | 0 |
| 0 | 1 | 2 | 1 | 0 | 0 | 0 | 1 | 2 | 3 | 4 | 0 | 2 | 1 | 1 | 0 | 0 | 3 | 3 |
| 0 | 2 | 2 | 3 | 1 | 3 | 3 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 1 | 1 | 0 | 2 | 3 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 3 | 2 |
| 3 | 3 | 3 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 3 | 2 | 3 | 2 |

artificial populations for different values of $\rho_{XY}$ and $\rho_{YZ}$. Following the same technique as discussed above, we have calculated the PREs against all the 6 populations with different pair of sample sizes $n$ and $n'$. Findings of the same are demonstrated in Table 6.

7.2. Numerical illustration through real data analysis

The ‘Acid Test’ for the efficacy of any estimator is possible only when it is applied on a real life situation. Keeping this in mind, we have taken a number of real life situations and tested our estimator on the data. The data (kept as medical bulletins) are taken from different websites maintained by different organizations under Govt. of India. We have considered three states in three geographical areas of India, which are far apart from one another, namely:

West Bengal (north-east region), Tamilnadu (extreme southern region) and Punjab (extreme north-west region). The details of the sources are furnished below.
Table 4
Data simulated on number of death records (z).

| 1  | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 0  | 0  | 1  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| 1  | 0  | 0  | 0  | 0  | 0  | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | 0  | 0  | 0  |
| 0  | 0  | 0  | 1  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  |
| 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 1  |
| 0  | 1  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 1  | 2  | 0  | 0  | 0  | 0  |
| 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | 0  | 0  | 0  | 0  |
| 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 1  |
| 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| 0  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |
| 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 1  |
| 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |

Table 5
Showing parameters considered and Network size and average Y generated in first, current occasion and fresh sampling (Artificial population 1).

| Variable parameters computed from population 1 N = 400 | Initial units (first occasion = 8) | Initial units (2nd occasion = 5) | Initial units (fresh occasion = 3) |
|------------------------------------------------------|------------------------------------|----------------------------------|-----------------------------------|
| \( \tau_X, \tau_X \) per unit (variable)             | \( \psi_i, m_i \) Mean SD          | \( \psi_i, m_i' \) Mean SD        | \( \psi_i, m_i'' \) Mean SD       |
| \( \tau_Y, \tau_Y \) per unit (variable)             | \( \psi_i, m_i \) Mean SD          | \( \psi_i, m_i' \) Mean SD        | \( \psi_i, m_i'' \) Mean SD       |
| \( \tau_Y \) (fresh), \( \tau_Y \) per unit fresh (variable) | \( \psi_i, m_i \) Mean SD          | \( \psi_i, m_i' \) Mean SD        | \( \psi_i, m_i'' \) Mean SD       |
| \( \tau_Z, \tau_Z \) per unit (variable)             | \( \psi_i, m_i \) Mean SD          | \( \psi_i, m_i' \) Mean SD        | \( \psi_i, m_i'' \) Mean SD       |
| \( \rho_{XY} \) (variable)                          | \( \psi_i, m_i \) Mean SD          | \( \psi_i, m_i' \) Mean SD        | \( \psi_i, m_i'' \) Mean SD       |
| \( \rho_{XZ} \) (variable)                          | \( \psi_i, m_i \) Mean SD          | \( \psi_i, m_i' \) Mean SD        | \( \psi_i, m_i'' \) Mean SD       |
| 1st occasion zero Count (variable)                   | \( \psi_i, m_i \) Mean SD          | \( \psi_i, m_i' \) Mean SD        | \( \psi_i, m_i'' \) Mean SD       |
| 2nd occasion zero Count (variable)                   | \( \psi_i, m_i \) Mean SD          | \( \psi_i, m_i' \) Mean SD        | \( \psi_i, m_i'' \) Mean SD       |
| \( \lambda_1 \)                                     | 19600                              | 5.4                               | 1.2                               |
| \( \lambda_2 \)                                     | 31600                              | 5.4                               | 1.2                               |
| \( \lambda_3 \)                                     | 52933.33                           | 5.4                               | 1.2                               |

