Journal of Statistics Applications & Probability Letters
An International Journal

http://dx.doi.org/10.18576/jsapl/100101

Estimation of COVID-19 cases using robust ratio type estimators: An application of ACS design

Rajesh Singh[1], *Sheela Misra*[2] *and Rohan Mishra*[1,*]

[1]Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi, India
[2]Department of Mathematics and Statistics, University of North Carolina at Greensboro, USA

Received: 12 Jun. 2022, Revised: 12 Aug. 2022, Accepted: 24 Sep. 2022
Published online: 1 Jan. 2023

Abstract: In survey sampling, presence of outliers in the data collected has been one of the biggest concerns. This problem becomes severe when dealing with estimation of communicable diseases. Further, there are several cases where the population under study is hidden clustered or clumped and non-adaptive sampling designs in these cases fail to give a reliable estimate like in case of COVID-19 during the initial days. To deal with these two problems, in this article we have developed eight robust estimators to estimate the unknown population mean of a rare or hidden clustered population in presence of outliers. In order to assess the performance of the proposed estimators against similar existing robust ratio type estimators, simulation studies have been conducted to estimate the average of COVID-19 cases in the Indian union territory of Andaman & Nicobar Islands and the state of Goa. The results of the study show that the developed estimators perform much better as compared to other similar existing robust ratio type estimators presented in this article.

Keywords: COVID-19, Robust type estimator, Robust ratio type, Auxiliary information, Simulation, Adaptive cluster sampling, SARS-COV-2.

---

1 Introduction

The type of sampling design to be used depends on the population under study and when that population is rare or highly clustered, the conventional sampling designs cannot be used to estimate the population's parameter of interest as in such a case, most of the sampled units will provide a heavily biased estimate of the population's parameter of interest. This has been, one of the major issues in survey sampling theory. However, Thompson (1990) proposed a sampling design that gives an edge to the researcher, allowing them to fix a criterion or a condition according to which the units will be selected in the sample. This sampling design is called Adaptive cluster sampling (ACS). Due to its need and flexibility, ACS designs have been used in various disciplines such as Ecological science e.g.,Achrya et al. (2000), Environmental science (e.g.,Correll (2001)), Epidemiological study and Social science (Thompson & Collins (2002); Thompson (1997)).

As per WHO, COVID-19 is caused by the SARS-CoV-2 virus. The virus spreads in various ways, some of them include, spreading between people who are in close proximity with each other or in a poorly ventilated indoor setting or getting infected by touching a surface contaminated by the virus when touching their eye, nose or mouth without cleaning their hands WHO (2020).

As a result, localities followed by cities and then entire states become a hotspot of the virus putting an immense burden on the healthcare system of a country. At first, when the virus starts to spread, the cases are highly clustered or rare and at different places depending on the carrier of the virus, in such a situation using classical sampling designs to estimate the average number of cases would result in extremely biased estimates. To deal with this, adaptive cluster sampling becomes the only viable option.

---

* Corresponding author e-mail: i.rohanskmishra@gmail.com

The data collected, frequently contains one or more typical observations known as outliers. These outliers are well separated from the majority or bulk of the data or in some way deviate from the general pattern of the data (Ricardo et al. (2019)). The problem with the outliers is that they adversely influence the estimates like the sample mean, sample standard deviation and correlation coefficient to name a few. In such situations, researchers have to use robust measures to reach a reliable estimate.

Using known auxiliary information many ratio estimators (Singh & Kumar (2011); Singh et al. (2013); Audu et al. (2021)) are proposed under in SRS. In ACS, Yadav et al.Yadav et al. (2016) proposed improved ratio estimators using known auxiliary information. Qureshi et al. (2018) and Zaman et al. (2021) using some robust and non-robust measures of auxiliary variable proposed robust ratio estimators to estimate unknown population mean of a finite population.

In this article, we propose eight robust ratio type estimators to estimate the average new COVID-19 cases, using a combination of robust and non-robust measures of auxiliary variable viz., Mid-range (MR), Tri mean (TM), Hodges Lehmann (HL) and kurtosis.

The methodology of ACS design is presented in Section 2. In Section 3 the proposed robust ratio type estimators along with their derivations of bias and MSEs up to first order of approximations are presented. In Section 4 a simulation study to estimate average number of new COVID-19 cases in the Indian union territory of Andaman and Nicobar Islands and state of Goa between 27^{th} of March 2020 to 4^{th} of July 2020 PRS India (2020) is conducted. A discussion on the results obtained is presented in section 5. The final concluding remarks of the article are presented in Section 6.

2 Methodolgy of ACS design

ACS is an adaptive sampling design in which, the units in the final sample depends on all the units which have been observed during the survey. Initially, a sample of size n_1 is drawn from the population of size N using any conventional sampling design (usually SRSWOR) and if these selected units satisfy some researcher-specific condition C, then additional units are drawn from a pre-defined neighbourhood.
So, before conducting the survey, two things should be clearly defined:

– the neighbourhood of a unit (or observation)
– the researcher-specific condition (C)

This researcher-specific condition for selecting the observation on survey variable y is usually $y_i > 0$. In ACS, the used choice of neighbourhood is 4 unit first order in which, if any i^{th} unit selected in the initial sample is greater than 0, the units adjacent to this i^{th} unit in its East, West, North and South directions are also selected. This process of selecting the neighbourhood keeps on going until no further additional unit satisfies the condition C.

