DOI: https://doi.org/10.4038/sljastats.v24i3.8114

Sri Lankan Journal of Applied Statistics, Vol (24-3)

Estimating COVID-19 Prevalence in Sri Lanka: A Dynamic Sampling Model Approach

J. D. T. Erandi[1,*], **U. P. Liyanage**[1], **and A. Gunawardana**[2]

[1]Department of Statistics and Computer Science, University of Kelaniya, Sri Lanka

[2]Department of Decision Sciences, University of Moratuwa, Sri Lanka

*Corresponding author: tenushierandi@gmail.com

Received: 9th November 2023/ Revised: 23rd November 2023/ Published: 30th December 2023

ABSTRACT

Over history, human has had to face various crises and diseases, and among them, the COVID 19 virus stands out as one of the most deathful diseases ever. It has brought numerous challenges to all the fields worldwide. Despite efforts to control its spread, the virus persists globally with varying intensity. Addressing this challenge requires an effective and precise control measure. The progression of the virus in different sub-regions is influenced by factors such as population density, public mobility, and healthcare infrastructure. Consequently, the prevalence of the virus varies across sub-regions. This study proposes an adaptive sampling design that modifies the stratified sampling technique to capture the changing prevalence of COVID-19, considering the dynamic nature of infected populations. This adaptation is essential as the increase of infected cases boosts the virus spread, and the standard sampling techniques do not address such dynamic population conditions in determining the sample size. The study aims to narrow the gap between reported and actual daily infections, providing more accurate estimates of virus distribution. The weighted allocation method incorporates the skewed pattern of coronavirus progression, with weights determined based on the first derivative of reported infected cases. This derivative information is based on the recent dynamics of the infected cases. Thereby larger weights were assigned when the virus progression increased, and smaller weights were assigned when the virus progression decreased. The resulting sample sizes for each sub-region are calculated using the modified stratified sampling method. Further, to illustrate the accuracy of the sampling design, simulated data from different *epidemic scenarios, such as community spread, cluster spread, and border spread was used. This simulation allowed us to test the robustness of the techniques for the different states of the virus progression based on the infected cases. The sample size obtained through this dynamic sampling technique exhibits a direct correlation with the fluctuations in the number of infected cases, increasing as the infection cases rise and decreasing as they decline. In conclusion, the study introduces a novel sampling technique that accommodates the dynamic nature of population sizes, and it can be straightforwardly applied for the real-world data as well. Thus, this modified stratified sampling method emerges as a precise approach for capturing the actual prevalence of COVID-19.*

Keywords: Community spread, COVID-19, Epidemic scenarios, Stratified sampling technique, Virus progression

1 Introduction

The COVID-19 pandemic has taken the world by storm. Every nation on earth has joined in the fight against this vicious virus, and every little contribution matters. Originating as a novel coronavirus in late 2019, it quickly became a big health crisis worldwide, causing a lot of sickness, loss, and changes in society. World leaders became ill, entire countries were locked down to avoid the spread of infection and international travel ceased Wickramaarachchi et al. (2020).

The governments of all the countries have implemented different strategies to slow down the outbreak, for example, complete or partial lockdown in some areas, have taken emergency actions, have provided portable facilities, and temperature measuring facilities in all public places, vaccinations Knipl et al. (2009). However, the precise identification of individuals infected with COVID-19 stands as a paramount strategy in managing the crisis. Because diagnostic testing is our window onto the pandemic and how it is spreading. Therefore, adequate diagnostic tests should be conducted. Hence, an effective and accurate sampling technique is essential to capture the true prevalence of COVID-19 with the limited facilities of RT-PCR and RAPID-ANTIGEN tests. But, as of our current knowledge, there is no adequate mechanism in place for collecting samples for diagnostic testing. Many previous studies have been conducted to explore about COVID-19 pandemic in the world as well as in Sri Lanka. Thus far, no study has been identified that addresses building a dynamic sampling technique to capture the prevalence of COVID-19.

The profile of the coronavirus progression in a sub-region can be changed due to numerous factors such as population density, public mobility, and available health facilities Vasireddy et al. (2021). Accordingly, the determination of preventing strategies based on the entire profile may not be equivalently valid for each sub-region. With the limited testing facilities, it becomes necessary to allocate the number of tests among various sub-regions in a purposeful manner. Hence, the importance of employing a suitable sampling technique is highlighted. But the main problem is the standard sampling techniques do not address such dynamic population conditions in determining the sample size. Consequently, this study decided to employ a novel sampling technique that adapts to the dynamic population.

2 Methodology

The main objective of this study is to develop a novel sampling technique to capture the actual prevalence of COVID-19 with the limited available testing facilities.