- Dept. of Health & Family Welfare, Govt. of West Bengal, India: West Bengal Covid-19 Health Bulletin, May’2020 (url: https://www.wbhealth.gov.in/pages/corona/bulletin)
- Directorate of Public Health and Preventive Medicine Health and Family Welfare Department, Government of Tamil Nadu, April’2020 (url: https://stopcorona.tn.gov.in/daily-bulletin).
- Department of Health & Family Welfare, Punjab, May’2021 (url: https://nhm.punjab.gov.in/media-bulletin_Sept.htm).
Table 6
PRES of the proposed estimator against $\tau_{YX}$ for six sets of values of $\rho_{YX}$, $\rho_{YZ}$.

| Natural Population 1: $\rho_{YX} = 0.085$ and $\rho_{YZ} = 0.8$ | Natural Population 2: $\rho_{YX} = 0.085$ and $\rho_{YZ} = 0.7$ | Natural Population 3: $\rho_{YX} = 0.075$ and $\rho_{YZ} = 0.8$ |
|---|---|---|
| $n$ | $n'$ | PRE | $n$ | $n'$ | PRE | $n$ | $n'$ | PRE |
| 8 | 5 | 622.83 | 8 | 5 | 846.31 | 8 | 5 | 632.82 |
| 8 | 4 | 642.12 | 8 | 4 | 885.32 | 8 | 4 | 649.48 |
| 10 | 7 | 754.24 | 10 | 7 | 935.47 | 10 | 7 | 832.11 |
| 10 | 6 | 801.77 | 10 | 6 | 988.69 | 10 | 6 | 897.34 |
| 12 | 9 | 824.14 | 12 | 9 | 1058.29 | 12 | 9 | 954.38 |
| 12 | 7 | 874.91 | 12 | 7 | 1289.76 | 12 | 7 | 159.67 |

| Artificial Population 4: $\rho_{YX} = 0.075$ and $\rho_{YZ} = 0.7$ | Artificial Population 5: $\rho_{YX} = 0.065$ and $\rho_{YZ} = 0.8$ | Artificial Population 6: $\rho_{YX} = 0.065$ and $\rho_{YZ} = 0.7$ |
|---|---|---|
| $n$ | $n'$ | PRE | $n$ | $n'$ | PRE | $n$ | $n'$ | PRE |
| 8 | 5 | 846.441 | 8 | 5 | 637.64 | 8 | 5 | 846.44 |
| 8 | 4 | 865.47 | 8 | 4 | 655.98 | 8 | 4 | 897.65 |
| 10 | 7 | 875.23 | 10 | 7 | 777.68 | 10 | 7 | 1025.74 |
| 10 | 6 | 899.54 | 10 | 6 | 829.88 | 10 | 6 | 1257.25 |
| 12 | 9 | 905.47 | 12 | 9 | 900.38 | 12 | 9 | 1358.24 |
| 12 | 7 | 945.69 | 12 | 7 | 1100.59 | 12 | 7 | 1587.27 |

7.2.1. Adaptation of available data

In the hypothetical population a small area within a municipal corporation was considered and the area was conceived as a grid where each square is a sampling unit (taken as household) in ACS. To apply the same in case of a total state, the basic parameters are to be properly scaled up. We would like to refer the Indian Council of Medical Research, New Delhi (ICMR) recommendation in this connection (vide ICMR containment plan, version 2 updated on 16.05.2020). In Article 5 (cluster containment and definition of cluster), we have considered a district as a gridded area. The sampling unit should be then a conglomerate of houses in that case to cover most of the households. One may start forming networks starting from a number of randomly chosen initial units in the district and then add them up to get the net network size for the district. This is justifiable because COVID tests are done at different medical centers by the medical staffs. These centers are randomly spread in the entire district. More tests are carried out in the pockets identified or suspected to be a hotspot by means of data available on contact tracing in different state govt. organizations of India. The total new positive cases detected on two dates (as explained in Article 2.3.1) against each of the districts of a state are furnished in Table 7. We assumed that at least 7 days are necessary to carryout the ACS process and consolidating the data and therefore we have recorded the deaths on an initial date (treated as an auxiliary variable stable over occasions) and recorded the no. of tested positives on a date 7 days after the initial date (1st occasion data X) and thereafter on a date 14 days after (current occasion Y). The PRES of the said populations are furnished in Table 8.