The units satisfying condition C form a network, and units not satisfying it are called edge units and are considered to be a network of size 1. The selection of any unit of a network leads to the selection of the entire network. These networks and edge units together form a cluster (Fig. 1).
The clusters are obviously not disjoint due to overlapping edge units but the units of a network are non-overlapping and thus the entire population can be partitioned as a set of networks and edge units.
Once there are no more additional units satisfying condition C, ACS terminates and the sample obtained consists of units selected in the initial sample and adaptively selected units.
Once the population is divided into networks and edge units, we make a transformed population by assigning the average value of a network to all the units of this network but edge units stay the same. Once the transformed population is obtained, and we consider averages of networks then ACS can be regarded as either SRSWOR or SRSWR.

3 Proposed robust ratio type estimators

The main objective of this research is to develop robust ratio type estimators that give minimum MSE as compared to existing robust ratio type estimators. Motivated by Singh & Mishra (2022); Singh & Mishra (2022) and Qureshi et al. (2018) we propose the following robust ratio type estimators:

$$t_1 = \bar{w}_y \frac{\mu_{w_x} MR + \beta_2(w_x)}{\bar{w}_x MR + \beta_2(w_x)}, \tag{1}$$

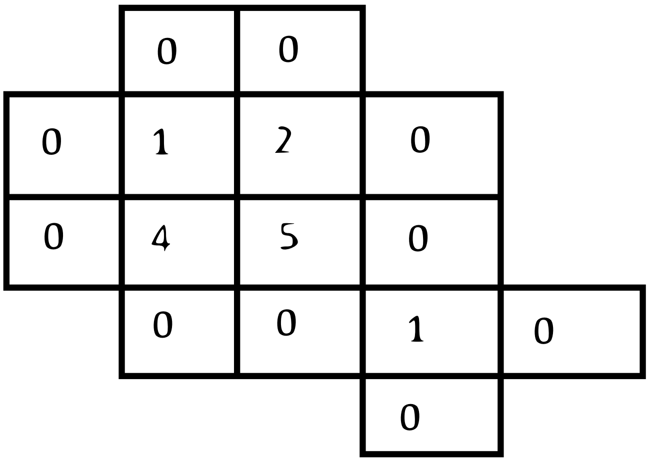

Fig. 1: An example of a hypothetical cluster with pre-defined condition (C) $y_i > 0$. The units having y-values 1, 2, 4, 5 and 1 form a network of size five. The edge units are the units with y values 0 and are adjacent to the y values greater than 0. Together they form a cluster.

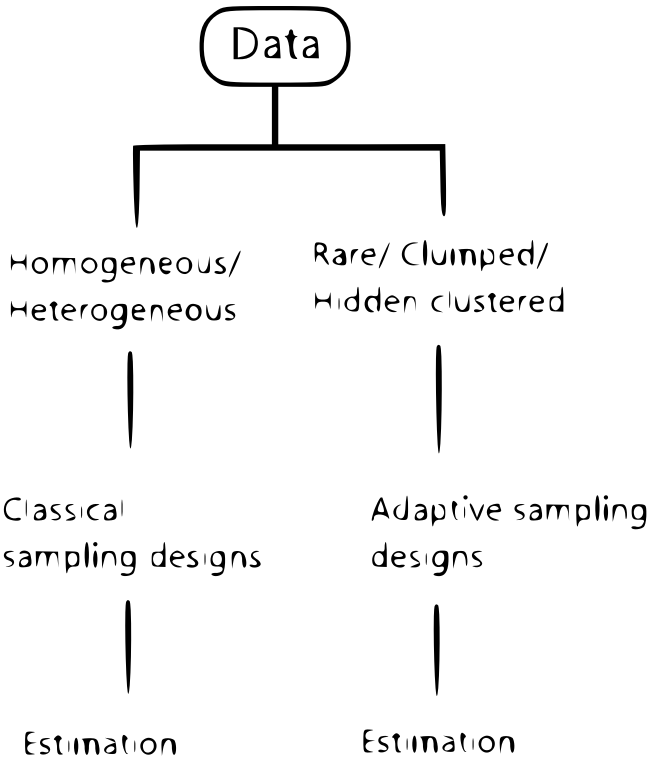

Fig. 2: Flow diagram describing when to use Adaptive designs

$$t_2 = \bar{w_y} \frac{\mu_{w_x} TM + MR}{\bar{w_x} TM + MR}, \quad (2)$$

$$t_3 = \bar{w_y} \frac{\mu_{w_x} HL + \beta_2(w_x)}{\bar{w_x} HL + \beta_2(w_x)}, \quad (3)$$

$$t_4 = \bar{w_y} \frac{\mu_{w_x} TM + HL}{\bar{w}_x TM + HL}, \tag{4}$$

and

$$t_5 = \bar{w_y} \frac{\mu_{w_x} TM + \beta_2(w_x)}{\bar{w}_x TM + \beta_2(w_x)}, \tag{5}$$

$$t_6 = \bar{w_y} \frac{\mu_{w_x} TM + M_d}{\bar{w}_x TM + M_d}, \tag{6}$$

$$t_7 = \bar{w_y} \frac{\mu_{w_x} HL + M_d}{\bar{w}_x HL + M_d}, \tag{7}$$

and

$$t_8 = \bar{w_y} \frac{\mu_{w_x} MR + M_d}{\bar{w}_x MR + M_d}, \tag{8}$$

where TM is tri-mean, MR is mid-range, HL is Hodges-Lehman, M_d is the median, $\beta_2(w_x)$ is the coefficient of skewness of transformed auxiliary variable,

$$w_{y_i} = \frac{1}{m_i} \sum_{j \varepsilon \Psi_i} (y_j),$$

$$w_{x_i} = \frac{1}{m_j} \sum_{k \varepsilon \Psi_j} (x_k),$$

and m_i and m_j are number of units in the network ψ_i and ψ_j repectively.
For obtaining the expression of bias and MSE of t_1, we rewrite (1) in error terms as:

$$t_1 = \mu_{w_y}(e_{w_y}+1)\left(\frac{MR\mu_{w_x} + \beta_2(w_x)}{\mu_{w_x}(e_{w_x}+1)MR + \beta_2(w_x)}\right) \tag{9}$$