2.1 Sampling

To answer the research question, it is doubtful that a researcher should be able to select a sample from an entire population. The population refers to the entire group that is the subject of the study. According to our study, it would be a larger community, region, or demographic that is affected by or vulnerable to COVID-19. A sample is a subset of the population selected for the study. The sample serves as the basis for drawing inferences about the larger population. Here, the selection of the correct sample is crucial with the available testing facilities. Therefore, a suitable sampling technique is essential to solve this problem. Sampling is the process of selecting a subset of elements from a larger population and there are diverse sampling methods available. Mainly it divides into two categories: Probability Sampling Technique and Non – Probability Sampling Technique.

2.1.1 Probability Sampling Technique

This Sampling technique uses randomization to make sure that every element of the population gets an equal chance to be part of the selected sample. It is alternatively known as random sampling Brown et al. (2001). Some common probability sampling techniques are: Simple Random Sampling, Stratified Random Sampling, Systematic Random Sampling, Cluster Sampling, Multi-Stage Sampling.

2.1.2 Non – Probability Sampling Technique

Non-Probability sampling can also be defined as a judgment or a non-random sampling. Every member or unit of the population does not get an equal chance to participate in the investigation. No random selection is made. The selection of the sample is made based on the subjective judgment of the investigator. These non-probability sampling techniques need not the population to be very precisely defined Taherdoost (2016). Some common probability sampling techniques are: Convenient Sampling, Snowball Sampling, Quota Sampling, Judgmental Sampling.

2.2 Stratified Random Sampling

Stratified sampling is where the population is divided into a number (L) of non-overlapping strata (or subgroups) and a random sample is taken from each subgroup. A subset is a natural set of items. Stratified sampling is often used where there is a great deal of variation within a population. Its purpose is to ensure that every stratum is adequately represented Kadilar et al. (2005). An advantage of this sampling design is that when a stratum is homogeneous, the measurements of the study variable (y) vary little from each other and the precise estimate of y can be obtained from a small sample in that stratum. Thus, combining these estimates from all L strata, the design produces a gain in the precision of the estimate of the variable in the entire population.

In this study, the stratified sampling technique was chosen as the foundational method for a novel sampling approach because the sample sizes across distinct sub-regions vary. However, the traditional stratified sampling technique may pose limitations in capturing infected samples with the dynamic nature of the population. Further, the target population of this study was the infected patients of all strata. Hence, the convenient stratified sampling technique was enhanced to capture the true prevalence of COVID – 19.

2.3 The Novel Dynamic Sampling Approach

With the limitations of the stratified sampling technique to capture the actual COVID – 19 prevalence, an extension of the traditional stratified sampling methodology was implemented. This extension is essential as the increase of infected cases boosts the virus spread, and the standard sampling techniques do not address such dynamic population conditions in determining the sample size. In alignment with the progression of the COVID – 19 viruses within each stratum, the allocation of samples varies. Subsequently, the larger number of samples is assigned for strata exhibiting increasing infection pattern, whereas a comparatively smaller number is assigned to sub – regions indicating a declining infection trend. As a results, a decision was made to introduce weighting factor for each sub – region, accounting the corresponding infection patterns.

2.3.1 Weight Allocation

The virus progression over a region exhibits a skewed pattern and this pattern can be used as a constrain to capture the true prevalence of the virus over the sub – regions. Thus, the weights are determined based on the first derivative of the infection curve. This derivative information is based on the recent dynamics of the infected cases.

The larger weights(ω_i) were assigned when the virus progression increased, and smaller weights (ω_i) were assigned when the virus progression decreased