Parametric values computed from different populations

- Natural Population 1 (West Bengal): $\rho_{YX} = 0.994$, $\rho_{XZ} = 0.993$ and $\rho_{YZ} = 0.958$
- Natural Population 2 (Punjab): $\rho_{YX} = 0.695$, $\rho_{XZ} = 0.821$ and $\rho_{YZ} = 0.791$
- Natural Population 3 (Tamilnadu): $\rho_{YX} = 0.959$, $\rho_{XZ} = 0.913$ and $\rho_{YZ} = 0.922$

8. Conclusion

8.1. Findings from simulation studies

On careful inspection through Tables 1 to 6, the following observations are evident:
Table 7
District wise new positive cases and deaths confirmed in three states.

| State          | District          | New cases confirmed during May’2020 | Death cases in May’2020 | District          | New cases confirmed during April’2020 | Death cases in April’2020 | District          | New cases confirmed during May’2021 | Death cases in May’2021 |
|----------------|-------------------|-------------------------------------|-------------------------|-------------------|---------------------------------------|--------------------------|-------------------|-------------------------------------|--------------------------|
| West Bengal, India | Alipurduar       | 0                                   | 0                       | SBS Nagar         | 19                                   | 1                        | 1                 | Ariyalur                            | 207                      | 267                      |
|                | Bankura           | 17                                  | 0                       | SAS Nagar         | 56                                   | 1                        | 1                 | Chengalpattu                        | 2 225                    | 2 092                    | 11                      |
|                | Birkamoor         | 49                                  | 0                       | Hoshiarpur        | 7                                    | 1                        | 1                 | Chennai                            | 6 538                    | 6 073                    | 47                      |
|                | Coochbehar        | 0                                   | 0                       | Jalandhar         | 24                                   | 0                        | 0                 | Coimbatore                         | 3 197                    | 3 335                    | 1                      |
|                | Dakshin           | 1                                   | 0                       | Amritsar          | 11                                   | 1                        | 0                 | Cuddalore                          | 645                      | 854                      | 2                      |
|                | Dinajpur          | 1                                   | 0                       | Ludhiana          | 11                                   | 16                       | 1                 | Dharmapuri                         | 271                      | 362                      | 0                      |
|                | Darjeeling        | 1                                   | 8                       | Patiala           | 2                                    | 31                       | 0                 | Dindigul                           | 429                      | 389                      | 4                      |
|                | Hoogly            | 99                                  | 114                     | Pathankot         | 22                                   | 24                       | 0                 | Erode                              | 781                      | 1 505                    | 5                      |
|                | Jalpaiguri        | 0                                   | 4                       | Mansa             | 11                                   | 11                       | 0                 | Kallakurichi                       | 179                      | 398                      | 1                      |
|                | Jhargram          | 3                                   | 3                       | Moga              | 4                                    | 4                        | 0                 | Kancheepuram                       | 889                      | 858                      | 6                      |
|                | Kalimpong         | 0                                   | 0                       | Faridkot          | 3                                    | 3                        | 0                 | Kanyakumari                        | 1 025                    | 1 096                    | 4                      |
|                | Kolkata           | 614                                 | 919                     | Ropar             | 3                                    | 3                        | 0                 | Karur                              | 319                      | 297                      | 1                      |
|                | Malda             | 19                                  | 87                      | Barnala           | 2                                    | 2                        | 0                 | Krishanagiri                       | 712                      | 729                      | 1                      |
|                | Murshidabad       | 4                                   | 66                      | FG Sahib          | 2                                    | 2                        | 0                 | Madurai                            | 1 250                    | 1 269                    | 0                      |
|                | Nadia             | 1                                   | 42                      | Kapurthala        | 2                                    | 3                        | 0                 | Nagapattinam                       | 396                      | 711                      | 0                      |
|                | North 24 Parganas | 181                                 | 345                     | Sangrur           | 2                                    | 3                        | 0                 | Namakkal                           | 560                      | 381                      | 0                      |
|                | Other State       | 13                                  | 36                      | Muktsar           | 1                                    | 1                        | 0                 | Nilgiris                           | 306                      | 325                      | 0                      |
|                | Paschim Parganas  | 15                                  | 10                      | Gurdaspur         | 1                                    | 1                        | 0                 | Perambalur                         | 140                      | 196                      | 0                      |
|                | Paschim Bardhaman | 5                                   | 30                      | Ferozepur         | 0                                    | 1                        | 0                 | Pudukottai                         | 342                      | 442                      | 0                      |