$e_{w_y} = \frac{\bar{w_y}}{\bar{\mu_{w_y}}} - 1$, $e_{w_x} = \frac{\bar{w_x}}{\bar{\mu_{w_x}}} - 1$, $E(e^2_{w_x}) = (\frac{1}{n} - \frac{1}{N})C^2_{w_x}$, $E(e^2_{w_y}) = (\frac{1}{n} - \frac{1}{N})C^2_{w_y}$,

$E(e_{w_x} e_{w_y}) = (\frac{1}{n} - \frac{1}{N})\rho_{w_x w_y} C_{w_x} C_{w_y}$,

$C^2_{w_x} = \frac{S^2_{w_x}}{\bar{\mu_{w_x}}^2}$, $C^2_{w_y} = \frac{S^2_{w_y}}{\bar{\mu_{w_y}}^2}$, $S^2_{w_x} = \frac{1}{N-1}\sum_{i=1}^{N}(w_{x_i} - \bar{\mu_{w_x}})^2$,

$S^2_{w_y} = \frac{1}{N-1}\sum_{i=1}^{N}(w_{y_i} - \bar{\mu_{w_y}})^2$, $\rho_{w_x w_y} = \frac{S_{w_x w_y}}{S_{w_x} S_{w_y}}$,

and $S_{w_x w_y} = \frac{1}{N-1}\sum_{i=1}^{n}(w_{x_i} - \bar{\mu_{w_x}})(w_{y_i} - \bar{\mu_{w_y}})$

Simplifying (9), we get

$$t_1 = \mu_{w_y}(e_{w_y}+1)\left(1 + \frac{\beta_2(w_x)}{MR\mu_{w_x}}\right)\left(1 + \frac{\beta_2(w_x)}{MR\mu_{w_x}} + e_{w_x}\right)^{-1} \tag{10}$$

Taking

$$\gamma_1 = 1 + \frac{\beta_2(w_x)}{MR\mu_{w_x}}$$

and using it (10), we get

$$t_1 = \mu_{w_y}(e_{w_y}+1)\gamma_1\left(\frac{1}{\gamma_1}-\frac{e_{w_x}}{\gamma_1^2}+\frac{e_{w_x}^2}{\gamma_1^3}\right) \tag{11}$$

$$t_1 = \mu_{w_y}-\mu_{w_y}\frac{e_{w_x}}{\gamma_1}+\mu_{w_y}\frac{e_{w_x}^2}{\gamma_1^2}+\mu_{w_y}e_{w_y}-\mu_{w_y}\frac{e_{w_x}e_{w_y}}{\gamma_1} \tag{12}$$

Subtracting μ_{w_y} from both sides in (12) and taking expectation we get

$$Bias(t_1) = \frac{N-n}{Nn}\mu_{w_y}(\left(\frac{MR\mu_{w_x}}{MR\mu_{w_x}+\beta_2(w_x)}\right)^2 C_{w_x}^2-\left(\frac{MR\mu_{w_x}}{MR\mu_{w_x}+\beta_2(w_x)}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}) \tag{13}$$

Similarly, subtracting μ_{w_y} from both sides in (12), squaring and taking expectation we get

$$MSE(t_1) = \frac{N-n}{Nn}\mu_{w_y}^2(C_{w_y}^2+\left(\frac{MR\mu_{w_x}}{MR\mu_{w_x}+\beta_2(w_x)}\right)^2 C_{w_x}^2-2\left(\frac{MR\mu_{w_x}}{MR\mu_{w_x}+\beta_2(w_x)}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}) \tag{14}$$

For obtaining the expression of bias and MSE of t_2, we rewrite (2) in error terms as:

$$t_2 = \mu_{w_y}(e_{w_y}+1)\left(\frac{TM\mu_{w_x}+MR}{\mu_{w_x}(e_{w_x}+1)TM+MR}\right) \tag{15}$$

Simplifying (15), we get

$$t_2 = \mu_{w_y}(e_{w_y}+1)\left(1+\frac{MR}{TM\mu_{w_x}}\right)\left(1+\frac{MR}{TM\mu_{w_x}}+e_{w_x}\right)^{-1} \tag{16}$$

Taking

$$\gamma_2 = 1+\frac{MR}{TM\mu_{w_x}}$$

and using it in (16), we get

$$t_2 = \mu_{w_y}(e_{w_y}+1)\gamma_2\left(\frac{1}{\gamma_2}-\frac{e_{w_x}}{\gamma_2^2}+\frac{e_{w_x}^2}{\gamma_2^3}\right). \tag{17}$$

On further simplification we get

$$t_2 = \mu_{w_y}-\mu_{w_y}\frac{e_{w_x}}{\gamma_2}+\mu_{w_y}\frac{e_{w_x}^2}{\gamma_2^2}+\mu_{w_y}e_{w_y}-\mu_{w_y}\frac{e_{w_x}e_{w_y}}{\gamma_2} \tag{18}$$