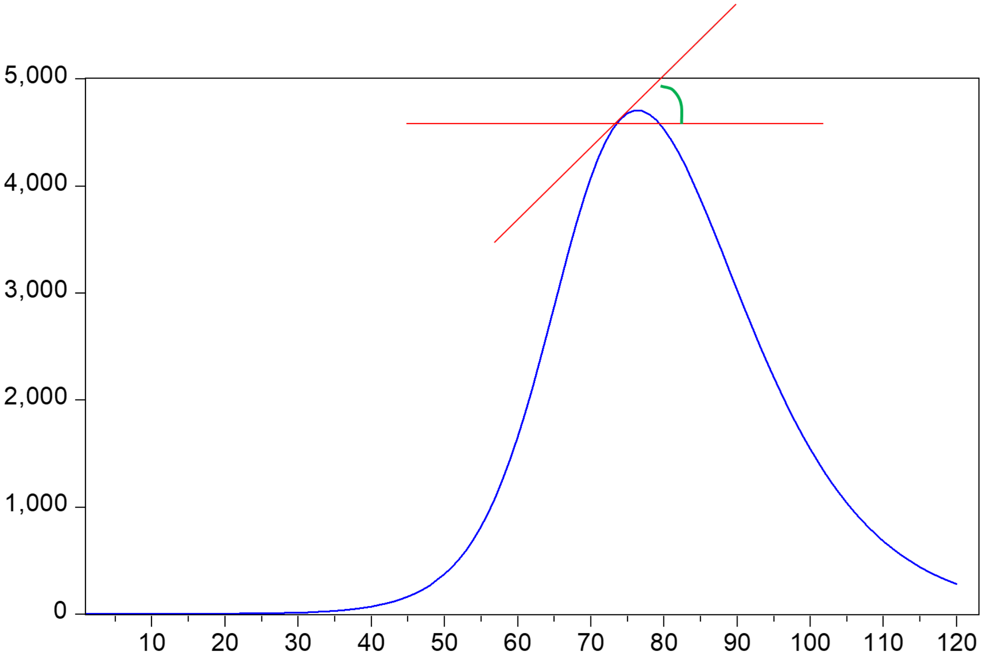


Fig. 1: General pattern of COVID-19 virus progression and slop calculation

using following candidate functions:

$$\omega_i = \begin{cases} 1+\tan\frac{\theta}{10} & ;0\le\theta<\frac{\pi}{2} \\ \frac{1}{\omega_i;\quad 0\le\theta<\frac{\pi}{2}} & ;\frac{\pi}{2}<\theta\le\pi \end{cases} \tag{1}$$

where θ is the slope of the infection curve varies from 0 to π.

2.3.2 Determination of Optimum Sample Size

After allocating the optimum weights for each distinct sub – region, the optimum sample size for diverse sub – regions were calculated. Our empirical sampling model was derived by enhancing the Neyman Optimum Allocation of Stratified Sampling Approach.

The best choice of sample size ($m_i = n_i\omega_i$) such that it minimizes the variance of $\hat{P}$ for a fixed total sample size n and the weights ω_i is,

$$m_i = n\frac{\sum_{i=1}^{L} N_i\sqrt{\omega_i P_i(1-P_i)}}{\sum_{i=1}^{L} N_i\sqrt{\omega_i P_i(1-P_i)}} \tag{2}$$

Where N_i = Number of sampling units in i^{th} stratum, n= Total sample size or Number of available testing kits per day, n_i = Number of sampling units to be drawn from i^{th} stratum, A_i = Number of COVID – 19 positive cases of i^{th}

stratum, $P_i = \frac{A_i}{N_i}$ = Population proportion of i^{th} stratum, ω_i = Weight for i^{th} stratum, m_i = The optimum sample size of i^{th} stratum.

2.4 Data Simulation

To illustrate the accuracy of the sampling design, simulated data generated using Matlab software for different epidemic scenarios, such as community spread, cluster spread, and border spread was utilized. Without loss of generality the data were simulated, partitioning the region Ω by placing a m×n spatial grid, where m, n ϵ N, and identify a sub-region as a grid cell. Assume that there are $N_{(i,j)}$ number of individuals in the cell $(\Omega_{i,j})$ for $1 \leq i \leq m$, $1 \leq j \leq n$ and i, j ϵ $\mathbb{N}$.