(continued on next page)
| State | West Bengal, India | Punjab, India | Tamilnadu, India |
|-------|-------------------|---------------|-----------------|
| District | New cases confirmed during May'2020 | Death cases in May'2020 | New cases confirmed during April'2020 | Death cases in April'2020 | Districts confirmed during May'2021 | Death cases in May'2021 |
| Date | 12th May (X) | 28th May (Y) | 5th May (Z) | Date | 14th Apr (X) | 21st Apr (Y) | 1st Apr (Z) | Date | 14th May (X) | 20th May (Y) | 1st May (Z) |
| Purba Bardhaman | 7 | 49 | 0 | Total | 183 | 257 | 4 | Ramanathapuram | 310 | 350 | 0 |
| Purba Medinipur | 20 | 39 | 0 | Ranipet | 596 | 719 | 2 |
| Purulia | 0 | 1 | 0 | Salem | 798 | 667 | 14 |
| South 24 Parganas | 51 | 104 | 1 | Sivagangai | 242 | 237 | 1 |
| Uttar Dinajpur | 3 | 83 | 0 | Tenkasi | 443 | 452 | 3 |
| Total | 1637 | 2571 | 61 | Thanjavur | 618 | 679 | 1 |
| | | | | Theni | 447 | 689 | 2 |
| | | | | Thirupathur | 328 | 565 | 7 |
| | | | | Thiruvallur | 1410 | 1791 | 14 |
| | | | | Thiruvannamalai | 397 | 674 | 2 |
| | | | | Thiruvanur | 324 | 734 | 1 |
| | | | | Thoothukudi | 885 | 1004 | 2 |
| | | | | Tirunelveli | 831 | 589 | 6 |
| | | | | Tiruppur | 771 | 1581 | 0 |
| | | | | Trichy | 1224 | 1375 | 5 |
| | | | | Vellore | 555 | 567 | 4 |
| | | | | Villupuram | 679 | 532 | 0 |
| | | | | Virudhunagar | 623 | 795 | 0 |
| Total | 31892 | 35579 | 147 |
Table 8
PREs of the proposed estimator against τ_{YHH} for the real data.

| Natural population 1: West Bengal | Natural population 2: Punjab | Natural population 3: Tamil Nadu |
|-----------------------------------|-------------------------------|----------------------------------|
| n  | n' | PRE | n  | n' | PRE | n  | n' | PRE |
| 15  | 10 | 1904.85 | 13  | 7  | 1741.24 | 18  | 10 | 4671.36 |
| 10  | 7  | 818.65  | 10  | 7  | 990.94  | 15  | 10 | 2720.18 |
| 10  | 5  | 1371.30 | 10  | 5  | 1490.15 | 15  | 8  | 4109.06 |
| 8   | 6  | 547.18  | 8   | 6  | 641.45  | 12  | 9  | 1756.85 |
| 8   | 4  | 1099.88 | 8   | 4  | 1199.00 | 12  | 6  | 3540.49 |

(i) PRE of proposed estimator against τ_{YHH} is high compared to conventional Hansen Hurwitz (ACS version) estimator. The result clearly justifies the combined approach of ACS and successive sampling strategy to estimate total count of infection in a pandemic situation.