Subtracting μ_{w_y} from both sides in (18) and taking expectation we get

$$Bias(t_2) = \frac{N-n}{Nn}\mu_{w_y}(\left(\frac{TM\mu_{w_x}}{TM\mu_{w_x}+MR}\right)^2 C_{w_x}^2-\left(\frac{TM\mu_{w_x}}{TM\mu_{w_x}+MR}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}) \tag{19}$$

Similarly, subtracting μ_{w_y} from both sides in (18), squaring and taking expectation we get

$$MSE(t_2) = \frac{N-n}{Nn}\mu_{w_y}^2(C_{w_y}^2+\left(\frac{TM\mu_{w_x}}{TM\mu_{w_x}+MR}\right)^2 C_{w_x}^2-2\left(\frac{TM\mu_{w_x}}{TM\mu_{w_x}+MR}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}) \tag{20}$$

Similary, we rewrite (3) in error terms to obtain its expression of bias and MSE as:

$$t_3 = \mu_{w_y}(e_{w_y}+1)\left(\frac{HL\mu_{w_x}+\beta_2(w_x)}{\mu_{w_x}(e_{w_x}+1)HL+\beta_2(w_x)}\right) \tag{21}$$

Simplifying (21) we get

$$t_3 = \mu_{w_y}(e_{w_y}+1)\left(1+\frac{\beta_2(w_x)}{HL\mu_{w_x}}\right)\left(1+\frac{\beta_2(w_x)}{HL\mu_{w_x}}+e_{w_x}\right)^{-1} \tag{22}$$

Taking

$$\gamma_3 = 1+\frac{\beta_2(w_x)}{HL\mu_{w_x}}.$$

On siplifying (22) we get

$$t_3 = \mu_{w_y} - \mu_{w_y}\frac{e_{w_x}}{\gamma_3} + \mu_{w_y}\frac{e_{w_x}^2}{\gamma_3^2} + \mu_{w_y}e_{w_y} - \mu_{w_y}\frac{e_{w_x}e_{w_y}}{\gamma_3}. \tag{23}$$

Subtracting μ_{w_y} from both sides in (23) and taking expectation we get

$$Bias(t_3) = \frac{N-n}{Nn}\mu_{w_y}(\left(\frac{HL\mu_{w_x}}{HL\mu_{w_x}+\beta_2(w_x)}\right)^2 C_{w_x}^2 - \left(\frac{HL\mu_{w_x}}{HL\mu_{w_x}+\beta_2(w_x)}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}). \tag{24}$$

Similarly, subtracting μ_{w_y} from both sides in (23), squaring and taking expectation we get

$$MSE(t_3) = \frac{N-n}{Nn}\mu_{w_y}^2(C_{w_y}^2 + \left(\frac{HL\mu_{w_x}}{HL\mu_{w_x}+\beta_2(w_x)}\right)^2 C_{w_x}^2 - 2\left(\frac{HL\mu_{w_x}}{HL\mu_{w_x}+\beta_2(w_x)}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}). \tag{25}$$

Similary, to obtain the expression of Bias and MSE of proposed robust type estimator t_4 we rewrite (4) in error terms to obtain its expression of bias and MSE as:

$$t_4 = \mu_{w_y}(e_{w_y}+1)\left(\frac{TM\mu_{w_x}+HL}{\mu_{w_x}(e_{w_x}+1)TM+HL}\right) \tag{26}$$

Simplifying (26) we get

$$t_4 = \mu_{w_y}(e_{w_y}+1)\left(1+\frac{HL}{TM\mu_{w_x}}\right)\left(1+\frac{HL}{TM\mu_{w_x}}+e_{w_x}\right)^{-1} \tag{27}$$

Taking

$$\gamma_4 = 1+\frac{HL}{TM\mu_{w_x}}.$$

On siplifying (28) we get

$$t_4 = \mu_{w_y} - \mu_{w_y}\frac{e_{w_x}}{\gamma_4} + \mu_{w_y}\frac{e_{w_x}^2}{\gamma_4^2} + \mu_{w_y}e_{w_y} - \mu_{w_y}\frac{e_{w_x}e_{w_y}}{\gamma_4} \tag{28}$$

Subtracting μ_{w_y} from both sides in (28) and taking expectation we get

$$Bias(t_4) = \frac{N-n}{Nn}\mu_{w_y}(\left(\frac{TM\mu_{w_x}}{TM\mu_{w_x}+HL}\right)^2 C_{w_x}^2 - \left(\frac{TM\mu_{w_x}}{TM\mu_{w_x}+TM}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}) \tag{29}$$

Similarly, subtracting μ_{w_y} from both sides in (28), squaring and taking expectation we get

$$MSE(t_4) = \frac{N-n}{Nn}\mu_{w_y}^2(C_{w_y}^2 + \left(\frac{TM\mu_{w_x}}{TM\mu_{w_x}+HL}\right)^2 C_{w_x}^2 - 2\left(\frac{TM\mu_{w_x}}{TM\mu_{w_x}+HL}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}) \tag{30}$$

The expressions of Bias and MSE are obtained similarly for the proposed robust ratio type estimators $t_5 - t_8$ and are presented in Table 1 and Table 2 respectively while Table 3 and Table 4 consists of the ratio estimator in case of SRS and the existing robust ratio estimators along with their expressions of Bias and MSE.