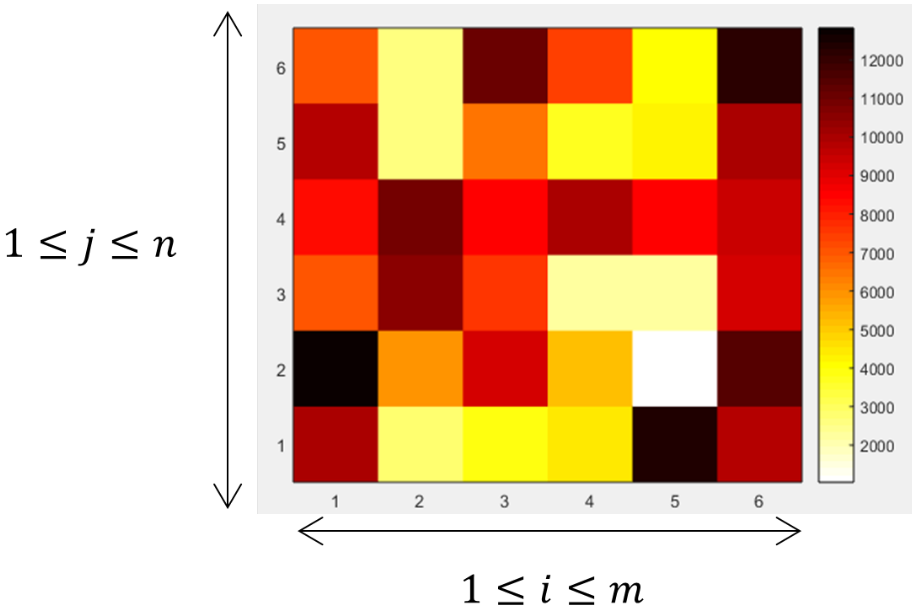
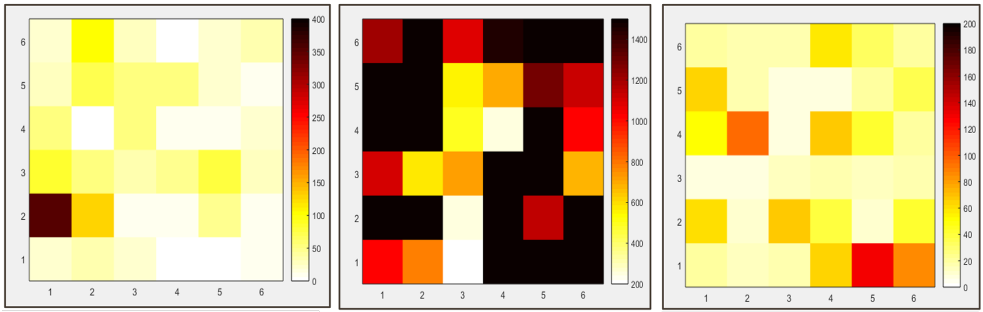


Fig. 2: Illustration of grid

Throughout the simulation studies, the 120 days of infections and 36 sub-regions (6×6 grid) were used and assumed that a total of 2,819,027 individuals (N) populate the 6×6 grid. Further, these simulated data have been obtained from an epidemiological model (SEIRD model).

2.4.1 Community Spreading Scenario

Community spreading of COVID-19 refers to the transmission of the virus within a local community where the source of infection is unclear or not directly linked to travel or known cases. It indicates that the virus is circulating among people in a particular area, making it challenging to trace and control its spread Arti et al. (2020).



(a)

(b)

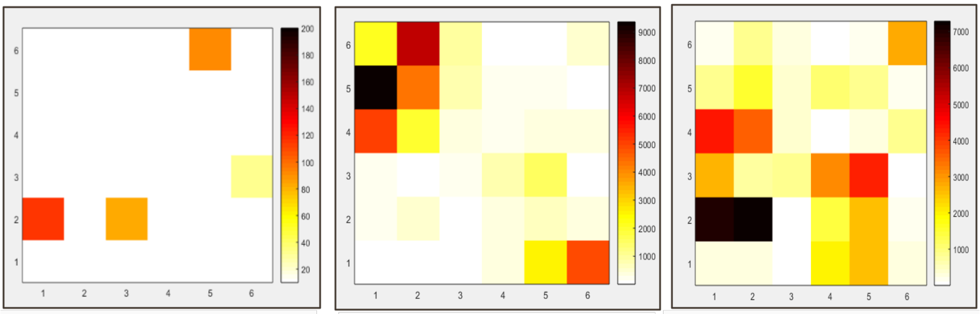
(c)

Fig. 3: Community Spread scenario: (a) Initial Infections (b) Infected individuals after 75 days (c) Infected individuals after 120 days

The initial phase, denoted by color plot (a), illustrates the distribution of the number of initially infected individuals at time 1. At this point, the virus has permeated the entire community. Subsequent color plot (b) depicts a progressive escalation in the spread of infection. Then the virus spread decreases and finally the spread has vanished as shown in color plot (c).

2.4.2 Cluster Spreading Scenario

Cluster spreading in the context of infectious diseases refers to the occurrence of localized outbreaks or clusters of cases within a specific geographic area or community. Unlike widespread community spreading, cluster spreading involves a more concentrated and identifiable grouping of infections Masrur et al. (2020).



(a)

(b)

(c)

Fig. 4: Cluster Spread scenario: (a) Initial Infections (b) Infected individuals after 80 days (c) Infected individuals after 120 days

In color plot (a) illustrates the primary phase COVID – 19 infections and they indicate that the virus is spread only to some particular sub-regions. Subsequently, the colour plot (b), illustrates the highest infection stage of the actual infections of COVID – 19 and wherein clusters propagate the virus throughout the societal landscape. Finally, as colour plot (c) the spreading of the virus is vanishing solely from the society.