(ii) It is also observed that ρ_{YX} and ρ_{YZ} have similar effect on PRE. When ρ_{YX} is fixed, PRE rapidly increases with decrease of ρ_{YZ}. When ρ_{YZ} is fixed, PRE very slowly increases with decrease of ρ_{YX}. The phenomena suggest that growth of infection is not strongly correlated with death rate which has been observed in many parts of world in recent situation of COVID-19.

(iii) Also with the increasing values of n, PRE is increased and when n is kept fixed, PRE is increased if more units is chosen from fresh occasion.

8.2. Findings from real data studies

After studying Tables 7 and 8, the following information is conspicuous:

(i) In real data, the computed correlation coefficients are high. Therefore, the PREs obtained is in general high.

(ii) Since in real life the only thing varies is the sample size i.e. number of initial starting units for ACS for first and current occasion (n and n'). Table 8 shows the change PRE for various pairs of values of n and n'. We observe that as the initial sample size increases, the PRE is increased steadily. Moreover, for a fixed n, if n' is decreased the PRE is further increased. This is in tune with the common experience that larger sample size yields better estimation. Also taking fresher sample in Y makes the estimation better.

It may be observed that, findings obtained in simulation studies are reflected in real life population studies. Therefore, we found the above estimation procedure is particularly effective for situations like COVID-19 where the infection does not spread uniformly and rapidly in a specific area. Hence, we recommend the procedure to survey statistician for practical use.

Acknowledgments

For the real life data, we are indebted to the following organizations for making their covid data freely available over the internet:

- Dept. of Health & Family Welfare, Govt. of West Bengal, India:
- Directorate of Public Health and Preventive Medicine Health and Family Welfare Department, Government of Tamil Nadu
- Department of Health & Family Welfare, Punjab)

We are also obliged to the reviewer and chief editor for their patient reading and offering valuable suggestions to ameliorate the quality of paper.

References

Dryver, A.L., Chao, C.T., 2007. Ratio estimators in adaptive cluster sampling. Environ. Off. J. Int. Environ. Soc. 18 (6), 607–620.

Hansen, M.M., Hurwitz, W.N., 1943. On the theory of sampling from finite populations. Ann. Math. Stat. 14, 333–362.
Jessen, R.J., 1942. Statistical investigation of a survey for obtaining farm facts. Iowa Agricul. Expt. Stat. Res. Bull. 304, 1–104.

Kiregyera, B., 1980. A chain ratio-type estimator in finite population double sampling using two auxiliary variables. Metrika 27, 217–223.

Kiregyera, B., 1984. Regression-type estimators using two auxiliary variables and the model of double sampling from finite populations. Metrika 31, 215–226.

Patterson, H.D., 1950. Sampling on successive occasions with partial replacement units. J Roy Stat. Soc. 12 (B), 241–255.

Rao, J.N.K., Graham, J.E., 1964. Rotation design for sampling on repeated occasions. J. Amer. Stat. Assoc. 59, 492–509.

Reddy, V.N., 1978. A study on the use of prior knowledge on certain population parameters in estimation. Sankhya, Series C 40, 29–37.

Salehi, M., Seber, G.A., 1997. Adaptive cluster sampling with networks selected without replacement. Biometrika 84 (1), 209–219.

Smith, D.R., Conroy, M.J., Brakhage, D.H., 1995. Efficiency of adaptive cluster sampling for estimating density of wintering waterfowl. Biometrics 77, 7–788.

Smith, D.R., Villella, R.F., Lemarie, D.P., 2003. Applications of adaptive cluster sampling to low-density populations of freshwater mussels. Environ. Ecol. Stat. 10 (715).

Thompson, S.K., 1990. Adaptive cluster sampling. J. Amer. Statist. Assoc. 85 (412), 1050–1059.

Tikkiwal, B.D., 1951. Theory of Successive Sampling (Thesis for Diploma). ICAR, New Delhi.

Tracy, D.S., Singh, H.P., Singh, R., 1996. An alternative to the ratio-cum-product estimator in sample surveys. J. Stat. Plan. Inference 53, 375–387.

Yates, F., 1949. Sampling Methods for Censuses and Surveys, 1 Edition. Charles Griffin and Company Limited, London.