Table 1: Proposed robust ratio type estimators with corresponding expression of bias

Estimator	Form	Bias
t_1	$\bar{w}_y\left(\frac{MR\mu_{w_x}+\beta_2(w_x)}{MR\mu_{w_x}+\beta_2(w_x)}\right)$	$\frac{N-n}{Nn}\mu_{w_y}\left(\left(\frac{MR\mu_{w_x}}{MR\mu_{w_x}+\beta_2(w_x)}\right)^2 C_{w_x}^2-\left(\frac{MR\mu_{w_x}}{MR\mu_{w_x}+\beta_2(w_x)}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}\right)$
t_2	$\bar{w}_y\left(\frac{TM\mu_{w_x}+MR}{TM\mu_{w_x}+MR}\right)$	$\frac{N-n}{Nn}\mu_{w_y}\left(\left(\frac{TM\mu_{w_x}}{TM\mu_{w_x}+MR}\right)^2 C_{w_x}^2-\left(\frac{TM\mu_{w_x}}{TM\mu_{w_x}+MR}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}\right)$
t_3	$\bar{w}_y\left(\frac{HL\mu_{w_x}+\beta_2(w_x)}{HL\mu_{w_x}+\beta_2(w_x)}\right)$	$\frac{N-n}{Nn}\mu_{w_y}\left(\left(\frac{HL\mu_{w_x}}{HL\mu_{w_x}+\beta_2(w_x)}\right)^2 C_{w_x}^2-\left(\frac{HL\mu_{w_x}}{HL\mu_{w_x}+\beta_2(w_x)}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}\right)$
t_4	$\bar{w}_y\left(\frac{TM\mu_{w_x}+HL}{TM\mu_{w_x}+HL}\right)$	$\frac{N-n}{Nn}\mu_{w_y}\left(\left(\frac{TM\mu_{w_x}}{TM\mu_{w_x}+HL}\right)^2 C_{w_x}^2-\left(\frac{TM\mu_{w_x}}{TM\mu_{w_x}+HL}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}\right)$
t_5	$\bar{w}_y\left(\frac{TM\mu_{w_x}+\beta_2(w_x)}{TM\mu_{w_x}+\beta_2(w_x)}\right)$	$\frac{N-n}{Nn}\mu_{w_y}\left(\left(\frac{TM\mu_{w_x}}{TM\mu_{w_x}+\beta_2(w_x)}\right)^2 C_{w_x}^2-\left(\frac{TM\mu_{w_x}}{TM\mu_{w_x}+\beta_2(w_x)}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}\right)$
t_6	$\bar{w}_y\left(\frac{TM\mu_{w_x}+M_d}{TM\mu_{w_x}+M_d}\right)$	$\frac{N-n}{Nn}\mu_{w_y}\left(\left(\frac{TM\mu_{w_x}}{TM\mu_{w_x}+M_d}\right)^2 C_{w_x}^2-\left(\frac{TM\mu_{w_x}}{TM\mu_{w_x}+M_d}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}\right)$
t_7	$\bar{w}_y\left(\frac{HL\mu_{w_x}+M_d}{HL\mu_{w_x}+M_d}\right)$	$\frac{N-n}{Nn}\mu_{w_y}\left(\left(\frac{HL\mu_{w_x}}{HL\mu_{w_x}+M_d}\right)^2 C_{w_x}^2-\left(\frac{HL\mu_{w_x}}{HL\mu_{w_x}+M_d}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}\right)$
t_8	$\bar{w}_y\left(\frac{MR\mu_{w_x}+M_d}{MR\mu_{w_x}+\beta_2(w_x)}\right)$	$\frac{N-n}{Nn}\mu_{w_y}\left(\left(\frac{MR\mu_{w_x}}{MR\mu_{w_x}+M_d}\right)^2 C_{w_x}^2-\left(\frac{MR\mu_{w_x}}{MR\mu_{w_x}+M_d}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}\right)$

Table 2: Proposed robust ratio type estimators with corresponding expression of MSE

Estimator	MSE
t_1	$\frac{N-n}{Nn}\mu_{w_y}^2\left(C_{w_y}^2+\left(\frac{MR\mu_{w_x}}{MR\mu_{w_x}+\beta_2(w_x)}\right)^2 C_{w_x}^2-2\left(\frac{MR\mu_{w_x}}{MR\mu_{w_x}+\beta_2(w_x)}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}\right)$
t_2	$\frac{N-n}{Nn}\mu_{w_y}^2\left(C_{w_y}^2+\left(\frac{TM\mu_{w_x}}{TM\mu_{w_x}+MR}\right)^2 C_{w_x}^2-2\left(\frac{TM\mu_{w_x}}{TM\mu_{w_x}+MR}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}\right)$
t_3	$\frac{N-n}{Nn}\mu_{w_y}^2\left(C_{w_y}^2+\left(\frac{HL\mu_{w_x}}{HL\mu_{w_x}+\beta_2(w_x)}\right)^2 C_{w_x}^2-2\left(\frac{HL\mu_{w_x}}{HL\mu_{w_x}+\beta_2(w_x)}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}\right)$
t_4	$\frac{N-n}{Nn}\mu_{w_y}^2\left(C_{w_y}^2+\left(\frac{TM\mu_{w_x}}{TM\mu_{w_x}+HL}\right)^2 C_{w_x}^2-2\left(\frac{TM\mu_{w_x}}{TM\mu_{w_x}+HL}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}\right)$
t_5	$\frac{N-n}{Nn}\mu_{w_y}^2\left(C_{w_y}^2+\left(\frac{TM\mu_{w_x}}{TM\mu_{w_x}+\beta_2(w_x)}\right)^2 C_{w_x}^2-2\left(\frac{TM\mu_{w_x}}{TM\mu_{w_x}+\beta_2(w_x)}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}\right)$
t_6	$\frac{N-n}{Nn}\mu_{w_y}^2\left(C_{w_y}^2+\left(\frac{TM\mu_{w_x}}{TM+M_d}\right)^2 C_{w_x}^2-2\left(\frac{TM\mu_{w_x}}{TM\mu_{w_x}+M_d}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}\right)$
t_7	$\frac{N-n}{Nn}\mu_{w_y}^2\left(C_{w_y}^2+\left(\frac{HL\mu_{w_x}}{HL\mu_{w_x}+M_d}\right)^2 C_{w_x}^2-2\left(\frac{HL\mu_{w_x}}{HL\mu_{w_x}+M_d}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}\right)$
t_8	$\frac{N-n}{Nn}\mu_{w_y}^2\left(C_{w_y}^2+\left(\frac{MR\mu_{w_x}}{MR\mu_{w_x}+M_d}\right)^2 C_{w_x}^2-2\left(\frac{MR\mu_{w_x}}{MR\mu_{w_x}+M_d}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}\right)$