2.4.3 Border Spreading Scenario

Border spreading in the context of COVID-19 typically refers to the transmission of the virus across national or regional borders. This transmission can occur through various means, including travel, trade, or other forms of cross-border interactions Zhu et al. (2021).

The colour plot (a) illustrates the initial stage of the infections of COVID – 19 and they indicate that the virus is spread only to the sub-regions within the border. Other sub-regions with white color do not show any infected patients. The high infection rate within the border stage of the infections of COVID – 19 is illustrated by the colour plot (b) and now the spread of the virus within the border is very high and the infection rate of that particular sub-regions is increased according to the time. The colour plot (c) illustrates the decreasing infection rate within the border stage of the infections of COVID – 19 and now the spread of the virus within the border is decreasing and the infection rate of the sub-regions beyond the boundary is increasing according to the time. Finally, the spread of the virus within the border is decreasing and the infection rate of the sub-regions beyond the boundary is increasing according to the time and it illustrates by the colour plot (d).

3 Results and Discussion

This simulation allowed us to test the robustness of the technique for the different states of the virus progression based on the infected cases. The color plots were used to illustrate the accuracy of the developed dynamic sampling technique. If the sample size obtained through this dynamic sampling technique exhibits a direct correlation with the fluctuations in the number of infected cases, increasing as the infection cases rise and decreasing as they decline, it can be deduced that this sampling technique works better with the dynamic nature of the population.

3.1 Simulation Results of Community Spreading Scenario

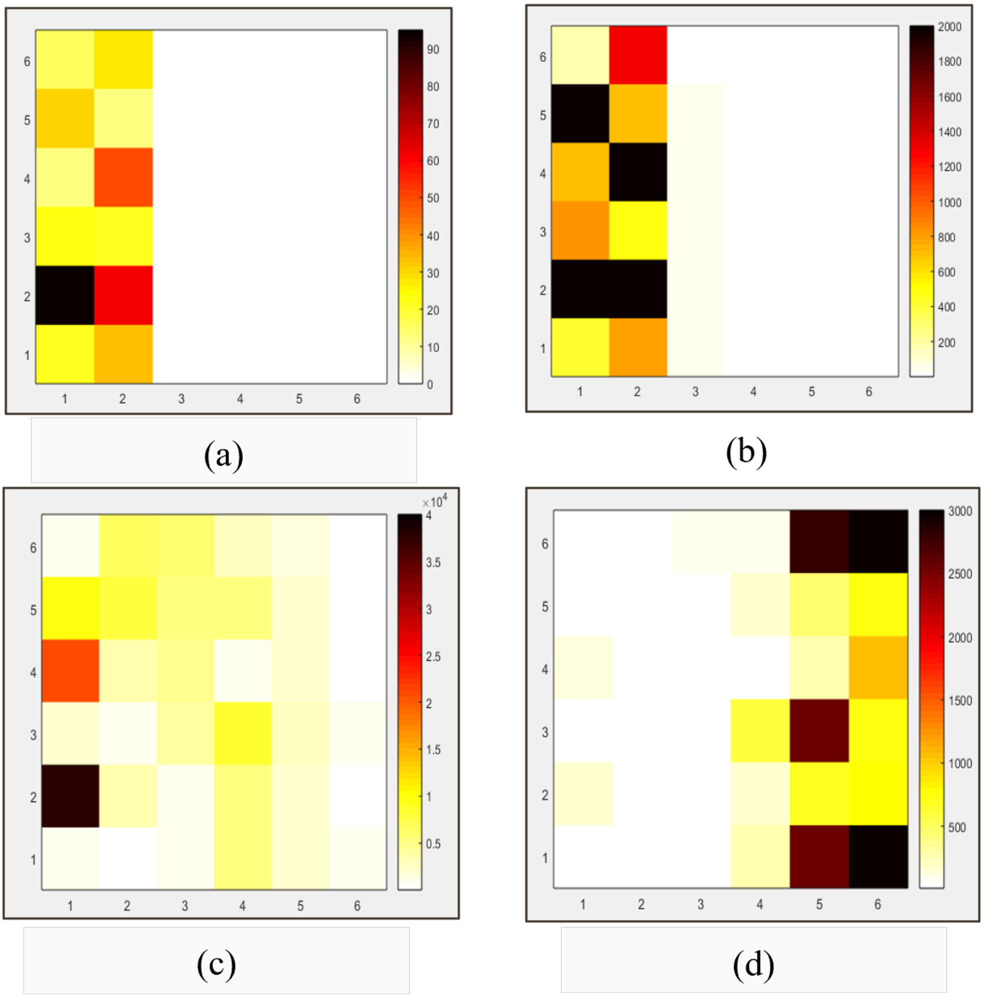
The color plots depicted in Figure 6 illustrating the efficacy of the novel dynamic sampling technique in the context of a community spreading scenario, compared to the actual infection rate. This technique demonstrates a heightened level of precision in yielding simulated data across all stages, encompassing the initial phase, the peak infection stage, and the subsequent decrease in infection rates.

3.2 Simulation Results of Cluster Spreading Scenario

In Figure 7, conveyed through color plots, illustrates the representation of a novel dynamic sampling technique designed for scenarios involving the spreading of clusters. The developed methodology captured the actual infection rate accurately for all three stages of cluster spreading scenario and it proved that this novel technique sensitive for the dynamic nature of the population.

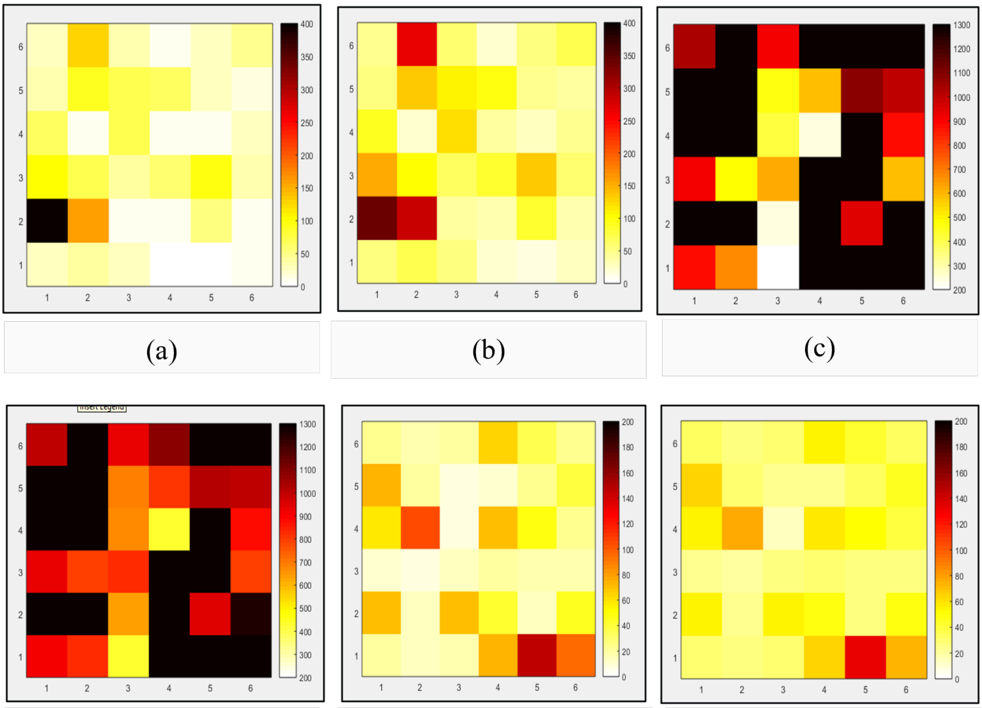
3.3 Simulation Results of Border Spreading Scenario

Based on the illustration of Figure 3, the actual infection of the border spreading scenario and the estimated infection pattern is very similar to each other for all four stages. It has been proved that this developed sampling technique suitable for this border spreading epidemic scenario as well.



(a) (b) (c) (d)