4 Simulation study

In this section, we have conducted two simulation studies on daily new cases of COVID-19 in the Indian union territory of Andaman and Nicobar islands and in the state of Goa from 27^{th} of March 2020 to 4^{th} of July 2020 PRS India (2020). Average cases of COVID-19 in this 100 days duration is estimated for the union territory of Andaman and Nicobar

Table 3: Existing ratio and robust ratio estimators with corresponding values of bias

Estimator	Form	Bias
t_{srs}*Dryver&Chao* (2007)	$\mu_{w_x}\frac{\bar{w_y}}{\bar{w_x}}$	$\frac{N-n}{Nn}\mu_{w_y}(C^2_{w_x}-\rho_{w_xw_y}C_{w_x}C_{w_y})$
t_{Q_1} Qureshi et al. (2018)	$\bar{w_y}\left(\frac{MR\mu_{w_x}+\beta_1(w_x)}{MR\mu_{w_x}+\beta_1(w_x)}\right)$	$\frac{N-n}{Nn}\mu_{w_y}\left(\left(\frac{MR\mu_{w_x}}{MR\mu_{w_x}+\beta_1(w_x)}\right)^2C^2_{w_x}-\left(\frac{MR\mu_{w_x}}{MR\mu_{w_x}+\beta_1(w_x)}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}\right)$
t_{Q_2} Qureshi et al. (2018)	$\bar{w_y}\left(\frac{HL\mu_{w_x}+\beta_1(w_x)}{HL\mu_{w_x}+\beta_1(w_x)}\right)$	$\frac{N-n}{Nn}\mu_{w_y}\left(\left(\frac{HL\mu_{w_x}}{HL\mu_{w_x}+\beta_1(w_x)}\right)^2C^2_{w_x}-\left(\frac{HL\mu_{w_x}}{HL\mu_{w_x}+\beta_1(w_x)}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}\right)$
t_{Q_3} Qureshi et al. (2018)	$\bar{w_y}\left(\frac{HL\mu_{w_x}+TM}{HL\mu_{w_x}+TM}\right)$	$\frac{N-n}{Nn}\mu_{w_y}\left(\left(\frac{HL\mu_{w_x}}{HL\mu_{w_x}+TM}\right)^2C^2_{w_x}-\left(\frac{HL\mu_{w_x}}{HL\mu_{w_x}+TM}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}\right)$

Table 4: Existing ratio and robust ratio estimators with corresponding expression of MSE

Estimator	MSE
t_{srs}*Dryver&Chao* (2007)	$\frac{N-n}{Nn}\mu^2_{w_y}\left(C^2_{w_y}+C^2_{w_x}-2\rho_{w_xw_y}C_{w_y}C_{w_x}\right)$
t_{Q_1} Qureshi et al. (2018)	$\frac{N-n}{Nn}\mu^2_{w_y}\left(C^2_{w_y}+\left(\frac{MR\mu_{w_x}}{MR\mu_{w_x}+\beta_1(w_x)}\right)^2C^2_{w_x}-2\left(\frac{MR\mu_{w_x}}{MR\mu_{w_x}+\beta_1(w_x)}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}\right)$
t_{Q_2} Qureshi et al. (2018)	$\frac{N-n}{Nn}\mu^2_{w_y}\left(C^2_{w_y}+\left(\frac{HL\mu_{w_x}}{HL\mu_{w_x}+\beta_1(w_x)}\right)^2C^2_{w_x}-2\left(\frac{HL\mu_{w_x}}{HL\mu_{w_x}+\beta_1(w_x)}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}\right)$
t_{Q_3} Qureshi et al. (2018)	$\frac{N-n}{Nn}\mu^2_{w_y}\left(C^2_{w_y}+\left(\frac{HL\mu_{w_x}}{HL\mu_{w_x}+TM}\right)^2C^2_{w_x}-2\left(\frac{HL\mu_{w_x}}{HL\mu_{w_x}+TM}\right)\rho_{w_xw_y}C_{w_x}C_{w_y}\right)$

Islands and the state of Goa, taking daily occurrences of new COVID-19 cases of Andhra Pradesh as an auxiliary variable for all the estimators.

The rationale behind taking the daily occurrence of new cases of the virus of Arunachal Pradesh as an auxiliary variable for estimation of average cases of Goa and Andaman and Nicobar islands is that the pattern of spread was to an extent same. Moreover, the correlation coefficient of daily new cases of Arunachal Pradesh with that of Goa and Andaman and Nicobar islands is greater than 0.5.

For simulation study, the MSE is,

$$MSE(t_*)=\frac{1}{4000}\sum_{i=1}^{4000}(t_*-\mu_{w_y})^2, \tag{31}$$

where t_* represents all the estimators presented in this study respectively and 10000 is the number of replications or the number of repetative samples drawn for sample of size 5, 7, 9, 11 and 13 days.

Table 5: MSE of all the estimators in case of Andaman & Nicobar Islands

n	5	7	9	11	13
t_{Q1}	0.6645	0.5335	0.4598	0.4113	0.3719
t_{Q2}	0.6646	0.5336	0.4597	0.4113	0.3719
t_{Q3}	0.6644	0.5334	0.4595	0.4114	0.3718
t_1	0.6644	0.5334	0.4597	0.4112	0.3718
t_2	0.6634	0.5330	0.4594	0.4110	0.3716
t_3	0.6644	0.5334	0.4591	0.4112	0.3718
t_4	0.6634	0.5326	0.4591	0.4107	0.3713
t_5	0.6644	0.5334	0.4597	0.4112	0.3718
t_6	0.6641	0.5331	0.4595	0.4110	0.3716
t_7	0.6641	0.5331	0.4595	0.4110	0.3716
t_8	0.6636	0.5328	0.4592	0.4108	0.3714

Table 6: MSE of all the estimators in case of Goa

n	5	7	9	11	13
t_{Q1}	58.3166	42.6051	34.9747	26.1612	20.7570
t_{Q2}	58.3141	42.6034	34.9733	26.1602	20.7563
t_{Q3}	58.2750	42.5565	34.9550	26.1405	20.7671
t_1	58.2963	42.5924	34.9647	26.1540	20.7513
t_2	58.2426	42.5588	34.9383	26.1348	20.7361
t_3	58.3037	42.5971	34.9684	26.1566	20.7534
t_4	58.1747	42.5164	34.9049	26.1104	20.7170
t_5	58.3037	42.5971	34.9684	26.1566	20.7534
t_6	58.2563	42.5674	34.9451	26.1397	20.7400
t_7	58.2563	42.5674	34.9451	26.1397	20.7400
t_8	58.2020	42.5334	34.9183	26.1202	20.7246

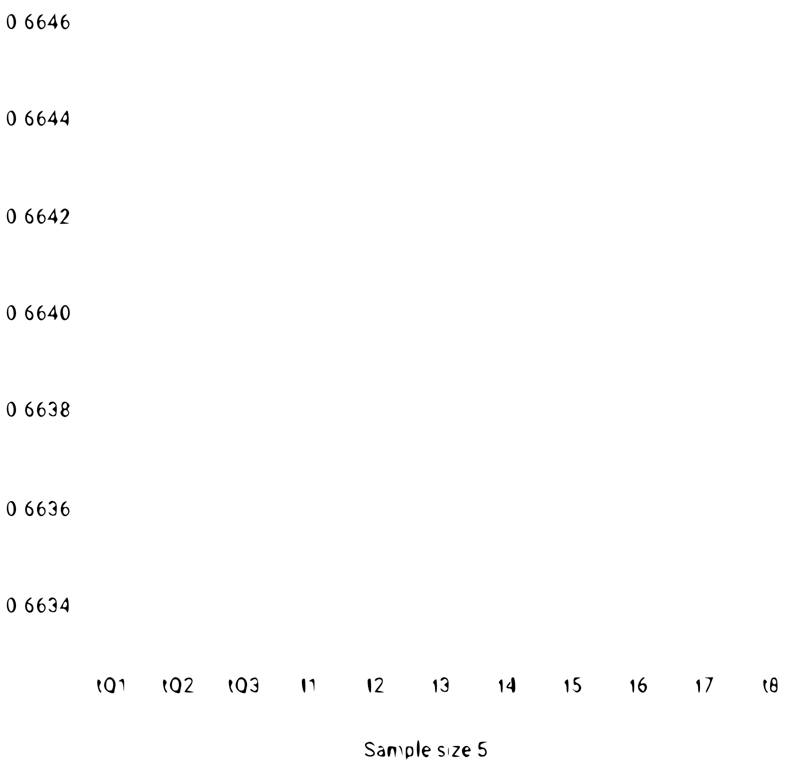

Fig. 3: MSE of all estimators for sample size 5 in case of Andaman & Nicobar Islands