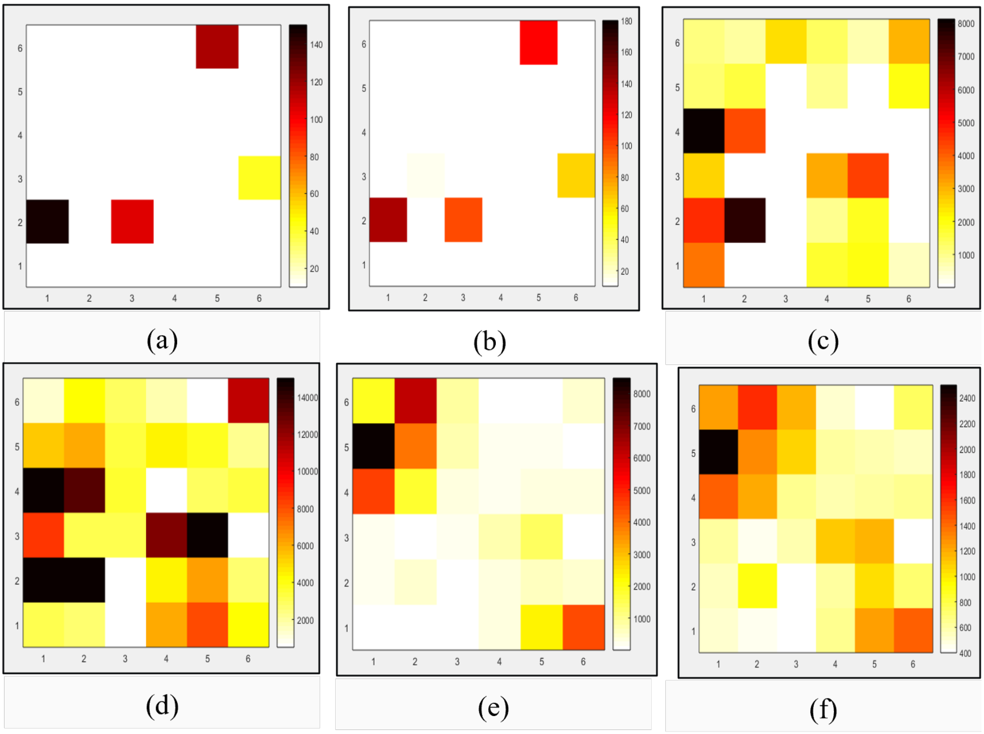
Fig. 5: Border spread scenario: (a) Initial infected individuals' distribution when border cross is not allowed. (b) Infected at time (T = 65) when border cross is allowed., (c) Infected at time 75 (T = 75) when border cross is allowed, (d) Infected individuals at time 120



(a) (b) (c)

(d) (e) (f)

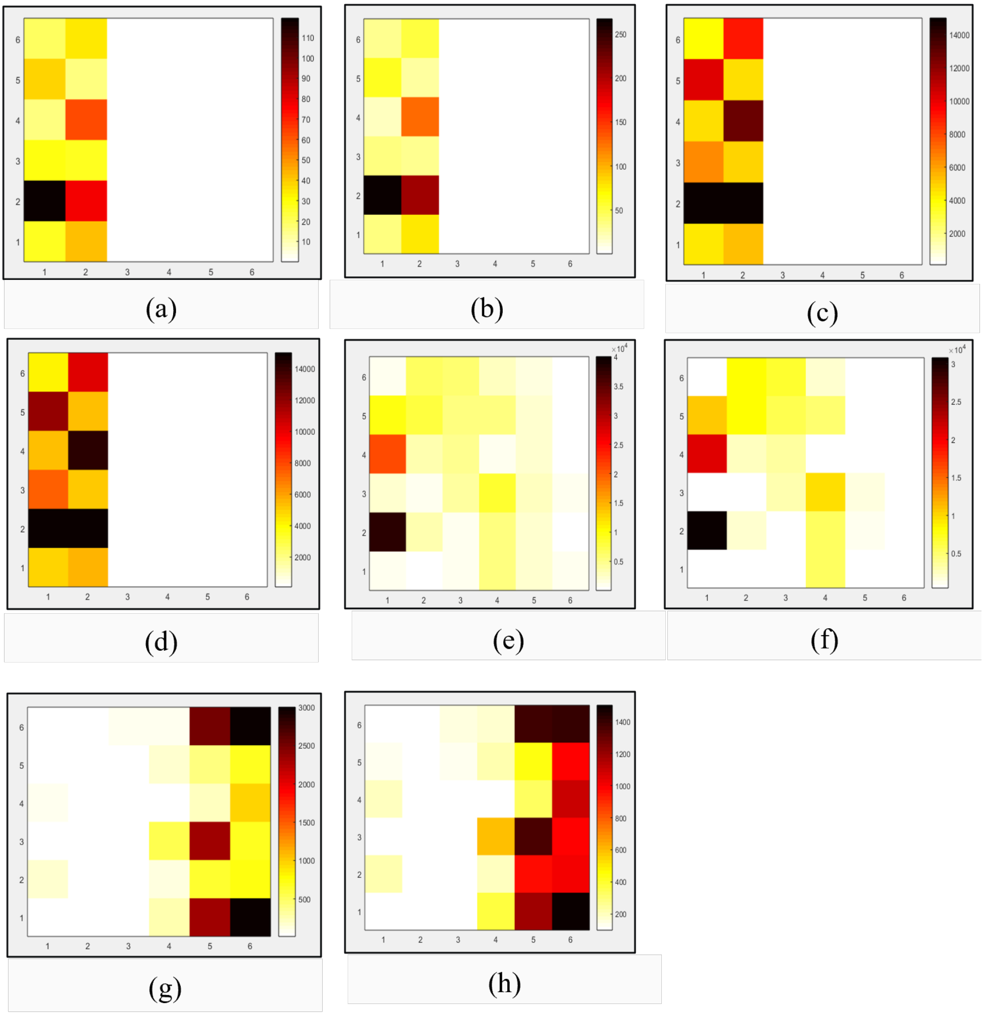
Fig. 6: (a) Actual Infection of Community Spread in initial stage (b) Estimated Infection of Community Spread in initial stage (c) Actual Infection of Community Spread in high infection stage (d) Estimated Infection of Community Spread in high infection stage (e) Actual Infection of Community Spread in infection rate decreasing stage (f) Estimated Infection of Community Spread in infection rate decreasing stage



(a) (b) (c)

(d) (e) (f)

Fig. 7: (a) Actual Infection of Cluster Spread in initial stage (b) Estimated Infection of Cluster Spread in initial stage (c) Actual Infection of Cluster Spread in high infection stage (d) Estimated Infection of Cluster Spread in high infection stage (e) Actual Infection of Cluster Spread in infection rate decreasing stage (f) Estimated Infection of Cluster Spread in infection rate decreasing stage



(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

Fig. 8: (a) Actual Infection of Border Spread in initial stage (b) Estimated Infection of Border Spread in initial stage (c) Actual Infection of Border Spread in high infection stage within the border (d) Estimated Infection of Border Spread in high infection stage (e) Actual Infection of Cluster Spread in infection rate decreasing stage (f) Estimated Infection of Cluster Spread in infection rate decreasing stage (e) Actual Infection of Cluster Spread in infection rate increasing stage of outside the border (f) Estimated Infection of Cluster Spread in infection rate increasing stage of outside the border

4 Conclusion

The COVID-19 pandemic has had a significant global impact, prompting every nation to join forces in combating this formidable virus. In light of this, even the smallest contributions hold immense value. Consequently, the primary objective of this study is to develop a dynamic sampling technique capable of capturing the actual prevalence of the virus within a dynamic population. The proposed sampling technique incorporates weight allocation based on the slope of the infection curve that means proportion to the current infection dynamics, resulting in more precise outcomes when applied to simulated data encompassing various epidemic scenarios: such as community spread, cluster spread, and border spread. It is important to note that this study did not employ real-world data, instead utilizing simulated data from the SEIRD (Susceptible–Exposed–Infected–Recovered–Deaths) model. This simulation allowed us to test the robustness of the techniques for the different states of the virus progression based on the infected cases. Further all the results were proved that this proposed sampling technique sensitive to the dynamic nature of population sizes and it can be straightforwardly applied for the real-world data as well. Additionally, all the results were visually represented using color plots, with accuracy assessments also relying on these plots. It should be acknowledged that there currently lacks a definitive mechanism for accurately determining the effectiveness of this dynamic sampling technique. However, at the end of the day, this novel dynamic sampling design can be considered as a precise and accurate technique to capture the actual prevalence of COVID-19 because it's sensitive for the dynamic population.

References

Knipl, D., and Rost, G., 2009. Modelling the strategies for age specific vaccination scheduling during influenza pandemic outbreaks, arXiv preprint arXiv, 0912.4662.

Brown, L. D., Cai, T. T., and DasGupta, A. ,2001. Interval estimation for a binomial proportion, *Statistical science*, 16(2): 101-133.

Wickramaarachchi, W. P. T. M., Perera, S. S. N., and Jayasinghe, S. , 2020. COVID-19 epidemic in Sri Lanka: A mathematical and computational modelling approach to control. *Computational and mathematical methods in medicine*.

Vasireddy, D., Vanaparthy, R., Mohan, G., Malayala, S. V., and Atluri, P. , 2021. Review of COVID-19 variants and COVID-19 vaccine efficacy: what the clinician should know?. *Journal of Clinical Medicine Research*, 13(6): 317.

Taherdoost, H., 2016. Sampling methods in research methodology; how to choose a sampling technique for research.

Kadilar, C., and Cingi, H., 2005. A new ratio estimator in stratified random sampling. *Communications in Statistics—Theory and Methods*, 34(3): 597-602.

Arti, M. K., and Bhatnagar, K., 2020. Modeling and predictions for COVID 19 spread in India. *ResearchGate, DOI* , DOI, 10.

Masrur, A., Yu, M., Luo, W., and Dewan, A., 2020, July. Space-time patterns, change, and propagation of COVID-19 risk relative to the intervention scenarios in Bangladesh. *International journal of environmental research and public health*, 17(16), 5911.

Zhu, Z., Weber, E., Strohsal, T., and Serhan, D., 2021. Sustainable border control policy in the COVID-19 pandemic: A math modeling study. *Travel medicine and infectious disease*, 41, 102044.