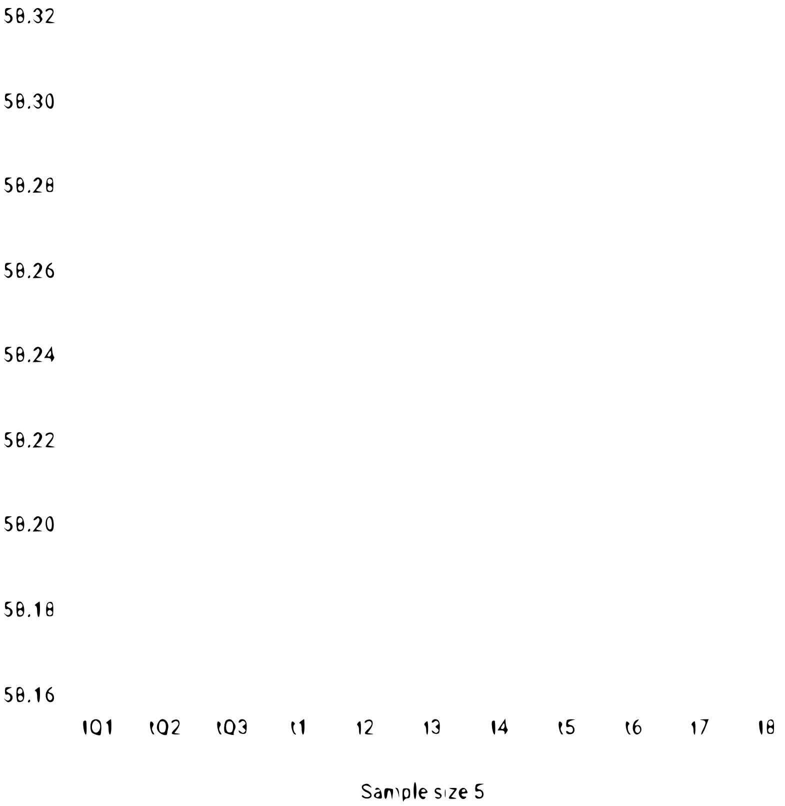

Fig. 4: MSE of all estimators for sample size 5 in case of Goa

5 Discussion

The aim of this article was to develop some robust ratio type estimators of finite population mean in the ACS design. In this article, we proposed eight robust ratio type estimators and derieved the expressions of bias and MSE of few of them up to first order of approximation. We used the proposed and existing robust ratio type estimators in estimating the average Covid-19 cases in the Indian Union territory of Andaman & Nicobar Islands and the state of Goa. We compared the results of all the estimators on the basis of MSE. From the results of the simulation studies presented in Table 5 and Table 6, we see that almost all the raobust ratio type estimators resulted in low MSE but after carefull observation we can see that in case of Andaman & Nicobar Islands proposed estimator t_4 results in lowest MSE. Same result is seen in case of Goa as well.

6 Conclusion

The existing and proposed robust ratio type estimators resulted in low MSEs and thus their use is encouraged when the population is rare and contains outliers but in real surveys when cost is to be minimized, it is paramount to get highest possible efficiency without increasing the sample size and in such a case it is important to have various estimators at hand which are highly precise and efficient. Thus when population under study is rare or clumped, as initial spread of SARS-COV-2 is, the proposed estimator t_4 is advised to be applied.

References

Acharya, B., Bhattarai, G., de Gier, A., & Stein, A. (2000). Systematic adaptive cluster sampling for the assessment of rare tree species in Nepal. Forest Ecology and Management, 137(1?3), 65?73. https://doi.org/10.1016/s0378-1127(99)00318-7.

Audu, A., Singh, R., & Khare, S. (2021). Developing calibration estimators for population mean using robust measures of dispersion under stratified random sampling. Statistics in Transition New Series, 22(2), 125?142. https://doi.org/10.21307/stattrans-2021-019.

Correll, R. L. (2001). The use of composite sampling in contaminated sitesa case study. Environmental and Ecological Statistics, 8(3), 185?200.

Dryver, A. L. & Chao, C.T., 2007. Ratio estimators in adaptive cluster sampling, Environmetrics: The official journal of the International Environmetrics Society, Wiley Online Library 18(6), 607-620.

PRS INDIA (2020). PRSINDIA ORGANISATION. https://prsindia.org/.

Qureshi, M. N., Kadilar, C., Noor Ul Amin, M., & Hanif, M. (2018). Rare and clustered population estimation using the adaptive cluster sampling with some robust measures. Journal of Statistical Computation and Simulation, 88(14), 2761?2774.

Robust Statistics: Theory and Methods (with R). (2019). Wiley.

Sing, R. & Kumar, M. (2011). A note on transformations on auxiliary variable in survey sampling. Model Assisted Statistics and Applications, 6(1), 17?19.

Singh, R., Kumar, M., & Singh, H. P. (2013). On estimation of poulation mean using information on auxiliary attribute. Pakistan Journal of Statistics and Operation Research, 363?371.

Singh, R. & Mishra, R. (2022). Improved Exponential Ratio Estimators in Adaptive Cluster Sampling. J Stat Appl Probab Lett, 9(1), 19-29.

Singh, R. & Mishra, R. (2022). Transformed ratio type estimators under Adaptive Cluster Sampling: An application to COVID-19. J Stat Appl Probab Lett, 9(2), 63-70.

Thompson, S. K. (1990). Adaptive cluster sampling. Journal of the American Statistical Association, 85(412), 1050?1059.

Thompson, S. K. (1997). Adaptive sampling in behavioral surveys. NIDA Research Monograph, 167, 296?319.

Thompson, S. K. & Collins, L. M. (2002). Adaptive sampling in research on risk-related behaviors. Drug and Alcohol Dependence, 68, 57?67.

WHO (2020). https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19-how-is-it-transmitted.

Yadav, S. K., Misra, S., Mishra, S. S., & Chutiman, N. (2016). Improved ratio estimators of population mean in adaptive cluster sampling. J Stat Appl Probab Lett, 3(1), 1?6.

Zaman, T., Dunder, E., Audu, A., Alilah, D. A., Shahzad, U., & Hanif, M. (2021). Robust regression-ratio-type estimators of the mean utilizing two auxiliary variables: A simulation study. Mathematical Problems in Engineering, 2021.